TEHNIČKI GLASNIK

TEHNIČKI GLASNIK / TECHNICAL JOURNAL – GODIŠTE / VOLUME 19 – BROJ / ISSUE 2

LIPANJ 2025 / JUNE 2025 – STRANICA / PAGES 173-340

/

SVEUČILIŠTE SJEVER / UNIVERSITY NORTH – CROATIA – EUROPE

ISSN 1846-6168 (PRINT) / ISSN 1848-5588 (ONLINE)

TECHNICAL JOURNAL

ISSN 1846-6168 (Print) ISSN 1848-5588 (Online)

TEHNIČKI GLASNIK - TECHNICAL JOURNAL

Scientific-professional journal of University North

Volume 19 Issue 2
Varaždin, June 2025 Pages 173-340

Editorial Office:
Sveučilište Sjever / University North – Tehnički glasnik / Technical journal
Sveučilišni centar Varaždin / University Center Varaždin
Jurja Križanića 31b, 42000 Varaždin, Croatia
Tel. ++385 42 493 338, Fax.++385 42 493 336
E-mail: tehnickiglasnik@unin.hr
https://tehnickiglasnik.unin.hr
https://www.unin.hr/djelatnost/izdavastvo/tehnicki-glasnik/
https://hrcak.srce.hr/tehnickiglasnik

Founder and Publisher:
Sveučilište Sjever / University North

Print:
Centar za digitalno nakladništvo, Sveučilište Sjever

All manuscripts published in journal have been reviewed.
Manuscripts are not returned.

The journal is free of charge and four issues per year are published
(In March, June, September and December)
Circulation: 100 copies

Journal is indexed and abstracted in:
Web of Science Core Collection (Emerging Sources Citation Index - ESCI), Scopus, EBSCOhost Academic Search Complete, EBSCOhost – One Belt, One Road Reference Source Product, ERIH PLUS, CITEFACTOR – Academic Scientific Journals, DOAJ – Directory of Open Access Journals, Hrčak – Portal znanstvenih časopisa RH

Registration of journal:
The journal "Tehnički glasnik" is listed in the HGK Register on the issuance and distribution of printed editions on the 18th October 2007 under number 825.

Preparation ended:	Published (online):	Published (print):
March 25, 2025	April 7, 2025	June 14, 2025

Legend:
(1) University North, (2) University of Slavonski Brod, (3) Faculty of Graphic Arts Zagreb, (4) Faculty of Civil Engineering Osijek, (5) Faculty of Engineering Rijeka, (6) Faculty of Mechanical Engineering and Naval Architecture Zagreb, (7) Faculty of Metallurgy Sisak, (8) Tomas Bata University in Zlín, (9) Department of Physics of the University of Josip Juraj Strossmayer in Osijek, (10) Faculty of Humanities and Social Sciences Osijek, (11) Karlovac University of Applied Sciences, (12) University of Applied Sciences Velika Gorica, (13) Department of Polytechnics - Faculty of Humanities and Social Sciences Rijeka, (14) Faculty of Electrical Engineering and Computer Science - University of Maribor, (15) Faculty of Civil Engineering - University of Maribor, (16) University College of Teacher Education of Christian Churches Vienna/Krems, (17) Faculty of Mechanical Engineering - Poznan University of Technology (Poland), (18) Mechanical Engineering Faculty Sarajevo, (19) University of Travnik - Faculty of Technical Studies, (20) Higher Education Technical School of Professional Studies in Novi Sad, (21) University of Novi Sad - Faculty of Technical Sciences, (22) Faculty of Mechanical Engineering - University of Montenegro, (23) Brno University of Technology, (24) Odessa State Academy of Civil Engineering and Architecture, (25) Faculty of Civil Engineering - University of Mostar, (26) Faculty of Manufacturing Technologies with the seat in Prešov - Technical University in Košice, (27) Faculty of Mechanical Engineering - University of Maribor, (28) College of Engineering, IT & Environment - Charles Darwin University, (29) Universite Libre de Bruxelles, (30) Vishwakarma Institute of Information Technology (Pune, India), (31) AISSMS Institute of Information Technology (Pune, India), (32) Permtech Research Solutions (India), (33) University of Belgrade, (34) National Dong Hwa University - Taiwan, (35) Faculty of Mechanical Engineering - Opole University of Technology (Poland), (36) TU Berlin - Germany, (37) Shahid Chamran University of Ahvaz - Iran, (38) University of Bologna - Italy, (39) University of Defence in Belgrade - Military Academy – Serbia, (40) Obuda University Budapest - Hungary

ISSN 1846-6168 (Print), ISSN 1848-5588 (Online)
https://doi.org/10.31803/tg-20230804071648

Original scientific paper
Received: 2023-08-04, Accepted: 2024-04-15

Theoretical Study of Various Fluids on the Transient Performance of a Two-Phase Closed Thermosyphon (TPCT)

Hesham Elkhatib*, Loula A. Shouman, Mohamed Abdel Aziz Salem

Abstract: Passive cooling heat exchangers are important for a nuclear reactor. Due to many advantages of passive cooling mechanism it may be used in nuclear reactors by employing two-phase closed thermosyphon (TPCT) during shutdown regime. Many thermosyphons can be penetrated in cooling system so that, the condenser section is exposed to air and the evaporator (evaporator) section immersed into water pool. Our study is proposed to predict the thermal behaviour of TPCT with different fluids, Water, Toluene, Acetone and Heptane. The fluid which realizes better heat transfer characteristics prediction is our study objective. The mathematical model is developed to predict thermal behaviour of these fluids to classify the most appropriate fluid. Water, acetone, heptane, and toluene are selected as working fluids in this study. Engineering Equation solver is used to formulate our model with finite difference numerical method. The results show that, Acetone has the highest heat transfer coefficient and the lowest thermal resistance in evaporator section. Water has the highest critical heat flux and highest evaporator surface temperature. Water as well has higher heat transmission coefficient in the evaporator section after Acetone. It has also low wall temperature and highest heat transfer coefficient in condenser section. Water heat transfer coefficient is three times more than the other fluids and more output cooling capacity. For the same size of TPCT, while the Acetone is the lowest thermal resistance, but water is more preferable working fluid because it is more available with low thermal resistance as well and highest cooling capacity which enables the TPCT to have more cooling capacity Validation of the proposed theoretical model is performed and showed a good evaporator and condenser surface temperature agreement with the empirical results.

Keywords: boiling; condensation; saturation pressure; saturation temperature; thermosyphon

1 INTRODUCTION

Heat pipes are the most common passive, capillary-driven of the two-phase systems. The liquid-to-vapour phase transition (boiling/evaporation and condensation) of a working fluid is a component of two-phase heat transfer. Simple two phase closed thermosyphon is a natural circulation heat pipe that, consists of a sealed pipe charged with working fluid under a vacuum condition without wick material [1]. Thermosyphon is passive two-phase heat transfer loops increase the thermal conductivity between a heat source and a heat sink by utilizing the very efficient thermal transport mechanism of evaporation and condensation. Additionally, thermosyphon heat pipes are found to have a number of general advantages that make them suitable for a variety of uses, including heating buildings with solar energy [2], water heaters, furnaces, and boilers, as well as cooling applications (such as cooling nuclear reactors, turbine blades, transformers, electronics, and internal combustion engines). In the last few decades, the thermosyphon systems were used as a passive cooling system for the spent fuel pool and nuclear reactors. Many academics have conducted theoretical and experimental research on the fundamental characteristics of two-phase closed thermosyphons. Thermosyphon functioning is affected by a number of variables, including fill ratio, working fluid, tilt, geometry, heat input, cooling water flow rate, and others. Patil Aniket D., and Yarasu Ravindra B. [3] in which they focused on aspects such as filling ratio, aspect ratio, heat load, mass flow rate, and inclination angle, all of which affect thermosyphon thermal performance, publish a review study. Many investigations were implemented in order to analyse and improve the thermosyphon's thermal performance, theoretically and/or experimentally to study the effect of mathematical Modelling, working fluid, surface geometry and mechanical modification. New heat transfer augmentation approaches such as resurfacing, ultrasonic wave utilization, and CFD analysis are also enlightening. Due to the extended application of heat pipe in the recent years, noticeable interest in the study of thermosyphon design and performance. These studies were carried out (but not limited to) by Thanya, et al [4], Filippo Cataldo and Yuri Carmelo Crea [5], M. G. Hammouda, et al [6], Ghasem, et al [7], and Myeongjin Kim, et al [8]. Several investigations were presented to study the effect of working fluids on the thermosyphon. Mozumder, et al. [9] designed, built, and tested a tiny heat pipe using the working fluids acetone, methanol, and water. Thermal resistance and total heat transfer coefficient were used to evaluate the heat pipe's performance. H. Jouhara, et al. [10] utilizing water and the dielectric heat transfer liquids FC-84, FC-77, and FC-3283 carried out a thermosyphon thermal performance experiment. S. M. Sadrameli, et al. [11] looked into the impact of working fluid occupancy inside heat pipes and the amount of heat input on thermosyphon performance. The operating fluid in a steel thermosyphon with a 16 mm diameter was toluene. Evaporator, adiabatic, and condenser section lengths were, respectively, 10, 23, and 17 cm. Russo, et al [12] investigated experimentally the thermal performance of various working fluids in thermosyphons that can be employed in thermal control of electronic equipment. Acetone, water, ethanol, and methanol were used as working fluids. A. Jay, D. Pingale, et al [13] constructed a two-phase closed thermosyphon (TPCT) with a 1 m length and outer diameter 19mm that is filled with different fluids (Propylene Glycol and Ethylene Glycol). Different fill volume ratios (40 and 60 %), and heat inputs (60 to 80 °C) were used to study the performance of thermosyphon. Rafal Andrzejczyk [14] investigated how several factors affected a wickless heat pipe's performance. With different filling ratio values (0.32, 0.51, 1.0), three

working fluids—Ethanol, Water, and SES36 (1, 1, 1, 3, 3-Pentafluorobutane) were investigated. Abdelrahim Abusafa and Aysar Yasin [15] investigated a two-phase closed thermosyphon system for cooling high heat flux electronic equipment. Using R-134a as a working fluid, the performance of the thermosyphon was studied. The influence of heat flux and the refrigerant pressure were investigated on the boiling section heat transfer coefficient. H. Z. Abou-Ziyan et al. [16] investigated the thermal performance of a two-phase closed thermosyphon using water and R134a as the working fluid in both stationary and vibratory situations. The thermal performance of an inclined two phase closed thermosyphon with distilled water and aqueous solution of n-Butane as a working fluid was explored by M. Karthikeyan, et al. [17]. Analytical calculations were made by Shwetabh Singh, et al. [18] to determine the contribution of the working fluid's thermo physical parameters to the formation of entropy. A calculation of entropy has been made using acetone, pentane, heptane, ammonia, pentane, and water as working fluids. Hussain Saad Abd, et al. [19] studied the wickless heat pipe's performance properties, including thermal power, working fluids, temperature change with varying input powers, thermal resistance, and heat transfer coefficient, which were shown experimentally. Various heat loads (from 20 to 60 W) and a filling ratio of 50 % were employed with the working fluids methanol and ethanol. M. M. Sarafraz, et al. [20], conducted a series of tests to assess the thermal performance and heat transmission of n-pentane-acetone and n-pentane-methanol mixtures inside a gravity-assisted thermosyphon heat pipe. The impact of a number of variables, including the input heat to the evaporator section, the carrying fluid filling ratio, the tilt angle of the heat pipe, and the carrying fluid type, on temperature distribution and heat pipe performance was examined. D. Mishkinis, et al. [21] carried out a review of prospective solutions for the development program "High Temperature LHP" and selected potential working fluids (including two-component combinations) and LHP materials for the envelope and wick. The typical working fluids for LHP are ammonia and propylene. Results show that if the temperature is close to 125 °C, organic liquids including acetone, ethanol, and methanol are incompatible with Ni wick due to a highly severe breakdown process. In several experiments and theoretical studies, the transient behaviour of closed two-phase thermosyphons was investigated to determine its impact on thermosyphon performance. The following list includes a few of these studies. Hichem Farsi, et al. [22] performed an experimental and theoretical investigation to examine the two-phase closed thermosyphon (TPCT) behaviour in transient regimes. Two types of TPCT responses were shown by experimental data. The mathematical model had been constructed to obtain an analytical expression of the system response time. In double-phase closed thermosyphon, Zhongchao Zhao and colleagues [23] studied thermal performance of phase change heat and mass transport in transient state by using computational fluid dynamics (CFD). To simulate condensation and evaporation in the thermosyphon at various heating inputs, a computational fluid dynamic model based on the fluid volume approach was created. This thermosyphon's operating fluid is deionized water, as stated. R. Parand, et al. [24] gave a theoretical analysis of the thermosyphon behaviour in the transient domain. To determine the temperature of the thermosyphon and the time required to attain steady state condition, a computer simulation programme based on the lumped technique has been created. Jiao Yonggang, et al. [25] presented a mathematical model to predict the unsteady state start-up process of heat pipe. The temperature, velocity and pressure distribution in the working trap was solved by FLUENT. According to the preceding review, several experimental and theoretical studies look into the effect of working fluid, while others look into the transient reaction. The goal of this research is to predict the transient response of various operating working fluids on the capability of (TPCT).

2 THEORETICAL MODEL

The two-phase closed thermosyphon (TPCT) is constructed up of an evacuated sealed tube that contains a small volume of liquid and is separated axially into three basic regions: evaporator (heating), adiabatic (thermally insulated), and condenser (cooling), as shown in Fig. 1. The heat from the evaporator component is transferred to the liquid in the thermosyphon through the pipe wall, causing it to boil. In the form of latent heat of vaporization, the heat is absorbed by the working fluid. The vapour moves upward because the evaporator zone has a higher pressure than the condenser zone. In the cooler condenser zone, the latent heat absorbed in the evaporator condenses and is released. The heat then exits the thermosyphon through the tube wall and into the outside, passing through the thin liquid coating. The flow circuit inside the tube is completed when the liquid is gravity-forced back to the evaporator section as a thin layer. The theoretical behaviour of a two-phase closed thermosyphon (TPCT) in the transient phase is investigated. The transient model was used to mimic the thermosyphon's response to various working fluids.

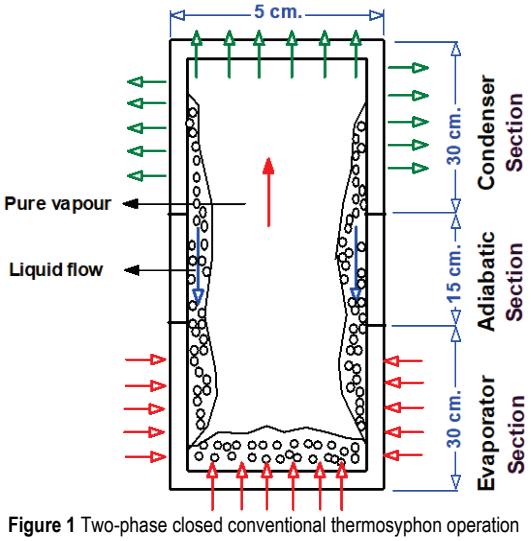


Figure 1 Two-phase closed conventional thermosyphon operation

2.1 Model Description and Assumptions

2.1.1 Thermosyphon Evaporating Section (Boiling)

A closed container makes up a thermosyphon, which collects energy along its evaporating part and extracts it along its condensation part. Look at a diameter D thin-mechanically polished wall material of thermosyphon. When heat is applied to the thermosyphon, saturated fluid is boiled at a specific internal pressure of evaporation part of length L_e. The vapour rise to condenser and condenses and falls by gravity down the wall of the condensation part of length L_c back into the evaporation part. When evaporation occurs at the surface that represents a liquid solid interface, it is termed boiling and the process occurs when the surface temperature in the boiling section T_{sb} exceeds the saturation temperature T_{sat} at adjusted pressure inside evaporation section. Heat energy is transferred from the outer surface to inside surface to the liquid by conduction and convection, which Newton's law of cooling [26] can be applied:

$$Q_{sb} = \overline{h}_{sb} A_{sb} \left(T_{sb} - T_{sat} \right) = \overline{h}_{sb} A_{sb} \Delta T_e \tag{1}$$

We can see vapour bubbles in this process, which propagates and subsequently separates from the surface. The dynamics and formation of vapour bubbles are intricately dependent on the excess temperature, the kind of surface, and the fluid's thermophysical characteristics like surface tension. The dynamics of vapour bubble production also affects the mobility of liquids close to the surface, which has a significant impact on the coefficient of heat transfer. The liquid is free of motion, and its movement at the surface is caused by free convection and mixing triggered by bubble development and detachment. Pool boiling can happen under a variety of circumstances.

There is nucleate boiling in the temperature range of 30 °C > T_e > 5 °C. Two distinct flow regimes may be observed in this range. At nucleation sites in the area, individual bubbles develop and detach from the surface. This separation causes a significant amount of fluid mixing close to the surface, significantly raising the heat transfer coefficients h and q_{sb}. In this regime, the majority of heat exchange occurs directly from the surface to a liquid that is moving at the surface rather than through the surface-born vapour bubbles.

Eq. (2) is developed by Rohsenow [27] and nucleate boiling uses this correlation widely. All liquid properties are evaluated at saturation temperature.

$$q_{sb} = \mu_l h_{fg} \left[\frac{g\left(\rho_l - \rho_v\right)}{\sigma_l} \right]^{0.5} \left(\frac{c_{p,l} \Delta T_e}{C_{s,f} h_{fg} \, Pr_l^n} \right)^3 \tag{2}$$

Critical heat flux represents an important characterization of boiling curve. It is necessary to operate evaporation process near to this point, but we as well consider the danger of dissipating heat in excess of this amount. Zuber [28] developed and, obtained an expression, which can be approximated as:

$$q_{max} = C_{s,f} h_{fh} \rho_v \left[\frac{\sigma_l g \left(\rho_l - \rho_v\right)}{\rho_v^2} \right]^{0.25} \tag{3}$$

At excess temperatures, a vapour film blankets and covers the surface until separation between the surface and liquid. Because conditions of laminar film condensation, it is common to consider film, boiling correlations on results obtained from condensation phenomena. One such result, which applies to film boiling on a cylinder diameter D, is of the form [26].

$$\overline{N}u_{sb} = \frac{\overline{h}_{sb} D}{k_v} = C_{s,f} \left[\frac{g(\rho_l - \rho_v) h_{fg} D^3}{\nu_v k_v (T_{sb} - T_{sat})} \right]^{0.25} \tag{4}$$

Where $\Delta T_e = T_{sb} - T_{sat}$.

The transient analysis takes place according to energy balance between boiling heat flux to the surface of boiling section and heat transferred to interior fluid by convection and conduction. This transient transfer occurs according to Eq. (5) [26].

$$(\rho V C_p)_{sb} \left(\frac{dT_{sb}}{dt} \right) = C_{s,f} h_{fg} A_{sb} \left[\frac{\sigma_l g(\rho_l - \rho_v)}{\rho_v^2} \right]^{0.25} - \\ - \mu_l h_{fg} \left[\frac{g(\rho_l - \rho_v)}{\sigma_l} \right]^{0.5} \left[\frac{c_{p,l}(\Delta T_e)}{C_{s,f} h_{fg} \, Pr_l^n} \right]^3 \tag{5}$$

The 60 % of maximum heat flux is our assumption [26] to be absorbed by heating section and it is assumed sufficient to boil the fluid at assigned vacuum pressures.

2.1.2 Condensing Section of Thermosyphon

Condensation happens when the temperature of a vapour is decreased below its saturation temperature. In industrial equipment like thermosyphon, the process results from direct contact between a cool surface and vapour. Its latent energy is released, heat energy is transferred to the surface, and the formation of condensate transforms the vapour into liquid phase.

The total heat transfer to the surface may be obtained by using Eq. (6) with and expressed as:

$$Q_{sc} = \overline{h}_{sc} A_{sc} (T_{sat} - T_{sc}) \tag{6}$$

The average Nusselt number in condensing section is:

$$\overline{N}u_{sc} = \frac{\overline{h}_{sc} L}{k_l} = 0.943 \left[\frac{\rho_l g(\rho_l - \rho_v) h_{fg} L^3}{\mu_l k_l (T_{sat} - T_{sc})} \right]^{0.25} \tag{7}$$

When the flow in tube is in the laminar, wavy regime, Kutateladze [29] advises a correlation:

$$\bar{N}u_{sc} = \frac{\bar{h}_{sc}\left(\frac{\nu_l^2}{g}\right)^{1/3}}{k_l} = \frac{Re_\delta}{1.08\,Re_\delta^{1.22} - 5.2} \tag{8}$$

Where,

$$Re_\delta = \frac{4q_{sc}}{h'_{fg}\mu_l \pi D} \tag{9}$$

And h'_{fg} is the modified latent heat of vaporization and expressed by:

$$h'_{fg} = 0.68\,h_{fg}cp_l\,(T_{sat} - T_{sc}) \tag{10}$$

And Re is Reynolds number based on liquid film thickness δ and its range for laminar flow is $30 \le Re \le 1800$.

The heat removed by condensing section depends on the heat absorbed by boiling section, so the transient equation yields the form [26]:

$$\left(\rho V C_p\right)_{sc}\left(\frac{dT_{sc}}{dt}\right) = \mu_l h_{fg}\left[\frac{g\left(\rho_l - \rho_v\right)}{\sigma_l}\right]^{0.5} * \left(\frac{C_{p,l}\Delta T_e}{C_{s,f}h_{fg}\,Pr_l^n}\right)^3 \pi D L_{sb} - \bar{h}_{sc}\pi D L_{sc}\left(T_{sat} - T_{sc}\right) \tag{11}$$

2.2 Solution Methodology

The solution of the previous Eqs. (5), (11) is modelled by finite difference method (Euler) using Engineering Equation Solver (EES) [31] which is used to formulate the model. The solution illustrates the average change in evaporator and condenser wall temperatures with time. The finite difference form of above equations yield to:

For evaporator section:

$$(\rho V C_p)_{sb}\left(\frac{T_{sb}^{t+1} - T_{sb}^t}{dt}\right) = C_{s,f}h_{fg}A_{sb}\left[\frac{\sigma_l g(\rho_l - \rho_v)}{\rho_v^2}\right]^{0.25} - -\mu_l h_{fg}\left[\frac{g(\rho_l - \rho_v)}{\sigma_l}\right]^{0.5}\left[\frac{c_{p,l}(T_{sb}^{t+1} + T_{sb}^t/2)}{C_{s,f}h_{fg}\,Pr_l^n}\right]^3 \tag{12}$$

For Condenser section:

$$(\rho V C_p)_{sc}\left(\frac{T_{sc}^{t+1} - T_{sc}^t}{dt}\right) = = \mu_l h_{fg}\left[\frac{g(\rho_l - \rho_v)}{\sigma_l}\right]^{0.5}\left(\frac{c_{p,l}\Delta T_e}{C_{s,f}h_{fg}Pr_l^n}\right)^3 \pi D L_{sb} - -\bar{h}_{sc}\pi D L_{sc}\left[T_{sat} - (T_{sc}^{t+1} + T_{sc}^t/2)\right] \tag{13}$$

Initial and boundary conditions: Temperature of the evaporator surface is assumed initially (Time=0) at the same temperature of the surrounding reactor pool temperature, while the condenser wall temperature is assumed as the surrounding air temperature in the reactor hall. The limit values of temperatures are the saturation temperature values corresponding to each fluid distinct vapour pressure which are selected at 40 °C. The TCPT is under different vacuum pressure related to each fluid type and that realizes 40 °C saturation temperature. This value keeps the pool temperature at 40 °C and enables TCPT to absorb accumulated residual heat. Tab. 1 illustrates these pressures.

Table 1 Vapour pressure for various types of fluids

Fluid type	Vapour pressure (Pa)	Saturation temperature (°C)
Water	7375	40
Acetone	56520	
Toluene	7903	
Heptane	12325	

2.3 Model Validation

The model validation is illustrated in Fig. (2), and compared with experimental results by Mohamed S. El-genk [30] while employing heat pipe to absorbing heat load of 600 watt via evaporating section and dissipating it by natural convection through condensing section to air. The pressure inside horizontal thermosyphon is kept at atmospheric pressure 1 bar and the material is copper. The model prediction shows reasonable results for evaporator and condenser wall surface temperatures with the measured values. The figure illustrates that; the difference between experimental results and model predictions do not exceed 5%.

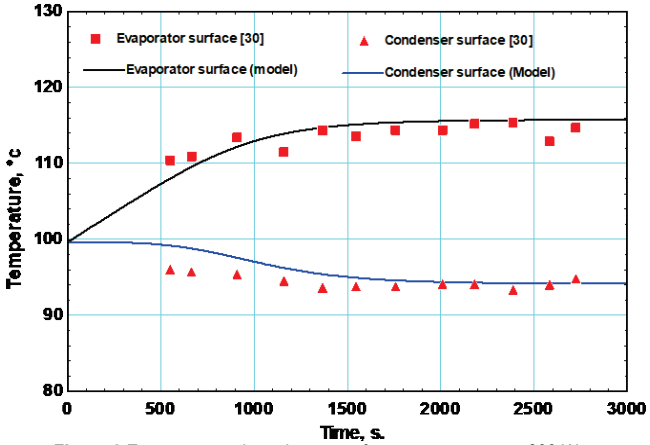


Figure 2 Evaporator and condenser surface temperatures at 600 Watt

3 RESULTS AND DISCUSSION

The theoretical model is design to study the transient behaviour of thermosyphon and the effect of the various working fluids on the thermosyphon feature. Saturation temperatures are adjusted to be the same for four fluids type by varying vacuum pressure inside heat pipe to satisfy the

same saturation temperature. The used thermosyphon is characterized by this information below:
Saturation temperature: 40 °C
Dimension:
Evaporator length: $30 + \pi d$ cm
Condenser length: $30 + \pi d$ cm
Adiabatic length: 15 cm
Diameter (d): 5.0 cm
Thickness: 3.0 mm
Material: Aluminium 6061
Nucleate boiling flux: 10 % of critical heat flux $C_{s,f} = 0.013$, and $n = 1.0$.

The results of the analytical modelling are primarily addressed and examined in order to evaluate the performance of a closed two-phase thermosyphon with different working fluids in transient operation.

3.1 Critical Heat Flux

The critical heat flux of considered fluids is shown in Fig. (3), which is calculated from Eq. (3) at saturation temperatures 40 °C and dependent on thermodynamic properties of each fluid. As shown in the figure; for the same condition of operation; the water has the greatest critical heat flux followed by Acetone, while Heptane and Toluene are the lowest and are relatively the same values. This behaviour is returned to the latent heat of the fluids, where it is the major parameter appears on the values.

It is assumed in model that the nucleate boiling heat flux is to be maintained at 10 % of the critical heat flux. Consequently this flux will be greater in Water and Acetone but will be lower in case of Heptane and Toluene which are the same.

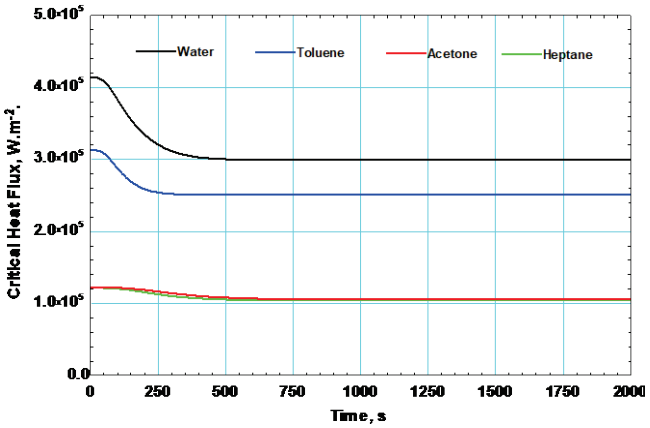


Figure 3 Critical heat flux of various fluids

3.2 Transient Evaporator and Condenser Temperature

The Fig. 4 illustrates the temperatures predicted by the mathematical model for evaporator wall surface versus time using various working fluids. It is shown from the figure that, at the beginning of calculation (when the wall surface temperature is reached approximately to the saturation temperature of the fluid about 40 °C) the evaporator surface wall temperature increases rapidly. This increasing in the temperature is due to the increase of heat flow from available temperature difference. However, the rise in temperature decreases until steady state condition is realized. This is an obvious fact due to the decrease in the driving force, namely the temperature difference. It is concluded from the Fig. 5 that, the wall temperature in the evaporator in case of water realizes higher temperature at steady state condition than the other fluids, this behaviour referred to the highest latent heat of vaporization of water and it means that water absorbs more heat and has more cooling capacity.

Also shown from the Fig. 5, the evaporator surface temperature achieves 15 °C higher than the saturation temperature for water while Heptane record only 7 °C rise due to the higher thermal conductivity of water than Heptane. The transient temperatures calculated by the mathematical model for condenser wall surface is illustrated in Fig. 5 for various working operating fluids. The figure demonstrates that, the condenser wall surface temperature decreases below the saturation temperature due to condenser cooling heat dissipation. With time, the rate of temperature decrease, however, slows until a steady state condition is established. This is a clear reality because the driving force, notably the temperature difference between wall and cooling air, has decreased. The condenser wall temperature in the Heptane reaches the steady state condition faster and has a lower temperature compared to the other fluids.

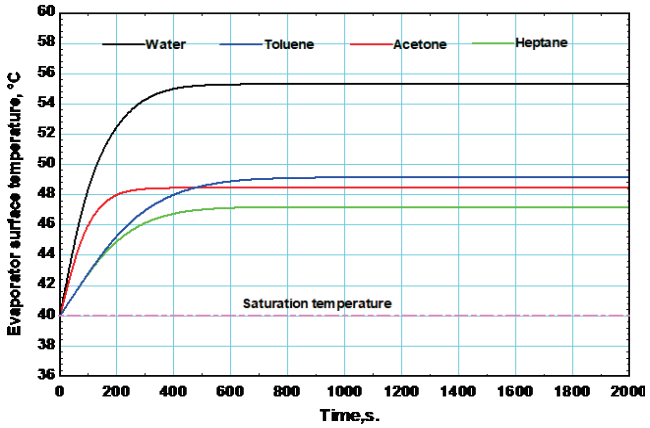


Figure 4 Transient Evaporator Surface Temperature for different working fluids

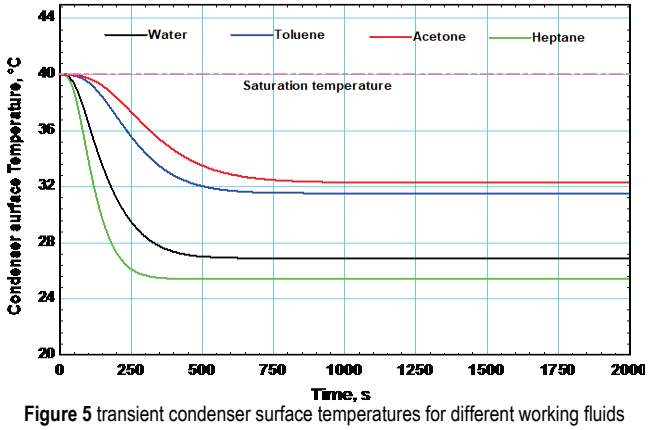


Figure 5 transient condenser surface temperatures for different working fluids

3.3 Evaporator and Condenser Heat Transfer Coefficient of Thermosyphon

The average heat transfer coefficients (h_e) and (h_c) in the evaporator and condenser are used to determine the degree of heat transfers within the thermosyphon. The average heat transfer coefficients are calculated using a mathematical model for both boiling, condensation and steady-state situations.

The transient heat transfer coefficient of the evaporator for the various thermosyphon working fluids is plotted in Fig. 6. At steady state, acetone has a higher heat transfer coefficient relative to other fluids and Toluene has a lowest heat transfer coefficient. As indicated in the figure, the evaporator heat transfer coefficient of Acetone has the greatest value of 2965 W/m^2·K, while the heat transfer coefficients for other operating fluids range from 1950, 1449 and 1152 W/m^2·K for Water, Heptane and Toluene respectively.

Coefficients are calculated using Newton's law of cooling. The values of condenser heat transfer coefficient are predicted from Eq. (8) based on some physical properties like kinematic viscosity, Reynolds number and thermal conductivity.

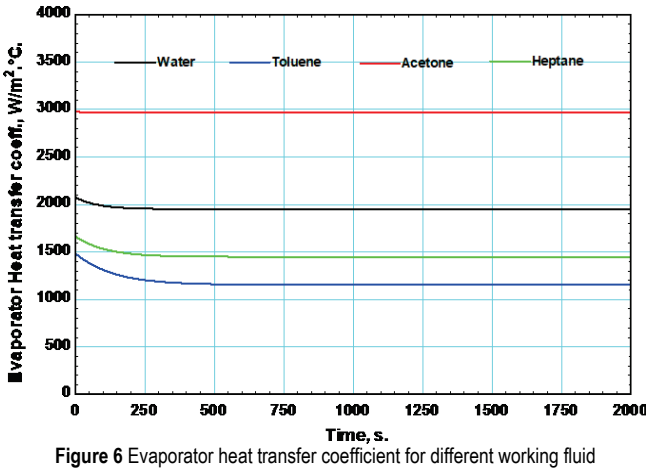


Figure 6 Evaporator heat transfer coefficient for different working fluid

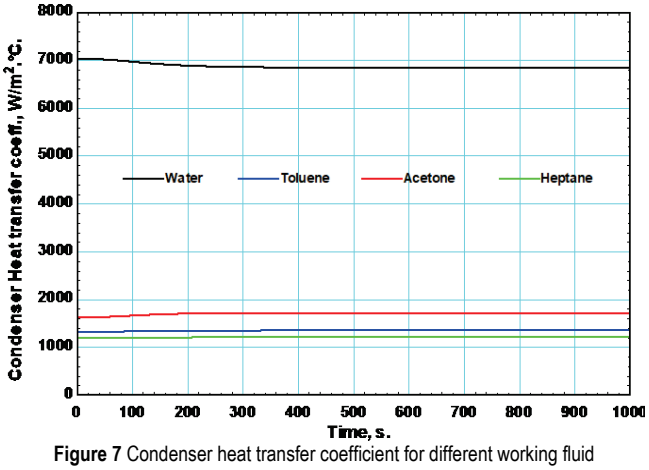


Figure 7 Condenser heat transfer coefficient for different working fluid

The transient heat transfer coefficient of the condenser in TPCT is shown in Fig. 8 for the various operating working fluids. As indicated in the diagram, the condenser heat transfer coefficient of Water has the greatest value then Toluene, Acetone and heptane respectively. The heat transfer coefficient of water is more than three times that of other working fluids.

3.4 Thermosyphon Resistance

Thermal resistance is an important parameter to the thermosyphon performance. The thermal resistance could be calculated using the below formula: $R = \frac{T_{sb} - T_{sc}}{Q}$, where Q is the heat absorbed by the heat pipe.

Fig. 8 shows the transient resistance of TPCT for the various operating fluids. As illustrate in the figure, the Toluene has the highest resistance than other fluids and the Acetone has the lowest. While the Acetone is the lowest resistance, the water is more used due to that, it is more available with no cost.

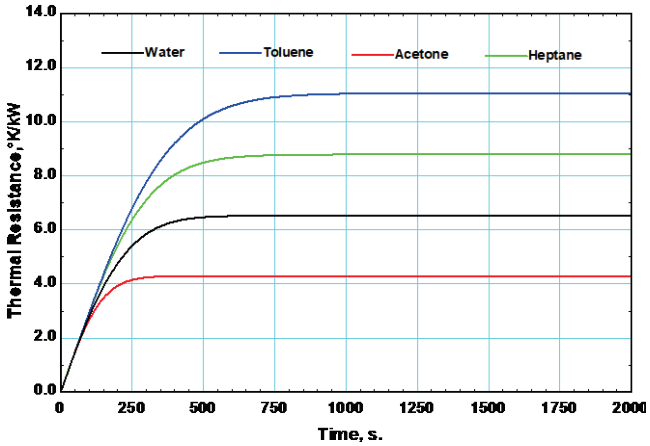


Figure 8 Heat pipe resistances for different working fluid at constant heat load

4 CONCLUSION

A mathematical model is developed to describe and study transient thermal behaviour of the two phase closed thermosyphon (TPCT) with different working fluids. The best fluid performance with TPCT is the main objective of the study. Evaporator and condenser surface temperature are the most important key results. Water, Toluene, Acetone and Heptane are used as working fluids in the present study. A validation of the mathematical model is presented using water as a working fluid by comparing the surface temperature of evaporator and condenser which are predicted from the model with experimental results [30]. The performance of TPCT with different fluids is presented. According to the results, Acetone has the highest heat transmission coefficient in the evaporator section. Its heat transfer coefficient is twice of Heptane and is slightly closer to water. Water has the highest critical heat flux, highest evaporator surface temperature and has relatively high evaporator heat transfer coefficient which means that water

absorbs more heat from surrounding heat load. In condenser, water has the maximum heat transfer coefficient, which is three times more than the other fluids. Water has lower heat resistance than heptane and Toluene and little higher than Acetone. Water is preferable working fluid in TPCT because it transfers more heat flux in the evaporator, dissipates more heat in condenser, has low thermal resistance and it is available with no cost.

5 REFERENCES

[1] Abdullahi, B., Al-dadah, K. R. & Mahmoud, S. (2019). *Thermosyphon Heat Pipe Technology*. IntechOpen. https://doi.org/10.5772/intechopen.85410

[2] Poulad, E. & Fung, A. (2012). Potential benefits from Thermosyphon-PCM (TP) integrated design for buildings applications in Toronto. *Proceedings of ESim: The Canadian Conference on Building Simulation*.

[3] Patil Aniket, D. & Yarasu Ravindra, B. (2012). Factors Affecting the Thermal Performance of Two phase closed thermosyphon A Review. *International Journal of Emerging Technology and Advanced Engineering, 2*(9).

[4] Parametthanuwat, T., Pipatpaiboon, N., Bhuwakietkumjohn, N. & Sichamnan, S. (2021). Heat transfer characteristics of closed-end thermosyphon (CE-TPCT). *Engineering Science and Technology, 27*, 101020. https://doi.org/10.1016/j.jestch.2021.05.024

[5] Cataldo, F. & Crea, Y. C. (2021). Experimental Analysis and Modeling of a Novel Thermosyphon System for Electronics Cooling. *ASME Journal of Electronic Packaging, 143*(4), 041110. https://doi.org/10.1115/1.4052670

[6] Hammouda, M. G., Ewing, D., Zaghlol, A. & Ching, C. Y. (2021). Heat Transfer Performance in a Vertical Grooved Thermosyphon. *International Journal of Thermofluids, 11*(23), 100107. https://doi.org/10.1016/j.ijft.2021.100107

[7] Beiginaloo, G., Mohebbi, A. & Afsahi, M. M. (2022). Combination of CFD and DOE for optimization of thermosyphon heat pipe. *Heat Mass Transfer, 58*, 561-574. https://doi.org/10.1007/s00231-021-03130-w

[8] Kim, M. & Moon, J. H. (2021). Numerical investigation and modeling of thermal resistance and effective thermal conductivity for two-phase thermosiphon. *Case Studies in Thermal Engineering, 27*, 101358. https://doi.org/10.1016/j.csite.2021.101358

[9] Mozumder, A. K., Akon, A. F., Chowdhury, M. S. H. & Banik, S. C. (2010). Performance of Heat Pipe for Different Working Fluids and Fill Ratios. *Journal of Mechanical Engineering, 41*(2). https://doi.org/10.3329/jme.v41i2.7473

[10] Jouhara, H., Martinet, O. & Robinson, A. J. (2008). Experimental Study of Small Diameter Thermosyphons Charged with Water, FC-84, FC-77 & FC-3283. *The 5th European Thermal-Sciences Conference*, Netherlands.

[11] Sadrameli, S.M., Forootan, D. & Farajimoghaddam, F. (2021). Effect of working fluid inventory and heat input on transient and steady state behavior of a thermosyphon. *J Therm Anal Calorim, 143*, 3825-3834. https://doi.org/10.1007/s10973-020-09294-7

[12] Russoa, G., Krambeck, L., Nishidaa, F., Santos, P. & A. A. Thiago. (2018). Thermal Performance of Thermosyphon for Different Working Fluids. *Revista de Engenharia Térmica, 15*(1). https://doi.org/10.5380/reterm.v15i1.62150

[13] Pingale, A. D., Patil, C. M. & Marathe, P. B. (2017). Experimental Investigation of Thermosyphon for Different Parameters. *Journal of Mechanical and Civil Engineering, 14*(1), 76-85. https://doi.org/10.9790/1684-1401067685

[14] Andrzejczyk, R. (2019). Experimental Investigation of the Thermal Performance of a Wickless Heat Pipe Operating with Different Fluids: Water, Ethanol, and SES36. Analysis of Influences of Instability Processes at Working Operation Parameters. *Energies, 12* (1), 80. https://doi.org/10.3390/en12010080

[15] Abusafa, A. & Yasin, A. (2020). Effect of Pressure on the Performance of Passive Two-Phase Closed Thermosyphon System Using R-134a. *Journal of Engineering Research and Technology, 7*(1). https://doi.org/10.33976/JERT.7.1/2020/1

[16] Abou-Ziyan, H. Z., Helali, A., Fatouh, M. and Abo El-Nasr, M. M. (2001). Performance of stationary and vibrated thermosyphon working with water and R134. *Applied Thermal Engineering, 21*, 813-830. https://doi.org/10.1016/S1359-4311(00)00089-2

[17] Karthikeyan, M., Vaidyanathan, S. & Sivaraman, B. (2010). Thermal performance of a two phase closed thermosyphon using aqueous solution. *International Journal of Engineering Science and Technology, 2*(5), 913-918.

[18] Singh, S. (2013). *The Effect of Working Fluid on Entropy Generation in a Heat Pipe System*. Department of Mechanical Engineering, National Institute of Technology Karnataka Surathkal, p.o. Srinivasnagar-575025, Mangalore, India.

[19] Abd, H. S., Assi H. H. & Abdulmunem, A. R. (2019). Thermal Performance of Wickless Heat Pipe with Methanol and Ethanol Working Fluids. *International Journal of Mechanical Engineering and Technology, 10*(1), 277-286.

[20] Sarafraz, M. M., Zhe, T., Tlili, I., Kazi, S. & Goodarzi, M. (2019). Thermal evaluation of a heat pipe working with n-pentane-acetone and n-pentane-methanol binary mixtures. *Journal of Thermal Analysis and Calorimetry, 139*(3). https://doi.org/10.1007/s10973-019-08414-2

[21] Mishkinis, D., Prado, P., Sanz, R., Radkov, A., Torres, A. & Tjiptajardja, T. (2009). Loop Heat Pipe working fluids for intermediate temperature range: from –40ºC to +125ºC. https://www.researchgate.net/publication/257859443.

[22] Farsi, H., Joly, J.-L., Miscevic, M. & Platel, V. (2003). An experimental and theoretical investigation of the transient behavior of a two-phase closed thermosiphon. *Applied Thermal Engineering, 23*, 1895-1912. https://doi.org/10.1016/S1359-4311(03)00147-9

[23] Zhao, Z., Zhang, Y., Zhang, Y., Zhou, Y. & Hu, H. (2018). Numerical Study on the Transient Thermal Performance of a Two-Phase Closed Thermosyphon. *Energies, 11*, 1433. https://doi.org/10.3390/en11061433

[24] Parand, R., Rashidian, B., Ataei, A. & Shakiby, K. (2009). Modelling the transient response of the thermosyphon heat pipes. J*ournal of Applied Sciences, 9*, 1531-1537. https://doi.org/10.3923/jas.2009.1531.1537

[25] Yonggang, J., Guodong, X., Dan, W., Wie, W. & Suge, H. (2012). Transient Characteristics of a two-phase Closed Thermosyphon with Isothermal Characteristics. *CIESC Journal, 63*(11).

[26] Incropera F. P., DeWitt, D. P., Bergmana, T. L. & Lavine, A. S. (2011). *Fundamentals of Heat and Mass Transfer*. 6th Edition, by John Wiley & Sons.

[27] Rohsenow, W. M. (February 21, 2022). Heat Transfer and Temperature Distribution in Laminar-Film Condensation. *ASME. Trans. ASME. November 1956, 78*(8), 1645-1647. https://doi.org/10.1115/1.4014125

[28] Zuber, N. (January 25, 2022). On the Stability of Boiling Heat Transfer. *ASME. Trans. ASME. April 1958, 80*(3), 711-714. https://doi.org/10.1115/1.4012484

[29] Kutateladze, S. S. (1963). *Fundamentals of Heat Transfer*. Academic Press, New York.
[30] El-Genk, M. S. & Lianmin, H. (1990). An experimental investigation of the transient response of a water heat pipe. *Int. J. Heat Mass Transfer, 36*(15), 3823-3830. https://doi.org/10.1016/0017-9310(93)90062-B
[31] Klein, S. A. & Alvarado, F. L. (1999). *EES manual*. Middleton, Wisconsin, USA
[32] Çengel, Y. A. (2002). *Heat Transfer a Practical Approach*. New York: McGraw-Hill, second edition.

Authors' contacts:

Hesham Elkhatib, Assistant Professor
(Corresponding author)
Reactors Department, Nuclear Researches Center,
Egyptian Atomic Energy Authority, Cairo, Egypt
heskhatib1966@yahoo.com

Loula A. Shouman, Assistant Professor
Reactors Department, Nuclear Researches Center,
Egyptian Atomic Energy Authority, Cairo, Egypt
loulashouman@yahoo.com

Mohamed Abdel Aziz Salem, Assistant Professor
Department of Nuclear Safety Researches and Radiological Emergencies,
NCRRT Center, Egyptian Atomic Energy Authority, Cairo, Egypt
bidosalem2@yahoo.com

Nomenclature:

A Area, m^2
C_p Specific heat, J/kg·°C
D Diameter, m
L Length, m
g Acceleration of gravity, 9.81 m/s^2
h Heat transfer coefficient, W/ m^2·°C
h_{fg} Latent heat of vaporization, J/kg·°C
k Thermal conductivity, W/m·°C
q Heat flux, W/m^2
Q Heat load, W
t Time, s
dt Time difference, s
T Temperature, °C
dT Temperature difference, °C
C Nucleate boiling coefficient
n Nucleate boiling exponent

Greek Symbols
μ Dynamic viscosity, N·s/m^2
β Thermal expansion coefficient, K^{-1}
v kinematic viscosity, m^2/s
ρ Density, kg/m^3
σ Surface tension, N/m
Δ Film thickness, m

Dimensionless Groups
Pr Prandtl number,
Re Reynolds number,
Nu Nuselt Number
Subscripts

a Adiabatic,
b Boiling
c Condenser,
e Evaporator
f Fluid,
l Liquid
s Surface
sat Saturation
v Vapour
sb Surface of boiling
sc Surface of condensation
s,f Surface-fluid combination

ISSN 1846-6168 (Print), ISSN 1848-5588 (Online)
https://doi.org/10.31803/tg-20230822132129

Original scientific paper
Received: 2023-08-22, Accepted: 2024-04-16

Enriched I-Q Active Filter Based Harmonics Mitigation in Rectifier-Inverter Fed Induction Motor Drive

Samsher Kadir Sheikh*, Manish Jain, Pawan Chandrakant Tapre

Abstract: In recent decade, an introduction of innovative technological development in power electronics sector is to perform significantly for compensation of power quality issues. The sugar industry's widespread implementation of variable-frequency induction-motor drives cause the electrical distribution systems to become worse. Greater harmonics produced by semiconductor current converters affects the sinusoidal voltage and increase reactive component of current by voltage. To overcome the aforementioned issues, the research paper examines issues of power quality in an induction motor (IM) drive system attached to utility grid system. The addition of filters, which reduces the power quality, issues in induction motor drives by providing lower THD. The IM drive system with an enhanced filter is employed to reduce harmonics produced by power semiconductor switches to keep the three-phase power supply in good condition. So, the instantaneous imaginary controller-based shunt active power filter (I-Q active filter) is employed to lessen the harmonic content that exists in the current at the point of source. To enhance the electrical network system's power quality, the proposed filter is injected parallel within the source and the IM drive system. The outcomes of simulations are analysed and it is executed by MATLAB/Simulink.

Keywords: harmonics; induction motor drive; I-Q active filter; power quality; sugar industry

1 INTRODUCTION

The necessity for electronic systems and devices has risen substantially over the past few decades, primarily as a consequence of growth in the economy and population [1]. As a consequence, power electronics and AC machine drives are making progress in a number of industries, such as manufacturing, electronic goods, electric vehicles, non-conventional energy and power generation plants [2, 3]. Because of the variable speed, high efficiency, and affordability, three phase IM drives are frequently utilized in manufacturing applications. The progress in solid-state power conversion has opened up opportunities for the flexible deployment of variable frequency induction motor drives (VFIMDs) in various fields such as air conditioning, blowers, rolling mills, and others. With increasing demand, industries are now providing solutions with power ratings reaching up to MW for purposes like electric propulsion, rolling mills, traction systems, and hybrid electric vehicles, highlighting the escalating significance of VFIMDs in contemporary industrial and transportation sectors [32]. The IM drive must function effectively on both the machine and the drive section. A diode bridge rectifier is used by Voltage Source Inverters (VSI) to supply the dc link voltage for three-phase induction motor drives. Huge currents are drawn from the supply line as a result of the power electronic switches' nonlinear behavior that raises the power requirement and machine rating [4]. This raises the total expense of operating the system. Induction motor drives are examined in various ways to monitor their efficacy in various scenarios. Subsequently enhancing power quality reduces the price of electricity, power quality assessment of induction motor drive systems is a highly significant research field [5]. In many recent applications, such as the drive of variable frequency induction motors, AC-DC converters are now widely used. Although certain applications, such as AC drives, occasionally call for DC power during transitional stages, they typically require DC power to operate. The aforementioned AC-DC converters, commonly referred to as rectifiers, are utilized in medium power applications and are powered by a three phase AC supply with a 33 kV or 11 kV voltage along with a step-down transformer [6]. The AC-DC converters, however, possess low input power factor, raised DC voltage ripple, increased THD, and poor input power quality. Various approaches have been developed to overcome these restrictions [7]. The filters are usually employed to eliminate harmonics that can be either passive or active. These filters raise losses and complexity of circuits while being inefficient. Due to their remarkable efficiency, higher reliability, and simple structure, multi pulse rectifiers have become increasingly popular for enhancing the input power quality and decreasing harmonics. In addition, multi-pulse rectifiers do not need LC filters, removing the potential for LC resonances. Additionally, the common mode voltage generated by rectifiers is removed through the use of a phase-shifting transformer [8]. The research investigation examined difficulties with reliability of power in a distributed multi-drive system linked to an uncontrolled rectifier. The primary use of an uncontrolled rectifier is as a front-end converter to supply direct current to inverters driving induction motors. So, a shunt active power filter is used to lessen the harmonic content that is present in the source current. To enhance the grid system's power quality, an active filter is parallel injected among the source and the multi-drive mechanism. The primary limitation associated with an active filter is its pricey, intricate control mechanism. Both back-to-back converter actions are subject to an ongoing research into predictive model control, which serves as an active power filter to compensate for harmonics and supply sinusoidal grid current. Power quality and harmonics removal are considered in this optimization. The main disadvantage of model predictive control is that it was designed to handle output disturbances and may struggle with input issues [9]. The study offers a technique for creating hybrid power filters that minimize distortion caused by harmonics and thus enhance power factor and lessen the harmonics issues. The primary

drawback this kind of filter are high computational expenditure [10]. By employing the harmonic injection method, an induction motor's harmonics that are present in its current, voltage, space, or time parameters have been attempted to be reduced. In order to achieve harmonic distortion levels up to or less than 3% THD, harmonics injection method first cancels up to the 50th harmonic [11]. The installation of an active rectifier to a frequency converter is one of the promising methods for enhancing power quality while an AC VFIMD is operating. In mining equipment with VFIMD, frequency converters with active rectifiers are an efficient tool for improving power quality and reducing consumption of energy [12]. The investigation examined the reference value of the total harmonic factor, requiring an AR filter and considering feed line length when configuring AR current controllers, ensuring compliance with recommendations for transient and steady state processes. [13]. Numerous studies on PWM methodologies' impact on VSI in Induction motor drives, however, only address specific parameters like torque ripple and line current THD. [14]. The research explores a selective harmonic reduction approach in a three-level NPC inverter-fed induction motor drive to minimize torque harmonics and mechanical and electrical resonance. [15]. certain torque and current harmonics which could cause disturbances as well as mechanical and electrical resonance are avoidable because of the different modes of operation. In a three-level NPC inverter-fed induction motor drive, the research investigated the selective harmonic reduction approach to minimize the determined torque harmonics [16]. For [17] enabling the drive to absorb or inject reactive power and deliver current supply that has minimal harmonics, a field-oriented control scheme has been established. The industries, which utilize numerous induction motor drives concurrently, require greater, less expensive approaches to enhancing the electrical performance. Existing Research proposes a control method for a dynamic voltage restorer (DVR) that protects an adjustable speed drive (ASD) from harmonics, imbalance, drop, and swell [18]. The sources of current harmonics are the various non-linear loads. Waveform changes in current and voltage are commonly caused by the harmonics in the current. As a consequence of the various loading scenarios, distorted current is injected from the point of common coupling to supply [19]. The addition of filtering circuit with enhances power quality with lower THD, enhancing the converter's efficiency and the induction motor's dynamic performance [20]. The IMD system with an enriched filter is used to reduce harmonic produced by power semiconductor switches. It also offers reactive power compensation to keep the three-phase power supply's quality significant that is necessary for preserving the functionality of other devices linked to the converter circuit.

The proposed research primary contribution consists comprising the following:

- For the minimization of harmonics, the power supply from the grid system, in sugar industry, an enriched I-Q active filter has been implemented.
- To instantaneous active and reactive power theory employed in the suggested design to extract the features
- The drive performance analysed in terms of power quality such as THD analysis and voltage and current maintenance.

The accompanying would be the format of this paper: Section 1 provides an introduction of drives. Section 2 elaborate the drive system of sugar industry. Section 3 depicts generation of harmonics in varying load situation. Section 4 describes the proposed methodology and system modelling. Section 5 explain the proposed methodology outcomes. The conclusions of this work are described in Section 6, and the references are given in subsequent section.

2 DRIVE IN SUGAR INDUSTRY

In a sugar mill, there are many different driven machineries, but generally, there are merely three main kinds of drives:

- Constant speed
- Variable speed over a limited range
- Variable speed over an extensive range.

The adjustable speed drives (ASD) known as variable frequency drives (VFDs) have been used in industrial settings to regulate the machinery's two major operating parameters, namely torque and speed. Fig. 1 depicts the block diagram for a three-phase motor drive system. The ASD method is used in industrial applications because distinct equipment in various sectors operates at distinct speeds [21]. VFD contributes a significant function in reducing the amount of harmonics.

The selection of AC and DC drives depends on the needs of the industrial application. While DC drives are employed to have high starting torque and constant speed. The objective for utilising AC drives is to have good control processes and minimize energy consumption, for example, they are used in boiler feed pumps and power generators. In the present research, the induction motor's speed and torque are controlled by an AC drives system. The impact of harmonics is greater in VFD. The nonlinear behaviour of the load may also be responsible for the VFD's poor power factor, which decreases with a reduction in motor speed and results in a significant quantity of induction harmonics being supplied again to the electric power supply [22].

2.1 Induction Motor Drives

The three-phase induction motor is ideal for constant speed drives in sugar factories, while the slip ring and squirrel cage motors are suitable for high torque and long run-up times, with easy-to-understand control gear. Approximately 400 electrical drives, varying in size about a few kW to a few MW, are used in a typical contemporary sugar factory for various operations. Solely three or four of these locations use SRIMs, which account for about 30% of their overall electrical usage. The aforementioned SRIMs are set up at pre-

crushing equipment that essentially uses large rotating knives to chop the cane, followed by heavy rotating hammers to shred it, to prepare the cane for effective squeezing the highest-performance motor in the industry, which uses nearly 15% of the installed power, is used for cutting operations.

Sugarcane industries commonly employ cogeneration plants, with induction machines for braking functions. Flywheels are fitted to mitigate load swings, particularly during breaking. Switched reluctance induction motors.

(SRIMs) with high inertia are recommended for heavy knife operations. However, SRIMs require external rheostats for starting and maintain slip resistance to minimize current fluctuations. Despite load stability benefits, significant energy is wasted in slip resistance. [23].

A variable frequency drive (VFD) is an electrical system that connects an electric motor to a supply system, supplying it with a variable frequency alternating supply voltage, allowing variable motor torque and speed. Modern VFDs consist of a rectifier, DC-link, and inverter sections. [24]. Three phase induction motor drives as shown in Fig. 1

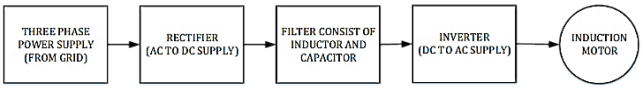


Figure 1 Three phase induction motor drive system

3 HARMONIC ANALYSIS IN INDUCTION MOTOR DRIVE

Power electronic converters are now frequently used to regulate the flow of electricity for digitization and energy conservation in both manufacturing and household usages. These kinds of converters frequently consume harmonic current as well as reactive power through the AC distribution, which negatively affects power quality [25, 26]. Rapid electricity rises in either a positive or negative direction, creating harmonics. As a consequence, the wave form of an electrical converter voltage supply's output is non-sinusoidal. In this sort of wave form, square waves and pulse waves create a swift and unexpected rise. Harmonic currents are caused by unpredictable masses that are rigid and produce a current wave that is completely distinct from the employed voltage wave.

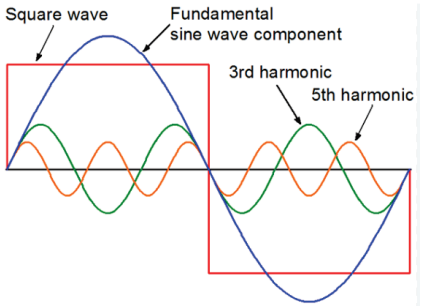


Figure 2 Electrical Waveforms Harmonic

Solid state power changers like diodes, thyristors, SCRs, and transistors transform DC power, making up most electronic gadgets. Harmonic frequencies in power systems are created by non-linear electronic and electrical devices. [27]. Fig. 4 illustrates the way the fundamental frequency produces non-sinusoidal distorting wave forms when combined with other order harmonics. The electric wave form harmonics depicts in Fig. 2. Harmonic current arises from rigid non-linear masses, causing a different current wave shape than applied voltage. Solid state power change devices, however, possess distinct voltage and current shapes.

Harmonic current arises from rigid non-linear masses, causing a different shape than the applied voltage wave. Modern technological masses have distinct voltage and current shapes, with electric potential appearing undulating but current waves moving slowly. Harmonics can be produced on the supply or load side depending on the load type. Current harmonics [28] are usually produced by voltage provides. The result could be the amount of distortion created as the facility line's current flows.

4 PROPOSED METHODOLOGY AND SYSTEM MODELLING

Voltage and current waveform degradation is being caused by the usage of power conversion devices and non-linear loads in consumers' and industries' applications. Increased distribution power losses, communication system interference, and malfunctions in delicate electronic equipment are all brought on by harmonics in electric lines. International electrical power quality standards impose limitations on supply voltage distortion and prohibit machinery from producing harmonic contents above specific thresholds. While they solve harmonic current problems, passive filters have disadvantages. The latest initiatives have focused on the creation of active filters to address these drawbacks.

4.1 Active Filters

Active power filters (APF) can be divided into two basic categories: shunt type and series type. It is possible to come across active filters acting in concert with passive filters and/or other active filters. The circuitry layout of a shunt active filter, that can correct both the harmonics of current and power factor in a three-phase generator with neutral wire, is illustrated in Fig. 3. Additionally, it enables load distribution, which cuts down on electricity in the neutral wire. Since the active filter doesn't need an internal power distribution, the power stage is essentially a voltage-source inverter (VSI) which regulates itself to behave like a current source. It has just one a capacitor on the DC side of the circuit.

The controller determines the reference currents I_{fa}, I_{fb}, I_{fc} needed through the inverter for generating the compensation currents I_{ca}, I_{cb}, I_{cc} based on the results of the measurement of the phase voltages V_{ca}, V_{cb}, V_{cc} as well as currents at load I_a, I_b, I_c There is no requirement for compensation for the current in neutral wire for balanced loads lacking third order current harmonics. The series active filter operates as a high impedance to the current harmonics generated by the power source side however does not compensate for harmonics in the load current. It ensures that

passive filters subsequently installed at the load input will prevent harmonic currents from draining into the remaining components of the power system. Utilizing a shunt active filter to make sure that both the voltage at the load and the supply currents have sinusoidal wave forms is another way to address the load current harmonics. The basic structure of shunt active filter shown in Fig. 3.

A non-linear load current that is divided into two distinct elements can identify a current harmonic. They are the fundamental nonlinear load current and the harmonic nonlinear load current utilized for harmonic injection into the utility grid system. Eqs. (1) - (5) depicts the grid voltage operation, non-linear current at the load, fundamental non-linear load magnitude, and load current.

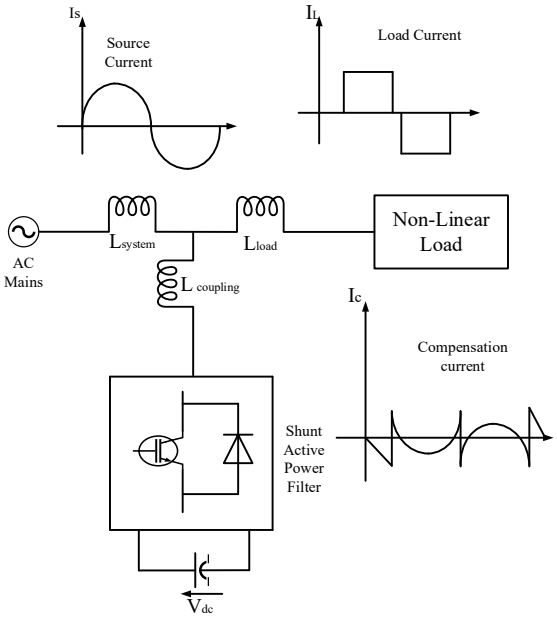


Figure 3 Basic scheme of shunt active power filter

The current harmonics can be determined using Eq. (6).

$$V_{\mathrm{g}}(t) = V_{\mathrm{g}} \cos(\omega t) \tag{1}$$

$$I_{\mathrm{nl}}(t) = I_1 \cos(\omega t + \alpha_1) + \sum_{k=2}^{\infty} I_{\mathrm{kh}} \cos(\omega t + \alpha_{\mathrm{k}}) \tag{2}$$

$$I_{\mathrm{fundm,\,nl}} = \frac{1}{\pi} \int_{-\pi}^{\pi} (I_1 \cos(\omega t + \alpha_1) \cos(\omega t) \mathrm{d}\omega t \tag{3}$$

$$I_{\mathrm{fundm,\,nl}} = I_1 \cos \alpha_1 \tag{4}$$

$$I_{\mathrm{fundm,\,nl}}(t) = I_1 \cos \alpha_1 \cos(\omega t) \tag{5}$$

$$I_{\mathrm{kh}}(t) = I_{\mathrm{nl}}(t) - I_1 \cos \alpha_1 \cos(\omega t) \tag{6}$$

Where $V_{\mathrm{g}}(t)$ denotes the voltage function at grid, $I_{\mathrm{nl}}(t)$ represents the nonlinear current at load, $I_{\mathrm{fundm,\,nl}}(t)$ indicates the fundamental nonlinear current at load. $I_{\mathrm{kh}}(t)$ specifies current harmonic function and nonlinear current function magnitude.

4.1.1 Mathematical Model of Active Power Filter

To investigate the DC-link potential response and current monitoring ability, condensed analytical designs of the APF was developed. When the suggested control approach is employed, harmonics in the current are quickly compensated for and variations in the voltage at the DC-link during fluctuating and stable conditions are efficiently reduced.

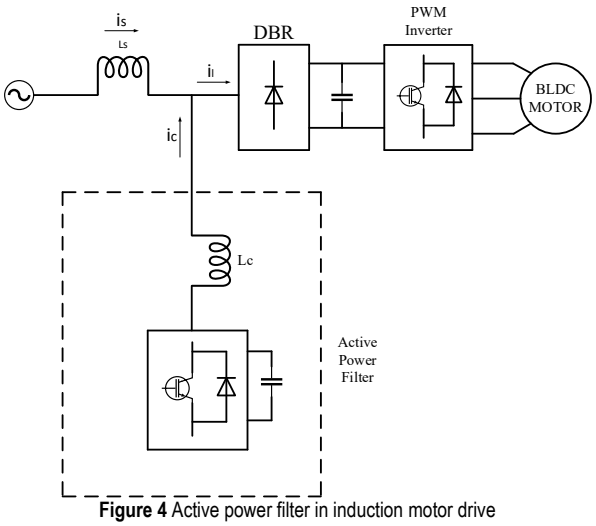


Figure 4 Active power filter in induction motor drive

The schematic diagram for the APF is shown in Fig. 4. As shown in Fig. 5, one can obtain the equivalent circuit of the APF. The switch status represented as d_{s} in Eq. (7).

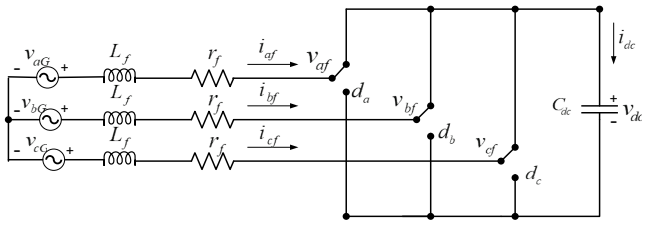


Figure 5 Active power filter equivalent circuit

$$d_{\mathrm{s}} = \begin{cases} 1 \text{ if } T_{\mathrm{s}}^{+}\text{: on, } T_{\mathrm{s}}^{-}\text{: off} \\ 0 \text{ if } T_{\mathrm{s}}^{+}\text{: off, } T_{\mathrm{s}}^{-}\text{: on} \end{cases} \tag{7}$$

Where T_{s}^{+} and T_{s}^{-} denotes the power switching transistor and s represents phase terminal a, b, c. The switching voltage for phase representation is denoted as V_{af1}, V_{bf1}, V_{cf1}. The voltage V_{dc} indicates the voltage at DC link and it can be expressed in below equations.

$$\begin{bmatrix} V_{\mathrm{af}} \\ V_{\mathrm{bf}} \\ V_{\mathrm{cf}} \end{bmatrix} = \frac{V_{\mathrm{dc}}}{3} \begin{bmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{bmatrix} \begin{bmatrix} d_{\mathrm{a}} \\ d_{\mathrm{b}} \\ d_{\mathrm{c}} \end{bmatrix} \tag{8}$$

The APF differential equations are expressed from the Fig. 5, which is expressed below.

$$L_{\mathrm{fl}}\frac{\mathrm{d}}{\mathrm{d}t}i_{\mathrm{af}} = V_{\mathrm{a1M}} - r_{\mathrm{fl}}i_{\mathrm{af}} - V_{\mathrm{af}}$$
$$L_{\mathrm{fl}}\frac{\mathrm{d}}{\mathrm{d}t}i_{\mathrm{bf}} = V_{\mathrm{b1M}} - r_{\mathrm{fl}}i_{\mathrm{bf}} - V_{\mathrm{bf}} \quad (9)$$
$$L_{\mathrm{fl}}\frac{\mathrm{d}}{\mathrm{d}t}i_{\mathrm{cf}} = V_{\mathrm{c1M}} - r_{\mathrm{fl}}i_{\mathrm{cf}} - V_{\mathrm{cf}}$$

$$C_{\mathrm{dc}}\frac{\mathrm{d}}{\mathrm{d}t}v_{\mathrm{dc}} = d_{\mathrm{a}}i_{\mathrm{af}} - d_{\mathrm{b}}i_{\mathrm{bf}} - d_{\mathrm{c}}i_{\mathrm{cf}} \quad (10)$$

Induction motor phase voltage denotes as V_{a1M}, V_{b1M}, V_{c1M}. Three phase power converter input voltage represent as V_{af}, V_{bf}, V_{cf}. The filter resistance and inductance indicate as r_{f1}, L_{f1}.The dc link capacitance is denoted as C_{dc}.

4.2 Enriched Active Filters

Extensive use of power converters and dynamic loads causes waveform degradation, resulting in harmonics and voltage drops. This leads to efficiency losses, increased power distribution losses, and equipment malfunctions. I-Q Active Filter is proposed to mitigate these issues. It transforms three-phase potential and current into orthogonal coordinates for instantaneous power assessment, applicable to various operations and waveforms.

$$\begin{bmatrix} V_0 \\ V_\alpha \\ V_\beta \end{bmatrix} = \sqrt{\frac{3}{2}} \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ 1 & \frac{-1}{2} & \frac{-1}{2} \\ 0 & \frac{\sqrt{3}}{2} & \frac{-\sqrt{3}}{2} \end{bmatrix} \begin{bmatrix} v_{\mathrm{a}} \\ v_{\mathrm{b}} \\ v_{\mathrm{c}} \end{bmatrix} \quad (11)$$

$$\begin{bmatrix} I_0 \\ I_\alpha \\ I_\beta \end{bmatrix} = \sqrt{\frac{2}{3}} \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ 1 & \frac{-1}{2} & \frac{-1}{2} \\ 0 & \frac{\sqrt{3}}{2} & \frac{-\sqrt{3}}{2} \end{bmatrix} \begin{bmatrix} I_{\mathrm{a}} \\ I_{\mathrm{b}} \\ I_{\mathrm{c}} \end{bmatrix} \quad (12)$$

Instantaneous zero sequence electrical power, real power derived, Instantaneous reactive power derived can be expressed as

$$p_0 = v_0 i_0 \quad (13)$$
$$p_i = v_\alpha i_\alpha + v_\beta i_\beta \quad (14)$$
$$q_i = v_\alpha i_\alpha - v_\beta i_\beta \quad (15)$$

The electrical power representation p_{i} and q_{i} associated with and β electric potential and current and it can be derived as

$$\begin{bmatrix} p_{\mathrm{i}} \\ q_{\mathrm{i}} \end{bmatrix} = \begin{bmatrix} v_\alpha & v_\beta \\ -v_\beta & v_\alpha \end{bmatrix} \begin{bmatrix} i_\alpha \\ i_\beta \end{bmatrix} \quad (16)$$

The reference compensating current equation for orthogonal coordinates represented as

$$\begin{bmatrix} i_\alpha \\ i_\beta \end{bmatrix} = \frac{1}{v_\alpha^2 + v_\beta^2} \begin{bmatrix} v_\alpha & v_\beta \\ -v_\beta & v_\alpha \end{bmatrix} \begin{bmatrix} \tilde{p}_{\mathrm{i}} \\ \tilde{q}_{\mathrm{i}} \end{bmatrix} \quad (17)$$

To determine the compensation reference current in α and β the power compensation is expressed as $\tilde{p}_{\mathrm{i}} - \overline{p}_0$ and q_{i}.

$$\begin{bmatrix} i_{\mathrm{c}\alpha}^* \\ i_{\mathrm{c}\beta}^* \end{bmatrix} = \frac{1}{v_\alpha^2 + v_\beta^2} \begin{bmatrix} v_\alpha & v_\beta \\ v_\beta & -v_\alpha \end{bmatrix} \begin{bmatrix} \tilde{p}_{\mathrm{i}} - \overline{p}_0 \\ \tilde{q}_{\mathrm{i}} \end{bmatrix} \quad (18)$$

Where:

$\overline{p}_0$ Denotes the mean value of instantaneous zero-sequence power transmitted by means of zero-sequence current and voltage components from the power source to the load.

$\tilde{p}_0$ Represents instantaneous zero-sequence power alternate value. The three-phase systems with neutral wire that have zero-sequence power are those.

$\tilde{p}_{\mathrm{i}}$ Denotes alternate instantaneous real power value, which is transmitted via the a-b-c coordinates among the electrical power source as well as the load.

$\tilde{q}_{\mathrm{i}}$ Represents instantaneous imaginary power. This includes the electrical energy which is transferred across the load phases.

In a system consisting of three phases with balanced sinusoidal voltages, the supply currents must also be sinusoidally balanced, in phase with the electrical potential and compensated for any undesirable power elements. Compensation of current based on coordinates shown in Fig. 6.

The zero sequence current should be compensated. The reference current compensation for zero coordinates is denotes as 0 and it can have expressed as

$$i_{\mathrm{c}0} = i_0 \quad (19)$$

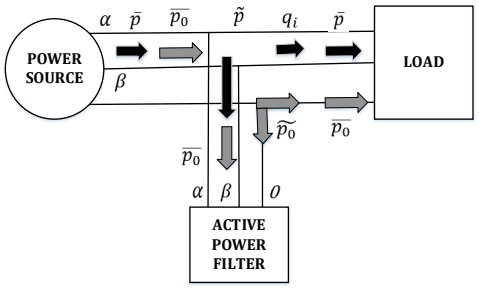


Figure 6 Power component current compensation-based coordinates

The reference compensated current in orthogonal coordinates α and β, the Eq. (20) in inverted and the compensated power are utilized

$$\begin{bmatrix} i_{c\alpha}^{*} \\ i_{c\beta}^{*} \end{bmatrix} = \frac{1}{v_{\alpha}^{2}+v_{\beta}^{2}} \begin{bmatrix} v_{\alpha} & -v_{\beta} \\ v_{\beta} & v_{\alpha} \end{bmatrix} \begin{bmatrix} p_{y} \\ q_{y} \end{bmatrix} \tag{20}$$

$$p_{y} = \tilde{p}_{\mathrm{i}} - \Delta\overline{p}_{\mathrm{i}} \tag{21}$$

$$\Delta\overline{p}_{\mathrm{i}} = \overline{p}_{0} \tag{22}$$

$$q_{y} = q_{\mathrm{i}} \tag{23}$$

$$q_{\mathrm{i}} = \overline{q} + \tilde{q} \tag{24}$$

The inverse of transformation is expressed to determine the reference current for compensation

$$\begin{bmatrix} i_{\mathrm{ca}}^{*} \\ i_{\mathrm{cb}}^{*} \\ i_{\mathrm{cc}}^{*} \end{bmatrix} = \sqrt{\frac{2}{3}} \begin{bmatrix} \frac{1}{\sqrt{2}} & 1 & 0 \\ \frac{1}{\sqrt{2}} & \frac{-1}{2} & \frac{\sqrt{3}}{2} \\ \frac{1}{\sqrt{2}} & \frac{-1}{2} & -\frac{\sqrt{3}}{2} \end{bmatrix} \begin{bmatrix} i_{\mathrm{c0}}^{*} \\ i_{\mathrm{c}\alpha}^{*} \\ i_{\mathrm{c}\beta}^{*} \end{bmatrix} \tag{25}$$

$$i_{\mathrm{cn}} = -(i_{\mathrm{ca}}^{*} + i_{\mathrm{cb}}^{*} + i_{\mathrm{cc}}^{*}) \tag{26}$$

4.3 Proposed Methodology for Harmonic Mitigation

The selection of AC and DC drives depends upon the needs of the application in the industry. While dc drives are utilized to have higher starting torque and constant speed, they are usually employed in motors. The reason for utilising AC drives is to achieve excellent control over processes and minimize consumption of energy. In the present investigation, the induction motor speed and torque are controlled by an AC drives mechanism [29].

Induction motor drive systems include a voltage source inverter, diode bridge rectifier, and speed control circuit. The intermittent power switching of electronic devices creates harmonics, leading to power quality issues. The block diagram of proposed investigation is illustrated in Fig. 7.

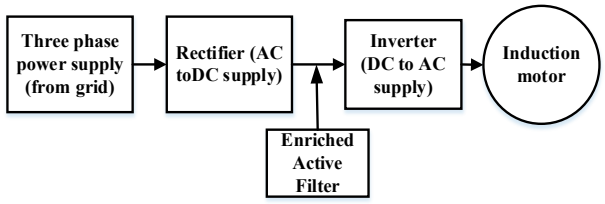


Figure 7 Block diagram of proposed system

A typical shunt active filter comprises two main components: a hysteresis current control section for providing gate pulses to the inverter, and a shunt active power filter controller unit. The hysteresis control processes power signals and synthesizes compensation current from the electrical grid, while the active filter controller continuously provides compensating current references to the hysteresis converter for real-time determination. This setup, illustrated in Fig. 3, consists of a voltage-fed converter, an active filter controller, and a hysteresis current controller, implementing an almost instantaneous control technique. Closed-loop operation is utilised by the shunt active filter controller used to determine the instantaneous values of the current compensation reference i_{c}^{*} for the hysteresis converter while continually detecting the current of load I_{l}.

Shunt active filter control is accomplished in three distinct phases.

- At initial stage, the respective current signals are sensed which is utilized to rectify the harmonics.
- In next stage, the instantaneous imaginary controller technique is implemented to derive compensating demands in relation to current.
- At final stage of controlling function, the hysteresis current control method is used to generate the gating signals for the three-phase inverter.
- The imaginary controller examines reference current via shunt active filter load current, employing frequency and time-domain adjustments to generate compensation commands. The imaginary power method transforms electrical potential and current into α, β, 0 coordinates, determining instantaneous power. The real and imaginary portion of power is represented as p^{*} as well as q^{*}. The instantaneous reference current $i_{\mathrm{ca}}^{*}, i_{\mathrm{cb}}^{*}, i_{\mathrm{cc}}^{*}$ determined by the inverse transformation of α, β, 0 to a, b, c. subsequently obtaining the reference values and actual quantities based on the evaluations, hysteresis control is implemented to create the transferring instructions for the VSI switching circuits. Hysteresis current control systems employ two-level comparators within a feedback loop. Switching commands activate when errors surpass designated tolerance bands.
- Total Harmonic Distortion (THD) is a metric utilized to evaluate the fidelity of electrical systems by measuring the discrepancy between an actual signal and an ideal sinusoidal reference. It is computed by normalizing the root mean square (rms) value of the signal's harmonic content (V) to the total rms value of the signal (V_{total}).

$$THD = \frac{\sqrt{\sum_{k=2}^{k_{\mathrm{m}}} (V_{\mathrm{K},k})^{2}}}{V_{\mathrm{total}}} \tag{27}$$

- The *THD* rate, known as the rms or effective *THD*, is commonly employed in power system applications.

5 SIMULATION RESULT

The proposed work is modelled, along with the impact of current and voltage at the load and source side is analysed to demonstrate the current and voltage behaviour when drives mechanisms are implemented. The induction motor drive system is the three-phase non-linear system that is under assessment for investigation of proposed research for harmonic analyzation and it is modelled in MATLAB/Simulink, which is shown in Fig. 8.

The system parameter depicts in below Tab. 1.

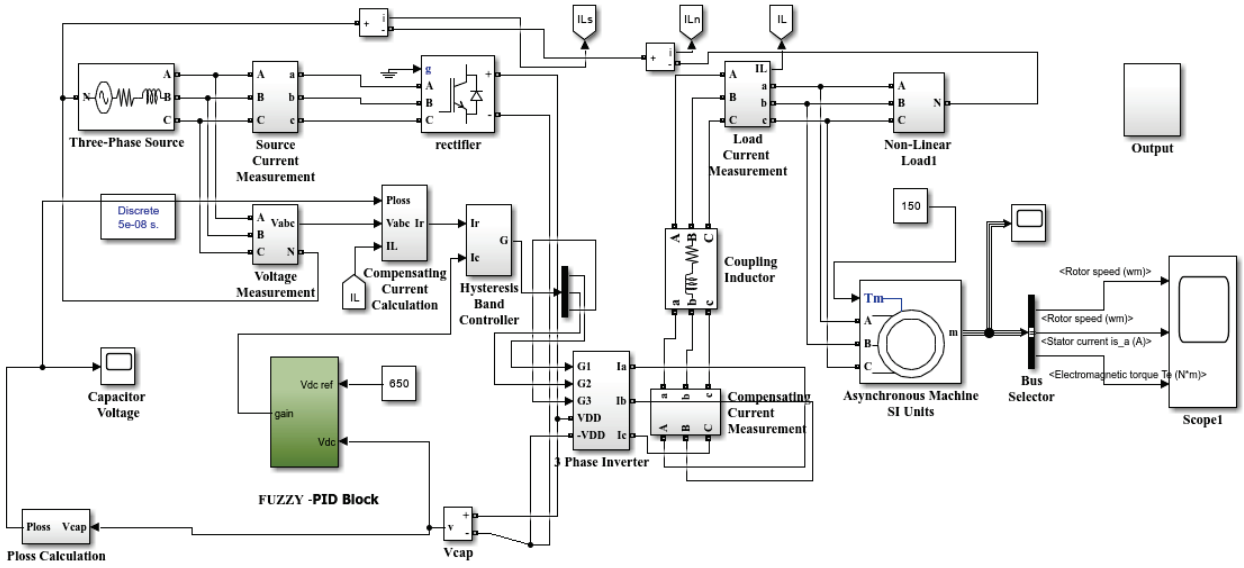


Figure 8 Proposed work MATLAB/Simulink model

Table 1 Simulation specifications

System Parameter	Values
Source side	415 V, f = 50 Hz, R_s = 0.89 Ω, L_s = 0.165 mH
ad Side	R_L = 0.73Ω, L_L = 0.3 mH
Filter Side	R_L = 0.8 Ω, L_L = 0.6 mH, C_{DC} = 1000 µF
Diode Bridge Rectifier	R = 15 Ω, L = 80 mH
DC Side Capacitor	1000 µF
PI controller gains	K_p = 7.5, K_i = 9
VSI parameters	4700 µF, V_{dbus} = 560 V
Induction Motor	10 HP, 400 V, 50 Hz, 1440 rpm

Simulation runs for 1 second pre-compensation, then enriched I-Q active filter provides compensation current for immediate waveform analysis, with compensating currents and dc link capacitor voltage at zero before compensation. The source voltage wave form shown in Fig. 9.

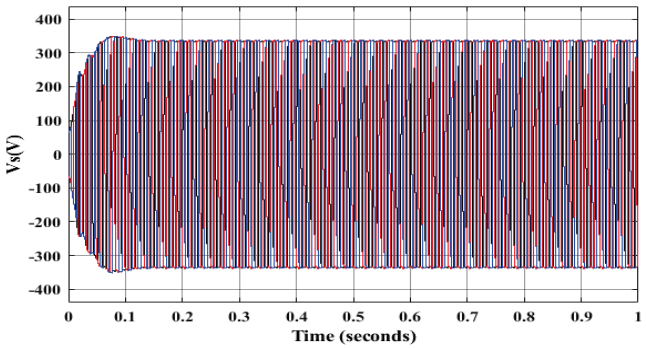


Figure 9 Source voltage output wave form

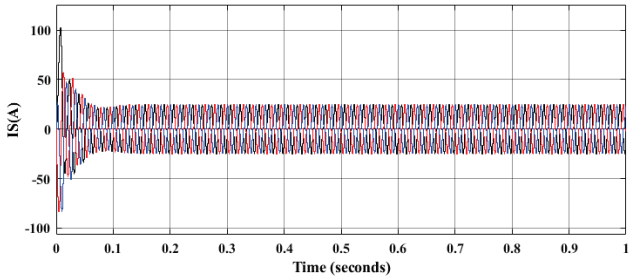


Figure 10 Source current

Since it can be observed, enriched filter compensation leads to the current from the source to evolve into sinusoidal. The source current output waveform depicts in Fig. 10. The generation of compensation current from the filter controller shown in Fig. 11 and the current delivered to the load current is indicates in Fig. 12.

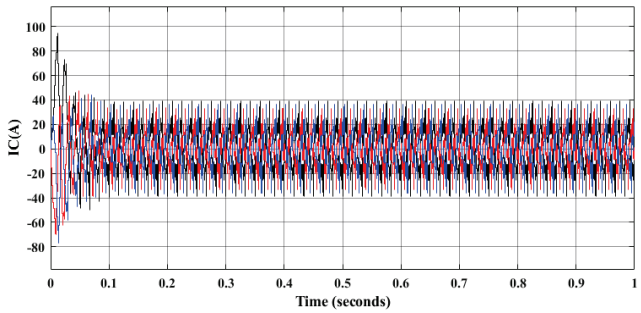


Figure 11 Compensation current

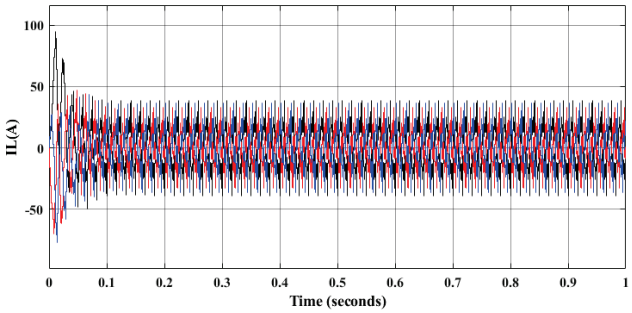


Figure 12 Load current

When an induction motor load is attached, the electrical current is distorted because of the harmonics in the load. Hysteresis control approach is subsequently utilized to produce switching signals after reference currents for the active filter have been generated through the instantaneous imaginary power strategy. The capacitor voltage shown in

Fig. 13. The induction motor stator current representation waveform is denoted in Fig. 14.

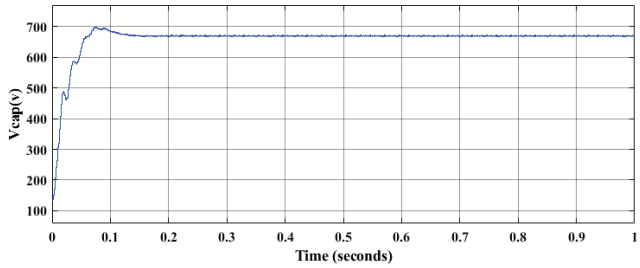


Figure 13 Capacitor voltage waveform

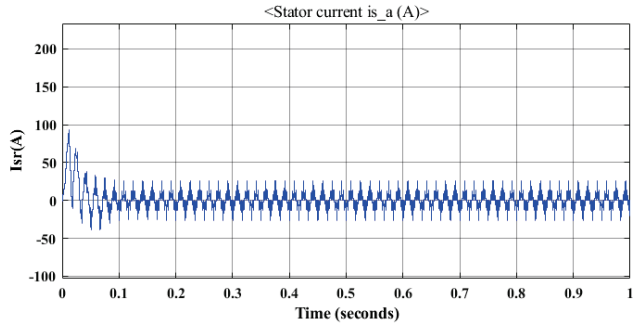


Figure 14 Induction motor stator current waveform

A three-phase rotor speed, and electromagnetic torque are illustrated as shown in Fig. 15 and Fig. 16. The rotor current oscillates up to 0.35 seconds and it attains a stable state and maintains steady after 0.35 seconds.

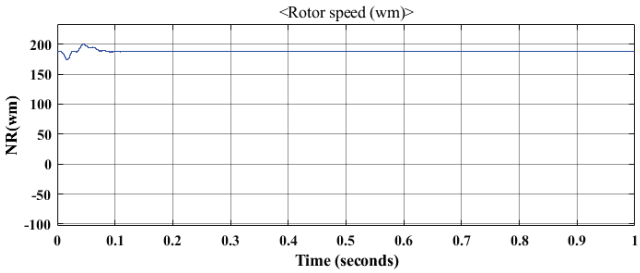


Figure 15 Rotor speed representation wave form

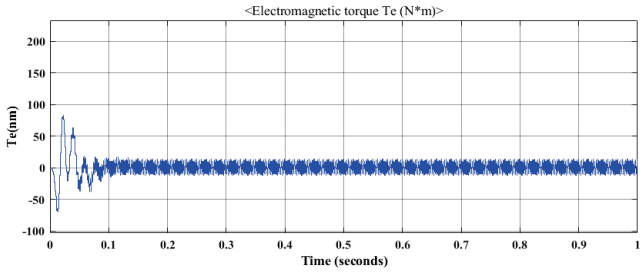


Figure 16 Electromagnetic torque wave form

The dynamic torque applied to the load, transitioning from full load to zero at specific time intervals (T_{fl} = 49.7 Nm, $T_{fl/2}$ = 24.86 Nm, $T_{fl/4}$ = 12.43 Nm, T_0 = 0 Nm) at t = 0.2 s, 0.5 s, and 0.7 s, as depicted in Fig. 17, poses a challenge for maintaining stable rotor speed. However, the proposed model effectively compensates for these fluctuating load torque conditions, ensuring a consistent speed of 150 rad/s which is shown in Fig. 18. Our proposed model dynamically responds to load torque fluctuations, stabilizing rotor speed and enhancing system reliability, crucial for unpredictable real-world conditions, ensuring consistent operation and minimizing instability risks.

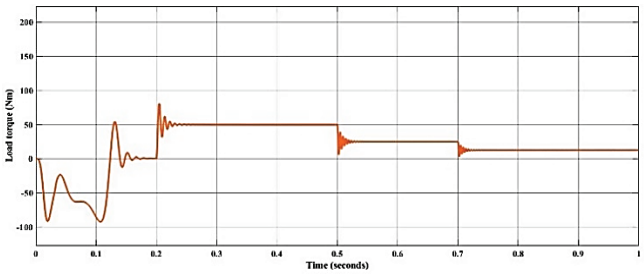


Figure 17 Varying load torque

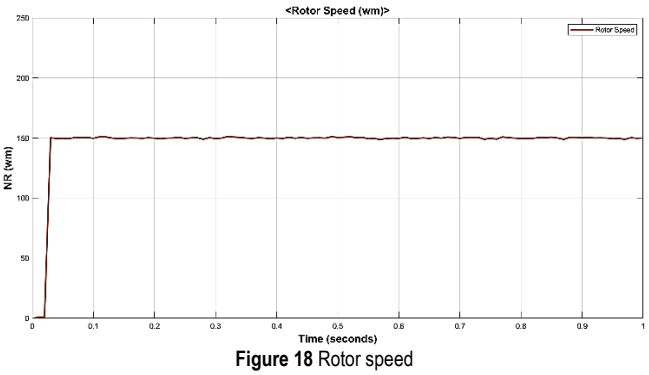


Figure 18 Rotor speed

When a single drive technique's THD is investigated, it should be observed that the single drive system's THD percentage is 25.68% without the presence of filter and it is shown in Fig. 19. At these circumstances, the THD% does not fall within the acceptable range as specified by the IEEE requirement. In order to limit the THD an enriched shunt active power filter is attached to the system in between the electrical grid and drive systems.

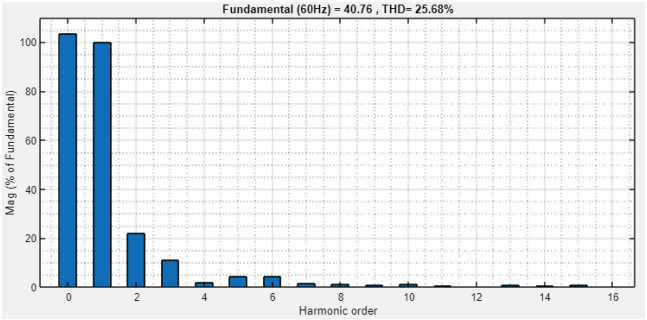


Figure 19 Percentage of THD without considering filter

The filter gets used for harmonics in the current based on variations on the load side. The most effective approach for driving motors with the fewest harmonics present, which subsequently enhances effectiveness and generation, is to add filters that minimize THD by nearly 1.75% for single drive systems, respectively. When the voltage at as well as the current from the source are in phase during compensation, the harmonics are removed from the source current, that

improves the power quality of entire grid connected drive system. The percentage of THD shown in Fig. 20.

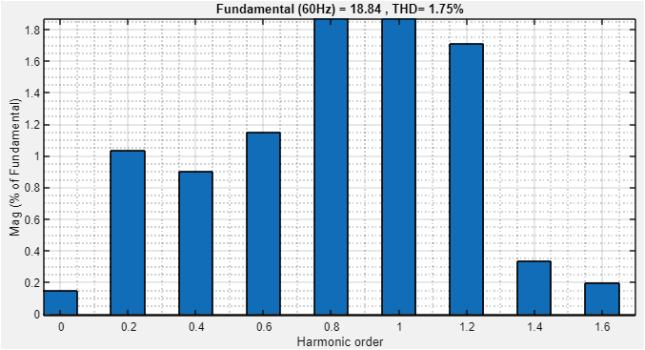


Figure 20 Percentage of THD with filter

The percentage of THD analysed for induction motor drive system in both without filter and with filter scenario and the respective percentage is shown in Tab. 2.

Table 2 Proposed filter THD percentage

THD analysis	Percentage of THD for load current without filter (%)	Percentage of THD for load current with filter (%)
Steady state condition at load current	25.68	1.75

The performance of proposed filter is analysed with the existing shunt active filter [30, 31] for single drive induction motor system. The percentage of THD is illustrated in below Tab. 3 and the comparative analysis shown in Fig. 21.

Table 3 THD analysis comparison

THD analysis	SAF	PI-SAF	PID-SAF	FLC-SAF	Proposed enriched SAF
Percentage of THD for single drive system without filter (%)	46.47	20.29	16.51	7.52	25.68
Percentage of THD for single drive system with filter (%)	5.45	13.60	3.86	2.66	1.75

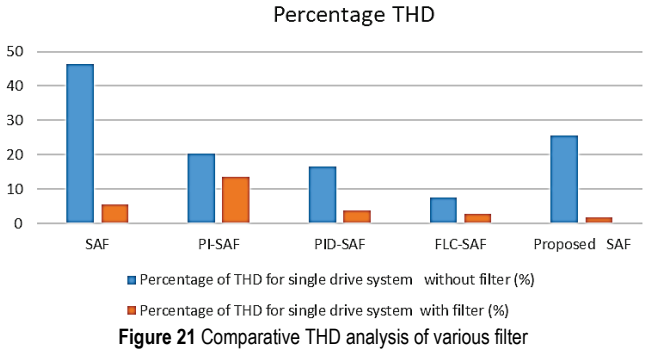


Figure 21 Comparative THD analysis of various filter

When the absence of shunt active filter with induction motor drive system, the occurrence of harmonics is about 46.47%. While implementation of filter with the drive circuit it will reduced as 5.45%. The absence of PI based SAF with motor drive system produce 20.29% THD. While implementing filter, it reduces the THD value to 13.60%. The percentage of THD while absence of PID-SAF is about 16.51%. Then the presence of PID-SAF eliminate the harmonics and decrease the THD to 3.86%. Without FLC-SAF the drive circuit produces the THD value of 7.52%. FLC-SAF with drive circuit rectifies the harmonics and its THD value is about 2.66%. The proposed filter initially not connected with the drive circuit and it induces the harmonics. The percentage of THD without filter is about 25.68%. Then the filter is connected with the drive circuit and its percentage THD is decreased to 1.75%. The comparison analysis of THD of different filter conclude that the proposed filter performs outstanding in terms of elimination of harmonics.

6 CONCLUSION

Maintaining the quality of power in the electrical distribution network is vital, particularly when irregular loads are linked into three phase AC lines. When nonlinear loads are connected through the three-phase distribution system from the utility grid, it is feasible that the critical loads are going to be impacted. Poor power quality conditions on the load side of the sugar processing plant create vulnerable condition in performance of induction motor drive when it serving as load. The proposed investigation provides an efficient way to enhance the power quality of the three-phase supply which is delivered into induction motor drive using the improved shunt active power filter. The phenomenon of hysteresis control technique as well as an enriched instantaneous imaginary controller based shunt active power filter (I-Q active filter) is implemented for obtaining the compensation current. The present research simulates a VSI fed induction motor drive and analyses output voltage and current waveforms. The variable frequency drive system normally employs an induction motor. In this instance, the harmonics caused by the variable frequency system are eliminated through a shunt active power filter with an improvised controller. The proposed work employed a single drive system and generation of harmonics are eliminated by proposed filter. The percentage of THD in single drive system without an implementation of filter obtained as 25.68%. While implementing the proposed filter, the percentage THD is reduced significantly and it is reduced as 1.75%. The proposed improvised filter performance is compared with existing shunt active filter. From the investigation it is concluded that the proposed enhanced controller based filter provide excellent outcomes in terms of harmonic reduction in induction motor used in sugar industry that adhere to harmonic control norms set by IEEE-519. The presented approach has been examined by the simulations result obtained from MATLAB/Simulink and the power quality criteria comply with the necessary standards.

7 REFERENCES

[1] Carpintero-Rentería, M., Santos-Martín, D. & Guerrero, J. M. (2019). Microgrids literature review through a layers structure. *Energies, 12*(22), 4381. https://doi.org/10.3390/en12224381

[2] Volosencu, C. (2021). Reducing energy consumption and increasing the performances of AC motor drives using fuzzy pi speed controllers. *Energies, 14*(8), p. 2083. https://doi.org/10.3390/en14082083

[3] Huh, N., Park, H. S., Lee, M. H. & Kim, J. M. (2019). Hybrid PWM control for regulating the high-speed operation of BLDC motors and expanding the current sensing range of DC-link single-shunt. *Energies, 12*(22), 4347. https://doi.org/10.3390/en12224347

[4] Nahin, N. I., Nafis, M., Biswas, S. P., Hosain, M. K., Das, P. & Haq, S. (2022). Investigating the input power quality of multi-pulse AC-DC power converter fed induction motor drives. *Heliyon, 8*(12), e11733. https://doi.org/10.1016/j.heliyon.2022.e11733

[5] Biswas, S. P., Anower, M. S., Sheikh, M. R. I., Islam, M. R. & Muttaqi, K. M. (2020). Investigation of the impact of different PWM techniques on rectifier-inverter fed induction motor drive. *Australasian Universities IEEE Power Engineering Conference (AUPEC)*, 1-6.

[6] Kant, P. & Singh, B. (2020). Multiwinding transformer fed CHB inverter with on-line switching angle calculation based SHE technique for vector-controlled induction motor drive. *IEEE Transactions on Industry Applications*, 56(3), 2807-2815. https://doi.org/10.1109/TIA.2020.2980131

[7] Biswas, S. P., Haq, S., Islam, M. R., Hosain, M. K. & Shaw, R. N. (2021). Advanced Level Shifted Carrier Based Bus Clamping PWM Technique for a 54 Pulse AC to DC Converter Fed MLI Based Induction Motor Drive. *The IEEE 6th International Conference on Computing, Communication and Automation (ICCCA2021)*, 587-592. https://doi.org/10.1109/ICCCA52192.2021.9666334

[8] Biswas, S. P., Anower, M. S., Haq, S., Islam, M. R., Rahman, M. A. & Muttaqi, K. M. (2021). A new level shifted carrier based PWM technique for a 5-level multilevel inverter used in induction motor drives. *IEEE Industry Applications Society Annual Meeting (IAS2021)*, 1-6. https://doi.org/10.1109/IAS48185.2021.9677179

[9] Ammar, A. (2019). Power Quality Improvement of PWM Rectifier-Inverter System Using Model Predictive Control for an AC Electric Drive Application. *International Conference on Electrical Engineering and Control Applications*, 427-439. Singapore: Springer Nature Singapore. https://doi.org/10.1007/978-981-15-6403-1_29

[10] Ahmed, G. E., Mohamed, F. N. & Nosier, M. A. (2019). Power quality improvement of sugar factories dc motor drive using hybrid filter. *Egyptian Sugar Journal, 12*, 112-129. https://doi.org/10.21608/esugj.2019.223913

[11] Ranga, P. & Mittal, M. (2020). Harmonic reduction in induction motor using harmonic injection method and passive filter: A comparison. *The First IEEE International Conference on Measurement, Instrumentation, Control and Automation (ICMICA2020)*, 1-5. https://doi.org/10.1109/ICMICA48462.2020.9242750

[12] Munoz-Guijosa, J. M., Kryltcov, S. B. & Solovev, S. V. (2019). Application of an active rectifier used to mitigate currents distortion in 6-10 kV distribution grids. *Journal of Mining Institute, 236*, 229-238. https://doi.org/10.31897/pmi.2019.2.229

[13] Shevyreva, N. Y. (2021). Effects of active rectifiers on power quality in supply systems in mineral mining industry. *Eurasian mining, 1*, 70-74. https://doi.org/10.17580/em.2021.01.14

[14] Wang, Y., Hu, C., Ding, R., Xu, L., Fu, C. & Yang, E. (2018). A nearest level PWM method for the MMC in DC distribution grids. *IEEE Transactions on Power Electronics, 33*(11), 9209-9218. https://doi.org/10.1109/TPEL.2018.2792148

[15] Allal, A. & Khechekhouche, A. (2022). Diagnosis of induction motor faults using the motor current normalized residual harmonic analysis method. *International Journal of Electrical Power & Energy Systems, 141*, 108219. https://doi.org/10.1016/j.ijepes.2022.108219

[16] Selvarasu, R., Kannan, C., Priyadharsini, S. & Shiferaw, D. A. (2022). Modeling and Control of Cascaded Multilevel Inverter for Harmonics Mitigation of Induction Motor Drive. *Proceedings of International Conference on Power Electronics and Renewable Energy Systems (ICPERES2021)*, 39-48. Springer Singapore. https://doi.org/10.1007/978-981-16-4943-1_5

[17] Mansour, A. S. (2017). Three-phase induction motor drive with reactive power injection to supply. *The Nineteenth IEEE International Middle East Power Systems Conference (MEPCON2017)*, 421-429. https://doi.org/10.1109/MEPCON.2017.8301215

[18] Khergade, A., Satputaley, R. J., Borghate, V. B. & Raghava, B. V. S. (2020, January). Harmonics reduction of adjustable speed drive using multi-objective dynamic voltage restorer. *The IEEE International Conference on Power Electronics, Smart Grid and Renewable Energy (PESGRE2020)*, 1-6. https://doi.org/10.1109/PESGRE45664.2020.9070589

[19] Sanatkar-Chayjani, M. & Monfared, M. (2017). High-order filter design for high-power voltage-source converters. *IEEE Transactions on Industrial Electronics*, 65(1), 49-58. https://doi.org/10.1109/TIE.2017.2714136

[20] Aciego, J. J., Duran, M. J., Gonzalez-Prieto, I. & Barrero, F. (2017). Analysis of the Power Quality in Six-phase Induction Motor Drives with Arbitrary Winding Spatial Shifting. *Renewable Energy and Power Quality Journal*, 479-484. https://doi.org/10.24084/repqj15.359

[21] Ullah, A., Sheikh, I. U. H., Arshad, S. & Saleem, F. (2019). Digital active power filter controller design for current harmonics in power system. *The 16th International Bhurban Conference on Applied Sciences and Technology (IBCAST2019)*, 384-388. https://doi.org/10.1109/IBCAST.2019.8667169

[22] Holmukhe, R. M., Gandhar, A., Chauhan, S. R. & Patil, M. N. (2022). Power quality analysis of a mechanical industry–*A case study. Journal of Information and Optimization Sciences, 43*(3), 615-627. https://doi.org/10.1080/02522667.2022.2083315

[23] Kumar, V. & Kumar, S. (2019). A 3-level Inverter based Induction Motor Drive for Cane Preparation in Sugar Industry. *The 2nd IEEE International Conference on Power Energy, Environment and Intelligent Control (PEEIC2019)*, 190-195. https://doi.org/10.1109/PEEIC47157.2019.8976857

[24] Scheuer, G., Schmage, T. & Krishnan, L. C. (2007). Medium voltage drives in the sugar industry. *International sugar journal*, 1301, 303.

[25] Fuchs, E. F. & Masoum, M. A. (2011). *Power quality in power systems and electrical machines*. Academic press.

[26] Zijiang, W. A. N. G., Qionglin, L. I., Yuzheng, T. A. N. G., Shuming, L. I. U. & Shuangyin, D. A. I. (2019). Comparison of harmonic limits and evaluation of the international standards. *In MATEC Web of Conferences*, 277, 03009. EDP Sciences. https://doi.org/10.1051/matecconf/201927703009

[27] Krishnasamy, B. & Ashok, K. (2022). Assessment of harmonic mitigation in v/f drive of induction motor using an ANN-based hybrid power filter for a wheat flour mill. *Processes, 10*(6), 1191. https://doi.org/10.3390/pr10061191

[28] Agrawal, S., Palwalia, D. K. & Kumar, M. (2022). Performance analysis of ANN based three-phase four-wire shunt active power filter for harmonic mitigation under distorted supply voltage conditions. *IETE Journal of Research, 68*(1), 566-574. https://doi.org/10.1080/03772063.2019.1617198

[29] Lee, K., Carnovale, D., Young, D., Ouellette, D. & Zhou, J. (2016). System harmonic interaction between DC and AC adjustable speed drives and cost effective mitigation. *IEEE Transactions on Industry Applications, 52*(5), 3939-3948. https://doi.org/10.1109/TIA.2016.2562006

[30] Thentral, T., Rathakrishnan, R., Anbalagan, V., Dhandapani, K., Sengamalai, U. & Ramasamy, P. (2022). Mitigation of current harmonics in multi-drive system. *International Journal of Power Electronics and Drive Systems, 13*(1), 113. https://doi.org/10.11591/ijpeds.v13.i1.pp113-121

[31] Sivakumar, A., Thiyagarajan, M. & Kanagarathinam, K. (2023). Mitigation of supply current harmonics in fuzzy-logic based 3-phase induction motor. International *Journal of Power Electronics and Drive Systems, 14*(1), 266. https://doi.org/10.11591/ijpeds.v14.i1.pp266-274

[32] Arumalla, R. T., Figarado, S. & Harischandrappa, N. (2020). Dodecagonal voltage space vector based PWM techniques for switching loss reduction in a dual inverter fed induction motor drive. *IEEE Journal of Emerging and Selected Topics in Industrial Electronics, 1*(2), 182-191. https://doi.org/10.1109/JESTIE.2020.2999583

[33] Arranz-Gimon, A., Zorita-Lamadrid, A., Morinigo-Sotelo, D. & Duque-Perez, O. (2021). A review of total harmonic distortion factors for the measurement of harmonic and interharmonic pollution in modern power systems. *Energies, 14*(20), 6467. https://doi.org/10.3390/en14206467

Authors' contacts:

Samsher Kadir Sheikh, PhD Scholar
Department of Electrical and Electronics Engineering, Mandsaur University,
Mandsaur, (M.P.), 458001, India
sheikhsam115@gmail.com

Dr. **Manish Jain**, Associate Professor
Department of Electrical and Electronics Engineering, Mandsaur University,
Mandsaur, (M.P.), 458001, India
manish.jain@meu.edu.in

Dr. **Pawan Chandrakant Tapre**, HoD in Department of Electrical Engineering
SND College of Engineering & Research Center,
Yeola Dist Nasik, (M.S.), 423401, India
pawan.tapre73@gmail.com

ISSN 1846-6168 (Print), ISSN 1848-5588 (Online)
https://doi.org/10.31803/tg-20230912130335

Original scientific paper
Received: 2023-09-12, Accepted: 2024-02-26

Artificial Neural Networks Application for the Croatian School Maintenance Cost Estimation

Ksenija Tijanić Štrok

Abstract: The quality of education is associated with the condition of the infrastructure in which the educational process occurs, necessitating the continuous maintenance of these facilities. Limited and often insufficient maintenance funds pose a challenge in this context. Current cost estimates are inaccurate, and data on school and maintenance costs are constrained. More accurate maintenance cost plans would contribute to a better understanding of budget distribution and more efficient financial management. This study aims to investigate the application of artificial neural networks (ANNs) in planning annual maintenance costs for schools in the Republic of Croatia (Primorje-Gorski Kotar County). Using a database and DTREG software, three different ANN models were developed: a multilayer perceptron (MLP), a generalized regression neural network (GRNN), and a radial basis function neural network (RBFNN). Comparisons of the results showed that the GRNN is optimal and achieves the highest accuracy in estimating school maintenance costs. These findings can benefit educational institutions and public bodies in budget planning and decision-making regarding maintenance.

Keywords: artificial neural network; cost estimation; maintenance cost; school buildings

1 INTRODUCTION

Schools are the centres of the educational (and upbringing) system, and the school building is an important indicator of its quality [1-4]. The quality of the school building is mainly ensured by adequate maintenance [5]. School maintenance includes all those activities that retain the desired functions and purposes of the school building during its lifetime [6].

School maintenance is usually divided into ongoing and investment [7]. Ongoing maintenance involves repairing damage caused by the daily use of the building. It includes inspecting the damage, repairing it, and implementing preventive and protective measures and interventions after unplanned events. Investment maintenance involves the implementation of (mostly planned) construction-craft works to improve the conditions of use of the building in the exploitation phase. Usually, it requires larger financial resources [8].

School maintenance is the responsibility of the public authority (state, county, city, municipality), which is in the role of its founder. Public authorities also finance maintenance, and the necessary funds are shown in the financial plans (cost plans). A review of [7, 9], found that these plans are not precise enough and that there is a high level of uncertainty in planning maintenance works and related costs. More accurate estimating of maintenance costs would make it possible to determine the target costs for planned maintenance works and provide information on the limits of available funds. It would give valid cost information to help maintainers make appropriate maintenance decisions [10]. Cost plans are mandatory for efficient cost management [7] and would allow building owners to get value for money spent on maintenance [10].

The public sector covers the costs of educational facilities; therefore, the planning and optimal distribution of these costs, as well as increasing the efficiency of maintenance management, is a significant and valuable goal of broader social interest [7]. Effective management of school maintenance implies planned, organized, and high-quality implementation of maintenance activities with optimal consumption of resources, primarily financial. Effective maintenance ensures the satisfaction of school staff and students by creating conditions that ensure their health and safety, facilitate teaching and learning, and improve school outcomes [11].

In Croatia, there is insufficient comprehensive research addressing the analysis of the expenses associated with maintaining primary and secondary schools, as stated in the available literature [7, 9, 12-14]. School maintenance is often neglected, characterized by a lack of coordination and high financial expenses. Estimates of maintenance costs show insufficient precision [7].

Given all the above, the need to develop a model for estimating the cost of school maintenance is very pronounced. The literature highlights the potential of machine learning for these purposes, using artificial neural networks (ANNs) [7, 15-19]. By using a database and applying computer models, there is the possibility of producing more accurate cost estimates [20].

The main goal of this paper is to investigate the possibility of developing a model for estimating the costs of maintaining school buildings using ANN in the Republic of Croatia. The model will be developed based on a database of school buildings with insight into historical data on actual maintenance costs. The model is intended to provide a more accurate, quick, simple, and structured assessment. The research is limited to Primorje-Gorski Kotar County.

The following research questions were defined;

- Is it possible to create a database on school buildings from Primorje-Gorski Kotar County with insight into historical data on maintenance costs?
- Based on the created database, is it possible to develop an ANN model for estimating the maintenance cost of school buildings with satisfactory accuracy ($R^2 > 0.64$, $MAPE < 30$ %)?
- Is there currently a pronounced inaccuracy in the estimation of maintenance costs, and will the application of the ANN model developed here reduce this inaccuracy?

The application of the developed ANN model should certainly help increase the efficiency of the maintenance process.

The work is organized as follows. The first section is introductory, in which the motives for conducting the research are stated, the main points from the literature review are highlighted, and the goal and research questions are outlined. The second section describes the applied methodology in detail with all the necessary descriptions and explanations. In the third section, the obtained results are presented and discussed. The last fourth section presents the conclusions.

1.1 Theoretical Background

1.1.1 School Building Maintenance Costs and Plans

Authors of [21] define building maintenance as an investment activity throughout the life of the building that ensures a satisfactory level of service. The primary goal of maintaining a school building is to extend its life [22], i.e., to keep it in a satisfactory functional, structural, and aesthetic condition for as long as possible [23, 24]. The lifespan of a building includes all phases of its life cycle, from the design of the building to its demolition. Maintenance costs are often higher than costs incurred in other stages of a building's life and can account for an average of 60% of all costs incurred during its lifetime [25], so it is essential to maintain the building properly, as very uneconomic costs can arise [26].

Building maintenance costs include the costs of labour, materials, equipment, tools, and any other related costs that may be incurred in building maintenance [26]. Maintenance cost estimation includes operation analysis and maintenance cost forecasting.

In the Republic of Croatia, financial needs for the maintenance of primary and secondary education are met for the most part through the Government of the Republic of Croatia, i.e., the Ministry of Science and Education. The main criteria for determining funds per founder include the number of school buildings, departments, and students. A certain financial part can be covered by the founders from their income. The founders are responsible for the distribution of available funds to the schools within their territorial units. The schools themselves analyse their needs, draw up budgets, and submit them to the founders. The founders then consider the reported needs and draw up a maintenance plan with a cost plan. Maintenance costs in all mentioned plans are mostly determined by empirical methods instead of sophisticated techniques. These amounts often do not reflect the actual state of affairs, so continuous updating and revision are necessary to adapt to changes and ensure the proper channelling of funds to prioritized needs [7]. Looking at the average difference between initially planned and actual costs, according to [7], it is very large, often up to 60%. The difference mainly arises due to changes in the scope of works, abandonment of some planned works, or the occurrence of maintenance works that were not foreseen in the plan [7]. The expressed inaccuracy in cost planning can negatively affect maintenance decision-making and the efficient use of the maintenance budget [9]. Therefore, it is necessary to design and develop assessment models that would reduce the level of uncertainty in the maintenance cost planning process.

Cost estimation models are vital for adequate cost determination [10]. According to [18], it should be noted that the maintenance cost model is probabilistic and not deterministic. This highlights the importance of developing innovative models that can assist decision-makers in determining the optimal choice of maintenance activities given the available budgets [27]. A database with insight into historical records can be used as an effective alternative to complement and improve current forecasting approaches [7]. Historical databases for forecasting the maintenance costs of educational buildings can contain various types of important information and data, such as the age of the building, location, area, number of floors, number of elevators, type of founder, number of work shifts, number of students, number of employees, type of heating, last modernization, etc. [7, 12, 14, 16-18]. This information forms a set of independent (predictive) variables based on which costs can be estimated (dependent, target variable). The application of such an approach requires access to data from different cases related to maintenance [28]. ANNs are considered potentially important applied methods for solving cost estimation problems [15] based on historical databases.

1.1.2 ANNs Application for Buildings Maintenance Cost Estimation

ANNs are a simplified mathematical model of processes carried out by networks of nerve cells in living beings. They consist of interconnected artificial neurons that know how to solve a particular problem after a learning process (training) via a database [29]. An ANN is built by arranging neurons in multiple layers, usually comprising input and output layers and one or more hidden layers. The output of one layer serves as the input to the next layer, and the strength of the output is determined by the connection weights between two adjacent layers. ANNs can learn from examples and independently discover the relationships between inputs and outputs [30]. In developing an ANN model, the most critical step is training the network, i.e., learning about the data, followed by testing or validating the network on data that did not participate in the model training.

The ANN consists of network architecture (topology), neuron connection scheme, neuron transfer function, and learning law. The architecture represents a particular arrangement and connection of neurons in the form of a network [31]. Different ANNs differ in architecture, such as Multilayer Perceptron (MLP), Radial Basis Function Neural Network (RBFNN), Generalized Regression Neural Network (GRNN), Temporal Neural Network (TNN), Neuro-Fuzzy, and others [32].

ANNs have proven their applicability in construction over the last few decades and have shown very good solutions to many problems [32]. Several prominent studies focusing on cost estimation have been recognized in building maintenance. Authors of [17] use ANNs to predict maintenance costs for higher education buildings. In their

case, the age of the building and the number of floors and elevators are important factors (variables) based on which costs can be estimated very well. Authors of [15] use an ANN to estimate the running costs of high-rise buildings. They identified seven important variables related to the mentioned costs: building type, gross floor area, pitched roof area, flat roof area, external glazing area, and the number of floors above and below ground level. They find that ANNs are a very accurate tool for estimating running building costs. Authors of [19] forecast the maintenance costs of residential buildings based on bills of quantities collected from building authorities. According to them, ANNs are suitable for modelling complex problems of a probabilistic nature and can easily be used for predicting future maintenance projects. Authors of [16] proposed a framework for an ANN model that learns about the maintenance costs of educational institutions. The main finding of this study is that the proposed ANN modelling framework is effective in estimating the maintenance costs of educational facilities. The developed model learns and generalizes claim payout records on the maintenance and repair costs from sets of facility asset information, geographic profiles, natural hazard records, and other causes of financial losses. Authors of [18] use 18 variables to build an ANN model to predict the maintenance costs of hospitals. They found that the ANN model is adequate, and the leading indicator of maintenance costs is the size of the hospital (area, number of patients).

This work is based on the argument that the use of ANNs, as advanced machine learning methods, can help solve the problem of predicting school building maintenance costs. The question arises as to why to use ANNs when using a simpler regression analysis can also obtain satisfactory results [7, 12, 33-38]. Many sources, however, point out that when comparing regression models with ANNs, regression still gives somewhat weaker results [15-17, 19, 38]. Given that building maintenance requires significant financial resources, any increase in the accuracy of the estimation model is essential [38]. Therefore, the application of ANNs in these cases is justified. Modelling techniques, including regression analysis, as well as some more advanced tools (fuzzy logic, genetic algorithms, base reasoning), have a hard time dealing with imprecision, incompleteness, and uncertainty of data and other variables that affect costs and their combinatorial effects and interrelationships. In these problematic areas, ANNs are often very strong [39].

As a result of the theoretical analysis, it can be concluded that the ANN is a promising approach to estimating school maintenance costs. Adopting an ANN to estimate school maintenance costs is doable and can yield satisfactory results.

2 METHODOLOGY

2.1 Study Area Selection

The Primorje-Gorski Kotar County in Croatia was chosen as the study area. The county covers 3 587 km^2 of the inhabited area of Croatia, in which about 7% of the total Croatian population lives. The county seat is located in Rijeka, the third-largest city in Croatia. The county has 14 towns and 22 municipalities, with 536 inhabited localities [40]. In these localities, there are 85 public schools, namely 57 primary and 28 secondary schools. The selected county ranks fourth in terms of the number of school facilities in the country. The location of primary and secondary schools within the selected study area is shown in Fig. 1.

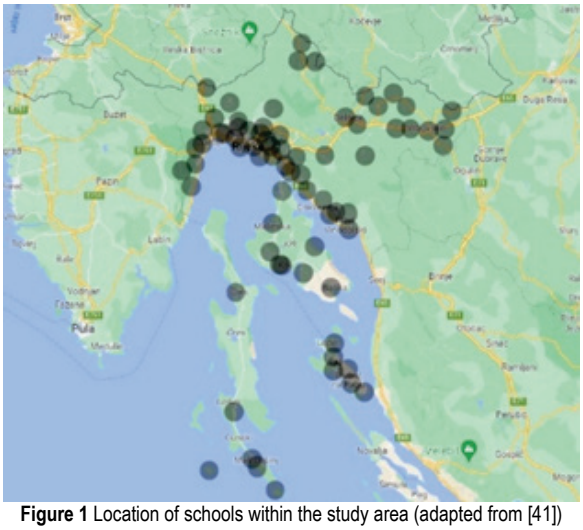


Figure 1 Location of schools within the study area (adapted from [41])

Primorje-Gorski Kotar County was chosen as the study area because of its size and importance for Croatia, a sufficient number of school institutions to create a representative data sample, and the data availability, as the paper's authors live and work in this area.

2.2 Data Collection and Processing

The collected database includes primary and secondary schools. Several documentation groups were used as data sources. Documentation was obtained from school representatives, i.e., from the founders (the City of Rijeka and Primorje-Gorski Kotar County) and the Ministry of Education of the Republic of Croatia. The mentioned documentation includes school work plans and programs, lists of institutions with their characteristics, lists of employees, students, etc. Annual financial reports of schools were also used, in which all receipts and expenses of the institution's operations were categorized. The data collected for this research from all the above documents include the following elements: type of school, type of founder, belonging to cultural property, type of heating, year of construction, total indoor area, number of employees, number of work shifts, number of students, number of classrooms, and realized annual maintenance costs.

In the Republic of Croatia, there was no such comprehensive database on both primary and secondary schools with a focus on maintenance and related costs. Based on the collected documentation, a review of previous research, and analysis, an extensive database was created that includes a set of information, i.e., independent variables considered relevant for defining the cost estimation model. Since there was no usable database, the aim of the research was to investigate a wider range of variables. Therefore, the database contains all those variables for which it was possible

to collect data from the available documentation, all with the purpose of determining the impact of individual variables on maintenance costs.

Maintenance costs refer to materials, parts, and services for ongoing and investment maintenance. The period of ten years, i.e., from 2013 to 2022, was observed. Available data on planned maintenance costs for 2022 were also collected to gain insight into the magnitudes of current cost estimation errors. Only schools that operate in one school building and do not share the building with another school (41 of them) were included in the research to avoid the problem of dividing or adding up maintenance costs from financial statements by the institution. If the school operates in several buildings, the problem is also information about the year of construction, belonging to cultural heritage, and type of heating, which can differ from building to building.

The database contains 11 predictive variables (X_1-X_{11}) and one target variable (Y), the average annual maintenance costs in the reference period. The reference period refers to the number of years for which cost data were collected for a specific building.

The average annual maintenance costs (*AAMC*) were obtained by following Ed. (1):

$$AAMC = \frac{1}{n}\sum_{i=1}^{n} PVMC_i. \quad (1)$$

In Ed. (1), *PVMC* refers to the present value of maintenance costs in a certain year, while n represents the reference period. The present value of costs was calculated according to the literature [36, 42, 43], with a discount rate of 3 % as per [44]. This is done because the value of money changes over time. All costs are expressed in euros (€). The set of predictive variables includes four qualitative variables and seven quantitative variables. The results of the statistical processing of the database are shown in Tab. 1 and 2.

Table 1 Characteristics of qualitative variables (sample proportions)

X_1 - School type	Primary	33
	Secondary	8
X_2 - Founder type	City	20
	County	21
X_3 - Cultural heritage	Yes	3
	No	38
X_4 - Heating type	Heating oil	24
	Gas	13
	Heating plant	4

Table 2 Descriptive statistics of quantitative variables

Variables	Mean	Standard Deviation	Min	Max
X_5	82.95	40.89	21	203
X_6	2969.16	1364.95	900	6778
X_7	50.68	16.7	23	93
X_8	1.27	0.45	1	2
X_9	8.15	1.8	5	10
X_{10}	330.17	209.79	30.4	919.8
X_{11}	17.48	8.08	7.33	39.8
Y	12403.08	5975.81	4018.7	25308.04

X_5 – Age to 2022; X_6 – Indoor area (m^2); X_7 – Number of employees; X_8 – Number of work shifts; X_9 – Reference period (RP); X_{10} – Average number of students in RP; X_{11} – Average number of class departments in RP; Y – Average annual maintenance costs (€) in RP

2.3 ANN Models Development

DTREG - *Predictive Modeling Software* was used to develop a model to predict maintenance costs using ANNs. The software offers the possibility to create several types of ANNs, of which MLP, GRNN, and RBFNN are very often used [32, 33, 38, 45, 46], as were employed in this paper. The greatest value of the aforementioned software is that it can self-optimize the parameters of the mentioned ANNs so that they yield the smallest estimation error. Moreover, the application of this software is straightforward and intuitive. Since the relationship between the target variable and the predictor is not known in advance, several ANN models need to be tested to select the best one so that the real data provide the highest accuracy [45]. By applying several types of ANNs, it was desired to test how suitable some of them are for estimating maintenance costs, and by trying several options, the best possible results should be obtained.

The MLP is a feed-forward ANN. It has an input layer of neurons that act as receivers, one or more hidden layers of neurons that compute the data and go through iterations, and finally, an output layer that predicts the output [47]. For training and validation of these networks, an algorithm with backpropagation of errors is used [48], which propagates through the network from the input to the output layer, then determines the error and propagates this error back to the input layer and includes it in the learning formula [49].

The GRNN consists of an input layer, a hidden layer (or pattern layer), an additional invisible layer for summation/division, and an output layer [50]. GRNN uses normalized Gaussian kernels in the hidden layer as activation functions. When training this network, it remembers each data pattern and does not require an iterative training procedure like the MLP network [46].

The RBFNN uses a radially symmetric and radially constrained transfer function in its hidden layer [51]. It consists of input, hidden, and output layers of neurons. The activation function of the hidden neurons is a Gaussian function. Learning the network is done in two steps. First, the weight coefficients of the network are determined from the input to the hidden layer and then from the hidden to the output layer [52].

During the development of the ANN model, different types of data representations were tried, such as originally recorded values and normalized values, to obtain a network model with the lowest estimation error.

For the collected school database to be used within the DTREG software, it first had to be formatted in a format suitable for analysis. The database must be in comma-separated value (CSV) format with values for one case per row and one column for each variable [53].

2.4 ANN Models Performance Comparison

The developed ANN models are evaluated and compared in terms of the magnitude of the estimation error they provide. The model error was measured using the coefficient of determination (R^2) and the mean absolute percentage error (*MAPE*). These two measures are among the most common

estimators of model accuracy [54] and have been used in numerous previous studies [8, 32, 33, 38, 45, 54].

The R^2 is a statistical measure that tests the overall fit of the predictive model, i.e., it indicates how much of the changes in the experimental values of the target variable are explained by the obtained model [36, 55]. R^2 values lie in the interval [0,1] [54], and the closer R^2 is to one, the more representative the model is. $R^2 = 0.900$ can be interpreted as follows: about 90.00% of the variation in the response can be explained by the predictor variables, while the remaining 10 % can be attributed to the unknown variables [54].

MAPE is a measure of predictive accuracy and is defined by the following formula [38]:

$$MAPE = \frac{1}{n}\sum\left|\frac{real\ value - estimated\ value}{real\ value}\right| (\%), \qquad (2)$$

where n is number of observations.

2.5 Best-fit Model Selection

The best-fit model will be selected based on the obtained values of R^2 and *MAPE*. According to Tab. 3 and Tab. 4, the interpretation of the selected error indicators is visible regarding the obtained sizes and according to Chaddock's scale [56] and Lewis's scale [57].

Table 3 Chaddock scale for interpreting the R^2 (adapted from [56])

R^2	Interpretation
<0.01	Absence of connection
0.01-0.24	Weak connection
0.25-0.64	Medium strength connection
0.65-0.99	Strong connection
>0.99	Full connection

Table 4 Lewis scale for interpreting the *MAPE* (adapted from [57])

MAPE (%)	Interpretation
<10.00	Very accurate prediction
10.00-20.99	Good prediction
21.00-50.00	Reasonable prediction
>50.00	Imprecise prediction

It is considered that models with R^2 over 0.64 describe the model well; that is, they provide a solid strong connection between actual and estimated values [58]. In cost prediction models, *MAPE* values up to 21% provide good prediction [59]. Some authors state that a *MAPE* of up to 30 % is considered a sufficiently good prediction [8, 38].

In this paper, an attempt will be made to obtain a model that gives $R^2 > 0.64$ with *MAPE* < 30 %, and it will be considered satisfactorily accurate. The goal is always to obtain models with as high an R^2 value as possible and as low a *MAPE* value as possible.

The obtained results of the optimal model will be additionally compared with modelling by regression analysis for the purpose of more significant statistical validation

The purpose of this study is to contribute to the maintenance management process in schools. Maintenance management encompasses a range of decisions, among which the decision-making regarding the maintenance tasks and associated costs holds significant importance. The objective is to effectively manage and optimally allocate these tasks and costs, for which the cost plan is of inestimable importance.

3 RESULTS AND DISCUSSION

The developed MLP, GRNN, and RBFNN models gave the results shown in Tabs. 5, 6, and 7. In addition to the results the developed networks gave based on the data in their original form, the results for R^2 and *MAPE* with normalized quantitative variables are shown in parentheses. The min-max normalization technique was used. The comparison shows that the results are somewhat closer with respect to R^2; however, *MAPE* shows significantly better performance in networks developed on the basis of the original data, so these will be presented in more detail. The reason for this may be the nature of the data. The actual data may already be in an acceptable shape and range, so additional normalization may lead to the loss of specific information or variations in the real data, resulting in reduced performance when applying networks with normalized data.

Table 5 Performance of MLP model

Parameters	MLP	
	Training	Validation
Mean target value for input data	12,331.55	12,698.11
Mean target value for predicted values	12,361.94	9,943.49
Correlation between actual and predicted	0.844	0.903
Proportion of variance explained by model (R^2)	0.713 (0.713)	0.614 (0.614)
MAPE	26.37 (69.29)	22.24 (22.27)
Analysis run time	00:02.84	

Table 6 Performance of GRNN model

Parameters	GRNN	
	Training	Validation
Mean target value for input data	12,331.55	12,698.11
Mean target value for predicted values	12,502.52	11,287.42
Correlation between actual and predicted	0.930	0.945
Proportion of variance explained by model (R^2)	0.862 (0.909)	0.811 (0.603)
MAPE	13.83 (43.74)	20.50 (50.55)
Analysis run time	00:00.67	

Table 7 Performance of RBFNN model

Parameters	RBFNN	
	Training	Validation
Mean target value for input data	12,331.554	12,698.11
Mean target value for predicted values	12,331.55	9,501.80
Correlation between actual and predicted	0.976	0.901
Proportion of variance explained by model (R^2)	0.953 (0.943)	0.542 (0.657)
MAPE	9.34 (31.39)	47.61 (72.50)
Analysis run time	00:01.28	

According to the results, the best model was provided by the GRNN, which demonstrated one of the highest R^2 values and the lowest *MAPE* values. These values indicate satisfactory accuracy during both training and validation. The obtained R^2 values during training and validation reveal a

strong relationship between actual and estimated costs. Regarding the *MAPE* value, it is an impressive 13.83 % during training. In validation, the *MAPE* is somewhat weaker, nearing the border, but still within the range of good prediction. The GRNN also exhibited the shortest time for data analysis and model creation. According to [53], it is not uncommon for the GRNN model to outperform MLP and RBFNN models in this type of estimation.

Slightly weaker results were observed with the MLP model, which, as per Tab. 3 and Tab. 4, fall within a reasonable prediction range. As for the RBFNN model, it achieved a commendable result during training, but its performance during validation was not as high. Therefore, among the three models, it may not be the most suitable for estimating maintenance costs. Possible reasons for these differences could be attributed to the nature, composition, and complexity of the database. Machine learning models exhibit diverse performance across different datasets, and the optimal model choice can vary based on the data characteristics and the specific problem at hand. Therefore, exploring multiple models is advisable to determine the most suitable one.

All presented results surpass current estimates. Specifically, during the database collection, data on planned and actual (realized) maintenance costs for the year 2022 were also collected. Planned cost values are reported in financial plans before the start of the fiscal year, while actual costs are reported in financial statements after the end of the observed year.

Such comprehensive data were collected for 20 schools. The comparison of initially planned and actual realized values for the specified accounting year resulted in the graph shown in Fig. 2.

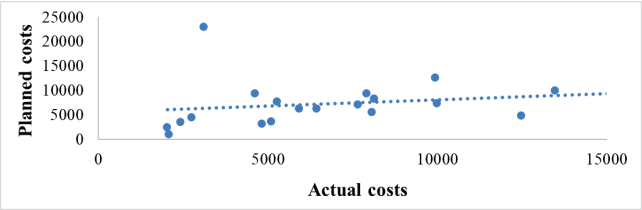
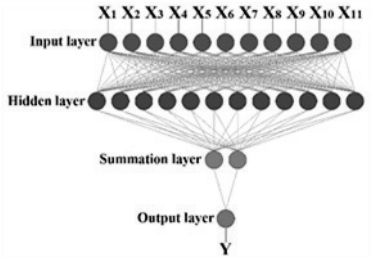


Figure 2 Relationship between actual and planned values of school maintenance costs for 2022

Comparing the planned cost values with the actual realized costs yields a *MAPE* value of 64.03 %, with an R^2 of 0.0484. In practice, planned cost values are typically adjusted based on needs, considering the available funds of the founder and often utilizing empirical methods. In the context of this research, which utilized a historical database for computer modelling with a particular focus on key predictor variables for assessing the target, it is not surprising that the obtained results significantly exceed the accuracy level relative to the actual operational scenario.

Considering the obtained results, the GRNN is acknowledged as the most suitable model for predicting actual annual maintenance costs. The parameters of the developed GRNN from DTREG are presented in Tab. 8. With the selected parameters; the model resulted in the smallest estimation error.

Table 8 Parameters of the developed GRNN

Type of model	General Regression Neural Network (GRNN)
Type of analysis	Regression
Number of neurons in the model	The minimum error occurred with 26 neurons in the model (12 in the hidden layer)
Sigma values for model	Sigma for each variable
Starting sigma search control	Min. sigma: 0.001 Max. sigma: 10 Search steps: 20
Model optimization and simplification	Remove unnecessary neurons Minimize error Retrain after removing neurons
Model testing and validation	Random sampling (20%)
How to handle missing values	Replace missing values with medians
Type of kernel function	Gaussian
Conjugate gradient parameters	Maximum total iteration: 5000 Iterations without improvement: 1000 Min. improvement delta: 1.000e-005 Absolute convergence tolerance: 1.000e-008 Relative convergence tolerance: 1.000e-004
Developed GRNN architecture	

The developed GRNN consists of 4 layers, with the optimal number of neurons in the hidden layer set at 12. DTREG conducts regression analysis within the ANN model, given the continuous nature of the target variable. The sigma values control the radius of influence of each point in the ANN model, and a separate sigma value is calculated for each predictor variable. This choice is recommended as it strikes a good balance between a single sigma and the possibility of a separate sigma for each target category. The initial sigma search control parameters determine the range of sigma values used during the initial search, and once the conjugate gradient method begins, the sigma values can extend beyond this range. DTREG utilizes the conjugate gradient algorithm to compute optimal sigma values and weights between neurons. To optimize and simplify the model, unnecessary neurons were removed, error was minimized, and retraining was performed after neuron removal [53].

Regarding model testing and validation, DTREG offers four options:

- A random percentage of rows is kept out when the model is created. After the model is created, this number of rows is run through the model, and the error is evaluated.
- The control variable is used to select which rows to hold out for validation.
- Cross-validation with the selected number of folds.

- Cross-validation with one omitted row in each model created [45, 53].

In this specific scenario, a random selection method was applied, involving the selection of 20 % of the data from the total database that did not participate in the ANN training process. This methodology, due to its simple implementation, quick performance, and reliability, is often used in similar research, as confirmed by the authorship of various scientists [15-19, 29, 39, 60, 61].

As part of this research, all the mentioned validation methods were checked, showing close results. Nevertheless, the random selection method resulted in slightly improved performance compared to the others, so it was chosen as authoritative.

In case certain values are missing in the database, the developed GRNN model replaces them with medians. However, there were no such values in this study.

The kernel function is used to take data as input and transform it into the required form of data processing. According to [53], the kernel function controls how the influence of a point decreases as the radius increases. In this case, the Gaussian function was applied. A Gaussian function causes the influence of a point to decrease according to the value of a Gaussian distribution cantered on the point. According to [53], Gaussian functions are almost always the best kernel.

The relationship between the actual and predicted values of annual school maintenance costs for the GRNN model is shown in Fig. 3.

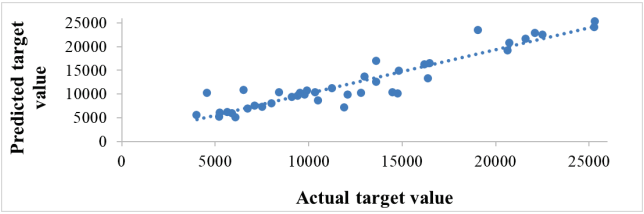


Figure 3 Relationship between actual and predicted values of the target variable (GRNN)

DTREG also calculates the relative importance of each predictor to model quality using sensitivity analysis. Below is the importance to three decimal places for the first five predictor variables:

1) Founder type (100.000 %)
2) Average number of class departments in RP (48.807 %)
3) Heating type (18.250 %)
4) Age to 2022 (8.848 %)
5) Indoor area (2.141 %).

The displayed values represent the percentage importance of each predictor in the model for predicting the target variable. It can be observed that the most crucial predictor for predicting actual maintenance costs is information about the type of school founder (the City of Rijeka or Primorje-Gorski Kotar County), followed by the number of departments, heating method, age, and indoor area. Other predictive variables from the database have an importance below 1 %.

The exceptionally significant importance of the variable "founder type" may reflect the complexity of financial, organizational, and structural factors associated with different types of school founders. The type of founder primarily affects the amount of financial support given their size, income, and maintenance policies. This financial support significantly influences the school's ability to maintain its infrastructure.

The average number of class departments is also crucial because it affects the size, scope, and frequency of maintenance. The number of departments influences the consumption of resources such as water, heating and cooling, use of equipment, etc., which can increase maintenance costs to ensure the proper functioning of all school spaces. Also, the number of departments directly affects the budget that the founders receive from the state to maintain their institutions.

The heating type also has a pronounced impact on school maintenance costs but to a lesser extent than the previous two factors. This variable may be more important than others because different heating types require different levels of attention and resources to maintain. The type of heating can affect the energy efficiency of the building as well as the infrastructure of the building, where all complex and more efficient systems require special maintenance and financial expenses. The specificity of the local environment is particularly important for heating, where local conditions and climate affect heating systems, which can have a significant impact on maintenance costs. Primorje-Gorski Kotar County's climate is characterized by warm summers but dry and extremely windy winters, which requires efficient heating systems.

The age of schools in the context of this research has a certain influence on maintenance costs. The age of the building can affect maintenance costs in terms of the need for modernization, improvement of energy efficiency, and adaptations to modern needs. Regular maintenance and planned investments can help extend the life of the building and reduce overall maintenance costs over time. According to the available information, the vast majority of schools in the observed research sample had a certain form of modernization intervention, so it is not surprising that the age of the building itself did not have a pronounced impact on maintenance costs. Unfortunately, more detailed information about the modernization of the facilities was not available.

According to the results, the indoor area of the school has a certain influence on maintenance costs, but it is the least significant of all the factors listed here. The size of the school affects maintenance costs primarily through the amount of maintenance work that needs to be done. It is possible that in this context, some other factors or specificities compensate or reduce this impact, such as the specificity and structure of buildings, energy efficiency, maintenance method, etc. Another important aspect may be that in the developed model for estimating maintenance costs, the school indoor area is reflected through the number of class departments, which means that the direct influence of the school area may be less pronounced in this specific case.

It should be noted that the obtained relationships of the variables may vary depending on the specifics of the data used in the model and the characteristics of the school facilities.

Additional statistical validation and robustness checks of the developed GRNN model were confirmed through a comparison with multiple regression analysis. Conducting the regression analysis in DTREG, the training results were $R^2 = 0.743$, $MAPE = 23.14$ %, and for validation, $R^2 = 0.440$ and $MAPE = 30.39$ %. These results indicate that the ANN model is a more appropriate choice when estimating maintenance costs and provides estimates with a smaller error.

Considering all the presented results, it can be concluded that the developed GRNN model for estimating the average annual costs of maintaining school buildings is suitable for use in schools in Primorje-Gorski Kotar County.

The developed cost estimation model contributes to increasing the efficiency of the maintenance process. The ANN model enables:

- Simple, quick, and structured planning of maintenance costs for specific time periods.
- A more accurate estimate that provides better insight into the budget, helps allocate funds, and aids decision-making and cost control.
- Reduction of financial losses and the possibility of creating a strategy for mitigating and reducing losses.
- By using the model, the value of the average annual maintenance costs is obtained, which remains constant for each year of maintenance in the observed reference period, allowing for planning over extended periods.
- Defining the characteristics of school buildings that affect maintenance costs and obtaining the necessary information for designing new schools concerning cost rationalization. In this case, the important characteristics are the type of founder, the number of class departments, the type of heating, the age, and the total indoor building area.

The magnitudes of the errors obtained from the GRNN model show that there is still room for progress and the development of more precise estimation models. The results prove that estimating maintenance costs is a demanding undertaking filled with numerous uncertainties. Still, any increase in accuracy in the estimations is of great importance, considering the amount of funds that need to be invested in schools. ANNs learn from input data; therefore, their quality and quantity directly affect the model's accuracy. The possibility of increasing the accuracy of the model can be seen in the expansion of the database with additional variables.

The results of this research can help school institutions and public bodies in planning costs and making maintenance decisions. Applying the ANN model opens up new possibilities for planning the necessary budget for school maintenance, making maintenance more efficient. A more accurate maintenance cost plan provides insight into the budget, helps allocate money, and helps control and monitor these costs. Founders managing school buildings can include in their budget the average required amount spent annually on school maintenance. For school buildings to be maintained, it is first necessary to consider the necessary maintenance activities and then to foresee certain funds for that purpose in the budget, which can be helped by the cost forecasting model developed here.

The main finding of this study is that the proposed ANN is effective in learning the maintenance costs of school facilities, comparable to other previous studies focused on a limited number of factors affecting maintenance costs [15-19].

Comparing the highlighted predictor variables for predicting maintenance costs, the results correlate with different studies where similar results have been obtained. In study [14], the type of founder of the institution and the area of the premises are also mentioned as important variables for estimating maintenance costs. In studies [16, 18, 37], the area of buildings is also highlighted, and in studies [16, 17], one of the important variables for estimating costs is the age of the examined buildings. Considering the quality and coverage of the database, this research further emphasizes the need to consider the number of departments and the type of building heating as relevant variables in the maintenance cost modelling process.

Looking at prominent studies from the Republic of Croatia where a model for estimating the costs of maintaining educational buildings is being developed, two studies stand out, in which, admittedly, regression analysis was used. The author of [12] estimated the costs of use and maintenance of faculty buildings and obtained a model with accuracy indicators of $R^2 = 0.669$ and $MAPE = 27.24$ %. The author of [14] developed a model for estimating primary school maintenance costs and obtained $R^2 = 0.759$, $MAPE = 21.17$ %. In this research, a larger database with both primary and secondary schools was used, and ANNs were applied as more advanced forecasting techniques. The obtained results are, therefore, more precise, indicating that the research is heading in the right direction and contributes to progress in the estimation modelling of the maintenance costs of educational buildings.

The first and basic limitation of the developed ANN model is its applicability to school buildings in Primorje-Gorski Kotar County. Also, since the database consisted only of schools operating in one building, the application is exclusively for such institutions. The limit is also related to the number of years over which school maintenance costs are estimated. The period for which the data was collected ranges from 5 to 10 years, and considering that, it is possible to estimate the costs for a maximum period of 10 years. The values obtained for R^2 and $MAPE$ suggest that there is room for creating models that could give even better results; expanding the database with additional variables is necessary.

Considering the highlighted limitations, it is recommended to expand the research to the entire territory of the Republic of Croatia (or beyond) and develop a model that will not have a regional character and potentially give better

results. For this reason, the database should be continuously expanded and updated.

4 CONCLUSION

The subject of this study is estimating the costs of maintaining primary and secondary schools in the Republic of Croatia using ANNs. The study successfully achieved the research goal and provided positive answers to all research questions.

Within the study, a database of school buildings in Primorje-Gorski Kotar County, Republic of Croatia, was created, containing historical data on maintenance costs. A large sample database with data for 41 schools, was utilized for further analysis and modelling. It encompasses 11 predictor variables (characteristics of primary and secondary schools) and one target variable (average annual maintenance costs). Data on maintenance costs were collected for the 5 to 10 years.

Based on the database, an ANN model was developed to estimate the maintenance costs of school buildings with satisfactory accuracy ($R^2 > 0.64$, $MAPE < 30$ %). Initially, three different ANNs were developed: MLP, GRNN, and RBFNN, using the software DTREG. The error of models is expressed using R^2 and *MAPE*, and the values obtained in all developed models are considered satisfactory for this type of assessment. Comparison of results showed that the highest estimation accuracy is achieved by applying GRNN, with a training *MAPE* of 13.83 % and validation *MAPE* of 20.50 %. R^2 values are 0.862 and 0.811, respectively, indicating that GRNN is optimal for estimating school maintenance costs.

The study demonstrated that the current inaccuracy in the estimation of school maintenance costs can be reduced by applying the developed ANN model. Specifically, by processing the actual data for the year 2022, it was found that, observing the actual planned and realized costs from the financial documents, the accuracy indicators are *MAPE* = 64.03 %, with $R^2 = 0.0484$, which is far below satisfactory values.

The application of the developed ANN model can contribute to increasing the efficiency of maintenance. The results can assist educational institutions and public bodies in planning costs and making maintenance decisions. By applying the developed model, it is possible to plan and calculate maintenance costs for a multi-year period in a simple, fast, and structured way. This approach can help reduce losses and develop strategies to mitigate them. Characteristics of schools that significantly affect maintenance costs were also identified, namely the type of founder, the number of school departments, the type of heating, age, and the total indoor area of the building, which can be useful when designing new school buildings.

No significant studies dealing with this topic in the manner approached here have been noted in the literature so far. In the Republic of Croatia, there was a lack of more extensive research analysing the costs of maintaining primary and secondary schools. Estimates of maintenance costs have shown inaccuracy, and the limited studies conducted were based on smaller databases of educational buildings, where cost prediction was reliant on simpler modelling techniques.

There are also some research limitations, such as the regional orientation, the limited database, and the reference period for which the data were collected. In future research, expanding the database and developing new estimation models is recommended to obtain even better results, as there is undoubtedly room for improvement considering the size of R^2 and *MAPE* obtained.

Acknowledgments

This work has been fully supported by the University of Rijeka project uniri-mladi-tehnic-22-66.

5 LITERATURE

[1] Abdelhamid, M., Beshara, I., & Ghoneim, M. (2013). Asset Management for Educational Buildings in Egypt. *World Academy of Science, Engineering and Technology, International Journal of Social, Behavioural, Educational, Economic, Business and Industrial Engineering, 7*(6), 1434-1443.

[2] Le, A. T. H., Domingo, N., Rasheed, E. & Park, K. S. (2022). Maturity model of building maintenance management for New Zealand's state schools. *Building Research & Information, 50(*4), 438-451. https://doi.org/10.1080/09613218.2021.1961573

[3] Tijanić Štrok, K., Car-Pušić, D. & Marenjak, S. (2023a). Elementary School Buildings Condition Assessment: Case of Primorje-Gorski Kotar County (Croatia). *Advances in Civil and Architectural Engineering, 14*(26), 95-117. https://doi.org/10.13167/2023.26.7

[4] Trachte, S. & De Herde, A. (2014). Sustainable refurbishment school buildings. Retrieved from https://www.iea-shc.org/data/sites/1/publications/subt.D.School%20Renovation.022015.pdf

[5] Tijanić Štrok, K., Car-Pušić, D. & Marenjak, S. (2023b). Priorities in Croatian School Building Maintenance: A Comparison of the Main Stakeholders' Views. *Sustainability, 15*(15), 1-21. https://doi.org/10.3390/su151511767

[6] Olanrewaju, A. L. & Abdul-Aziz, A. R. (2014). *Building Maintenance Processes and Practices: The Case of a Fast Developing Country*. New York, Springer. https://doi.org/10.1007/978-981-287-263-0

[7] Tijanić Štrok, K. (2021). Razvoj modela za učinkovito upravljanje održavanjem javnih obrazovnih građevina (Development of the model for efficient maintenance management of public educational buildings). *Doctoral dissertation*, Josip Juraj Strossmayer University of Osijek, Faculty of Civil Engineering and Architecture Osijek, Croatia. (in Croatian)

[8] Tijanić, K. & Car-Pušić, D. (2019) Procjena operativnih troškova škola primjenom umjetnih neuronskih mreža (Estimation of schools operating costs with the application of artificial neural networks). *Proceedings of Common Foundations 2019,* Rijeka, Croatia, 126-131. (in Croatian)

[9] Tijanić, K., Car-Pušić, D. & Marenjak, S. (2019). Contribution to increasing the cost efficiency in the exploitation phase of public education facilities. *Proceedings of the 14th International Conference OTMC and the 7th IPMA Research Conference,* Zagreb, Croatia, 9-22.

[10] Le, A. T. H., Domingo, N., Rasheed, E. & Park, K. S. (2018). Building Maintenance Cost Planning and Estimating: A

Literature Review. *Proceedings of the 34th Annual ARCOM Conference,* Belfast, United Kingdom, 697-706.
[11] Tijanić Štrok, K., Marenjak, S. & Car-Pušić, D. (2022). Analysis of the current maintenance management process in school buildings: Study area of Primorje-Gorski Kotar County, Republic of Croatia. *Frontiers in Built Environment, 8*, 912326. https://doi.org/10.3389/fbuil.2022.912326
[12] Krstić, H. (2011). Model procjene troškova održavanja i uporabe građevina na primjeru građevina Sveučilišta Josipa Jurja Strossmayera u Osijeku (Maintenance and Operation cost's prediction model based on the Facilities of the University of Josip Juraj Strossmayer in Osijek). *Doctoral dissertation*, Josip Juraj Strossmayer University of Osijek, Faculty of Civil Engineering Osijek, Osijek, Croatia. (in Croatian)
[13] State Audit Office (2018). *Izvješće o obavljenoj reviziji učinkovitosti kapitalnih ulaganja u osnovne i srednje škole na području Republike Hrvatske (Report on the Audit of the Effectiveness of Capital Investments in Primary and Secondary Schools in the Republic of Croatia)*. Sisak. (in Croatian)
[14] Tijanić Štrok, K. (2023). Estimation of Maintenance Cost in Elementary Schools of Primorje-Gorski Kotar County, Croatia. *Proceedings of the Interdisciplinary Management Research XIX*, Osijek, Croatia, 248-268.
[15] Alqahtani, A. & Whyte, A. (2016). Estimation of life-cycle costs of buildings: regression vs artificial neural network. *Built Environment Project and Asset Management, 6*(1), 30-43. https://doi.org/10.1108/BEPAM-08-2014-0035
[16] Kim, J. M., Yum, S., Son, S., Son, K. & Bae, J. (2021). Modeling deep neural networks to learn maintenance and repair costs of educational facilities. *Buildings, 11*(4), 165. https://doi.org/10.3390/buildings11040165
[17] Li, C. S., Chen, P. J. & Guo, S. J. (2010). Application of back-propagation artificial neural network to predict maintenance costs and budget for university buildings. *Proceedings of 2010 6th International Conference on Natural Computation*, New York, USA, 1546-1551. https://doi.org/10.1109/ICNC.2010.5583722
[18] Olanrewaju, A. L. & Ooi, Y. L. (2022). Determinants of maintenance cost of hospital buildings: an artificial neural network approach. *IOP Conference Series: Earth and Environmental Science, 1067*(1), 012083. https://doi.org/10.1088/1755-1315/1067/1/012083
[19] Otmani, A., Bouabaz, M. & Al-Hajj, A. (2020). Predicting maintenance and rehabilitation cost for buildings based on artificial neural network and fuzzy logic. *International Journal of Computational Intelligence and Applications, 19*(01), 2050001. https://doi.org/10.1142/S1469026820500017
[20] Naik, M. G. & Radhika, V. (2015). Time and cost analysis for highway road construction project using artificial neural networks. *Journal of Construction Engineering and Project Management, 5*(1), 26-31. https://doi.org/10.6106/JCEPM.2015.5.1.026
[21] Cerić, A. & Katavić, M. (2000). Upravljanje održavanjem zgrada (Building maintenance management). *Građevinar, 53*(00), 83-89. (in Croatian)
[22] Besiktepe, D., Ozbek, M. E. & Atadero, R. A. (2020). Identification of the criteria for building maintenance decisions in facility management: First step to developing a multi-criteria decision-making approach. *Buildings, 10*(9), 166. https://doi.org/10.3390/buildings10090166
[23] Bentley, T. (2012). *Learning beyond the classroom: Education for a changing world*. United Kingdom, Routledge. https://doi.org/10.4324/9780203201756
[24] Izobo-Martins, O. O., Ekhaese, E. N., Ayo-Vaghan, E. O. & Olotuah, A. O. (2018). Assessing Users' Perceptions of the Current Maintenance Disorder of Public Secondary School in Ogun, Nigeria. *Journal of Building Construction and Planning Research, 6*(2), 90-101. https://doi.org/10.4236/jbcpr.2018.62006
[25] Cerić, A., Završki, I., Vukomanović, M., Ivić, I. & Nahod, M. (2019). BIM implementation in building maintenance management. *Građevinar, 71*(10), 889-900. https://doi.org/10.14256/JCE.2730.2019
[26] Obradović, D. & Marenjak, S. (2017). Uloga održavanja u životnom ciklusu građevine (The role of maintenance in the life cycle of a building). *Proceedings of 26. Međunarodni znanstveno-stručni skup "Organizacija i tehnologija održavanja"*, Osijek, Croatia, 61-67. (in Croatian)
[27] Ghafoori, M. & Abdallah, M. (2022). Building Optimization Model for Minimizing Operation and Maintenance Costs. *Proceeding of Construction Research Congress*, Reston, Virginia, USA, 422-431. https://doi.org/10.1061/9780784483978.044
[28] Kwon, N., Song, K., Ahn, Y., Park, M. & Jang, Y. (2020). Maintenance cost prediction for aging residential buildings based on case-based reasoning and genetic algorithm. *Journal of Building Engineering, 28*, 101006. https://doi.org/10.1016/j.jobe.2019.101006
[29] Sušanj, I. (2017). Razvoj hidrološkog modela otjecanja s malih slivova temeljen na umjetnoj neuronskoj mreži (Development of the hydrological rainfall-runoff model based on artificial neural network in small catchments). *Doctoral dissertation*, University of Rijeka, Faculty of Civil Engineering, Rijeka, Croatia. (in Croatian)
[30] Bouabaz, M. & Hamami, M. (2008). A cost estimation model for repair bridges based on artificial neural network. *American Journal of Applied Sciences, 5*(4), 334-339. https://doi.org/10.3844/ajassp.2008.334.339
[31] Ilić, V. (1999). Neuronske mreže (Neural Networks). Retrieved from http://solair.eunet.rs/~ilicv/neuro.html (in Serbian)
[32] Petruseva, S., Zileska-Pancovska, V. & Car-Pušić, D. (2019). Implementation of process-based and data-driven models for early prediction of construction time. *Advances in Civil Engineering, 2019.* https://doi.org/10.1155/2019/7405863
[33] Car-Pušić, D., Tijanić, K., Marović, I. & Mlađen, M. (2020). Predicting buildings construction cost overruns on the basis of cost overruns structure. *Scientific Review Engineering and Environmental Studies, 29*(3), 366-376. https://doi.org/10.22630/PNIKS.2020.29.3.31
[34] Kim, J. M., Kim, T., Yu, Y. J. & Son, K. (2018). Development of a maintenance and repair cost estimation model for educational buildings using regression analysis. *Journal of Asian Architecture and Building Engineering, 17*(2), 307-312. https://doi.org/10.3130/jaabe.17.307
[35] Li, C. S. & Guo, S. J. (2012). Life cycle cost analysis of maintenance costs and budgets for university buildings in Taiwan. *Journal of Asian Architecture and Building Engineering, 11*(1), 87-94. https://doi.org/10.3130/jaabe.11.87
[36] Obradović, D. (2022). Doprinos povećanju učinkovitosti održavanja kanalizacijskih sustava primjenom modela procjene troškova održavanja (A contribution to increasing the efficiency of sewerage systems maintenance by applying a maintenance cost estimation model). *Doctoral dissertation*, Josip Juraj Strossmayer University of Osijek, Faculty of Civil Engineering and Architecture Osijek, Osijek, Croatia. (in Croatian)
[37] Plebankiewicz, E. & Grącki, J. (2023). Analysis and prediction of universities' buildings' renovation costs using a regression model. *Applied Sciences, 13*(1), 401. https://doi.org/10.3390/app13010401

[38] Tijanić, K., Car-Pušić, D. & Šperac, M. (2020). Cost estimation in road construction using artificial neural network. *Neural Computing and Applications, 32*, 9343-9355. https://doi.org/10.1007/s00521-019-04443-y
[39] Ahiaga-Dagbui, D. D. & Smith, S. D. (2012). Neural networks for modelling the final target cost of water projects. *Proceedings of the 28th ARCOM*, United Kingdom, 307-316.
[40] Primorsko-goranska županija (2020). O županiji (About the county). Retrieved from https://www.pgz.hr/o-zupaniji/ (in Croatian)
[41] Google Maps (2023). Primorje-Gorski Kotar County. Retrieved from https://www.google.com/maps/d/viewer?mid=1q6BOwisSY1nH5SkRZWce_CzCTcs&hl=en_US
[42] Čulo, K. (2010). *Ekonomika investicijskih projekata (Economics of Investment Projects)*. Osijek, Faculty of Civil Engineering Osijek. (in Croatian)
[43] Narodne novine (2017). *Odluka o kamatnim stopama, diskontnoj (eskontnoj) stopi i naknadama Hrvatske narodne banke (Decision on interest rates, discount rate and fees of the Croatian National Bank).* NN 94/2017-2189. (in Croatian)
[44] Hrvatska narodna banka (2022). *Aktivne kamatne stope NBH – HNB (Active interest rates of CNB*. (in Croatian)
[45] Car-Pušić, D., Petruseva, S., Zileska Pancovska, V. & Zafirovski, Z. (2020b). Neural network-based model for predicting preliminary construction cost as part of cost predicting system. *Advances in Civil Engineering, 2020*, 1-13. https://doi.org/10.1155/2020/8886170
[46] Ghaderzadeh, M., Rebecca, F. E. I. N. & Standring, A. (2013). Comparing performance of different neural networks for early detection of cancer from benign hyperplasia of prostate. *Applied Medical Informatics, 33*(3), 45-54. https://ami.info.umfcluj.ro/index.php/AMI/article/view/425
[47] Desai, M. & Shah, M. (2021). An anatomization on breast cancer detection and diagnosis employing multi-layer perceptron neural network (MLP) and Convolutional neural network (CNN). *Clinical eHealth, 4*, 1-11. https://doi.org/10.1016/j.ceh.2020.11.002
[48] Jurković, Z. & Marinković, R. (2012). Development of Neural Network Models for the Frequency of Use of Faculty Library. *Ekonomski vjesnik, 25*, 106-116. (in Croatian)
[49] Educational Repository for Intelligent Systems (2003a). Mreža "širenje unatrag" (Backpropagation). Retrieved from http://eris.foi.hr/11neuronske/nn-predavanje5.html (in Croatian)
[50] Educational Repository for Intelligent Systems (2003b). Ostali algoritmi neuronskih mreža - Neuronska mreža opće regresije (Other neural network algorithms - General Regression Neural Network). Retrieved from https://eris.foi.hr/11neuronske/nn-predavanje6-gr.html (in Croatian)
[51] Zekić-Sušac, M. & Musser, E. (2010). Neuronske mreže u predviđanju ponude i potražnje prehrambenih proizvoda (Neural networks in forecasting the supply and demand of food products). *Proceedings of "Kontinentalni gospodarski resursi u funkciji razvitka turizma Republike Hrvatske"*, Osijek, Croatia, 275-294. (in Croatian)
[52] Bullinaria, J. A. (2004). Radial Basis Function Networks: Introduction. Retrieved from http://www.cs.bham.ac.uk/~jxb/NN/l12.pdf
[53] Sherrod, P. (2014). *DTREG Predictive Modeling Software*. DTREG program manual.
[54] Petruseva, S., Car-Pušić, D. & Zileska Pancovska, V. (2016). Predicting construction time by using general regression neural network. *Proceedings of International Scientific Conference People, Buildings and Environment*, Brno, Czech Republic, 33-46.
[55] Serdar, V. (1966). *Udžbenik statistike (Textbook of statistics).* Zagreb, Školska knjiga. (in Croatian)
[56] Chaddock, R. E. (1925). *Principles and Methods of Statistics*. Cambridge, The Riverside Press.
[57] Lewis, C. D. (1982). *Industrial and business forecasting methods*. London, Butterworths.
[58] Moore, D. S., McCabe, G. P. & Craig, B. A. (2014). *Introduction to the practice of statistics*. New York, W.H. Freemen & Co.
[59] Moreno, J. J. M., Pol, A. P., Abad, A. S. & Blasco, B. C. (2013). Using the R-MAPE index as a resistant measure of forecast accuracy. *Psicothema, 25*(4), 500-506. https://doi.org/10.7334/psicothema2013.23
[60] Marović, I., Androjić, I., Jajac, N. & Hanák, T. (2018). Urban road infrastructure maintenance planning with application of neural networks. *Complexity*, *2018*, 1-10. https://doi.org/10.1155/2018/5160417
[61] Beljkaš, Ž. & Knežević, M. (2021). Use of artificial intelligence for estimating cost of integral bridges. *Građevinar*, *73*(03), 265-273. https://doi.org/10.14256/JCE.2831.2019

Author's contacts:

Ksenija Tijanić Štrok, PhD
University of Rijeka, Faculty of Civil Engineering
Ul. Radmile Matejčić 3, 51000, Rijeka, Croatia
+38551265920, ksenija.tijanic@uniri.hr

ISSN 1846-6168 (Print), ISSN 1848-5588 (Online)
https://doi.org/10.31803/tg-20231116101632

Original scientific paper
Received: 2023-11-16, Accepted: 2024-03-04

BSO-CNN: A BSO Pressure Optimized CNN Model for Water Distribution Networks

Waghmare Shwetambari Pandurang*, Renu Praveen Pathak, Imtiyaz Ahmad Wani

Abstract: Urban water distribution networks must use pressure management to reduce water leakage by modifying storage tank pressure levels in response to variations in water demand. Since each demand node usually restricts the maximum pressure that may be applied, addressing pressure issues at individual nodes is also crucial. To overcome these difficulties, a brand-new Convolutional Neural Network (CNN) Pressure Optimization Model is proposed. This model collects real-time data on water levels and pressure by utilizing level and pressure sensors, and a Backtracking Search Optimization (BSO) model is used to process the data. Subsequently, the optimized data is employed to execute accurate flow control protocols and detect possible leakage points. Major advantages are achieved with the BSO-CNN approach, including lower operating costs, more efficiency, and less water pressure needed. Furthermore, the model demonstrates a noteworthy decline in leakage rates, attaining a noteworthy reduction of roughly 31.5 % in the water distribution network. Improving the sustainability and performance of urban water distribution systems can be achieved through the proper integration of predictive modeling and advanced optimization techniques.

Keywords: Backtracking Search Optimization Algorithm (BSO); Convolution Neural Network (CNN); hydraulic reliability; leakage; pressure management; storage tank; Water Distribution Networks (WDN)

1 INTRODUCTION

An essential and renewable resource for maintaining life on Earth is water. It is of the utmost importance and a basic requirement for human existence [1, 2]. Water delivery firms provide one of the most basic necessities of modern life by facilitating the distribution of water to enterprises, industrialized communities, and metropolitan centers. Water delivery firms operate within a complete water supply system that includes multiple stages, including the procurement, treatment, and distribution of water to final users [3]. These systems are carefully planned and operated to guarantee a steady and dependable supply of clean water to fulfill the various demands of the community [4].

Pumping stations are essential to the operation of water supply systems because they keep the water pressure constant across the distribution network [5]. These stations pressurize the water network using sophisticated processes, frequently depending on raised reservoirs or storage tanks to enable effective distribution [6]. Water delivery companies, along with the infrastructure they support, such as pumping stations, are essential to contemporary civilization because they guarantee the availability of clean, drinkable water, which is necessary for the survival of enterprises, industries, and communities [7]. Water Distribution Networks (WDNs) are an integral part of industrialized societies' infrastructure, carrying out the crucial task of effectively transporting water to locations of consumption while preserving proper pressure and velocity levels [8]. WDNs are intricate networks made up of reservoirs, consumption nodes (many households, businesses, and industries), and a complex web of pipelines that connects these nodes. Ensuring the safety and dependability of the water supply is important to WDN operations [9]. This means that in addition to supplying water at the places of use, sufficient pressure levels must be maintained to enable its effective distribution. But WDNs face many difficulties, chief among them being energy consumption and water leakage [10]. For WDNs, water leakage is a major concern, typically accounting for 25 % to 30 % of the total amount of water provided. This inefficiency contributes to financial losses, negative environmental effects, and the loss of a valuable resource [11]. Furthermore, a large amount of WDN operating costs are related to energy usage, which emphasizes the need for efficiency gains. WDNs' scope and complexity are growing along with the population and size of metropolitan regions [12]. This increase highlights how crucial it is to have the best possible design, operation, and rehabilitation procedures in place to guarantee that customers receive satisfactory services. Strategic planning and management are crucial to successfully address difficulties and improve the resilience and sustainability of water supply systems in the face of changing urban landscapes and increasing demands, given the huge extent and complexity of WDNs [13].

Water loss is an inherent difficulty in almost all Water Distribution Systems (WDS), regardless of their age or design. The kind and amount of these losses varies based on a number of factors [14]. Water loss may be regarded as inevitable to some degree, but proactive steps can be done to lessen its effects and lower related expenses. Administrators are now concentrating on creating leakage management models to achieve the greatest possible reduction in leaks at the lowest possible cost as a result of efforts to address water loss [15]. This goal has sparked a lot of research projects that try to come up with novel tactics and tools to stop water leaks [16]. The adjustable pressure in the network pipes is one important factor that greatly affects the amount of leakage in a WDS. Water leakage, electricity consumption, and the general safety and dependability of the water supply are all significantly impacted by the pressure levels in the Water Distribution Network (WDN). Operators can minimize the amount of water loss by carefully controlling and regulating the pressure inside the network, which improves the system's sustainability and efficiency [17].

In this regard, decreasing pressure alone or in conjunction with other approaches is a feasible, practical, and

cost-efficient strategy for leakage management [18]. This force reduction is required to provide the sufficient pressure to satisfy the consumers demand throughout the day and night. Using Pressure Relief Valves in the WDS is one strategy for reducing excess pressure in the network. The installation of Pressure Reducing Valves in a WDS might be viewed as an optimization challenge. However, in order to minimize leakage, optimization methods can be employed to find the optimum number, placement, and settings of the valves [19], so if the pressure is excessive, the structure suffers with several catastrophic water leaks, enhanced energy usage, or even uncontrollable cylinder rupture rates. Owing to the difficulty of enormous WDNs as well as the time-varying features of water needs, universal management is essential to accomplish pressure management. The statistical properties of the research is given in Tab. 1.

Table 1 Statistical properties

Statistical properies	Value
Skewness	0.32
Coefficient of variation	0.15
Confidence Interval	[0.28, 0.36]
Distribution Characteristics	Normal
Minimum	10.4
Maximum	25.6
Median	18.2

The objective of this investigation is to create a pressure model which minimises system leakage while taking system robustness and difficulty in meeting demand into account by estimating hourly levels of water in storage facilities for four alternative seasonal usage methods. A BSA-based optimization approach was used for this objective. The paper's important focus is the development of a water level fluctuation mechanism in the water tank to prevent leakage.

The article is arranged as : Section 2 explores the literature on pressure optimization in WDNs. Section 3 describes a Novel Neural Network Based Pressure Optimization Model. Section 4 delves into the empirical results, comparisons, and analyses. Section 5 brings the paper to a close.

2 LITERATURE SURVEY

Much research on the pressure optimization of WDNs has been published. Several proposed solutions, such as the two-phase system in [20], are founded on the design and improvement of devices that lower pressure. Since adding pressure-lowering valves to existing WDNs is challenging, more effort is being focused on improving the control of existing WDN elements in order to maximize pressure.

Jafari-Asl et al. [21] are investigating the best pressure control for reducing leakage in water supply networks. In order to solve these problems, an optimisation strategy centred on a recently developed method called Cultural Algorithm (CA) was introduced. However, this technology for locating pressure-reducing valves and making an effort to regulate them decreased the system's average leakage rate by 10 % in the third top eras.

Mehdi et al. [22] employed a distinct VSI index and enhanced the NPRI index to create a novel, step-by-step method for determining the ideal state and arrangement of PRVs in the WDN. The PSO algorithm was employed in this method, and the outcomes demonstrate how effective it is to change nodal pressures with reduced leakage rates by strategically placing PRVs in WDN.

An inventive optimization technique called the Whale Optimization Algorithm (WOA) is utilized to build pipe networks as cheaply as possible by Riham et al. [23]. It employs a performance indicator based on generations and evaluations and spherical search agents to handle discontinuity in pipe sizes. WOA was found to be more cost-effective when compared to other optimization approaches in the study. For faster convergence, more investigation is required.

Manolis et al. [24] suggested using a unique heuristic technique to create the best looping Water Distribution Network (WDN). The foundation of this approach is choosing the best watercourses to include in a WDN. The new method does not need the use of a penalty function or the modification of any parameters. The optimization approach is driven by two distinct subroutines that undertake search space exploration and exploitation by attempting sequentially and globally targeted decreases in network pipe widths.

Sitzenfrei et al. [25] created a unique edge betweenness centrality (EBCQ)-based design technique for the Water Distribution network study using CNA (Complex Network study). Determining how successfully CNA could be used as a stand-in for pressure assessment in the absence of hydraulic models is the aim of this optimization technique. Consequently, a detailed analysis of the different weight improvements is conducted, along with an evaluation of the impact on pressure predictions.

Salcedo et al. [26] presented a methodology for efficient WDN synthesis using disjunctive probabilistic Mixed-Integer Nonlinear Programming (MINLP) that accounts for linked uncertainty in nodal requests. This research used approaches for removing nonconvex nonlinearities in equations to eliminate unnecessary complexity. The impact of various covariance matrices is investigated in order to understand how the system handles uncertainty. A case study was created in order to evaluate the concept and development suggestions. The results show that when uncertainty is present, the overall stochastic solution of WDN outperforms the predictable one, suggesting that ignoring uncertainty in the optimization problem could lead to a suboptimal or, in the worst case, unrealistic WDN design.

Diego et al. [27] suggested a strategy to produce near-optimal Pareto fronts (PFs) by combining domain information from energy-based methodologies with a genetic computation to increase convergence rate and lower overall processing needs. This method is broken down into three steps: pre-processing the data from the Optimal Power Usage Surface (OPUS), parametric calibration, and regular feedback in NSGA II utilizing OPUS. This method was tested in four benchmark systems with different features in order to reduce WDN costs while increasing reliability. The

results showed that the feedback system increases the effectiveness of the system, particularly during the first period of use.

Cassiolato et al. [28] proposed a precise mathematical programming method to reduce the cost of looping WDN while accounting for a discrete range of commercial sizes and the pipe lengths that are provided.

Mehzad et al. [29] optimized pump scheduling by taking three objective characteristics into account. Water velocity in the pipes and nodal pressures determine hydraulic reliability. On the other hand, we designed a three-objective optimization approach named Clustered Non-dominated Archiving Ant Colony Optimization (Clustered-NA-ACO) and assessed its efficacy using DTLZ test functions because the quality dependability at consumption nodes depends on water age.

Zhang et al. [30] offered a water supply strategy that combines the regulatory and financial cost functions via a weighting function to solve the procedure optimization problem. It also linked the price of threats with the controllable pressure based on customer dispersion. Conditions that are right can guarantee closed-loop stability. These approaches work well at different risk levels, which is essential for WDN operation.

However, in various current works, there will be limits such as it requires more cost even if it decreases the cost, they did not concentrate on other characteristics such as leakage, energy. To take into account all of the constraints, a new pressure control model must be constructed.

3 NEURAL NETWORK-BASED PRESSURE OPTIMIZATION MODEL

The operational costs of water distribution networks (WDNs) are directly impacted by variations in consumer demand. It is crucial to make the best use of available water resources in order to reduce these expenses. In this context, dynamic operational optimization strategies—including efficient pressure management—are essential. Nevertheless, despite its significance, there is a conspicuous lack of study on pressure management strategies designed to deal with water scarcity problems. After identifying this gap, it is necessary to formulate appropriate optimization objectives and control strategies. To ensure effective WDN operation in the face of water scarcity concerns, these factors must be addressed. To address this requirement, a novel neural network-based pressure optimization model for WDN water pressure regulation has been developed. The objective of this model is to optimize pressure management in order to improve operational efficiency and lessen the negative impact of water scarcity on WDNs. To begin, a Backtracking Search Optimization (BSO)-based approach is utilised to estimate the seasonal changes in a holding tank in order to prevent leaks. In the proposed study, the outcomes of a water distribution simulation study are consumed to train a CNN system. Lastly, the CNN model's output as a water pressure function is connected with BSO-based optimisation problem to adjust water pressure and leakage at every node of the water distribution network depending on storage tank water level, water usage, and node elevation. Fig. 1 depicts the proposed work's architecture.

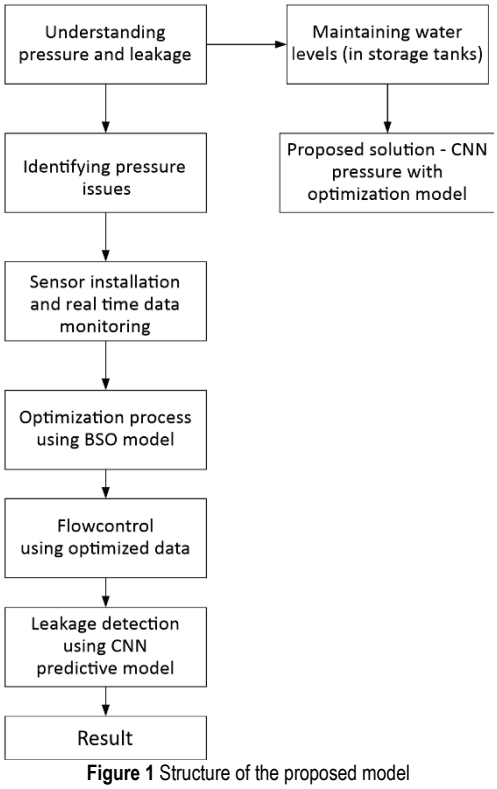


Figure 1 Structure of the proposed model

3.1 Backtracing Search Optimization BSO Algorithm based Pressure Management

Pipes, pumps, tanks, and valves comprise a system for distributing water. Each of these elements fail, but because to the great amount of pipes in water supply system, pipe leakage is far more common. Because leakage is strongly related to pressure, a pressure control system might be devised to reduce leakage. Consequently, in this work, the Backtracking optimization approach is utilised to calculate the ideal water value levels in the holding tank. BSAs are adaptive approaches that can be successfully applied to optimization issues. BSAs are most commonly used to solve issues with a significant level of complexity, non-linear behaviour, with a great amount of decision variables. The existing BSA models are modified and made compatible for our application of WDS. In Fig. 2 the water distribution system is organized as follows.

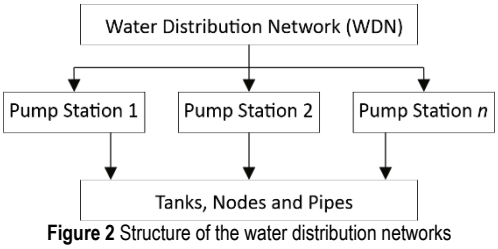


Figure 2 Structure of the water distribution networks

It is required to create a framework that could mirror the interaction among the controllable variable in the actuator as well as the controlled variable. Every constituent's pressure-flow relationship is examined. For tanks, the mass balance

interaction connecting the storage capacity *f* in the j^{th} tank is stated as

$$f_j(d_t+1)= f_j(d_t)+\Delta d_t\left(\sum_z b_{z,j}^{in}(d_t) - \sum_m b_{j,m}^{out}(d_t)\right) \quad (1)$$

where Δd_t signifies the sampling time and d represents the discrete-time instant. The inflow from the z^{th} element to the j^{th} tank is denoted by $b_{z,j}^{in}$, while the outflow from the j^{th} tank to the m^{th} element is denoted by $b_{j,m}^{out}$. The pressure supplied by the j^{th} tank is expressed in a_j, that incorporates the j^{th} tank's elevation and water level. The pressure a_j is then calculated as follows:

$$a_j(d) = \frac{f_j(d_t)}{C_j} + U_j \quad (2)$$

where C_j is j^{th} tank's cross-sectional area and U_j is the j^{th} tank's elevation. Flow is denoted as *b*, whereas pressure is denoted as *a* in the following equations. The definition rules for *b* and *a* subscripts are the same as for $b_{z,j}^{in}(d)$, $a_j(d)$. The formulation for mass conservation in the l^{th} node for nodes is

$$\sum_z b_{z,l}^{in}(d) = \sum_m b_{l,m}^{out}(d) \quad (3)$$

The inflow might be controlled by pump stations or uncontrolled by tanks or other networks. The outflow is not ever tampered with. The Hazen-Williams equation is used to define the pressure-flow relationship in pipes. A pipe connection could be categorized as follows:

$$a_z(d) - a_m(d) = \frac{10.67 N_{z,m}}{S_{z,m}^{1.852} T_{z,m}^{4.87}} b_{z,m}\left|b_{z,m}\right|^{0.852} \quad (4)$$

where $N_{z,m}$, $T_{z,m}$, and $S_{z,m}$ are the length, diameter, then roughness coefficient of the pipe, separately. Pump stations generate network pressure; flows thru the pump stations are also controlled variables. Water usage by consumers in a certain geographical region is represented by the water usage sectors. In this research, short-term consumption forecasts are used, which can be derived using appropriate forecasting techniques.

In the WDN concept, tank pressures constitute the state variable vector $\boldsymbol{g}$, while flow-through pump units create the control input vector $\boldsymbol{e}$. The non-controlled input vector $\boldsymbol{k}$ comprises all non-manipulated flow via the pipes, the pressures at nodes linking to the water usage sectors create the controlled output sequence *h*, while all additional pressures make up the non-controlled output sequence *i*. The water-consuming vector $\boldsymbol{t}$ is formed by the spent water from all water consumption sectors. e_1 reflects movement of pump stations linked to the tank, while e_2 represents the movement of everyone else. The system's cumulative weight balance is expressed as shown on Fig. 3.

The dynamic model (tank model) could be stated using Eqs. (1), (2), and (5) as where *P*, Q_1 and Q_2 are system matrices of appropriate size. Eqs. (2), (4), and (5) could be used to describe the model's dynamic component. Ehere *v*(.) and *x*(.) are nonlinear functions.

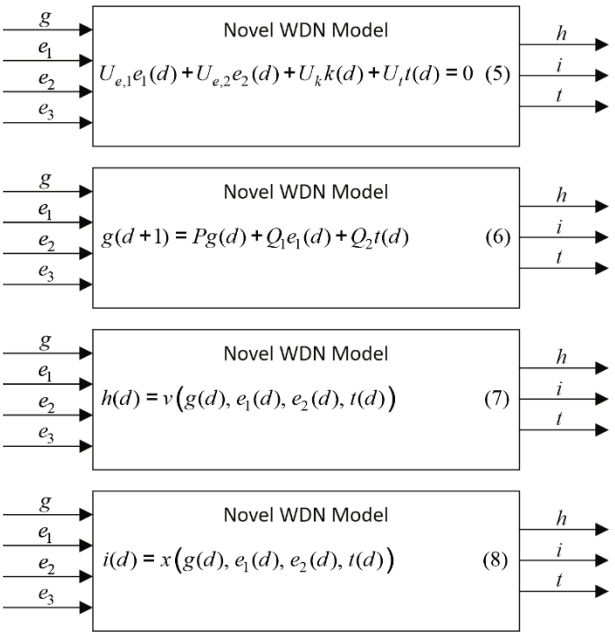


Figure 3 Novel WDN model

Overall, the WDN model Fig. 3 could be expressed as,

$$\boldsymbol{g}(d+1) = P\boldsymbol{g}(d) + Q_1 e_1(d) + Q_2 \boldsymbol{t}(d) \quad (9)$$
$$h(d) = v\left(\boldsymbol{g}(d), e_1(d), e_2(d), \boldsymbol{t}(d)\right) \quad (10)$$
$$i(d) = x\left(\boldsymbol{g}(d), e_1(d), e_2(d), \boldsymbol{t}(d)\right) \quad (11)$$
$$U_{e,1}e_1(d) + U_{e,2}e_2(d) + U_k\boldsymbol{k}(d) + U_t\boldsymbol{t}(d) = 0 \quad (12)$$

An optimization model is presented in the paper to decrease water distribution system leakage using a pressure control approach whereas.

3.2 Population based BSO

BSO is a population-based evolutionary technique employed in order to tackle numerical optimization issues with true values. The proposed model's objective function is given in Eq. (13)

$$\text{Minimize } Z = J\sum_{z=1}^{ls}\sum_{m=1}^{lh}\sum_{k=1}^{ln} S_k[A_{zmk}]^{l_k} \quad (13)$$

Where, *Z* is to minimize the leakage a pressure management scheme could be implemented and A_{zmk} is pressure of k^{th} node at *zm* instance. Where as, *zm* instance is refered as mth hour of z^{th} session, *J* is a penalty coefficient equal to 10^4 is used when the limitations on indices are violated. Otherwise, it is equivalent to one, S_k kth node's fixed leakage–pressure relation coefficient, ls number of seasons considered, lh number of hours considered in the performance period of analysis, ln total number of water distribution network nodes, *S* the number of water distribution network storage tanks, *z* the index of considered seasons ranges from 1 to ls, m in the performance period of

analysis, the index of hours ranges from 1 to lh, k index of nodes ranging from 1 to ln. The hydraulic simulation framework is employed to determine nodal pressure as a function of every storage tank's water table, water demand, and node height. Higher as well as lower water elevation limitations in every holding tank,at every nodal pressures, daily and session wise pumping capacity and resilience, but also failure indices constrain the procedure. k^{th} node pressure A_{zmk} is specified by Eq. (14).

$$A_{zmk} = f(U_k, b^*_{zmk}, UR_{1zm}, UR_{2zm}, \dots\dots, UR_{Dzm})\ \forall z, \forall m, \forall k \quad (14)$$

Where z is the index of considered seasons ranges from 1 to l s. m is In the performance period of analysis, the index of hours ranges from 1 to lh. k is index of nodes ranging from 1 to ln. U_k is elevation of the k^{th} node (m), b^*_{zmk} is water demand of the k^{th} node at the m^{th} hour of the z^{th} season (l/s), UR_{zm} is at the m^{th} hour of the season, the head of j^{th} tank (m). Commulative water level of all D water tanks is modeled in Eq. (15).

$$\sum_{j=1}^{D} C_{jzm+1} = \sum_{j=1}^{D} C_{jzm} + \sum_{l=1}^{L} CA_{lzm} - B_{zm}\ \forall z, \forall m \quad (15)$$

Where C_{jzm+1}water volume in the subsequent j^{th} tank at mth hour of z^{th} season (m^3), C_{jzm} is level of water in j^{th} tank at m^{th} hour of z^{th} season (m^3), D is Number of storage tank in the water distribution network, l is index of pumps, j is index of storage tank, B_{zm} is total water demand (m^3/h) at the m^{th} hour of the z^{th} season, L is number of pumps in the water distribution network, CA_{lzm} is water discharged by l^{th} pump at m^{th} hour of z^{th} season (m^3), z is the index of considered seasons ranges from 1 to ls. m is In the performance period of analysis. Eq. (15) gives timely continuity equation for saved volume of water in the water tank during every instance. Eqs. (16) and (18) explain the network's resilience and failure indices, which are used to provide a good smallest degree of WDN for all type of variation in water level. Resilience is described as the ability to handle including an incident or a breakdown in equipment. Due to the severity of rapid head loss, it's necessary to have more energy in every node than what is needed. As a consequence, once the system has failed, this extra power is used to compensate for greater head losses. The network's capacity for dealing with an emergency improves as its resilience index improves. In water supply structure, technical fault refers to a deviation from the system's lowest needed pressure. In this case, due to a lack of pressure, the complete required requirements were met, and the system is unable to sustain a minimum goal level. WDN Resiliance indices at zm instance is calculated as follows.

$$Ir_{zm} = \frac{\sum_{k=1}^{ln} b^*_{zmk}(A_{zmk} - A_{min})}{\sum_{j=1}^{D} BR_{jzm} UR_{jzm} + \sum_{l=1}^{L} (POW_{lzm}/\gamma) - b^*_{zmk} A_{min}}\ \forall z, \forall m \quad (16)$$

Where Ir_{zm} is index of resiliency of the water distribution network at the m^{th} hour of the z^{th} season, b^*_{zmk}, is water demand of the k^{th} node at the mth hour of the z^{th} season (l/s) A_{zmk} is k^{th} node pressure at the m^{th} hour of the z^{th} season (m), A_{min} is every node must have a minimal amount of pressure (m), BR_{jzm} is discharge from the j^{th} tank at the m^{th} hour of the z^{th} season (l/s), UR_{jzm} head of j^{th} tank at m^{th} hour of z^{th} season (m), POW_{lzm} is power input to the network by pump l at the m^{th} hour of the z^{th} season (W), γ is specific weight of water, z is the index of considered seasons ranges from 1 to ls. m is In the performance period of analysis, to estimate the failure index for the overall infrastructure, first use Eq. (17) to estimate the failure intensity at every node, then employ Eq. (18) to calculate the failure index for the entire structure. WDN hydraulic failure index If_{zmk} at zm instance for k^{th} node is given as Eq. (17).

$$If_{zmk} = \begin{cases} 0\ \forall k: A_{zmk} \geq A_{\min} \text{ and} \\ b^*_{zmk}(A_{\min} - A_{zmk})\ \forall k: A_{zmk} < A_{\min} \text{ and} \end{cases} \forall z, \forall m, \forall k \quad (17)$$

Where If_{zmk} is k^{th} node failure index during the m^{th} hour of z^{th} season, A_{zmk} pressure of the k^{th} node at the m^{th} hour of the z^{th} season (m), A_{min} is every node must have a minimal amount of pressure (m), b^*_{zmk} is water demand of the k^{th} node at the m^{th} hour of the z^{th} season (l/s)), z is the index of considered seasons ranges from 1 to ls. m is In the performance period of analysis k is index of nodes ranging from 1 to ln

$$If_{zm} = \frac{\sum_{k=1}^{ln} If_{zmk}}{\sum_{k=1}^{ln} b^*_{zmk} A_{min}} \forall z, \forall m \quad (18)$$

Where If_{zm} is hydraulic failure index of the system at the m^{th} hour of the z^{th} season, k is index of nodes ranging from 1 to ln, ln is total number of water distribution network nodes, If_{zmk} is k^{th} node failure index during the m^{th} hour of z^{th} season, A_{min} is every node must have a minimal amount of pressure (m), b^*_{zmk} is water demand of the k^{th} node at the m^{th} hour of the z^{th} season (l/s)), z is the index of considered seasons ranges from 1 to ls. m is in the performance period of analysis.

$$A_{\min} < A_{zmk} < A_{\max} < A_{j\,\max}\ \forall z, \forall m, \forall k \quad (19)$$

Where A_{zmk} pressure of the k^{th} node at the m^{th} hour of the z^{th} season (m), $A_{\min}$ is Every node must have a minimal amount of pressure (m), $A_{\max}$ maximum suggested pressure at each node (m), $A_{j\,\max}$ is maximum permissible pressure at each node (m), z is the index of considered seasons ranges from 1 to ls. m is In the performance period of analysis k is index of nodes ranging from 1 to ln. The following various condition are applicable for optimization and given in Eq. (20) to the Eq. (24).

$$UR_{j\,\min} < UR_{jzm} < UR_{j\,\max}\ \forall z, \forall m, \forall j \quad (20)$$
$$CA_{lzm} < CA_{l\,\max}\ \forall z, \forall m, \forall l \quad (21)$$
$$\sum_{l=1}^{L} \sum_{m=1}^{lh} CA_{lzm} < CA_{day}\ \forall z \quad (22)$$
$$Ir_{zm} > Ir_{\min}\ \forall z, \forall m \quad (23)$$
$$If_{zm} < If_{\max}\ \forall z, \forall m \quad (24)$$

Where, the remaining variables are $UR_{j\,\min}$ minimum tank water level j (m), $UR_{j\,\max}$ is maximum tank water level j (m), UR_{jzm} head of j^{th} tank at m^{th} hour of z^{th} season (m) , $CA_{l\,max}$ maximum hourly capacity of the pump l. (m^3), CA_{day} is maximum amount of daily pumping volume required (m^3), $Ir_{\min}$ minimal resilience index for water distribution networks, $If_{\max}$ is maximum allowable failure index for a water distribution network, z is the index of considered seasons ranges from 1 to ls , m is in the performance period of analysis k is index of nodes ranging from 1 to ln, l is index of pumps,L is number of pumps in the water distribution network, lh number of hours considered in the performance period of analysis

3.2.1 Pressure Management using BSO

BSO maintains a population of I and T are water consuming vectors for this purpose. Three major genetic algorithms are used to establish BSO experimental populations (selection, mutation, and crossover). It has a random mutations mechanism which uses a 1-direction distinct for every target individual, as well as a memory that keeps a population from a arbitrarily selected earlier generations to use when setting the search-direction matrix. BSO employs five evolutionary techniques like: initialization, 1st selection, mutation, crossover, and 2nd selection. Fig. 4 depicts the overall flowchart of BSO.

Objective Function Evaluation: BSO starts by randomising nodes inside their numerical range with a uniform random distribution function:Generalized pressure at a tank is given by Eq. (25)

$$A_{z,m} \sim U(low_m, up_m), z = 1,2, \dots \dots \dots I, m = 1,2, \dots \dots T \quad (25)$$

where I indicates the seasons and T represents the storage tank in the WDN. The uniform distribution of water level is denoted by U. $A_{z,m}$ denotes the z^{th} tank individual's pressure of the mth node. low_m, and up_m are the pressure at lower and upper boundaries, accordingly.

The dependent variable constraints are combined into objective function under consideration via quadratic penalty terms as shown in Eq. (26). As a consequence, infeasible solutions that break the limitations are unlikely to be handed onto future generations. As a consequence, the objective function f can be generalised and written in Eq. (26).

$$f = F + \lambda_f \sum_{Llim_f} \Delta f_{\text{load}}^2 + \lambda_B \sum_{L\,lim_b} \Delta b_x^2 + \lambda_{ac} \Delta a_c^2 + \lambda_{cv} \sum_{Llim_{cv}} \Delta c_v^2 \quad (26)$$

Where, F is a control factor of mutation operator; λ_f, λ_B, λ_{ac}, and λ_{cv} are the penalty factors , $Llim_f$ is the water level which has exceeded their limitations. $Llim_B$ denotes the storage tank whose water levels are beyond the limits, $Llim_{cv}$ denotes the set of overflow lines, and Δf_{load}, Δb_x, Δa_c, and Δc_v are denotes water flow, reactive water flow, active water flow, and apparent water flow which is defined in following Eq. (27) to Eq. (30).

$$\Delta f_{\text{load}} = \begin{cases} f_{\text{load}}^{\min} - f_{\text{load}} & \text{if } f_{\text{load}} < f_{\text{load}}^{\min} \\ f_{\text{load}}^{\max} - f_{\text{load}} & \text{if } f_{\text{load}} > f_{\text{load}}^{\max} \end{cases} \quad (27)$$

$$\Delta b_x = \begin{cases} b_x^{\min} - b_x \text{ if } b_x < b_x^{\min} \\ b_x^{\max} - b_x \text{ if } b_x > b_x^{\max} \end{cases} \quad (28)$$

$$\Delta a_c = \begin{cases} a_c^{\min} - a_c \text{ if } a_c < a_c^{\min} \\ a_c^{\max} - a_c \text{ if } a_c > a_c^{\max} \end{cases} \quad (29)$$

$$\Delta c_v = c_v^{\max} - c_v \text{ if } c_v > c_v^{\max} \quad (30)$$

In which the superscripts 'min' and 'max' represent the variable's minimal and highest numbers.

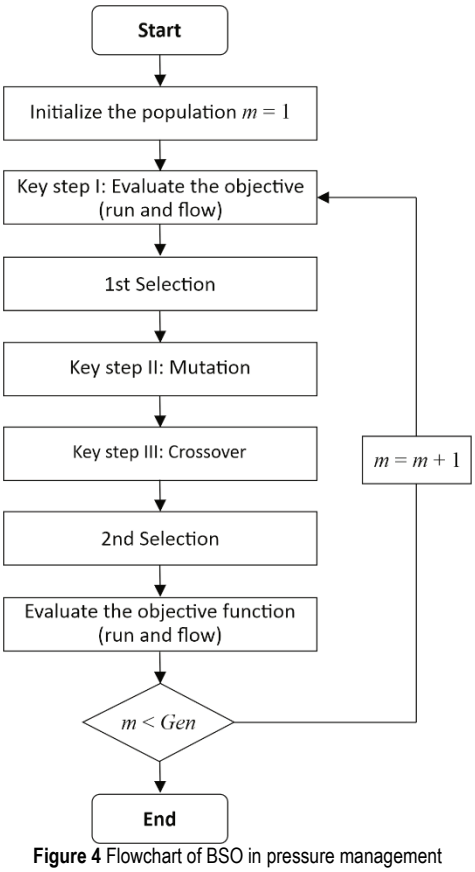


Figure 4 Flowchart of BSO in pressure management

1st Selection: At this step, BSO produces the historic demographic utilisation to determine the search direction, as shown in the formula given in Eq. (31).

$$A_{z,m}^{old} \sim U(low_m, up_m) \quad (31)$$

Where $z = 1, 2, \dots, I$, $m = 1, 2, \dots, T$, $A_{z,m}^{old}$ focuses on the historical population BSO enables consumers to rebuild the historic demographic at the start of every cycle by adhering to a rule given bellow.

$$A_{\text{old}} = \begin{cases} A & \text{if } p < q \\ A_{\text{old}} & \text{if } p \geq q \end{cases} \quad (32)$$

Where, p and $q \sim U(0,1)$ to detect if the historical population is collected from the prior generation. The shuffling algorithm is subsequently employed to reshuffle the members in the population as regards in Eq. (33) in which in

which the permuting () function is a random shuffling operator.

$$A_{\text{old}} = permutting\ (A_{\text{old}}) \tag{33}$$

Mutation: The mutation process employs the subsequent procedure to produce mutant BSO vector $\boldsymbol{V}$ at every generation and accordingly control factor F is formulated as bellow.

$$F = A + \boldsymbol{V}(A_{\text{old}} - A) \tag{34}$$

Where, V is an actual value that indicates the storing tank's level of water. By taking into consideration the values of the historical population, the BSO used previous experiences to establish the water level at the storage tank.

Algorithm 1 BSA Crossover Strategy

Input: *V*, *mix rate*, *I* and *T*
Output: *D*: Trial Population
map(1:*I*, 1:*T*) = 0
If *p* < *q* | *p*, *q* ~ *U*(0, 1) then
for *z* from 1 to *I* do
map$_{z,\ e(1:[mixrate.rnd.T])}$ = 1 | *e* = *permuting* ((1, 2, 3,…, *T*))
end
else
for *z* from 1 to *I* do, *map*$_{(z,\ randz(T))}$ = 1, *end*
end
D := Mutant
for *z* from 1 to *I* do
for *m* from 1 to *T* do
if *map*$_{(z,\ m)}$ = 1 then $D_{(z,\ m)} = A_{(z,\ m)}$
end
end

Crossover Process: Algorithm 1 depicts crossover process of BSA. This procedure is divided into two stages. The initial generates a binary integer-valued matrix called as a map directing crossover locations using a mix rate. The next permits the BSA's crossover method to be employed in order to generate the trial population's final form (Algorithm 1) by using the formula given in Eq. (35).

$$D_{z,m} = \begin{cases} A_{z,m} \text{ if } map_{z,m} = 1 \\ F_{z,m} \text{ if } map_{z,m} = 0 \end{cases} \tag{35}$$

Where $D_{z,m}$ are the trial members. In BSA's crossover process, the mix rate variable defines the amount of parameters of persons which will change in a trial population.

2nd Selection: The process is then employed in the final step to evaluate the fitness of the trial vector as well as the associated target vector and pick the parent that would live in the upcoming generations, as regards:

$$A_{z,m}^{\text{next}} = \begin{cases} D_{z,m}, if\ f(D_{z,m}) < A_{z,m} \\ A_{z,m} \text{ otherwise} \end{cases} \tag{36}$$

Where, *f*() denotes the fitness function. As a result, the population either improves or remains constant in terms of objective values. Several steps are again continued via generations, and the process ends whenever the optimum number of generations are reached or some specific stopping criteria is met.

3.2.2 The Head-Driven Simulation Method

HDSM ought to be utilized to study the network for reliable leakage estimation because leakage and pressure are inextricably linked. Moreover, the EPANET software was used to model network leakage in our research. The EPANET2.0 programme is utilized to model water network hydraulic systems. This program has various features that render it suited for hydraulic study of water distribution networks as compared to alternative methods []. The emitter could be employed in every node, is one of EPANET's effective capabilities. As demonstrated here, rate of flow of water with every emitter can be computed as a in term of pressure within a specific node is specified by Eq. (37).

$$B_k = S_k A_k^{l_k} \tag{37}$$

At node k, B_k is the rate of flow of water, A_k is the pressure, S_k is the intensity of flow, and l_k is a power term. This function is useful for simulating leaks. S_k and l_k parameters must be evaluated for this purpose. The assumption is that l_k is constant and is equal to 1.18, and S_k is determined by Eq. (38).

$$S_k = S \sum_{m=1}^{j} \frac{N_{km}}{2} \tag{38}$$

Where, N_{km} is the mth pipe at kth node, and S is a network-wide constant, as illustrated below.

$$S = \frac{Leak}{\sum_{k=1}^{ln}\left(\sum_{m=1}^{j} \frac{N_{km}}{2} \times A_k^{l_k}\right)} \tag{39}$$

Where $Leak$ denotes the amount of leaking but during weakest night flow. The pressure in the preceding equation is determined at the point of least night flow.

To calculate pipe leakage, the influence of pipe average pressure must be taken into account. This suggest changes that every half-length of a pipe possesses the identical stress as its end device, which was integrated into the leakage solution via the S calculation in Eq. (39). This technique overestimates leak during the initial half while underestimating leakage in the second half, balancing the others.

After computing the *S* value for the entire network, Eq. (38) can be applied to determine S_k for every node, that is utilized by the hydraulic simulation study as the emitters flow rate coefficient. For faster convergence, it is critical to have a valid initial guess when calculating the *S* factor. The initial estimate was calculated for this reason utilizing node pressure without accounting for leakage. The leakage is estimated after the initial event, then *S* is updated using Newton's method and relation between new S_{new} and old S_{old} is given by Eq. (40).

$$S_{\text{new}} = S_{\text{old}} - \frac{f(S_{\text{old}})}{f'(S_{\text{old}})} \quad (40)$$

where S_{new} and S_{old} represent the earlier also innovative forecasts of S, and f is a function of the distinction among expected and measured leakages. Newton's approach was utilized until the estimated leakage equals the observed leakage value. In this case, the leakage is estimated if many actual emitter coefficients are discovered at each nodes.

3.3 CNN Model for Pressure Management

To estimate pressure variations, a numerical solution must be combined with suggested leakage decrease optimization method, as indicated in Eq. (14). Hydraulic simulator techniques are increasingly being used to describe the complex and non - linear behavior of water distribution systems. Unfortunately, hydraulic simulation models had difficulties of projecting the dynamic consequences of various control settings in respect to the baseline circumstances with short-term demand fluctuations because the inflict to the computation complexity. Furthermore, When hydraulic simulation approach are described by an I/O connection, that could be mapped to use a multivariate functional in this scenario, there is bound to be a large chance for boosting computational performance. The proposed method employs the outcomes of a water supply modeling research to develop a CNN framework. The CNN architecture's output as a water pressure function is again linked with a BSO-based optimization technique to put on water pressure and leakage at every node of the distribution system for water depends on holding tank water level, water usage, and node height.

Because the input-output connection in the CNN framework is straightforward to estimate, it may be viewed as a basic approximation of input-output computations in every descriptive or optimization technique. The advantages involve decreased runtime and simple implementation, while the drawback is a loss of precision. The trade-off must be where and how the concept is used, in addition to who the end users are. The hydraulic simulation method's outputs are employed in this study in a CNN framework which may be linked to the optimisation problem. CNN algorithms have been taught to reproduce simulated findings over time. The processing order of the optimization method could be decreased by substituting the EPANET2.0 hydraulic simulation system with CNN. It is important to underline that diminishing precision. The processing challenges could be considerably decreased as a result.

The CNN algorithm has been trained utilizing EPANET 2.2 findings for various storage tank water levels and water demands within their variation limits. The strategy can be chosen after test different multilayer feed-forward CNNs.

4 RESULT AND DISCUSSION

This section addresses the overall performance of our proposed scheme. Moreover, previous work comparison findings are compaired.

Tool: EPANET 2.2 AND PYTHON
OS: Windows 7 (64 bit)
Processor: Intel Premium
RAM: 8 GB RAM

Table 2 Network table - nodes at 1:00 hrs

Node ID	Elevation (m)	Base demand (MLD)	Pressure (m)
Junc j1	15	0.005	17.54
Junc j2	15	0.005	17.54
Junc j3	15	0.005	17.54
Junc j4	15	0.005	17.54
Junc j5	15	0.005	17.54
Junc j6	15	0.005	16.83
Junc j8	15	0.005	16.83
Junc j9	15	0.005	17.54
Junc j10	15	0.005	17.55
Junc j11	15	0.005	17.55
Junc j12	15	0.005	17.55
Junc j13	15	0.005	17.55
Junc j14	15	0.005	17.55
Junc j15	15	0.005	17.56
Junc j16	15	0.005	17.56
Junc j17	15	0.005	17.56
Junc j18	15	0.005	17.56
Junc j19	15	0.005	17.56
Junc j20	15	0.005	17.56
Junc j21	15	0.005	17.56
Junc j22	15	0.005	17.58
Junc j23	15	0.005	17.60
Junc j24	15	0.005	17.61
Junc j25	15	0.005	17.57
Junc j26	15	0.005	17.56
Junc j27	15	0.005	17.55
Junc j28	15	0.005	17.54
Junc j29	15	0.005	17.58
Junc j30	15	0.005	17.58
Junc j31	15	0.005	17.64
Junc j32	15	0.005	17.64
Junc j33	15	0.005	17.70
Junc j34	15	0.005	17.70
Junc j35	15	0.005	17.70
Junc j36	15.5	0.005	17.22
Junc j37	15.5	0.005	17.22
Junc j38	15.5	0.005	17.21
Junc j39	15.5	0.007	17.33
Junc j40	15.5	0.007	17.65
Junc j41	15.5	0.007	17.65
Junc j42	15.5	0.007	17.67
Junc j43	15.5	0.007	17.87
Junc j44	15.5	0.007	18.31
Junc j45	15	0.007	18.81
Junc j46	15.5	0.007	18.31
Junc j47	15.5	0.007	18.31
Junc j48	15.5	0.007	18.31
Junc j49	15.5	0.007	18.31
Junc j50	15.5	0.007	18.31
Junc j51	15.5	0.007	18.31
Junc j52	15.5	0.007	18.31
Junc j53	15.5	0.007	18.31
Junc j54	15.5	0.007	18.31
Junc j55	8	0.013	24.71
Junc j56	8	0.013	24.71
Junc j57	8	0.013	24.71
Junc j58	8	0.013	24.71
Junc j59	8	0.013	24.70

4.1 Dataset Description

The details of water supply network from City and Industrial development Corporation of Maharastra limited

(CIDCO)is collected. This CIDCO contains details of drawings of Water Supply Network and Elevated Storage Reservoir (ESR) / High Storage Reservoir (HSR) and Master Balance Reservoir (MBR) of Kharghar Node. This water distribution system in Kharaghar requires effective water management for both supply and demand.Overall, the system's fundamental issues identified is leakage.

It is proposed that Pressure management is the easiest, shortest, but possibly least cheap approach for reducing leakage in kharghar water distribution systems. The drop in network pressure leads to a reduction in leakage rate. Added pressure beyond the absolute lowest value promotes systems resiliency, while pressure deficiency (underneath the lowest permissible value) contributes to request shortfall, in which customers cannot be fully provided. As a result, the optimum combination of numerous decision variables could be found to use a search strategy including BSA to calculate the appropriate level of pressure. As per the CIDCO rule, 650 litres of water have to be supplied per flat per day.

The overall amount of junction in the network is 840, number of reservoir is 1, number of tanks is 9, number of pipe is 831 of various diameter such as 100 mm, 150 mm, 200 mm, 250 mm, 300 mm, 350 mm, 400 mm, 450 mm, and above 500 mm. The network has a total number of PRV valve as 130 and roughness coefficient as 110. Moreover, the pressure value is taken for per hour, as a result, we have taken the 6 hour pressure value. The collected 1 hour pressure value data sample is tabulated as Tab. 2.

Then, the obtained EPANET 2.2 values are fed into the proposed Neural Network based Optimization model. The collected data's are processed through the PYTHON to find out the water pressure and leakage.The maximum pressure value as 327.68 is set, if the pressure value is exceed this limit it will be considered as a leakage. Lastly, the proposed method is contrasted with the current framework to demonstrate its effectiveness. The following is a Tab. 3 of the proposed CNN's structure.

Table 3 Architecture of the proposed CNN

Layer Type	Specifications
Convolutional layer 1	Kernel Size: 5×5; Filter: 64; Stride: 3×3; Padding: Same
Max pooling layer 1	Pool Size: 3×3; Stride: 3×3 Padding: Same
Convolutional layer 2	Kernel Size: 5×5; Filter: 128; Stride: 3×3; Padding: Same
Max pooling layer 2	Pool Size: 3×3; Stride: 3×3 Padding: Same
Convolutional layer 3	Kernel Size: 5×5; Filter: 256 Stride: 3×3; Padding: Same
Max pooling layer 3	Pool Size: 3×3; Stride: 3×3 Padding: Same
Fully Connected Layer	Size: 64
Output Layer	Size: 1

4.2 Performance of Proposed Model

This section assesses the effectiveness of the proposed system, where *Precision*, *Accuracy*, *Recall*, *F1 Score*, and loss metrics were used to assess the proposed prototype.

4.2.1 Loss vs Epochs Graph

The training loss of a proposed Neural Network Based Pressure Optimization Model, which measures the difference between the model's predictions and the actual target values in the training dataset, is used to assess the model's performance. More precise forecasts and efficient identification of underlying patterns in the training data are indicated by a smaller training loss. As the model goes through multiple epochs of iterative optimization, the training loss trend steadily decreases. In just 5 epochs, the model reached a training loss of less than 0.0682, demonstrating a rapid convergence to remarkably accurate predictions on the training set. Low training loss over a short period of epochs indicates successful data collection and learning.

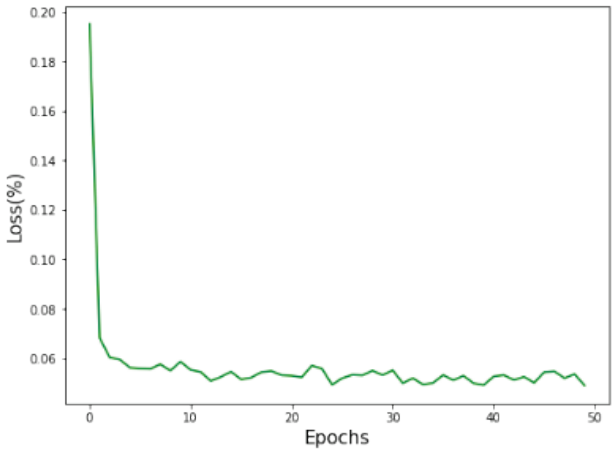


Figure 5 Loss vs Epochs Graph proposed approach

4.2.2 Performance Parameters

This chapter provides an overview of our proposed methods, in which several metrics including *Accuracy*, *F1 Score*, *Precision*, and *Recall* were used to evaluate the unique method's effectiveness in water distribution system pressure optimization. Various performance indicators, includes *Accuracy*, *Precision*, *Recall*, and *F1 Score*, were evaluated using the Eqs. (41)-(44).

$$Accuracy = \frac{True\ Positive + True\ Negative}{True\ Positive + True\ Negative + False\ Positive + False\ Negative} \quad (41)$$

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive} \quad (42)$$

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative} \quad (43)$$

$$F1\ Score = \frac{2*Precision*Recall}{Precision + Recall} \quad (44)$$

The novel approach of the model—which combines a Convolutional Neural Network (CNN) with the Backtracking Search Optimization algorithm—allows it to precisely identify positive cases and classify instances within the dataset with low misclassification rates. The suggested approach's excellent levels of *Accuracy*, *Precision*, *Recall*, and *F1 Score* attest to its applicability and promise for real-

world uses in a variety of fields where precise classification is critical. Overall, the results show that the suggested method is dependable and effective at handling classification jobs with a high degree of precision and accuracy.

4.3 Comparison Results

This chapter discussed the proposed method's comparative study, In this study, which the our technology is comaired with the methods like Nave Bayes, Logistic Regression, and Convolution Neural Network (CNN).

4.3.1 Accuracy Comparison

In terms of accuracy, the suggested method performs better than baseline models such as Naive Bayes, Logistic Regression, and conventional Convolutional Neural Network (CNN). With an accuracy rate of 99.80 %, the approach outperforms Logistic Regression by 98.23 %, Naive Bayes by 95.49 %, and standard CNN by 98.62 %. This improvement in performance highlights the usefulness and edge of the new strategy over traditional techniques. The results presented in Tab. 4 validate the effectiveness of the suggested methodology and show that it is more accurate than baseline models and sophisticated techniques. It also shows that it has the potential to outperform current methods in real-world applications.

Table 4 Overall Accuracy

Methods	Accuracy (%)
Naïve Bayes	95.49
Logistic Regression	98.23
Convolution Neural Network (CNN)	98.62
Proposed Method	99.80

4.3.2 Precision

A comparison of the precision levels between the suggested strategy and alternative techniques is presented in Tab. 5. The suggested strategy performs 98 % better than the BSO-based CNN method and 98.23 % better than the Naive Bayes model, demonstrating its efficacy and superiority over conventional techniques. The strategy significantly improves the accuracy and reliability of the classification task, achieving an excellent precision rate of 99.80 %. These results confirm the effectiveness of the new approach and highlight its potential to beat current approaches in real-world scenarios where great precision in classification tasks is required. The suggested method is a major improvement over traditional procedures.

Table 5 Overall Precision

Methods	Precision (%)
Naïve Bayes	98
Logistic Regression	98.23
Proposed Method	99.80

4.3.3 Recall Comparison

The recall performance of the suggested method is contrasted with other approaches in Tab. 6. The usefulness and superiority of the suggested strategy over conventional approaches are demonstrated by its 95.40 % and 99 % performance gains over the BSA-based CNN method and Logistic Regression method, respectively. The technique produces an astounding 100 % recall rate, which is a major improvement over earlier solutions and a notable improvement in the dataset's ability to detect pertinent cases. These results confirm the new strategy's effectiveness and highlight its potential to beat current approaches in real-world scenarios where good recall in classification tasks is needed.

Table 6 Overall Recall

Methods	Recall (%)
Naïve Bayes	95.40
Logistic Regression	99
Proposed Method	100

4.3.4 F1-Score Comparison

A comparison of the *F1 Scores* of the suggested strategy with other approaches is presented in Tab. 7. The suggested method shows its superiority in terms of *F1 Score*, outperforming both Logistic Regression and the BSA-based CNN method by 99.10 % and 97.65 %, respectively. With a remarkable *F1 Score* of 99.90 %, the approach significantly enhances the model's capacity to strike a balance between recall and precision. These results demonstrate the usefulness of the unique strategy and its superiority over conventional ways in producing better outcomes.

Table 7 Overall F1 Score

Methods	F1 Score (%)
Naïve Bayes	97.65
Logistic Regression	99.10
Proposed Method	99.90

5 CONCLUSION

The appropriate hour water level fluctuations in a water distribution storage tank with various seasons are explored in proposed research. To estimate the effective mechanism, the pressure that exists at various network nodes under various operating circumstances are evaluated. To acquire better precise pressure measurements at various nodes of network, the quantity of leakage quantity at each node is analysed. Because it requires awhile to integrate the hydraulic simulation and optimization algorithms, a CNN algorithm is trained to predict the system's hydraulic properties. Directly linking a hydraulic simulation with the BSO, design leads to enhance the accuracy of the output. While the optimization method is designed for a large collection of storage containers. The study's findings can be applied to other components of the structure. In this example, the optimal water volumes in over-all accessible storage facilities are utilised with CNN output in addition to BSO model decision factors. Whenever a software like this technique are utilised in a realistic situation, the network's associated to water supply could be planned better effectively by minimizing 31.5 % leakage. The findings show that there is a need of implementing a combined solution toolkit to ensure the efficient functioning of the water distribution system.

6 REFERENCES

[1] Jia, Y. H., Mei, Y. & Zhang, M. (2020, June). A memetic level-based learning swarm optimizer for large-scale water distribution network optimization. *Proceedings of the 2020 Genetic and Evolutionary Computation Conference*, 1107-1115. https://doi.org/10.1145/3377930.3389828

[2] Savic, D. A. & Walters, G. A. (1997). Evolving sustainable water networks. *Hydrological sciences journal*, *42*(4), 549-564. https://doi.org/10.1080/02626669709492053

[3] Liu, H., Shoemaker, C. A., Jiang, Y., Fu, G. & Zhang, C. (2020). Preconditioning water distribution network optimization with head loss–based design method. *Journal of Water Resources Planning and Management*, *146*(12), 04020093. https://doi.org/10.1061/(ASCE)WR.1943-5452.0001299

[4] Samani, H. M. & Mottaghi, A. (2006). Optimization of water distribution networks using integer linear programming. *Journal of hydraulic engineering*, *132*(5), 501-509. https://doi.org/10.1061/(ASCE)0733-9429(2006)132:5(501)

[5] Yazdani, A., & Jeffrey, P. (2011). Complex network analysis of water distribution systems. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, *21*(1). https://doi.org/10.1063/1.3540339

[6] Brdys, M. A. & Ulanicki, B. (1994). Operational control of water systems: structures, algorithms, and applications. *Environment International, 21*(3), 347. https://doi.org/10.1016/0160-4120(95)90073-X

[7] Giudicianni, C., Di Nardo, A., Di Natale, M., Greco, R., Santonastaso, G. F. & Scala, A. (2018). Topological taxonomy of water distribution networks. *Water*, *10*(4), 444. https://doi.org/10.3390/w10040444

[8] Zhao, W., Beach, T. H. & Rezgui, Y. (2015). Optimization of potable water distribution and wastewater collection networks: A systematic review and future research directions. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, *46*(5), 659-681. https://doi.org/10.1109/TSMC.2015.2461188

[9] Araujo, L. S., Ramos, H. & Coelho, S. T. (2006). Pressure control for leakage minimisation in water distribution systems management. *Water resources management*, *20*, 133-149. https://doi.org/10.1007/s11269-006-4635-3

[10] Jia, Y. H., Mei, Y. & Zhang, M. (2021). A two-stage swarm optimizer with local search for water distribution network optimization. *IEEE Transactions on Cybernetics*.

[11] Ulanicki, B., AbdelMeguid, H., Bounds, P. & Patel, R. (2008). Pressure control in district metering areas with boundary and internal pressure reducing valves. *Water Distribution Systems Analysis 2008*, 1-13. https://doi.org/10.1061/41024(340)58

[12] Mahdavi, M. M., Hosseini, K., Behzadian, K., Ardehsir, A. & Jalilsani, F. (2010). Leakage control in water distribution networks by using optimal pressure management: A case study. *Water Distribution Systems Analysis 2010*, 1110-1123. https://doi.org/10.1061/41203(425)101

[13] Lambert, A. (2001). What do we know about pressure-leakage relationships in distribution systems. *Conference: Proc. IWA System Approach to Leakage Control and Water Distribution Systems Management*.

[14] Pérez, R., Puig, V., Pascual, J., Quevedo, J., Landeros, E. & Peralta, A. (2011). Methodology for leakage isolation using pressure sensitivity analysis in water distribution networks. *Control Engineering Practice*, *19*(10), 1157-1167. https://doi.org/10.1016/j.conengprac.2011.06.004

[15] Kurland, N. B. (2011). Evolution of a campus sustainability network: a case study in organizational change. International *Journal of Sustainability in Higher Education, 12*(4), 395-429. https://doi.org/10.1108/14676371111168304

[16] Tabesh, M. & Homehr, S. (2006). Leakage management in water networks using regulation optimization of PRV faucets genetic algorithm. *Water Resource Management Conference of Isfahan Industrial University*, Society of Iran Water Resource Science and Engineering.

[17] García-Ávila, F., Zhindón-Arévalo, C., Valdiviezo-Gonzales, L., Cadme-Galabay, M., Gutiérrez-Ortega, H. & del Pino, L. F. (2022). A comparative study of water quality using two quality indices and a risk index in a drinking water distribution network. *Environmental Technology Reviews, 11*(1), 49-61. https://doi.org/10.1080/21622515.2021.2013955

[18] Jahandideh-Tehrani, M., Bozorg-Haddad, O. & Loáiciga, H. A. (2020). Application of particle swarm optimization to water management: an introduction and overview. *Environmental Monitoring and Assessment, 192*, 1-18. https://doi.org/10.1007/s10661-020-8228-z

[19] Liu, W. & Song, Z. (2020). Review of studies on the resilience of urban critical infrastructure networks. *Reliability Engineering & System Safety, 193*, 106617. https://doi.org/10.1016/j.ress.2019.106617

[20] Mosetlhe, T. C., Hamam, Y., Du, S. & Monacelli, E. (2021). Appraising the impact of pressure control on leakage flow in water distribution networks. *Water, 13*(19), 2617. https://doi.org/10.3390/w13192617

[21] Jafari-Asl, J., Sami Kashkooli, B. & Bahrami, M. (2020). Using particle swarm optimization algorithm to optimally locating and controlling of pressure reducing valves for leakage minimization in water distribution systems. *Sustainable Water Resources Management*, *6*, 1-11. https://doi.org/10.1007/s40899-020-00426-3

[22] Ezzeldin, R. M. & Djebedjian, B. (2020). Optimal design of water distribution networks using whale optimization algorithm. *Urban Water Journal*, *17*(1), 14-22. https://doi.org/10.1080/1573062X.2020.1734635

[23] Manolis, A., Sidiropoulos, E. & Evangelides, C. (2021). Targeted path search algorithm for optimization of water distribution networks. *Urban Water Journal*, *18*(3), 195-207. https://doi.org/10.1080/1573062X.2021.1877739

[24] Sitzenfrei, R., Wang, Q., Kapelan, Z. & Savić, D. (2020). Using complex network analysis for optimization of water distribution networks. *Water resources research*, *56*(8), e2020WR027929. https://doi.org/10.1029/2020WR027929

[25] Salcedo-Díaz, R., Ruiz-Femenia, R., Caballero, J. A. & Ravagnani, M. A. (2020). Water Distribution Network Optimization Considering Uncertainties in the Nodes Demands. *Computer Aided Chemical Engineering*, 48, 1183-1188. https://doi.org/10.1016/B978-0-12-823377-1.50198-1

[26] Páez, D., Salcedo, C., Garzón, A., González, M. A. & Saldarriaga, J. (2020). Use of energy-based domain knowledge as feedback to evolutionary algorithms for the optimization of water distribution networks. *Water*, *12*(11), 3101. https://doi.org/10.3390/w12113101

[27] Cassiolato, G., Carvalho, E. P., Caballero, J. A. & Ravagnani, M. A. (2021). Optimization of water distribution networks using a deterministic approach. *Engineering Optimization*, *53*(1), 107-124. https://doi.org/10.1080/0305215X.2019.1702980

[28] Mehzad, N., Asghari, K. & Chamani, M. R. (2020). Application of clustered-NA-ACO in three-objective optimization of water distribution networks. *Urban Water Journal*, *17*(1), 1-13. https://doi.org/10.1080/1573062X.2020.1734633

[29] Zhang, Y., Zheng, Y., Li, S. & Zou, Y. (2021). Economic model predictive control for the operation optimization of water distribution networks with risks. *Asian Journal of Control*, *23*(1), 128-142. https://doi.org/10.1002/asjc.2218

[30] Zhang, Y., Li, S., Zheng, Y. & Zou, Y. (2020). Multi-model based pressure optimization for large-scale water distribution networks. *Control Engineering Practice*, *95*, 104232. https://doi.org/10.1016/j.conengprac.2019.104232

Authors' contacts:

Waghmare Shwetambari Pandurang, Research Scholar
(Corresponding author)
Department of Mathematics, Sandip University,
Trimbak Road, Nashik 422213, Maharashtra, India
E-mail: shwetambari.deore@bvcoenm.edu.in

Renu Praveen Pathak, Associate Professor Dr. and Head
Department of Mathematics, Sandip University,
Trimbak Road, Nashik 422213, Maharashtra, India

Imtiyaz Ahmad Wani, Assitant Professor Dr.
Malla Reddy University,
Maisammaguda, Dulapally, Hyderabad, Telangana 500043, India

ISSN 1846-6168 (Print), ISSN 1848-5588 (Online)
https://doi.org/10.31803/tg-20240220084956

Original scientific paper
Received: 2024-02-20, Accepted: 2024-04-30

Effects of High Oxygen Concentration on Driving Performance and Cerebral Hemodynamic Response

Kyu-Beom Kim, Soon-Cheol Chung, Byung-Chan Min*

Abstract: The positive effect of a high oxygen concentration on human cognitive ability for single tasks has been proven in various ways; however, few studies have investigated its effect on complex cognitive tasks. The aim of this study was to investigate the effect of a high oxygen concentration in driving situations that require complex cognitive abilities. The experiment was conducted in a laboratory equipped with a car simulator with 10 volunteer subjects. Driving was performed under normal and high-oxygen-concentration conditions. Various driving parameters and oxyhaemoglobin (HbO) levels in certain brain areas were measured during the experiments. The results show that vehicle velocity was maintained more consistently in the high-oxygen-concentration condition, and the HbO concentration increased (indicating greater activation) in the prefrontal and parietal cortices.

Keywords: cerebral hemodynamic response; driving performance; fNIRS; high oxygen concentration; oxyhaemoglobin

1 INTRODUCTION

Oxygen is present in the atmosphere at a concentration of approximately 21 %, and it is consumed in various biological processes essential to the survival of plants and animals. The oxygen concentration in human blood is maintained above 95 %, enabling energy production within cells and enabling organs and tissues to perform their functions [1]. In particular, the brain consumes a large amount of oxygen (approximately 20 % of the human body's oxygen consumption) because of the dense concentration of nerve cells [2, 3]. It is also the most sensitive organ to oxygen deficiency [4]. Irreversible brain damage occurs when the brain does not receive sufficient oxygen, leading to a decline in high-level cognitive functions, including degradation of processing speed, memory, and executive function [5].

Because of the essential role of oxygen in the brain, research is being conducted on improving cognitive processing ability by artificially oversupplying oxygen. Some studies show that high oxygen concentration has a positive effect on cognitive processing ability [6-9]. Chung et al. [7] reported that visuospatial performance improved and blood oxygen saturation increased when 30 % oxygen was supplied compared to when 21 % oxygen was supplied. In addition, an increase in brain activity has been reported under 30 % oxygen concentration conditions using functional magnetic resonance imaging (fMRI). In addition, 30 % oxygen supply improves memory, verbal cognition, and working memory performance. However, these studies have been conducted only on a single task in a static environment. Therefore, the effects of high oxygen concentration on the performance of complex cognitive processing tasks must be verified.

Driving requires high levels of information processing because various factors, such as interactions with in-vehicle devices and changes in driving conditions (such as nearby vehicles, infrastructure, and weather) must be considered [10-13]. Driving requires high performance on complex cognitive processes, such as attention, visuospatial perception ability, memory, and problem-solving ability [14, 15].

Whereas previous studies have shown that high oxygen concentration affects cognitive processing for a single task, the aim of this study was to observe the effect of high concentrations of oxygen in a driving environment and verify the hypothesis that high oxygen concentration can activate the brain and improve performance on tasks requiring a high level of cognitive processing. For objective verification, brain activation via functional near-infrared spectroscopy (fNIRS) and driving performance were observed using driving performance data.

2 METHODS

2.1 Participants

Through a subject recruitment announcement, ten men, all licensed drivers in their 20s, voluntarily participated in the experiment (Tab. 1). Their average driving experience was 2.2 years. They had no history of brain disease or related surgery, and they were recommended to get more than 7 h of sleep the night before the experiment. On the day of the experiment, activities such as caffeine and drug consumption, smoking, and drinking that could affect fNIRS measurements were prohibited. Before starting the experiment, were explained to all participants, and written consent to participate was obtained.

Table 1 Experimental subject data

Gender	Number of subjects	Mean of age (SD)	Driving experience
Male	10	24.78 (±1.63)	2.20 (±1.23)

2.2 Oxygen Condition

Oxygen was supplied by connecting an oxygen respirator to an oxygen cylinder composed of oxygen and nitrogen. Cylinder 1 supplied 21 % oxygen and 79 % nitrogen, and Cylinder 2 supplied 30 % oxygen and 70 % nitrogen. The oxygen pressure was set at 68.95 kPa.

2.3 Driving Condition

The driving simulator used in the experiment was a GDS-3000s (Gridspace, South Korea) model, and the driving devices (accelerator, foot brake, and steering wheel) and display devices (speedometer, RPM meter, and trafficator) were implemented in the same manner as in an actual vehicle. A driving simulation environment was presented using three 32-in LCD monitors. The driving environment was a round trip on a six-lane highway; as shown in Fig. 1, a course of 3.4 km wide and 2.4 km long was driven in two lanes at a speed of 100 km/h.

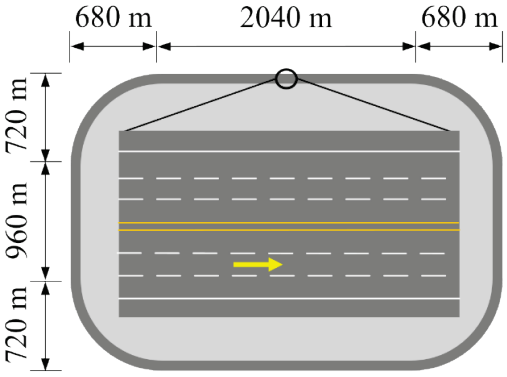


Figure 1 Driving environment

2.4 Experimental Process

The experimental process is shown in Fig. 2. The participants were asked to rest (Rest phase) for 20 min with their eyes closed. Then, oxygen was supplied for 20 min during the driving exercise (Driving phase). Subsequently, the same experiment was repeated for each subject with only a change in the oxygen concentration. At this time, the order of oxygen concentration was counterbalanced. In addition, the subjects were unaware of the oxygen concentration presented.

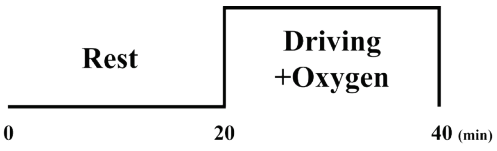


Figure 2 Experimental process

2.5 Measuring and Analysing Driving Performance Data

The parameters of the driving data were vehicle position, velocity, engine speed (RPM), percentage of the accelerator pedal manipulated (e.g., 100 %, the accelerator pedal fully depressed), and percentage of the steering wheel manipulated (100 %, the steering wheel fully rotated), and were measured at a sampling rate of 1 Hz. The last 5 min (15-20 min) of data in the driving phase were used for the analysis, and the parameters (Velocity, RPM, ACC, Steering) corresponding to the cornering section were extracted from the vehicle's location data (Tab. 2). The coefficient of variation was used as an indicator of driving performance. Therefore, the coefficients of variation for the velocity (CV_Velocity), RPM (CV_RPM), accelerator manipulation (CV_ACC), and steering wheel manipulation (CV_Steering) were calculated. Subsequently, a paired t-test was performed to compare the oxygen conditions (21 % and 30 %) using SPSS version 25 (IBM, USA).

Table 2 Driving performance data parameters

Parameter	Position		Velocity	RPM	ACC	Steering
	X	*Z*				
Unit	m	m	km/h	-	%	%

2.6 Cerebral Blood Flow Response Measurement and Analysis Method

Cerebral blood flow was measured using a NIRSport2 device (NIRx, Germany). This device consists of a source that emits near-infrared rays of two wavelengths (760 and 850 nm) and a light detector. Optical data (the amount of light detected) were measured at a sampling rate of 5.08 Hz. Thirty-two optodes (source: 16, detector: 16) were attached to the prefrontal, frontal, and parietal areas using the 10-20 system, and optical data from 49 channels were acquired (Fig. 3).

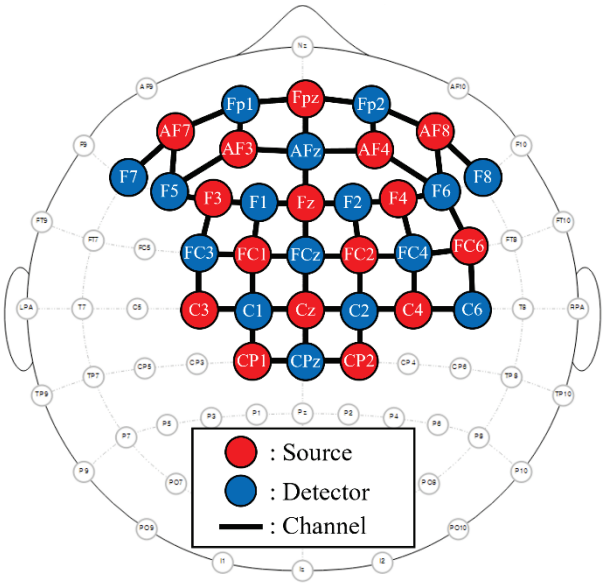


Figure 3 Optode attachment locations

The optical data obtained through fNIRS were converted to hemodynamic data using nirsLAB software (NIRx Medical Technologies, USA). First, bandpass filtering (0.01-0.2 Hz) was performed to remove physiological noises such as heartbeat and respiration. Subsequently, after adjusting the baseline of the entire dataset using data from the rest of the section, the oxyhaemoglobin (HbO) concentration was extracted using the modified Beer-Lambert law. The differential path length factor (DPF) was set to 7.25 (W1) and 6.38 (W2) [16-18].

The HbO concentration data extracted in the form of a time series were analysed using MATLAB version R2021b (MathWorks, USA). First, the average HbO concentration value in the last 5 min (15-20 min) of the driving phase was calculated for each channel. Subsequently, the brain was

divided into six areas (left prefrontal, right prefrontal, left frontal, right frontal, left parietal, and right parietal) and the HbO concentration of the channels corresponding to each area was averaged (Tab. 3). For the comparison of oxygen concentrations, a paired t-test was performed for the two conditions (21 % and 30 %) using SPSS.

Table 3 Channels corresponding to each brain area

Brain area	Channel (Source-Detector)
Left prefrontal	Fpz-Fp1, AF3-AFz, AF3-Fp1, AF3-F5, AF7-Fp1, AF7-F5, AF7-F7
Right prefrontal	Fpz-Fp2, AF4-AFz, AF4-Fp2, AF4-F6, AF8-Fp2, AF8-F6, AF8-F8
Left frontal	Fz-F1, F3-F1, F3-F5, F3-FC3, FC1-FCz, FC1-F1, FC1-FC3, FC1-C1, C3-FC3
Right frontal	Fz-F2, F4-F2, F4-F6, F4-FC4, FC2-FCz, FC2-F2, FC2-FC4, FC2-C2, C4-FC4
Left parietal	Cz-C1, C3-C1, CP1-C1, Cp1-Cpz
Right parietal	Cz-C2, C4-C2, CP2-C2, CP2-CPz

3 RESULTS

3.1 Driving Performance

A comparison of the driving data for the two oxygen conditions is shown in Fig. 4. CV_Velocity was 0.04 ± 0.02 under 21 % oxygen and 0.02 ± 0.01 under 30 % oxygen, which means that the variation under high oxygen concentration was significantly small ($p < 0.05$). CV_RPM and CV_ACC also significantly decreased under 30 % oxygen compared to those under 21 % oxygen ($p < 0.05$). In contrast, CV_Steering showed no significant difference between the two oxygen conditions (Tab. 4).

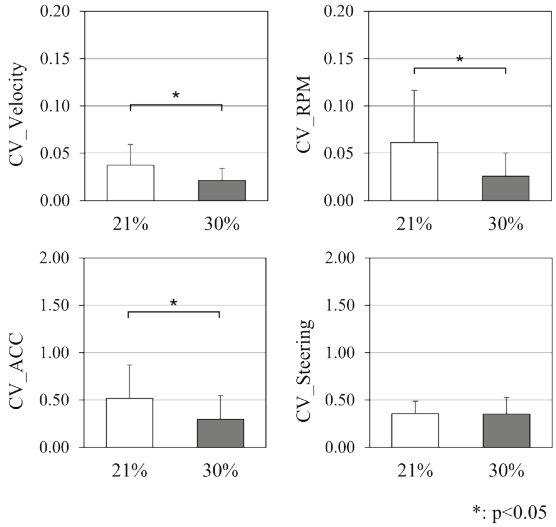


Figure 4 Results of driving performance data analysis

Table 4 Paired t-test results for oxygen conditions on driving performance

Parameter	Condition, %	Mean (*SD*)	*t*	*p*	*Cohen's d*
CV_Velocity	21	0.04 (± 0.02)	2.63	0.028	0.87
	30	0.02 (± 0.01)			
CV_RPM	21	0.06 (± 0.06)	2.66	0.026	0.69
	30	0.03 (± 0.02)			
CV_ACC	21	0.52 (± 0.35)	2.39	0.041	0.78
	30	0.30 (± 0.25)			
CV_Steering	21	0.36 (± 0.13)	0.03	0.980	0.01
	30	0.35 (± 0.17)			
Cohen's d = Mean difference / SD_{pooled}					

3.2 Cerebral Hemodynamic Response

A comparison of the HbO concentration by oxygen condition, as shown in Fig. 5, indicated that the HbO concentration increased under 30 % oxygen compared to 21 % oxygen. As shown in Tab. 5, statistically significant changes were observed in three of the six brain regions (left prefrontal, right prefrontal, and right parietal; $p < 0.05$). The HbO concentration in the left prefrontal increased by 1.90×10^{-5} mM, from 0.19×10^{-5} mM under 21% oxygen to 2.09×10^{-5} mM under 30 % oxygen ($t = 3.089$, $p = 0.013$). In the case of the right prefrontal region, it increased by 1.95×10^{-5} mM, from 0.21×10^{-5} mM under 21 % oxygen to 2.16×10^{-5} mM under 30 % oxygen ($t = 3.021$, $p = 0.014$). In the case of the right parietal region, it increased by 0.79×10^{-5} mM, from 0.50×10^{-5} mM under 21 % oxygen to 1.29×10^{-5} mM under 30 % oxygen ($t = 2.463$, $p = 0.036$).

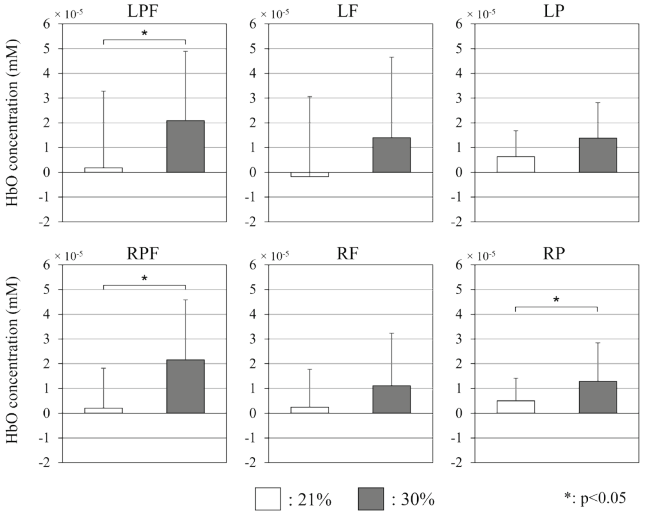


Figure 5 Hemodynamic response

Table 5 Hemodynamic response paired t-test results based on oxygen

Brain area	Condition, %	Mean (*SD*)	*t*	*p*	*Cohen's d*
Left prefrontal	21	0.19 (± 3.10)	−3.09	0.013	−0.61
	30	2.09 (± 2.80)			
Right prefrontal	21	0.21 (± 1.61)	−3.02	0.014	−0.89
	30	2.16 (± 2.43)			
Left frontal	21	−0.18(± 3.24)	−1.45	0.182	−0.46
	30	1.40 (± 3.25)			
Right frontal	21	0.25 (± 1.11)	−2.00	0.077	−0.44
	30	1.11 (± 2.12)			
Left parietal	21	0.63 (± 1.05)	−2.05	0.070	−0.57
	30	1.39 (± 1.43)			
Right parietal	21	0.50 (± 0.92)	−2.46	0.036	−0.58
	30	1.29 (± 1.55)			
Cohen's d = Mean difference / SD_{pooled}					

4 DISCUSSION

This study was conducted to investigate the effect of 30 % oxygen concentration on driving performance. HbO concentration was measured using the cerebral hemodynamic response as a physiological indicator, and fluctuations in velocity, RPM, ACC, and steering wheel were observed to assess driving performance.

The cerebral hemodynamic response analysis showed that the HbO concentration increased in the high oxygen concentration condition compared to that in the normal oxygen concentration condition. Clear differences were observed in the left prefrontal, right prefrontal, and right parietal regions. Driving performance analysis showed that the coefficients of variation of velocity, RPM, and ACC were smaller under high oxygen concentration conditions than under normal oxygen concentration conditions. In other words, the velocity was maintained at a constant value without unnecessary manipulation under high oxygen concentration conditions.

These results show that the effect of a high oxygen concentration is significant, even in complex driving situations that simultaneously require various cognitive resources, and the results suggest that improvement in a single cognitive ability by high-concentration oxygen also affects complex cognitive abilities.

In particular, the prefrontal cortex is associated with various factors, including memory and decision-making, and it is thought that high oxygen concentration activates the corresponding area, thereby increasing driving performance. Chung et al. observed that the prefrontal lobe was activated when 30 % oxygen concentration was present, and task performance also improved [8, 19]. In addition, Hannah reported that drivers with longer driving experience exhibited a higher prefrontal cortical activation than did novice drivers [20].

Among the cognitive abilities required for driving, the parietal cortex is related to visuospatial perception. Chung et al. observed that the parietal cortex was activated through the supply of high-concentration oxygen (30 %) when performing a visuospatial task. In addition, activation of the corresponding areas leads to improvements in visuospatial cognitive performance [8]. As an interpretation of this, it is judged that high concentration oxygen provides sufficient oxygen to the brain and activates the neural network, thereby improving cognitive ability.

In addition, Just et al. compared brain activation and driving performance in conditions where only driving was performed and where driving and additional work were performed simultaneously [21]. They found that the parietal area was more activated when only driving was performed compared to the dual-task condition, and driving performance improved.

Another interpretation of the effect of high oxygen concentration while driving is the reduction of driving fatigue. Sung et al. observed changes in driving performance under low, normal, and high oxygen concentration conditions [22]. They found that the reaction time for operating the brake pedal decreased under high-concentration oxygen conditions compared to low-concentration oxygen conditions. These results indicate that the supply of high-concentration oxygen reduced driving-induced mental fatigue, thereby improving driving ability. Their work differs from the present study in that driving ability was improved over a long period of time (120 min after starting driving); however, the improvement in driving ability was attributed to a reduction in fatigue 30 min after the start of driving. High oxygen concentration therefore reduces driving fatigue in a relatively short period of time, which improves driving performance.

Taken together, it is clear that high oxygen concentration causes activation by supplying sufficient oxygen to brain regions related to driving (prefrontal and parietal), thereby improving cognitive processing ability and thus driving performance. Additionally, the supply of high-concentration oxygen reduces driving fatigue, which also improves driving performance. In other words, high oxygen concentrations are believed to have a positive effect even in driving situations requiring complex cognitive abilities.

5 CONCLUSION

In conclusion, the findings of this study suggest a significant positive impact of high oxygen concentration (30 %) on driving performance, as evidenced by changes in cerebral hemodynamic response and driving-related parameters. The increased HbO concentration in key brain regions, such as the left prefrontal, right prefrontal, and right parietal areas, indicates enhanced cognitive processing abilities during driving tasks.

The study observed reduced variability in velocity, RPM, and ACC under high oxygen concentration conditions, highlighting a more stable and controlled driving performance. Notably, the prefrontal cortex, associated with memory and decision-making, and the parietal cortex, linked to visuospatial perception, were activated, contributing to improved driving abilities.

Furthermore, the study suggests that high oxygen concentration may play a role in mitigating driving fatigue, as reflected in decreased reaction times for brake pedal operation. This reduction in mental fatigue within a relatively short period could contribute to sustained improvements in driving performance.

This study has a limitation in that the demographic characteristics of the study participants were limited. According to previous studies, high oxygen concentrations differ in the amount of oxygen contained in the blood (blood oxygen saturation) depending on age. Driving performance also varies depending on age and sex; therefore, further research that considers these factors could lead to a more in-depth understanding of the effects of high-concentration oxygen. In addition, it is necessary to observe whether the effect of high concentration oxygen appears in actual driving situations rather than long-term driving situations or simulation situations. In summary, the positive effects of high oxygen concentration on driving performance, cognitive processing, and fatigue reduction indicate its potential as a promising intervention in enhancing safety and efficiency, even in complex driving scenarios requiring various cognitive resources.

6 REFERENCES

[1] Fujiwara, T. & Maeda, M. (2001). Effects of oxygen and refresh space for the elderly. *Journal of Human Life Engineering, 2*(3), 8-11.

[2] Masamoto, K. & Tanishita, K. (2009). Oxygen transport in brain tissue. https://doi.org/10.1115/1.3184694
[3] Zauner, A., Daugherty, W. P., Bullock, M. R. & Warner, D. S. (2002). Brain oxygenation and energy metabolism: part I—biological function and pathophysiology. *Neurosurgery, 51*(2), 289-302. https://doi.org/10.1097/00006123-200208000-00003
[4] Sharp, F. R. & Bernaudin, M. (2004). HIF1 and oxygen sensing in the brain. *Nature Reviews Neuroscience, 5*(6), 437-448. https://doi.org/10.1038/nrn1408
[5] Turner, C. E., Barker-Collo, S. L., Connell, C. J. & Gant, N. (2015). Acute hypoxic gas breathing severely impairs cognition and task learning in humans. *Physiology & Behavior, 142*, 104-110. https://doi.org/10.1016/j.physbeh.2015.02.006
[6] Moss, M. C. & Scholey, A. B. (1996). Oxygen administration enhances memory formation in healthy young adults. *Psychopharmacology, 124*, 255-260. https://doi.org/10.1007/BF02246665
[7] Chung, S. C., Lee, B., Tack, G. R., Yi, J. H., Lee, H. W., Kwon, J. H., ... & Sohn, J. H. (2008). Physiological mechanism underlying the improvement in visuospatial performance due to 30% oxygen inhalation. *Applied Ergonomics, 39*(2), 166-170. https://doi.org/10.1016/j.apergo.2007.05.008
[8] Chung, S. C., Tack, G. R., Lee, B., Eom, G. M., Lee, S. Y. & Sohn, J. H. (2004). The effect of 30% oxygen on visuospatial performance and brain activation: an fMRI study. *Brain and Cognition, 56*(3), 279-285. https://doi.org/10.1016/j.bandc.2004.07.005
[9] Chung, S. C. & Lim, D. W. (2008). Changes in memory performance, heart rate, and blood oxygen saturation due to 30% oxygen administration. *International Journal of Neuroscience, 118*(4), 593-606. https://doi.org/10.1080/00207450601067299
[10] Koo, T. Y., Park, K. J., Kim, B. Y., Kim, H. J. & Suh, M. W. (2009). A study on driver's workload of telematics using a driving simulator: A comparison among information modalities. *International Journal of Precision Engineering and Manufacturing, 10*, 59-63. https://doi.org/10.1007/s12541-009-0048-6
[11] De Waard, D. & Brookhuis, K. A. (1996). The measurement of drivers' mental workload.
[12] Foy, H. J. & Chapman, P. (2018). Mental workload is reflected in driver behaviour, physiology, eye movements and prefrontal cortex activation. *Applied Ergonomics, 73*, 90-99. https://doi.org/10.1016/j.apergo.2018.06.006
[13] Zhang, H., Guo, Y., Yuan, W. & Li, K. (2023). On the importance of working memory in the driving safety field: a systematic review. *Accident Analysis & Prevention, 187*, 107071. https://doi.org/10.1016/j.aap.2023.107071
[14] Kim, S. Y., Park, H. I., Yoon, H. J., Yoo, G. P., Lee, S. G., Lee, J. B. & Yoo, I. G. (2016). Effects of concentration training based on virtual reality on driving performance. *The Journal of Korean Society of Assistive Technology, 8*, 11-18.
[15] Broadbent, D. P., D'lnnocenzo, G., Ellmers, T. J., Parsler, J., Szameitat, A. J. & Bishop, D. T. (2023). Cognitive load, working memory capacity and driving performance: A preliminary fNIRS and eye tracking study. *Transportation research part F: traffic psychology and behaviour, 92*, 121-132. https://doi.org/10.1016/j.trf.2022.11.013
[16] Li, Y., Shen, H., Yu, Y. & Hu, D. (2021, May). Multi-classification of fNIRS signals in four body parts motor imagery tasks measured from motor cortex. In *the 4th IEEE International Conference on Artificial Intelligence and Big Data (ICAIBD2021)*, 500-504. https://doi.org/10.1109/ICAIBD51990.2021.9459089
[17] Yun, S. H., Jang, T. S. & Kwon, J. W. (2024). Cortical activity and spatiotemporal parameters during gait termination and walking: A preliminary study. *Behavioural Brain Research, 456*, 114701. https://doi.org/10.1016/j.bbr.2023.114701
[18] de Lima-Pardini, A. C., Morais, G. A. Z., Balardin, J. B., Coelho, D. B., Azzi, N. M., Teixeira, L. A. & Sato, J. R. (2017). Measuring cortical motor hemodynamics during assisted stepping–an fNIRS feasibility study of using a walker. *Gait & Posture, 56*, 112-118. https://doi.org/10.1016/j.gaitpost.2017.05.018
[19] Chung, S. C., Iwaki, S., Tack, G. R., Yi, J. H., You, J. H. & Kwon, J. H. (2006). Effect of 30% oxygen administration on verbal cognitive performance, blood oxygen saturation and heart rate. *Applied Psychophysiology and Biofeedback, 31*, 281-293. https://doi.org/10.1007/s10484-006-9023-5
[20] Foy, H. J. & Chapman, P. (2018). Mental workload is reflected in driver behaviour, physiology, eye movements and prefrontal cortex activation. *Applied ergonomics, 73*, 90-99. https://doi.org/10.1016/j.apergo.2018.06.006
[21] Just, M. A., Keller, T. A. & Cynkar, J. (2008). A decrease in brain activation associated with driving when listening to someone speak. *Brain Research, 1205*, 70-80. https://doi.org/10.1016/j.brainres.2007.12.075
[22] Sung, E. J., Min, B. C., Kim, S. C. & Kim, C. J. (2005). Effects of oxygen concentrations on driver fatigue during simulated driving. *Applied Ergonomics, 36*(1), 25-31. https://doi.org/10.1016/j.apergo.2004.09.003

Authors' contacts:

Kyu-Beom Kim
Department of Biomedical Engineering, School of ICT Convergence Engineering, College of Science & Technology, Konkuk University,
268, Chungwon-daero, Chungju-si, Chungcheongbuk-do, Republic of Korea
rlarbqja0507@kku.ac.kr

Soon-Cheol Chung
Department of Biomedical Engineering, School of ICT Convergence Engineering, College of Science & Technology, Konkuk University,
268, Chungwon-daero, Chungju-si, Chungcheongbuk-do, Republic of Korea
scchung@kku.ac.kr

Byung-Chan Min
(Corresponding author)
Department of Industrial Management Engineering, Hanbat National University,
125, Dongseo-daero, Yuseong-gu, Daejeon, Republic of Korea
bcmin@hanbat.ac.kr

ISSN 1846-6168 (Print), ISSN 1848-5588 (Online)
https://doi.org/10.31803/tg-20240228104040

Original scientific paper
Received: 2024-02-28, Accepted: 2024-05-21

A Study on Visualization of Digital Twin Electric Vehicles Based on Gamification: A Simulation Using Game Engines

Jung Tae Kim

Abstract: This study investigated a methodology for implementing and simulating a digital twin (metaverse) electric vehicle visualization with gamification using a game engine (Unity 3D). This paper is important because it is the first study to visually represent an electric vehicle using a game engine. The core of this study is to analyze the differences between internal combustion and electric vehicles by implementing them in a digital twin, and the KONA model of Hyundai Motor Company, which has two types of engines, was selected. To create the driving roads to compare the performance of the two cars, the terrain was created with Unity Component Terrain. After creating the terrain, configuring the digital twin assets, creating a fuel economy system, setting up object placement and collision settings, and animating, the team created a simple comparison drive that yielded meaningful results. Several gamification elements were applied to motivate players to use the implemented digital twin electric vehicle simulation, increasing curiosity and engagement. Further research will be conducted to add more gamification elements and augment the digital twin experience to make it more engaging and visualize the electric vehicle to increase driver engagement and understanding.

Keywords: digital twin; electronic car; gamification; game engine; visualization

1 INTRODUCTION

The purpose of this research is to develop a methodology to simulate the driving environment scenario of an electric vehicle on a digital twin by visualizing it with a gamification interface using the game engine Unity 3D. The results of this study are expected to quickly and effectively convey information necessary for the manufacture and operation of electric vehicles, and will be very useful for the design of electric vehicle designs, and in particular, gamification elements applied to vehicles in virtual space will contribute to providing a positive driving experience for drivers.

To implement both internal combustion and electric vehicles in the digital twin, products with two different engine types were selected. The KONA model of Hyundai Motor Company, which has both internal combustion and electric models, was selected to be implemented with the game engine Unity 3D. This study was conducted with the following three goals:

1) Differentiating from existing electric vehicle research, a digital twin driving environment based on the Unity 3D game engine was implemented to visualize and compare the differences between electric vehicles and internal combustion vehicles intuitively.
2) By utilizing digital twin technology to create a simulation similar to the actual driving environment of electric vehicles, gamification elements were applied to make the features of electric vehicles more vivid and fun to experience.
3) A visualization methodology is proposed to increase drivers' interest, interest, engagement, and understanding of EVs by applying gamification elements (points, levels, progressive bars, competitions, challenges, etc.).

2 RELATED WORKS

In recent years, many manufacturers have been conducting research and development on electric vehicles. In addition, as the parts and software of vehicles become more complex and diverse as autonomous driving functions become more advanced, it is time-consuming and costly to manually evaluate and verify them in the real world [1]. With the development of gamification and digital twin technologies, it has become possible to efficiently visualize electric vehicles using them. In this paper, an electric vehicle visualization was studied by combining gamification and digital twin technology through Unity 3D.

In this study, Unity3D, a game engine, was mainly utilized to create a digital twin environment. Unity3D game engine is one of the general-purpose engines utilized by most game production companies and game developers around the world. Unity Engine was developed by David Helgason, Joachim Ante, Nicholas Francis, and others. They founded Unity and developed the Unity Engine to help small game developers create games [2]. Unity3D can work on MS Windows and Mac OS X. Unity3D is available for Windows and Mac. Programs created in Unity3D run on windows, Mac OS, Xbox, PlayStation, Wii, IOS, and android [3].

Digital twin is a technology that implements a 3D model of a twin that identically reflects the physical features of a real object in the virtual world (Digital), simulates it in real-time synchronization with the real object, and uses it for real-world decision-making such as control, analysis, and prediction [4].

The twin of a digital twin, as used in the current concept of a digital twin, first appeared in NASA's Apollo program. In this program, the environments of two or more space vehicles were simulated under the same conditions to enable accurate flight condition prediction and real-time simulation [5]. Digital twin was first introduced in Grieves' work in 2003 and have continued to evolve to the present day [6].

By adopting a real-time ‘digital twin’ framework that bridges the virtual and physical worlds, projects can take full advantage of digital twin technology's ability to avoid technical defects or unexpected performance anomalies that arise later in the process. To improve the validation of any

finished product, using simulation from the earliest stages of prototype development to real-world testing can prevent costly design changes later in product development. Utilizing Digital twin for this simulation approach can save money and increase efficiency.

The core of digital twin is to integrate the real world and the virtual world in real time by implementing a model that reflects reality in the virtual world, and the process of "create → send → aggregate → analyze → understand → execute" of data [7].

Based on these benefits and procedures, this study aims to visualize an electric vehicle in a digital twin using Unity 3D. The process of visualizing an electric vehicle in a digital twin environment consisted of five steps.

① Set (create) goals - Set and create the goals you want to achieve with Digital twin in research. In this case, the project visualizes an electric vehicle and a conventional vehicle in the same virtual environment and compares various performance and post-driving status factors such as fuel efficiency, speed, and durability.

② Vehicle selection and data collection (transmission) - After selecting the vehicle model that best meets the project goals for visualization, all elements of the transmission, such as driving-related specifications, exterior shape, and functions of the model selected in this paper, are analyzed and information is collected for later production.

③ Digital twin driving environment planning (aggregation) - Based on the data collected and aggregated, the system, content, and user interface to be implemented in the Unity 3D space were planned.

④ Implementation (analysis and understanding) - Collect the same data as in real-world driving within the digital twin, analyze it to implement the 3D environment, and create a Unity environment that players can understand on the digital twin.

⑤ Driving test and drawing conclusions (execution) - Collect various data through driving test of electric vehicles in the implemented 3D digital twin driving environment, simulate the situation in the real environment based on this data, and draw conclusions.

Gamification is a new word made up of the words game +ification (a Latin noun –ification suffix) and refers to the application of game elements to non-video game contexts. Gamification is the recognition of problems in everyday political, economic, social, cultural, and industrial situations and relationships between people (users, consumers, receptors, etc.) that seem unrelated to games and the application of "game elements" as solutions [8]. Gamification can be used to make the driving experience of an electric vehicle more fun and enjoyable in a visualized space, and increase the driver's sense of engagement and understanding.

Examples of electric vehicle visualization implementations using gamification and Digital twin include Hyundai Motor Group's trial of Digital twin technology for managing electric vehicle battery performance [9] and BMW Group's use of NVIDIA's Omniverse artificial intelligence (AI) to create a virtual car production environment to virtually build cars [10]. The system adds a gaming element to engage the driver. Digital twin technology can not only monitor the status of electric vehicles in real time, but also synchronize information and functions from the manufacturing site to derive various analytical results to evaluate, analyze, optimize, and predict the situation on site [11].

3 RESEARCH METHODOLOGY

The methodology of this research is to create and implement a digital twin virtual space similar to the real driving environment, and to conduct simulation experiments in which the user selects a vehicle and controls and manipulates variables while driving.

First of all, to realize the digital twin driving environment of goal 1) presented in the introduction, three scenes were set up in Unity 3D: lobby, options, and driving. The digital twin environment was then built as a PC application (deliverable) for users to play. The application allows the user to select the options, driving, and exit scenes via three buttons, and provides a comparable and testable 3D driving environment for further research.

For Objective 2), the application was gamified to provide the driver with a fun and enjoyable experience of driving an electric vehicle, either an EV or a conventional car. Gamification technology can increase the player's engagement and understanding in the virtual space of the digital twin. Therefore, this research project was implemented by providing appropriate gamification elements to provide the driver with the experience of driving a car video game.

To implement Objective 3), the gamification was realistic enough to allow the driver to select a 'damage option' from the vehicle's options during the driving test, just like in a car video game. This damage option implements an event that dents the body of the vehicle when the vehicle collides with another object, to visualize the impact of damage to the body in an accident during real-world driving and provide the driver with a realistic sense of immersion.

Physical phenomena and basic vehicle dynamics that occur in real-world driving situations were also applied. A system was created to independently drive all four wheels of the car to check the status of all wheels, and the wheel drive system was reinforced and tested. The durability of the car body was implemented so that it can be adjusted through the mass variable of RigidBody.

In addition, unique performance indicators of electric vehicles (e.g., fuel economy, speed change, Regenerative Braking) were implemented on the Digital twin. The fuel system of each vehicle was implemented in the form of a progressive bar in gamification so that the performance of electric and conventional vehicles could be compared. In addition, gamification systems were implemented to change the details to differentiate between electric and internal combustion vehicles, such as the ABS system.

In conclusion, this paper presented a simulation experiment of gamification-based digital twin electric

vehicle visualization, implementing a real driving environment on a digital twin and allowing users to select an electric vehicle and an internal combustion vehicle to compare and analyze the differences in participation and interest levels, creating a basic environment for further research.

4 ELECTRIC VEHICLE DIGITAL TWIN VISUALIZATION RESEARCH PROJECT DETAILS

This research is a project that combines the Unity3D game engine, Digital twin, and gamification technology to visualize the driving environment of an electric vehicle and allow users to conduct simulation experiments. The purpose of the research is to develop a methodology for realizing gamification-based electric vehicles and their virtual driving environments in Unity 3D. In particular, the UI design was created with feedback from electric vehicle experts.

The results of this study will be very useful in providing information for the design, manufacturing, and operation of electric vehicles. Furthermore, the driving environment in the digital twin and the gamification techniques applied to the electric vehicle will provide users with an interesting driving experience and make a meaningful contribution to the electric vehicle industry.

There are many possible methods and tools for building a digital twin environment. In this study, the general-purpose game engine Unity3D (21.3.15f1 and 21.3.33f1) was mainly used, and the accompanying graphics tools (3D Max Studio, Zbrush, Photoshop, etc.) and programming language C# were applied. The recommended hardware specs for the build are CPU: AMD Ryzen 5-5 7600 (Raphael), Memory: SEC DDR5-4800 (16 GB), Graphic: Geforce RTX 3070 Miracle II D6 8GB, and the app should run fine on these specs.

The following five steps were used to guide the research. The research content for each of the five stages of the EV visualization process covered in the previous literature review is as follows.

4.1 Goal Setting

There are five main goals for this study:

- Visualize and communicate the differences between electric vehicles and internal combustion vehicles in an easy-to-understand way so that they can be intuitively understood and compared.
- Increase the attractiveness of the visualization by adding gamification elements by actively utilizing the features of electric vehicles.
- Increase drivers' interest and engagement with EVs through EV visualization to improve understanding.
- By utilizing digital twin technology to create a simulation that is similar to the actual driving environment of an electric vehicle, it is possible to more vividly experience the characteristics of an electric vehicle in a real driving environment.
- It is set to strengthen competitiveness in the field by producing technical results that are differentiated from existing electric vehicle visualization research.

4.2 Vehicle Selection and Data Collection

Since the main focus of the study is to implement an internal combustion and electric vehicle in a digital twin to analyze the differences between the two, it was necessary to select a product with two different engine types within the same vehicle. Therefore, the KONA model from Hyundai was chosen, which has both internal combustion and electric models.

The image below is of the 2021 Hyundai KONA model, which is very detailed in terms of exterior appearance and interior implementation (Fig. 1).

A car is divided into two parts: the body and the sash, and since the body is the most important part for driving, a detailed model of the body was selected. For the interior of the car, the project did not implement additional models, prioritizing the reproduction of the system through code. The interior model will be added in future years through enhancements. In addition to these parts, there are many other parts in the car, so the project was conducted by collecting as much data as possible.

Figure 1 2021 Hyundai Kona Electric 3D model (up), 2021 Hyundai Kona N-Line model (down)

4.3 Planning and Organizing the Unity3D Environment

For this study, in the game engine Unity 3D, three scenes were set up for this study: Lobby, Options, and Driving. The options scene was set up to be scalable, utilizing Unity's Additive Scene feature to open multiple scenes on top of a specific scene.

To apply the basic physics system to the vehicle, this paper introduces Unity's RigidBody. Also introduced at this stage is the MeshCollider, which implements collision events with external objects. There are many options for Colliders,

but are chose Mesh Collider to accurately fit the shape of the model. For the Rigidbody, it is added to the highest parent object, while for the Collider, it is added to the Body child object because it needs to follow the shape of the model (Fig. 2).

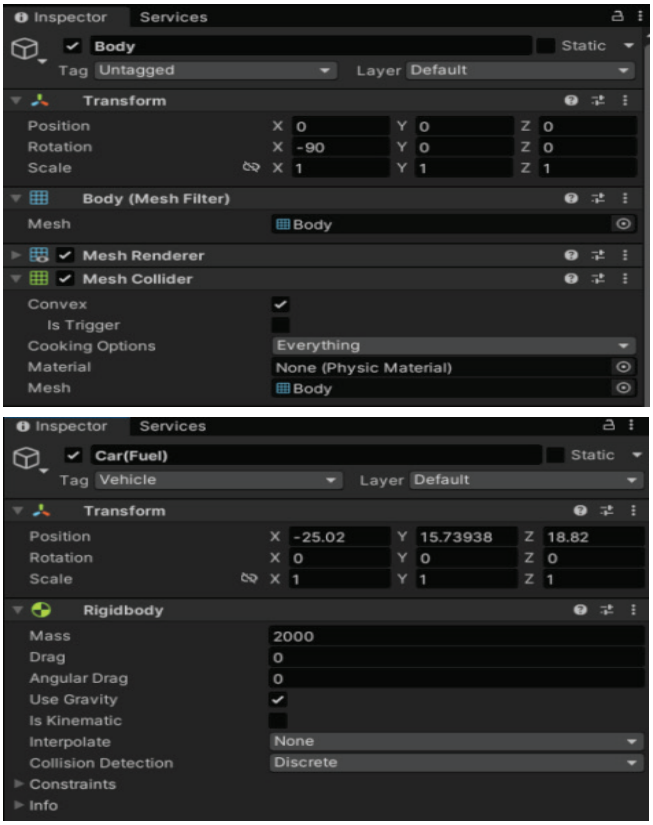


Figure 2 RigidBody and mesh collider. The game engine can be manipulated to adjust the physical behavior of the vehicle and adjust the values of collision event variables.

Figure 3 Lobby scene. When the driver launches the application, they are greeted with a video game-style lobby window.

When players launch the Digital twin program, they'll see three buttons to start, options, and exit. In addition, a 3D background has been created and placed in the background to show the title of the project and a preview of the car player will be driving (Fig. 3).

In order for the player to press the 'Start' button to make the car move, Unity Engine needs to receive input from the player. To do this, the variables that need to be input are summarized below (Tab. 1).

Table 1 Player input data

Steering-related	
Throttle	Float variables for accelerator inputs
Brakes	Bool variables for break input
Steering	Float variables for direction (negative 'left', 'center' for 0, positive 'right')
Clutch	Float variables for clutches
Handbrake	Bool variables for handbrake
ShiftUp	Bool variables that correspond to a gear up.
ShiftDown	Bool variables that correspond to downshifting
EngineStartStop	Bool variables corresponding to engine start
Light-related	
LeftBlinker	Bool variables for left turn signal
RightBlinker	Bool variables for right turn signal
LowBeamLights	Bool variables for high beams
HighBeamLights	Bool variables for high beams
HazardLights	Bool variables for emergency lights
ExtraLights	Bool variables related to additional indicator lights needed for other vehicles
Others	
Horn	Bool variables corresponding to klaxon
CruiseControl	Bool variables for cruise control

4.4 Create a Digital Twin

4.4.1 Implementing the EV Digital Twin Terrain and Basic UI

To create a driving surface for comparing the performance of internal combustion and electric vehicles, you can use Unity Component Terrain to create the underlying terrain. For example, mountains, hills, and holes for lakes are created using this feature (Fig. 4). The dirt road model is then overlaid with the driving road model, which is further modified and textured for added realism.

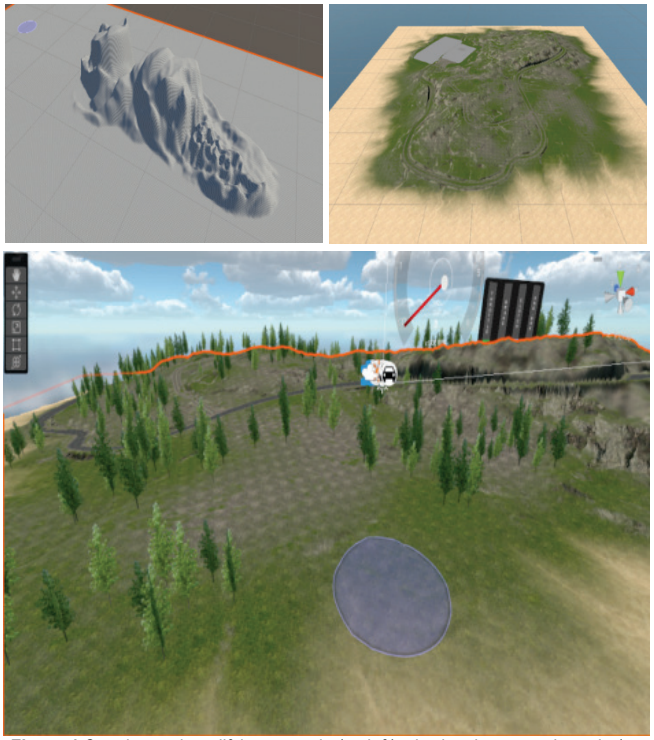


Figure 4 Creating and modifying a terrain (up left), viewing the created terrain (up right), and modifying the terrain texture (down)

Textures such as stone textures for hills, grass textures for grass, and sand textures for sandy ground are applied using the brush function of the terrain so that they can blend in naturally. After that, trees, bushes, and various structures are installed to create a sense of three-dimensionality.

To create the digital twin electric car, the UI for the driving screen was synthesized by using the player input data as the corresponding variables. On the left side of the driving UI, there is an analog instrument panel to check the current RPM and gear of the car, and on the bottom right, a slider gauge shows the values of Throttle, Brake, Clutch, Hand Brake, and Handle that the user inputs when driving (Fig. 5).

On the right side of the driving UI, there is an analog instrument panel that displays the current speed, turn signal, high beam, low beam, etc. and the current Torque, Lat Slip, Lng Slip, and Load values through the input data, a horizontal slider that shows the accumulated damage of the car, and a button to enable/disable damage and a button to repair. The integration with the UI was finalized by referencing the variables managed by the Scriptable Object to reference the material while driving. In future research, more road types, weather conditions, slopes, ground conditions, etc. will be implemented.

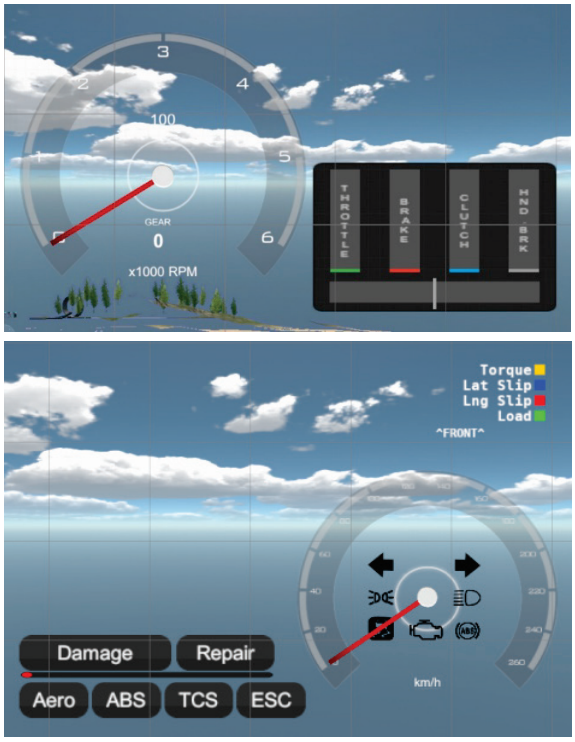


Figure 5 Driving UI left (up) and right (down). It is implemented so that the user can check the driving status and conditions by looking at the HUD in the front of the car.

The energy display for the car was implemented by creating a ScriptableObject that stores fuel based on data from the car based on user input, so that the maximum amount of energy that can be stored and the amount of energy currently remaining can be linked to the UI (Fig. 6).

To differentiate between electric vehicles and internal combustion vehicles, the UI for checking the fuel level was differentiated by using a battery-shaped UI to emphasize the fact that electricity is used for energy, and for internal combustion vehicles, the analog instrument panel that usually displays the remaining fuel in internal combustion vehicles was replaced with a vertical slider.

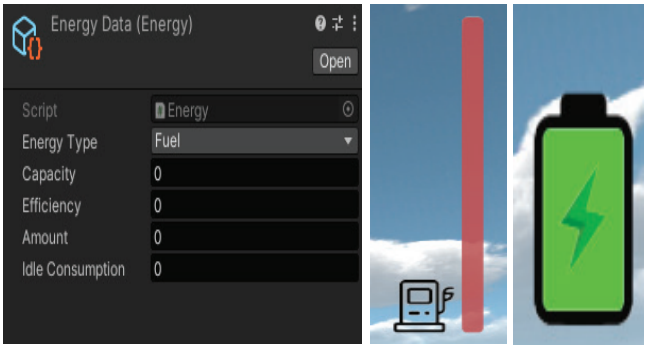


Figure 6 Car energy ScriptableObject (left), car energy display UI internal combustion (center) / electric (right)

Finally, regenerative braking, which is a unique characteristic of electric vehicles, is an electric braking technology that recovers kinetic energy (brake and acceleration) and stores it in the battery [11]. In this study, a visualization UI (Fig.7) that simulates the regenerative braking of an electric vehicle on a digital twin was implemented and the driving characteristics were compared with those of an internal combustion vehicle.

Figure 7 First Person View - Regenerative Braking UI Visualization (up) Third Person View - Regenerative Braking UI Visualization (down)

4.4.2 Animating an Electric Car Digital Twin

When the car is moving, turning the steering wheel causes the front wheels to rotate the steering shaft connected to them according to the amount of rotation of the steering wheel. The pinion gear at the end rotates and finally engages with the rack gear connected to it to finally turn the wheel to change the direction of the car. Fig. 8 shows the animation of the wheels using this principle.

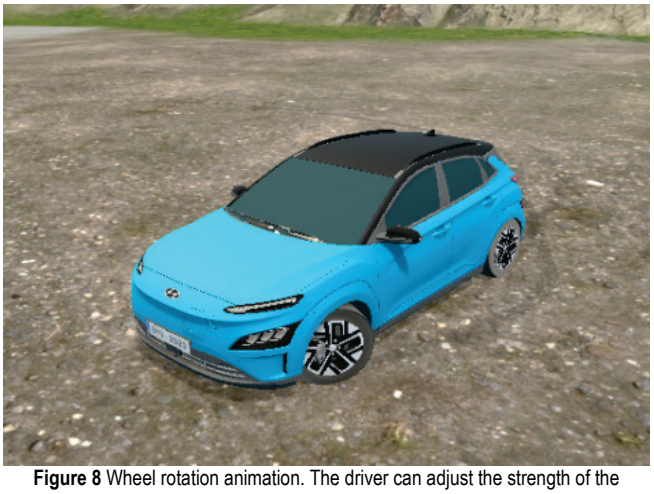


Figure 8 Wheel rotation animation. The driver can adjust the strength of the rotation animation by adjusting the value of a variable in Unity.

A car is a device that allows the left and right wheels of a car to change the number of revolutions through a differential gear to make it rotate smoothly without effort on bumpy roads and when turning. When a car runs on a flat road, the two wheels have the same rotational resistance, so they show the same number of revolutions, but when passing through bumpy roads such as mountain roads and hills, the left and right wheels have different rotational speeds (Fig. 9).

To realize this phenomenon, the coefficient of rolling resistance of the car's wheels was calculated and reflected in real time, and the wheel animation was applied to emphasize the realistic feeling.

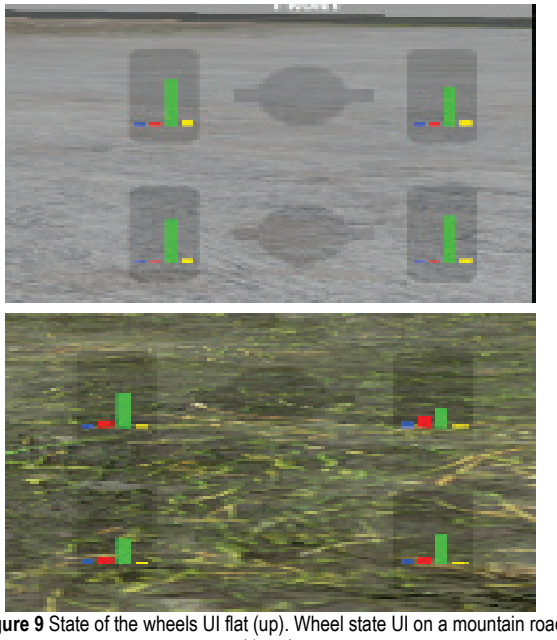


Figure 9 State of the wheels UI flat (up). Wheel state UI on a mountain road (down).

While the basic movement of a digital twin vehicle has been implemented so far, there are still some gaps in the vehicle's movement that need to be filled. In real-world driving, the suspension and inertia of a car cause the body to sway when the car hits a bump or curves. In this project, linear interpolation was used to realize these animations. In the case of a traditional car body, the model's position is fixed immediately after a collision, while interpolation allows it to move to the next position more naturally.

This allows us to represent the natural sway of the body when it collides or curves above a certain speed. The interpolation value was made a variable so that it could be adjusted to the vehicle's information if it changed in the future.

4.5 Test and Draw Conclusions

If users select the Damage option in the car's options during a drive test and the car collides with another object, they can implement an event that dents the car's body in the area of the collision to visualize the impact of damage to the car in a real-world accident (Fig. 10). The durability of the body can be adjusted through the mass variable of the RigidBody.

Figure 10 Implementing car body damage from collisions. Adjusting the parameter values in the Damage options can change the amount of damage.

In addition, fuel systems for electric and internal combustion vehicles, changes in details to distinguish between electric and internal combustion vehicles, and ABS systems were implemented (Fig. 11). A system that allows all four wheels of the car to drive independently and check the status of all four wheels was created, and the wheel drive system was reinforced and tested.

Figure 11 Track scene for electric vehicles. This was implemented to allow simulation experiments by adjusting the values of variables.

5 RESULTS AND DISCUSSION

In this paper, a digital twin environment was implemented based on a game engine (Unity3D) and gamification technology was applied to efficiently visualize electric vehicles. This paper aims to present an efficient

visualization method for electric vehicles by implementing a digital twin virtual environment based on the game engine Unity3D and applying gamification elements to the virtual space, virtual vehicles, and objects.

This study was a simulation project to visualize the driving comparison between electric and internal combustion vehicles using a gamification-based digital twin environment with a game engine to make it more visually appealing to drivers and increase their engagement and understanding. Hyundai Kona, which is in line with this research, was developed to implement a physical driving environment so that players can simulate playtesting. In addition to providing the driver with a visually appealing UI (with EV expert feedback) from electric vehicle experts. That is fun and enjoyable like a game, it was also visualized to monitor the status of the electric vehicle in real time. This paper can serve as an overview of how these factors affect driver decision-making and vehicle performance in digital twin.

However, this research project, which applied gamification elements to digital twin environments and vehicles, aims to provide a driving experience for electric vehicles by increasing driver engagement and understanding, but due to the limitations of the research period and budget, it is not possible to increase the driver experience to a sufficient level. Nevertheless, users and subsequent researchers who encounter the results of this project will have a positive perception of EVs and increased interest in the EV industry.

In the future, follow-up studies will be conducted to further analyze the funding and information of other EV models other than the Kona. Findings of this study. It is planned to expand the driving environment and variable conditions of the vehicle to implement more realistic damage conditions, physical conditions of the vehicle, and environmental factors (road type, weather conditions, etc.) to study the impact on vehicle performance. In addition, the follow-up research will provide a HUD that incorporates a gamification-based feedback system on the digital twin to help users learn how to drive more efficiently and highlight the benefits of electric vehicles. This expanded follow-up research will include user experience research by conducting playtesting surveys with actual drivers or potential EV buyers. This will allow for a comparative analysis of engagement between EVs and internal combustion vehicles and evaluate the effectiveness of gamification strategies applied more broadly to each vehicle on users in the EV driving environment.

This study and subsequent studies are expected to contribute to the realization of actual driving, fuel efficiency, and feelings of use by building environments such as factories that produce actual electric vehicles and repair shops that inspect vehicles into digital twin, and making electric vehicles in the digital twin in a suitable form.

Acknowledgement

This research was supported by "Regional Innovation Strategy (RIS)" through the National Research Foundation of Korea (NRF) funded by the Ministry Education (MOE) (Foundation Task Management Number: 2022RIS-006).

6 REFERENCES

[1] Park, C. W., Chung, S. H. & Lee, H. C. (2020). Vehicle-in-the-Loop in Global Coordinates for Advanced Driver Assistance System. *Appl. Sci., 10*(8), 2645. https://doi.org/10.3390/app10082645
[2] Lee, J. W. & Jung, J. P. (2021). Learn Unity by Making Games. 22-24.
[3] http://en.wikipedia.org/wiki/Unity_(game_engine)
[4] CHO Alliance. (2023). Digital twin technology development, application, and service commercialization strategy, 25-29.
[5] Rosen, R., Wicher, G. V., Lo, G. & Bettenhausen, K. D. (2017). About the Importance of Autonomy and Digital twin s for the Future of Manufacturing. *IFAC Papers on Line, 48*(3), 567-572. https://doi.org/10.1016/j.ifacol.2015.06.141
[6] Grieves, M. (2014). Digital twin: Manufacturing Excellence through Virtual Factory Replication. *White Paper, 1*, 1-7.
[7] R&D Information Center. (2022). Metaverse/Digital twin Technology Forecast and Virtual Convergence Technology (XR) Industry Analysis, 154-156.
[8] Kim, J. T. (2022). Introduction to Gamification. 5-12.
[9] https://www.hyundai.co.kr/news/CONT0000000000032186
[10] https://blogs.nvidia.co.kr/2021/04/16/nvidia-omniverse/
[11] https://en.wikipedia.org/wiki/Regenerative_braking

Author's contacts:

Jung Tae Kim, Professor
Department of Game Studies, Dongyang University,
2784 Pyeonghwa-ro, Dongducheon-si, Gyeonggi-do,
11307, Republic of Korea
kimj35063@gmail.com

ISSN 1846-6168 (Print), ISSN 1848-5588 (Online)
https://doi.org/10.31803/tg-20240623232615

Original scientific paper
Received: 2024-06-23, Accepted: 2024-10-31

Outlook on Methods Dealing with Ludonarrative

Enio Vešligaj, Andrija Bernik*

Abstract: Ludonarrative has become one of the most prevalent topics in video games. More and more games are judged or analysed in accordance with ludonarrative methods. This paper will analyse the work which has been done to guide ludonarrative methods. The methods will be presented so that we are able to summarise and find the gaps in today's game development process. Most of the works on ludonarrative methods are theoretical and few are practical. They focus on specific intricacies of ludonarrative and from there procedures are built. The purpose of this paper is to present and find where and how methods in relevance to ludonarrative can be used. With the review of those methods, it can be found that most of the methods are used to review finished products (games) in accordance with ludonarrative ideas. The authors will review and compare the various works in relation to ludonarrative.

Keywords: game development; ludonarrative; methods; video games

1 RELATED WORK

The term *Ludonarrative dissonance* coined by Hocking [1] has marked gaming studies. In his blog post, he sparked the thoughts of many authors by proposing ludonarrative dissonance. This idea stems from Hocking's playthrough of the BioShock game [1]. He went on to explain the discrepancy between ludic and narrative elements in the game. This term tries to explain how some games create dissonance between game elements and narrative. At first, ludonarrative studies consisted of mostly of researching what is ludonarrative and how does it affect the games or players. Aarseth [2] and Murray [3] were one of the first to propose ideas that later will be used to study ludonarrative. Aarseth suggest that games have separate elements (ludic and narrative) and that player needs to have some kind of interactivity with those elements. Murray says that this is more about the player and "the satisfying power to take meaningful action and see the results of our decisions and choices" (1997, p. 126). Later Toh [4] in his work offered a different way to look at the ludonarrative. He branched it into plethora of different subbranches. Each branch has its own relevance depending on the style that the game is taking. Approaches from authors such as Jenkins [5], Louchart et al. [6], Seraphine [7], Frasca [8] and Roth [9] go more into theoretical realm of ludonarrative while offering new ideas how to research it. Those ideas look at the different media (books, cinema, games). With different media in mind, the ideas proposed by them are objectively looked at and try to explain what is ludonarrative or how certain ludonarrative conditions are created. Donoghue [10] takes inspiration from Toh and creates *LAF model.* LAF model tries to break down games into components. Each component has different requirements and those requirements were combined to create a scale to understand and qualify ludonarrativity inside the game. Dodds [11] research through design offers a way to create ludonarrative consistent game. His *trait systems* show how each small design ideas, if combined correctly can have a great impact on the consistency of the game. Aarseth [12] builds upon his old works and proposes how certain game elements correspond with each other. Depending on the game you wish to create, it may be valuable to have a different approach to creating it. Similarly to Aarseth, Hunicke et al. [13] and Maraffi [14] created frameworks which try to find and understand what is and how the player – game connection is created. Hunicke et al. MDA framework shows how depending on the position of the person, outlook on the game can be different and that it needs to keep both views on the game while creating it (player and designer view). Maraffi's SGRplay framework uses MDA as its basis and creates a survey to analyze the connections between player and the game. Like MDA framework offered a way to create a game while following certain approaches, Purnomo et al. [15] *GAMING* formula and Koenitz [16] *IDN* framework look at the building blocks of games and offer ideas which need to keep in mind while creating the games.

While all the methods mentioned offer insights into ludonarrative and how to plan for it, it does not tell us much about how exactly to create it. Despain & Ash [17] go in depth while looking at the *ludonarrative harmony*. Instead of testing already existing published games, Despain & Ash went on to create their own model to achieve ludonarrative harmony. Their model offered an idea how to look at the games and their creation. This model offers a specific way to achieve harmony while creating a game. By following Despain & Ash model, Söderlund & Hedlund [18] created a demo game and surveyed it to see if it works. This method proves that there are specific steps you can take to create ludonarratively harmonious game. By reviewing various ludonarrative methods, the areas those methods cover are relatively small, mostly used for reviewing papers and some touch to add to games during their creation. After compiling the methods and understanding how they are used, this paper can present the gaps, which are left in ludonarrative approach and application of its methods.

2 LUDONARRATIVE: UNRAVELING THE NEXUS OF GAMEPLAY AND STORYTELLING

In the dynamic realm of video games, the fusion of gameplay and narrative—commonly referred to as **ludonarrative**—has emerged as a pivotal area of exploration. As players traverse virtual landscapes, their decisions and interactions shape the unfolding story. Let us

delve into the theories and methodologies that underpin this intricate relationship.

2.1 Murray's Vision: Games as Narrative Evolution

In her seminal, work *Hamlet on the Holodeck: The Future of Narrative in Cyberspace* [3], Murray posits that games represent a novel phase in narrativity. Players now wield agency, influencing the narrative trajectory through their choices. This interplay between player agency and storytelling defines the ludonarrative landscape.

2.2 Aarseth's Dichotomy: Ludic and Narrative Elements

Aarseth [2] dissects games into distinct components: the **ludic** (pertaining to gameplay mechanics) and the **narrative** (weaving the story). Crucially, Aarseth emphasizes player immersion—the act of becoming an integral part of the game. This perspective views players not merely as external observers but as active elements within the game's fabric.

3 METHODS AS A REVIEW OF LUDONARRATIVE

For this paper, all chosen ludonarrative methods comply with the following:

1) Give an understanding of what is ludonarrative
2) Present which approach was taken to address ludonarrative
3) Approach has a method which can be followed to review or avoid ludonarrative
4) Approach offers steps on how to avoid certain ludonarrative problems.

3.1 User Roles in Interactive Narrative Systems

Louchart et al. [6] delve into the practical implementation of ludonarrative. They categorize user roles within interactive narrative systems, aligning them with desired experiences. Let us explore these roles (Tab. 1):

1) **Spectator**: The passive observer, akin to an audience member at a theater.
2) **Author**: Empowered to shape the narrative, akin to a playwright.
3) **Explorer**: Curious and interactive, akin to an adventurer charting unexplored territories.
4) **Pawn**: Navigating predefined paths, akin to a chess piece following rules.

Narrative traditionally revolves around textual elements, but ludonarrative transcends mere words. It marries game mechanics (ludic aspects) with storytelling (narrative aspects). Even games predominantly narrative-driven fall under the ludonarrative umbrella. Roth [9] underscores that ludonarrative is inherently about meaning-making—a process where gameplay and narrative intertwine.

"1) when players interact with the game system and derive meaning purely based on mechanical level and 2) when they interpret the resulting narrative meaning of game events and their actions. [9]"

Table 1 User roles in interactive narrative systems by Louchart et al. [6]

Role of the user	Description	Interactivity
Spectator	In the sense of a reader or a passive audience. The user witness the work and creativity of the author without possibilities of intervention.	Extremely limited to no one.
Author	The user participates in the creation of story content and its articulation from an authorial perspective without taking part in its unfolding from a character or player viewpoint.	Interactivity is not an issue with this perspective of the user.
Spect-Actor	The user has a limited perception of the story unfolding and has limited interaction with characters concerning their decisions.	Interactivity present but limited by actor's desire to consult the audience.
Participant	As in video-games, the user is immersed in the story from a character perspective and only perceive what he as a character has access to within the limitations of its environment.	Interactivity is present but limited by the story environment and gameplay.

In certain games, player progression mirrors the prescribed narrative. Consider war games, where players assume the role of soldiers. The path forward—whether through combat or strategic choices—aligns seamlessly with the overarching story? Here, ludonarrative harmonizes player agency with the game's predefined trajectory. In the vast landscape of video games, two contrasting approaches emerge: those that align the story with player progression and those that diverge from it. Consider a game where the narrative portrays a pacifist protagonist, yet the player must resort to violence to advance. These divergent approaches yield distinct player experiences, as meaning-making remains subjective.

"Intuitively, we might understand narrative as located within a narrative product like a printed book or a movie. However, the cognitive perspective stresses the point that narrative resides within the human mind as a mental construct." [9]

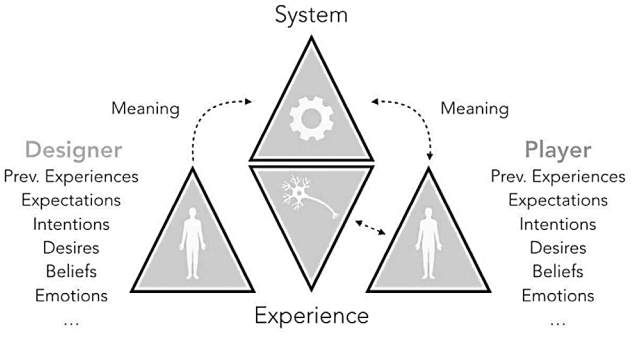


Figure 1 Factors influencing the player's ludonarrative meaning-making [9]

Roth's work (Fig. 1) sheds light on experience dynamics. It has not solely shaped by a player's gaming history but also influenced by their mindset and the game's design. Previous

gaming experiences, education, and real-life encounters all contribute. Meanwhile, a player's mentality—shaped by thought processes and moral compass—interacts with how the game system was crafted. These elements collectively mold the player's journey.

Seraphine [7] takes a different path, seeking harmony between narrative and gameplay. The game creator's direction plays a pivotal role. Understanding the distinction between **incentives** (what drives player actions) and **directives** (how the game guides players) is crucial. Additionally, players often perceive in-game avatars as extensions of themselves. Achieving **ludonarrative coherence**—where game elements seamlessly align with the story—requires a delicate balance. Developers must grapple with relinquishing authorship to empower players as co-authors.

3.2 Exploring Ludonarrative Models: Different Perspectives and Approaches

One of the most comprehensive takes on ludonarrative categories is by Toh [4]. Toh dissected ideas behind the ludonarrative and created a model that is based on a multimodal discourse analysis framework. This method uses a research approach where it is needed to carefully create questions with which you interview others and record the answers. This was done by using different games for which he had multiple sessions of research and interviews until the players concluded playing the games. After everything was done, all the results were gathered and analysed to create the model. The ludonarrative model for video game analysis was constructed by incorporating all the results from the testing. This model can be seen in Toh's work *A Multimodal Discourse Analysis of Video Games:A Ludonarrative Model* [4].

Toh's model is an idea of how to do the ludonarrative analysis for games. It highlights how different elements are tied together and are intertwined. The model was created by using a limited number of games and it needs to be tested and developed even further by creating a larger library of games, which are tested as well with a larger participant group.

A different take on the models has Donoghue [10] who proposes the LAF model (Ludonarrative Analytical Framework). This framework tries to break down games into components. The components were derived from multiple theories and frameworks (Tab. 2).

For easier understanding and visualisation, Donoghue created a ludonarrative scale. This scale is designed in a similar way as a Likert scale and it shows how much harmony or dissonance is occurring in different game components.

This scale was incorporated inside a web of all the components for an easier visual representation. The LAF model was prototyped further by using Close Playing (an adapted version of the close reading method) to gain more understanding of how the games are developed. Tests of the model were done with the Case Study method.

"The ludonarrative analytic framework (LAF) and ludonarrative scale provide an original method to understand, qualify and communicate the effective ludonarrativity within a game." [10]

The LAF method tries to show a way to understand the capacity of ludonarrativity in games as well as a basis for other frameworks to understand how game elements are connected to ludonarrative (Figs. 2 and 3).

Table 2 Donoghue LAF model components [10]

Ludonarrative Category	Provocations for Classification/Scale
Railroad/Sandbox	Is there an ordered set of objectives or goals? What is the player experience and player journey? How much variation would there be between different playthroughs and playstyles? What kind of agency does the player have? Is there a structured temporality? Are levels designed to allow for player choice?
Character Driven/World Driven	What kind of perspective is the player provided? Does it change at any point? If so, what does the change indicate? Where is the narrative focus? How is narrative/gameplay information delivered to the player? Visuals Are there multiple streams of temporality? Micronarrative
Emergent/Embedded	To what degree does the game have an authored narrative/ an authored play experience? How much space is there for player freedom and expression? Focalisation granularity etc. Agency
Actor/Avatar	Where is the player's perspective? How do the ludic systems interact with perspective? Are there diegetic elements to the game? Does the PC have a unique voice? How do they interact with NPCs?
Affect/Cognitive	Do the ludic and physical elements of the game develop an embodied sense of the emotions present in the game? Is the player invested in the emotions present in the game or the game systems? Are there particular moments of emotional or affective responses/triggers in the narrative or within the embodied ludic interface?

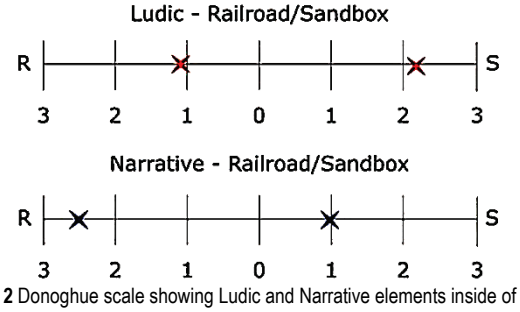


Figure 2 Donoghue scale showing Ludic and Narrative elements inside of Hollow Knight game [10]

A more creative process of creating a method is by research through design as seen in Dodds [11]. He went on to explore what are the design requirements of a ludo-narrative consistent game system. Through the creation and assessment, Dodds evaluated an idea of a ludonarrative consistent game. To accomplish this, he created a few

prototypes (demo game stages) which consist of small narratives influenced by the player's actions. The explored prototype gave a new approach to designing games and creating ludonarrative consistency but did not offer a set system to do this. What can be taken from this work is the approach to each piece of the game. The prototypes explored are Player intention, Narrative agency, Dialogue, Player-character, Story development, and Combat and movement. Those prototypes are a piece of what Dodds calls *trait systems*. One such example is the clumsy trait (Fig. 4).

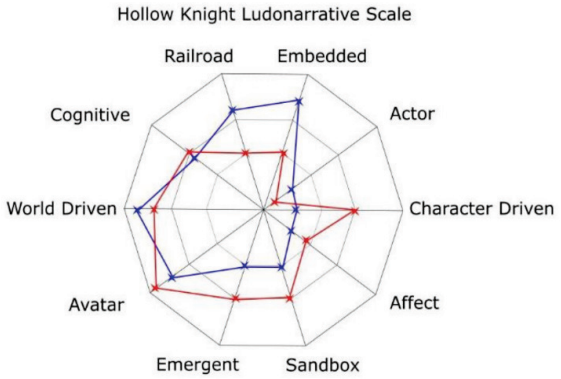


Figure 3 Donoghue LAF model web for the Hollow Knight game [10]

"An example from the prototype is the clumsy trait. The clumsy actions looked for by the system are: rolling in the dirt, rolling in water, breaking objects, 'accidentally' hitting unaggressive non-player-characters (NPCs), and 'accidentally' killing wildlife. Accidentally, in this case, is defined by whether the player continues the behaviour or not. The number of clumsy actions required for the clumsy trait to be assigned is two." [11]

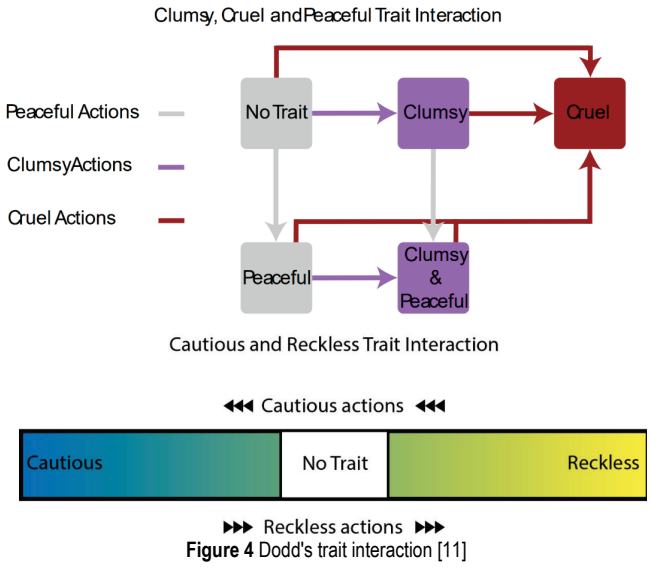


Figure 4 Dodd's trait interaction [11]

There are multiple traits inside the prototype, which are assigned in a specific way as explained in the aforementioned blockquote. It is important to be aware of the different requirements for the trait system. The requirements are narrative potential, mechanical cohesion and trait interaction.

Narrative potential describes how much traits can change and interact with the narrative. This influence can be a minor one like the use of colours in the game for different game choices, or it can be a major one like altering the story progression.

Mechanical cohesion explains how everything is connected, while primarily focusing on the connection between traits, story and game mechanics.

Trait interaction is the connection between multiple traits inside the same system. Some traits can coexist or even influence one another while others do not allow for this. The main basis for using those requirements is ludonarrative consistency. It is important to keep in mind all the different aspects while creating the story. It needs to be created simultaneously with narrative and gameplay in mind or it will be very difficult to incorporate one after the other is finished.

Because of how the trait interaction is made (mainly focusing on the gameplay and interaction), there is a lack of cohesion with the narrative. To fix this problem, the Weather system was created. The weather system used game development software, which allows switching between different light and weather conditions. This produced different times of day (morning, mid-day, evening) or different weather (sunny, cloudy) which would allude to time progression. This is an excellent example of how smaller things can make connections between the main components. To emphasize the weather system and even trait system, Dodds [11] uses artistic freedom to show the distinctions in choices and settings.

In later Aarseth's work [12], his model explains a common ground between narratology and ludology. He offers that games can go from pure story to pure game experience depending on how they are created. Their elements dictate how much story or game elements the game will contain and according to those elements, the game goes from narrative to ludic. This can be used to check the type of the game or certain progressions and see if there is a possible adaptation (connection) between the story and the game. Aarseth called these element events. Although those elements have different uses, they serve the same purpose inside the narrative and the games. Those common elements are:

- World - Ludic and extra ludic spaces where the plot/progress is happening.
- Objects, which determine the degree of player agency in a game.
  - Static (usable objects).
  - Destructible (e.g. buildings in RTS games, where RTS – Real Time Strategy; subgenre of Strategy games. https://blog.acer.com/en/discussion/117/what-is-an-rts-game accessed on 12/12/2023).
  - Changeable (e.g. weapons in Resident Evil 4).
  - Creatable (e.g. armour in World of Warcraft).
  - Inventible (creatures in Spore).
- Agents (or Characters) are classified in terms of their depth/shallowness and their malleability/potential for player control. Agents are divided into three types:
  - Bots (no individual identity).
  - Shallow characters (little or no personality, have names and individual appearance).
  - Deep characters.

- Events can be categorised by their status and the presence of kernels and satellites. They are split into:
- Fully plotted (pure story).
- Dynamic satellites (playable story).
- Dynamic kernels (multipath/quest games).
- No kernels (pure game).

Aarseth claims that every game and story has these 4 elements. It is important to note that some elements are more descriptive of ludical aspects while others are for narrative.

Taking the MDA as its basis, Maraffi [14] has created the SGRplay framework. The MDA framework by Hunicke et al. [13] focuses on game development stages instead of elements. Maraffi uses an idea of the components and creates the three-part survey in combination with the Venn diagram. Using the three types split from MDA, the SGRplay framework analyses three types of play. Maraffi states, "My framework seeks to address claims that gamers are not performers by showing how avatar acting affordances are increasing in many RPG games".

The survey from the framework serves as a connector between game avatars with the players. It tries to evaluate RPG experience as well as player understatement. This is done with a combination of questions, which are asked for each distinct part of the survey (types). Three parts of the survey are Narrative features, Competitive features and Performative features.

The survey then needs to be analysed following the number patterns (answers are numbers, which range from 1 to 10). These numbers are calculated and following a provided formula for the Game Playfulness Factor. When the survey and calculations are completed, the results are then split inside the Venn diagram as shown in Fig. 5.

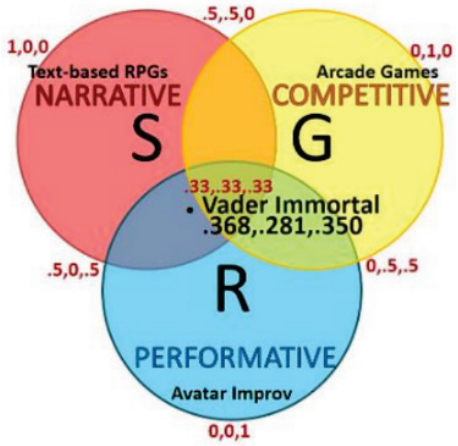


Figure 5 The SGRplay framework - Venn diagram of survey values [14]

In connection to the game avatars and players, Purnomo et al. [15] constructed a formula called GAMING (Gaming system, Attributes, Mechanics, Indexicalities, Narratives and Geosocial systems). One of the main inspirations for the GAMING formula is Hymes' SPEAKING formula ("Hymes' SPEAKING, which stands for Settings and Scenes, Participants, Ends, Act Sequences, Keys, Instrumentalities, Norms, and Genres; The formula is not intended to examine and explore the prosthetic nature of games" [15]). The GAMING formula classifies the game avatars into three categories:

- Gamer-centric: usually created through character creation tools, e.g. Elder Scroll V: Skyrim, Final Fantasy XIV, Soul Calibur VI, etc.
- Game-centric: possible limited customisations for avatars. Mostly possible slight modifications like costumes or growth options.
- Gaming-centric: has options like the previous two, but the avatar's function is different. It is used mostly for communication with other players instead of focusing on the game progression, e.g. Dragon Ball Fighter Z.

Avatar classification is an example of how different avatar types affects the game. Some are more relevant to the game and its progress while others are only a tool for the player to use. Depending on the games and the avatars, classifications through game elements vary. Not all aspects of avatar use are available or needed in different game genres (and game progressions).

	GAMING	SPEAKING
Worlds	**Geosocial systems**	**Settings**
		Scenes
Objects	Gaming systems	Instrumentalities
Agents	**Attributes**	**Participants**
	Mechanics	
Events	Narratives	Act Sequences
	Indexicalities	Ends
		Keys
		Norms
		Genres

Figure 6 Connection between Aarseth [12] method with GAMING and SPEAKING formulas

Fig. 6 presents the connection between the three ideas ([12], Hymes SPEAKING formula and the GAMING formula). It is shown how three different ways of describing certain elements are diversified. In Aarseth's [12] work he has fewer elements, while GAMING and SPEAKING formulas offer few expansions on the same ideas (e.g. Worlds > Geosocial systems > Settings and Scenes).

The diversity of elements shown for the GAMING formula is used because the uses of the different methods and formulas have different goals. If used in cohesion shown in Fig. 6, they serve as a detailed description of certain aspects, like the game avatar types. It stands to show how the communication between the player and avatar is done through game elements. Knowledge gained through the integration of different ideas shows specific relations that can be useful during the game development stage.

3.3 Methods as a Creation of Ludonarrative

In a similar way to Aarseth [12], Hunicke et al. [13] created the MDA framework. Instead of having elements of game, they used stages of game development as a basis for this framework. Because stages differ, there needs to be a way to connect those stages and show their correlations. This framework breaks down the game into components and their design counterparts (Fig. 7).

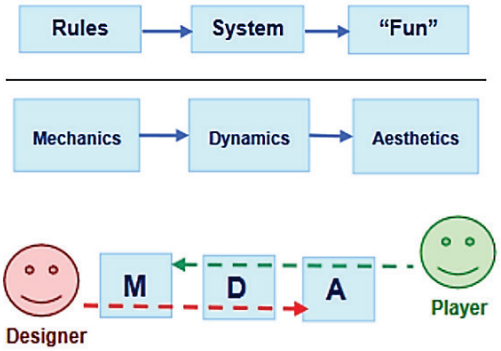


Figure 7 The MDA framework: components, counterparts and designer-player connection [13]

- Mechanics – the technical part of games, their components that are constructed from algorithms (the rules of the game).
- Dynamics – the correlation between the behaviour of the mechanics being influenced by players' inputs (how the game functions).
- Aesthetics – the emotional response in the player during gameplay (experience).

Games use different engaging factors. Some are more focused on cooperation, some on visual feedback while others on challenges. These main elements are usually catered towards a specific target audience and type of game.

Dynamics differ because they need to connect the player with the game. This is accomplished by giving the player an option of how the player can affect the game. If there are various options, different players can play in different ways (or the same player can play the game with different approaches). An example of this idea is Monopoly (1935) which leaves the decisions to the player while always having the same mechanics (throwing dice, buying-selling the properties). Although the rules are always the same, the game varies with the player's approach.

Mechanics are the backbone of games. Those are the various actions which will be allowed for the player to control together with the game's content. Well-structured mechanics create for better game experience.

Tunning is more like the hidden step of the MDA framework. Tunning stands for the acts of testing, analysing and adapting (tunning) the game. An act of tuning happens whenever we try to iteratively refine the game to achieve a certain balance. This can be accomplished if there is a certain idea of how the game should be played. In most cases, the most important thing is a balance between all game aspects. Alongside the components, Fig. 6 shows the connection between the developer sides of the game with the player side. Those two sides can be seen as separate components. The differences in these perspectives are important to grasp. The player's view (experience of the game) starts from the aesthetics part (visual feeling of the game, etc.) and then goes to the mechanics which is the last one explored. The designer's side is the opposite, starting from mechanics which are constructed and going to the aesthetics.

Koenitz [16] proposes IDN framework, which is used more as a tool for interactive storytelling than for game development. Even though it has different use, the application is similar. The main basis for this framework is how it is shown. The basic view is to show it in three parts:

- System – describes the digital artefact. This also includes codes, programs and hardware. (p. 4)
- Process – created with users' interaction with the system. Actions by which the system is used describe the process. (p. 7)
- Product – Narrative created through the act of process (instantiated narrative).

The system is the primary part. After the user starts to engage with the system, the process is created. Since this process can vary, this results in different products, which Koenitz calls instantiated narrative. IDN framework finds its use in the analysis. It can be applied to diverse works to understand the structure of the narrative and storytelling. To show this idea in a better way he created a subcategory of system called Protostory. This subcategory consists of 4 parts (Fig. 8); Environment Definitions, Assets, Settings and Narrative Design.

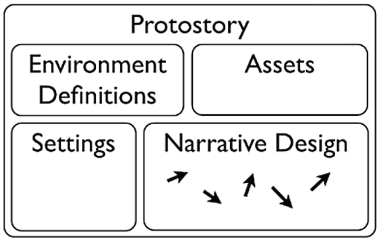


Figure 8 Koenitz IDN System elements [16]

To keep a flexible use of the system, the term narrative design is used. The narrative design describes the flexibility of narrative presentation while the vectors (arrows) show specific directions in those narratives.

One of the most practical methods is proposed by Despain & Ash [17]. They have created a method that is used for game development while also achieving ludonarrative harmony. The process of creating the model was done by establishing two different game designs for the two different directions (ludonarrative harmony and ludonarrative dissonance). While keeping the flow between the elements, the harmony should not be broken and if strayed from that, dissonance is created. Fig. 9 shows the two models.

The two models can be used together to keep track of possible deviations from harmony to dissonance.

Phases presented in Ludonarrative Harmony model are used for testing. Phase One analyses the 4 different outlines of the model (mechanic, narrative, context and emotional fulfilment) which are put in a loop to show their correlation. The 4 outlines are used to identify core aspects of each (actions taken by the player for mechanics; how the story is told for narrative and how it adds to the mechanics; context is the reason behind those actions and how it is communicated to the player why this story exists; emotional fulfilment shows connection and reason to continue using certain actions). For an easier understanding of how to explain those elements, Despain and Ash have created 4 questions:

- Mechanic: How is the player doing an action?
- Narrative: Why is the player doing the action?

- Context: Where is that action happening (broadly); Does the action fulfil the narrative; and why does the action matter?
- Emotional Fulfilment: What purpose does that action serve for the player?

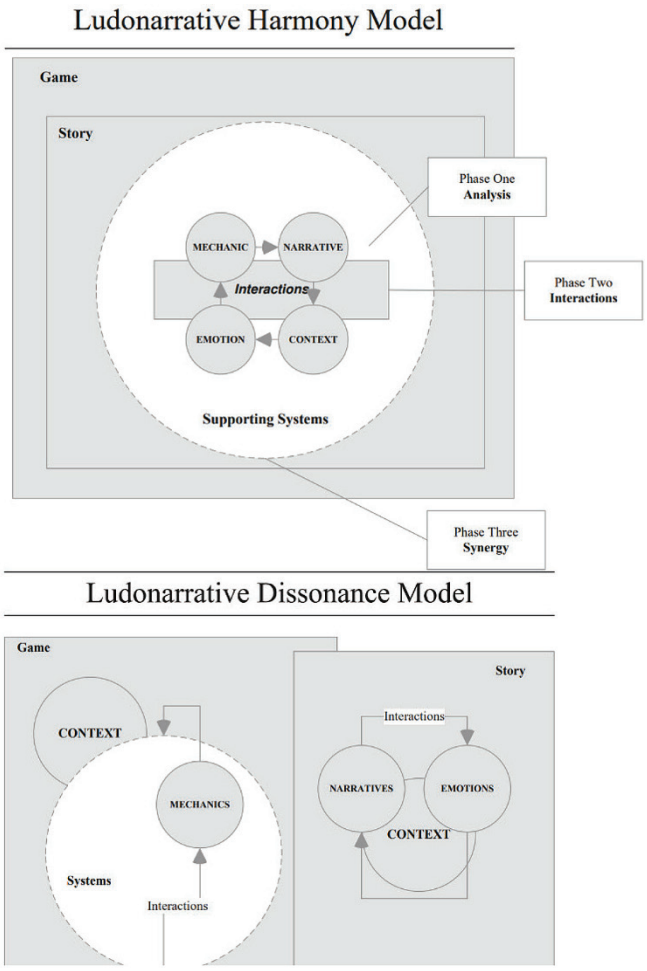


Figure 9 Ludonarrative harmony model and Ludonarrative dissonance model [17]

The models use careful iteration process during which the elements are explained, polished and structured. Proposed way of using this model is during the game creation. This way it is possible to always come back and iterate what works and what does not in the model for easier game development. Despain and ash with the proposal of the model also created a demo game, which they present through the model examples. Later Söderlund and Hedlund [18] created an aretfact using the same models and then tested it, proving the validity of Despain and Ash model.

4 DISCUSSION

Methods presented here are some of the works done regarding the topic of ludonarrative. They are created to be used as a way of studying or understanding ludonarrative and its possible problems. The table below shows methods reviewed in this paper and which ones have a practical or theoretical use.

Table 3 Summarization of methods

Offers… / Authors	An understanding of ludonarrative	A way to review or understand ludonarrative in a game	A way to create or avoid ludonarrative in a game
Aarseth	✓	✓	✗
Murray	✓	✗	✗
Toh	✓	✓	✗
Jenkins	✓	✗	✗
Louchart et al.	✓	✓	✗
Seraphine	✓	✓	✗
Roth	✓	✓	✗
Donoghue	✓	✓	✗
Dodds	✓	✓	✗
Hunicke et al.	✓	✗	✓
Maraffi	✓	✓	✗
Purnomo et al.	✓	✓	✗
Koenitz	✓	✗	✓
Despain and Ash	✓	✓	✓

As shown in Tab. 3, most of the methods reviewed in this paper provide a deeper understanding of ludonarrative. Even though most of the authors help us on how to find and understand ludonarrative in games, only some provide a practical way of creating or avoiding ludonarrative.

Jenkins [5] presents an idea that interaction between the player and the game needs to be catered and created. To create a narrative in the game it is important to portray the story the game wishes to tell the player. Some games are too abstract to tell that story, such as *Tetris* (1984). In addition, because of the abstract nature of some games, it is important to look at the whole concept of the game and not only the game itself. To see the whole concept, we try to structure the story and the narrative reasonably [19]. That is why it is important to look at the whole concept of the game and not only the game itself. Games are special because of the possible interaction between the player and the medium used. Games invite the player to interact with the storyline which in turn makes them more immersed in the game [20]. Jenkins [5] offers Environmental storytelling as a solution to connect the player and the game. Environmental storytelling is a notion of combining physical space and its elements with the story previously known to most.

Theoretical methods help us understand the problem, but practical ones can be more suited for not having the problem. Most of the work focus on understanding how something is created. Although there are proposals for how to avoid, for example ludonarrative dissonance, there are not plenty of works done on proposing and showing a clear way to do so. Authors like Aarseth, Murray and Toh present an understanding of ludonarrative. That understanding always goes further with each subsequent work from other authors going into details that are more specific. Each step is questioned, tested, reviewed and understood. Authors like Donoghue even propose ways on how to review the games and present the results. Dodds, by keeping all this in mind creates a demo game. Although the development of this demo game has valid steps, it is hard to reproduce. This leaves us with a question how to create a game while adhering to knowledge gained through all those works. Ludonarrative

dissonance, depending on how it is viewed, can be problematic for some games. By considering the games impact, there is a possibility to reduce the negatives while also further enhancing player's experience. As proposed by Despain and Ash [17] and Swords [21], there are approaches to game development which take into account game creation. Those methods are not meant to be used as a reviewing tool of the game after its release, but as a guidance tool. It could be more theoretical where certain steps are followed with open structure (Despain and Ash) or more structured where specific steps need to be satisfied (Swords). Swords has created *The Forest Paths* methods that is used for narrative design. It takes specific steps in which storytelling is enabled using game as a medium.

We suggest that by reviewing other methods that are used for checking ludonarrative instances in games, more structured methods can be created which don't allow for mistakes, which can lead to loss of game immersion, game interest or game playability. Those methods would need to be catered for beginning stages of game development when the structure is most important. Such methods can be followed as a guide to know exactly what needs to be developed and if it does not work, check if something new can be added while providing value to the game. One of the problems, which occur after the game is finished, is unintentional ludonarrative dissonance. This can be brought by not having a clear understanding of fundamental game elements and their cohesion to others. It often happens that some things do not work or could not be developed so others are quickly added to fill the gap. That not only methods made for early stages of game development can guide developers, but they can be used as a teaching tool for new people in the industry or even students. This paper serves a research purpose for the creation of the methodology, which will be created and presented. This method will focus on finding applications for the ideation stage of game development from already mentioned methods.

5 CONCLUSION

With deeper understanding of how games work, more value is presented to the developers of said games. Studies such as ones on ludonarrative provide an informative outlook on games as finished products and their impact on players. That information can be valuable for the new games being created. Although information is helpful, it takes a lot of time and effort to adapt it into creative processes.

Authors that are reviewed in this paper took that information and created methods that either help with reviewing which parts cause ludonarrative or offer structured ways on how to design games to avoid some of the ludonarrative pitfalls. Methods such as those can be used inside game development to grasp an idea on how to approach specific game elements during the development. Depending on game type, methods from authors such as Despain and Ash, Koenitz or Swords can provide tools or ideas on how to design and avoid certain ludonarrative misteps. Even though those methods can be used, it does not mean they are catered for various game genres and types. Methods presented are only a small piece of game development that needs to be used carefully as to not lock ourselves into a box on how you should go about creating a game. Ideally methods used in cohesion with the right approach and adaptability could be more suited for ideation stage of game development. If a method is created that adheres to those other methods and expands upon them with a broader perspective, it could help game developers to create better games while adhering to target audience.

6 REFERENCES

[1] Hocking. C. (October 07, 2007.) Ludonarrative Dissonance in Bioshock: The problem of what game is about [Blog post]. Click Nothing, TypePad. Retrieved April 15, 2024 from https://www.clicknothing.com/click_nothing/2007/10/ludonarrative-d.html

[2] Aarseth, E. J. (1997). Cybertext, Perspectives on Ergodic Literature. *Leonardo Music Journal, 8*, 77. https://doi.org/10.2307/1513408

[3] Murray, J. H. (1997). *Hamlet on the Holodeck: The Future of Narrative in Cyberspace*. Simon and Schuster.

[4] Toh, W. (2015). A Multimodal Discourse Analysis of Video Games: A Ludonarrative Model. *Thesis*.

[5] Jenkins, H. (2002). Game Design as Narrative Architecture. Computer, 44.

[6] Louchart, S., Aylett, R., Dias, J., Figueiredo, R. & Paiva, A. (2008). Authoring emergent narrative-based games.

[7] Seraphine, F. J. N. (2016). Ludonarrative Dissonance: Is Storytelling About Reaching Harmony?

[8] Frasca, G. (2013). Simulation versus Narrative: Introduction to Ludology.

[9] Roth, C. (2019). Video Game Violence from the Perspective of Cognitive Psychology. Role Identification and Role Distancing in A WAY OUT. In Video Game Violence from the Perspective of Cognitive Psychology. Role Identification and Role Distancing in A WAY OUT (pp. 53–62). Transcript Verlag. https://doi.org/10.1515/9783839450512-006

[10] Donoghue, A. (2021). The quest for ludonarrative: Designing a ludonarrative analytical framework for analysis and design of ludonarrative in games through analysed case studies: Hollow Knight, Celeste, and Brothers: A Tale of Two Sons

[11] Dodds, R. (2021). An Exploration of Ludonarrative Consistent Game Systems.

[12] Aarseth, E. (2012). A narrative theory of games. Proceedings of the International Conference on the Foundations of Digital Games, 129–133. https://doi.org/10.1145/2282338.2282365

[13] Hunicke, R., LeBlanc, M. & Zubek, R. (2004). MDA: A Formal Approach to Game Design and Game Research.

[14] Maraffi, C. (2022). Multimodal Framework for Enhancing RPG Playfulness through Avatar Acting Affordances (SSRN Scholarly Paper 4183800; Issue 4183800).

[15] Purnomo, Sf. L. A. Purnama, Sf. L. S., Untari, L., Asiyah, N. & Anggraini, N. (2019). Gamemunication: Prosthetic Communication Ethnography of Game Avatars. *Jurnal Komunikasi: Malaysian Journal of Communication, 35*(4), Article 4. https://doi.org/10.17576/JKMJC-2019-3504-01

[16] Koenitz, H. (2010). Towards a Theoretical Framework for Interactive Digital Narrative. In R. Aylett, M. Y. Lim, S. Louchart, P. Petta, & M. Riedl (Eds.), Interactive Storytelling (pp. 176–185). Springer. https://doi.org/10.1007/978-3-642-16638-9_22

[17] Despain, W. & Ash, L. (2016). Designing For Ludonarrative Harmony.

[18] Söderlund, S. & Hedlund, R. (2022). Mekaniker och narrativ i harmoni.
[19] Lindley, C. (2005). Story and Narrative Structures in Computer Games 1.
[20] Pearce, C. (2004). Towards a Game Theory of Game. First Person: New Media as Story, Performance, and Game.
[21] Swords, A. (2023, May 21). The Forest Paths Method for Narrative Design. Forest Paths Method.

Authors' contacts:

Enio Vešligaj
University North,
104. brigade 1, 42000 Varazdin, Croatia

Andrija Bernik
(Corresponding author)
University North,
104. brigade 1, 42000 Varazdin, Croatia
abernik@unin.hr

ISSN 1846-6168 (Print), ISSN 1848-5588 (Online)
https://doi.org/10.31803/tg-20240829131515

Original scientific paper
Received: 2024-08-29, Accepted: 2024-11-22

The Impact of Design Team Characteristics on Construction Project Performance with the Mediating Role of Construction Project Costs

Parisa Khorashadi Moghaddam*, Nasrin Izadian, Mehdi Haghighatjoo, Amirhossein Moghaddas Jafari, Marzieh Zahedi

Abstract: The aim of this research was to examine the impact of design team characteristics on construction project performance with the mediating role of construction project costs. A total of 212 project managers from construction projects participated in the study. Data were collected using a questionnaire and analyzed by PLS-SEM using SMARTPLS3 software. The results indicated that design team characteristics significantly and positively influence the performance of construction projects. Design team characteristics also have a significant and negative impact on construction project costs. The effect of construction project costs on construction project performance was found to be negative and significant. Furthermore, the mediating role of construction project costs in the relationship between design team characteristics and construction project performance was positive and significant. As a result, it can be inferred that the characteristics of the design team in construction projects lead to improved performance of these projects by reducing construction project costs.

Keywords: construction project costs; construction project performance; design team characteristics

1 INTRODUCTION

The construction industry, recognized for its complexity within project-oriented industries, has been the subject of extensive study by researchers in project management. This industry significantly contributes to development and the attainment of societal objectives. The construction industry represents a significant portion of the economy, accounting for around 10% of the gross national product in developed nations. This industry is characterized by its complexity, as it encompasses a variety of participants such as clients, contractors, consultants, stakeholders, and regulatory bodies. Additionally, the effectiveness of the construction sector is closely linked to the overall performance of the national economy. According to Yan [1], the contribution of the construction sector to the economy varies significantly, with advanced economies experiencing an added value between 7 and 10 percent, while developing nations see a contribution ranging from 3 to 6 percent. These figures highlight the crucial economic significance of the construction industry and its essential role in the growth and advancement of countries worldwide [2, 3]. Hence, enhancing processes and implementing effective project management strategies within this sector can lead to improved productivity, which in turn may positively influence the broader national economy.

Despite the significant contribution of the construction industry to the economies of developing countries and its crucial role in their development, the performance of this industry remains generally low. As Idoko and Ifediora [4] noted, many projects in developing countries face significant increases in time and costs, and may even become entirely abandoned either before or after completion, failing to achieve their intended benefits. Moreover, the growth of the construction sector in developing nations tends to be significantly slower than that of other industries within those nations, as well as in comparison to the construction sectors in more developed countries. Achieving success in construction projects is a key objective for project stakeholders, including both owners and contractors, which has led to considerable research focused on identifying the elements that contribute to project success. In addition, public project owners are exploring various delivery methods tailored to specific project characteristics to enhance the success rates of construction initiatives and are committed to selecting the most suitable contractor for each individual project [5]. Project success is an abstract concept, and no widely accepted general definition exists. Based on research literature, the following fundamental issues have been reported regarding construction project performance: low quality, budget overruns, lateness, unsafe construction, and customer dissatisfaction [6, 7]. Additionally, construction projects frequently encounter intricate issues involving various stakeholders. These challenges include disputes among team members, such as clients and contractors, as well as resistance from external groups like local communities affected by the projects [8]. In recent years, the landscape of construction projects has grown increasingly demanding for both contractors and clients due to stringent budgetary constraints and tight schedules. Consequently, to enhance the likelihood of successful project outcomes, it is essential to identify the factors that may contribute to either the success or failure of construction initiatives in order to optimize overall project performance.

The construction sector is characterized by its fragmented structure. Historically, the design phase has been viewed as a distinct entity, separate from the construction phase of a project. Construction teams typically start fresh for almost every new project. Research by Evbuomwan and Anumba [9] indicated that one of the contributing factors to subpar delivery outcomes within the construction industry is the lack of collaborative efforts among project stakeholders. The results of research studies highlight that effective process and team integration play a crucial role in facilitating the essential transformations needed for improved outcomes in the construction sector. Nonetheless, merely assembling individuals does not inherently result in effective teamwork. Teams in construction begin to form as they embark on new projects, aiming to showcase their worth through their performance and the strength of their collaborative dynamics. To achieve successful project outcomes, it is vital for

construction firms to focus on enhancing, assessing, and evaluating the efficiency of their teams [10].

In any construction project, it is essential to involve key stakeholders, which typically consist of the client, contractors, and the design team, including architects and engineers. Designers hold a pivotal position, as their responsibilities cover the entire project lifecycle, from inception to finalization [11] . Notably, critical decisions that influence overall costs are primarily made during the initial phases of the design process. Research studies have shown that key characteristics of design teams in construction projects are: procurement, skills, team experience, communication, collaboration, and motivation for innovation [12, 13]. Procurement encompasses the processes involved in selecting and engaging a team responsible for the design and construction of a project. It plays a critical role in shaping both the delivery and financing aspects of a project. Additionally, procurement impacts the level of trust among stakeholders, as well as facilitating open communication and interactions among team members [14, 15]. Furthermore, it significantly influences risk management strategies within construction projects. The term 'skills' pertains to the collective expertise possessed by the design team, with the exclusion of the client's abilities. Team experience pertains to the construction experience and overall project background of the team. Communication refers to the channels of communication during the project execution process. Collaboration pertains to the team members' feelings regarding teamwork and participation during the design phase. Motivation for innovation refers to the efforts of team members to introduce innovative ideas and solutions [12, 16, 17]. Research shows that the characteristics of the design team have an impact on project success [10, 18]. Some research studies highlight the necessity of integrating ecological considerations into the urban design process [19, 20] or municipal solid waste management [21] to foster sustainable development, thus underscoring the importance of stakeholder collaboration in achieving effective environmental outcomes within construction projects. Moreover, addressing the cognitive demands associated with urban design processes during construction projects could further mitigate risks and optimize performance across diverse stakeholder groups [22]. These findings suggest that augmenting these characteristics may facilitate the attainment of project objectives and enhance overall outcomes. Therefore, it is essential to conduct a comprehensive and earnest evaluation of these dimensions during the initial phases of design and planning for construction projects. Such an approach can lay a robust foundation for the future success of these endeavors.

As previously mentioned, the construction industry is a key factor in the development of any city and plays a significant role in job creation. However, it has often experienced stagnation due to various internal and external reasons. Among these, the most important issues are time delays and cost overruns, which severely impact construction project performance. Construction performance can be evaluated through completion scheduling, completion costs, productivity of completed tasks, and safety [23, 24]. This is the primary reason many projects remain unfinished, leading developers to move on to the next project. Understanding the impacts and factors influencing time delays and cost increases is crucial [25]. Research has also indicated that construction project costs significantly affect project success [23, 26]. In summary, the issue of construction project performance manifests itself in various ways. Many completed projects fail in terms of time performance, while others fail in cost performance or in other performance metrics. Projects have frequently faced setbacks throughout history, including client-related problems, material access issues, road closures, design modifications, added work, and delayed decisions. Consequently, examining the factors influencing the performance of construction projects has become a key issue in this field. A review of the research background indicates that, until now, no study has explored the impact of design team characteristics on the performance of construction projects with the mediating role of construction project costs within the framework of structural equation modeling (SEM). Hence, the fundamental objective of the present study is to investigate the effect of design team characteristics on the performance of construction projects, with construction project costs serving as a mediating variable. This aim is intended to enrich the research literature and empirical evidence in this area and to take a step towards improving the performance of construction projects. Drawing upon existing theoretical literature and the framework established through prior studies, the conceptual model for this research is illustrated in Fig. 1. As can be seen, the characteristics of the design team in construction projects are considered as the independent variable, construction project costs as the mediating variable, and construction project performance as the dependent variable.

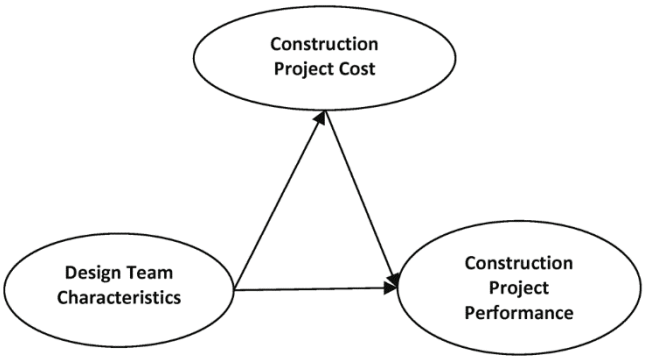


Figure 1 The conceptual model

Thus, the hypotheses of the research are as follows:

H1: The characteristics of the design team in construction projects influence the performance of construction projects.

H2: The characteristics of the design team in construction projects influence the costs of construction projects.

H3: The costs of construction projects influence the performance of construction projects.

H4: The costs of construction projects play a mediating role in the impact of the characteristics of the design team in construction projects on the performance of construction projects.

2 RESEARCH METHODOLOGY

This research employs a descriptive correlational design utilizing SEM with partial least squares (PLS) to investigate the intricate relationships between key variables. SEM is particularly advantageous in this study as it effectively captures complex interrelationships and assesses the impact of independent variables on dependent variables, even in the presence of small sample sizes and non-normally distributed data [13, 27]. By leveraging the strengths of SEM, this research provides a comprehensive and valid analysis of the relationships among the examined variables. This approach not only enriches our understanding of the underlying factors at play but also advances our knowledge of the social and economic phenomena being studied.

2.1 Population

The target population for this research consisted of construction project managers in Iran who possess at least five years of professional experience. To collect the requisite data, an initial phone contact was established to solicit participation, following which a questionnaire was distributed to 260 of these managers. A total of 212 questionnaires were completed and returned, yielding a response rate of 82%. This high level of participation underscores the willingness and commitment of project managers to contribute their insights and experiences, thereby enhancing the credibility of the findings derived from this research.

2.2 Measures

To measure the characteristics of the design team, the questionnaire developed by Hu & Skibniewski [12] was utilized. This questionnaire consists of 6 items. For assessing the performance of construction projects, the questionnaire by Nguyen and Watanabe [7] was employed, which contains 7 items. To measure the costs of construction projects, indicators introduced in the study by Olawale and Ming [28] were used with slight modifications. This questionnaire includes 6 items. The variables were assessed using a five-point Likert scale, with response options ranging from 1 (indicating a very low level) to 5 (indicating a very high level).

3 RESULTS

3.1 Measurement Model Test

To assess reliability, Cronbach's alpha and composite reliability were used, while validity was evaluated using factor loadings, average variance extracted (AVE), and the Fornell-Larcker criterion. The composite reliability index proposed by [29, 30] is superior to Cronbach's alpha because, in Cronbach's alpha, all observable variables in each measurement model have equal weights. This effectively equalizes their relative importance. In contrast, composite reliability does not make this assumption; it actually uses the factor loadings of the items during its calculation, thus making composite reliability values generally higher and more accurately reflective compared to Cronbach's alpha. The threshold for this index, similar to Cronbach's alpha, is a value of 0.7 or higher for internal consistency within the measurement model. Regarding the factor loading of each item, a factor loading of 0.6 or greater in confirmatory factor analysis indicates that the construct is well defined [31]. As shown in Tab. 1, the factor loadings for the research variables are above 0.6, thus confirming the factor loadings.

Table 1 Reliability

Variables	Item	Factor Loading	Cronbach's alpha	Composite reliability	AVE
Design Team Characteristics	1	0.826	0.918	0.936	0.710
	2	0.850			
	3	0.873			
	4	0.847			
	5	0.868			
	6	0.789			
Construction Project Cost	1	0.718	0.834	0.874	0.539
	2	0.823			
	3	0.818			
	4	0.678			
	5	0.675			
	6	0.677			
Construction Project Performance	1	0.746	0.880	0.907	0.582

A critical point here is that if, after calculating the factor loadings, we encounter values less than 0.6 between the construct and its indicators, we need to revise those indicators (questionnaire items) or remove them from our research model. To examine convergent validity, the AVE index was used. Alipour et al. [31] suggest that AVE values should reach at least 0.5, which signifies that a construct accounts for roughly 50% or more of the variance observed in its indicators [30]. In Tab. 1, the factor loadings, composite reliability, and AVE for the variables investigated in this study are displayed. The results presented in Tab. 1 demonstrate adequate and suitable reliability for the constructs examined.

The results of discriminant validity are reported in Tab. 2. It indicates that the square root of the AVE for each research variable exceeds their correlations with other variables. These findings indicate the appropriate validity of the measurement tools.

Table 2 Correlation Matrix and Square Root of AVE

Variables	Design Team Characteristics	Construction Project Cost	Construction Project Performance
Design Team Characteristics	0.84		
Construction Project Cost	–0.53**	0.73	
Construction Project Performance	0.59**	–0.63**	0.76
$^{**}p < 0.01$			

3.2 SEM Testing

To evaluate the performance of construction projects, a conceptual model was tested using PLS-SEM. The relationships among the research variables and the tested model are depicted in Fig. 2. It demonstrates that the

characteristics of the design team have a positive and significant effect on the performance of construction projects, whereas their influence on project costs is negative and significant. Additionally, the costs associated with construction projects have a negative and significant impact on overall project performance. The figures within the circles indicate the explained variances for each of the research variables. Tab. 3 presents the estimated path coefficients alongside the variance explained for the variables examined in the study.

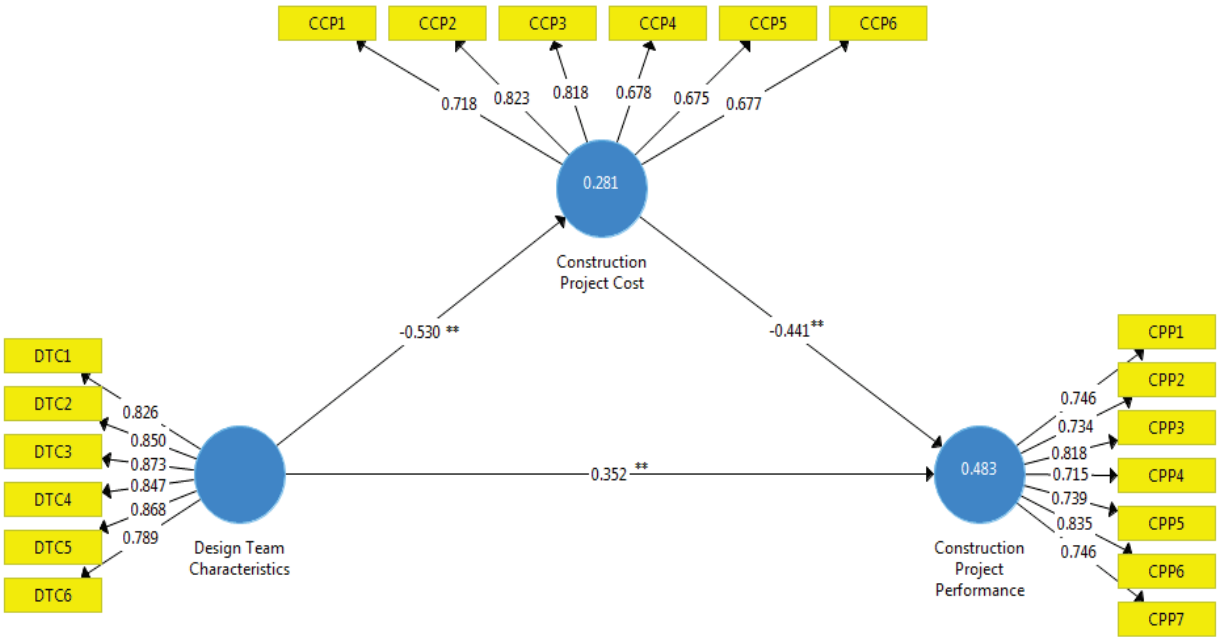


Figure 2 Tested model

Table 3 Path coefficients and explained variance

Variables	Path Coefficient	*t*-value	*p*-value	Explained Variance
On construction project performance via: Design team characteristics Construction project costs	 0.352** -0.441**	 4.135 6.389	 0.001 0.001	0.483
On construction project costs via: Design team characteristics	-0.53**	10.619	0.001	0.281
The mediating role of construction project costs	0.23**	5.711	0.001	-

*$p < 0.05$, **$p < 0.01$

Table 4 Hypotheses Results

Hypotheses	Results
H1: The characteristics of the design team in construction projects influence the performance of construction projects.	Confirmed
H2: The characteristics of the design team in construction projects influence the costs of construction projects.	Confirmed
H3: The costs of construction projects influence the performance of construction projects.	Confirmed
H4: The costs of construction projects play a mediating role in the impact of the characteristics of the design team in construction projects on the performance of construction projects.	Confirmed

According to Tab. 3, the mediating role of construction project costs in the relationship between design team characteristics and construction project performance is positive and significant. Furthermore, 48% of the variance in construction project performance and 28% of the variance in construction project costs can be explained by the variables in the research model. The results of the hypothesis testing are presented in Tab. 4.

The goodness-of-fit index (GOF) measures the model's predictive ability, particularly in relation to the endogenous latent variables. In this study, the calculated absolute GOF for the proposed model was 0.48, suggesting that the model demonstrates a good fit. A GOF value exceeding 0.36 is considered indicative of satisfactory model quality.

4 DISCUSSION

This research sought to establish a model that examines how the characteristics of design teams influence the performance of construction projects, considering project costs as a mediating factor through SEM. Findings revealed that the model aligns well with the collected data, accounting for 48% of the variability in construction project performance and 28% of the variability in project costs.

The results indicated that design team characteristics have a significant positive impact on the performance of construction projects; however, their impact on construction project costs is significant and negative. Therefore, design team characteristics lead to a reduction in construction project costs and, consequently, improve the performance of construction projects. This finding is consistent with the results of the research conducted by Radhakrishnan et al. [18] and Azmy [10]. To explain this finding, it can be stated that

if the selection and hiring process for the design team of a construction project is based on competency and capability, several key factors come into play. When the design team possesses both collective and specialized skills, has relevant construction and project experience, maintains essential communication channels, fosters collaboration and participation among team members, and is motivated to generate innovative ideas and solutions, it ultimately leads to a reduction in construction project costs. Consequently, this approach significantly improves the overall performance of the construction project. Teamwork is defined as "a group of individuals who are mutually dependent on each other to achieve a common goal." It encompasses commitment, collaboration, clear objectives, and goals. Commitment, accountability, and the skills of team members are some of the essential characteristics required for team members to reach their common goals. Therefore, for effective teamwork, it is essential for team members to establish a shared understanding of behavioral standards, project planning, role allocation, and task management, including scheduling and decision-making processes. Addressing these aspects is crucial for minimizing misunderstandings throughout the design phase of construction projects.

Teamwork is also important due to the increased efficiency it can provide, and since it resembles a structure where members work in a planned manner through both collaboration and communication, it is well-organized, and there is regular information exchange about ideas and products. For this reason, projects that aim to reduce costs and increase effectiveness prefer to utilize teamwork.

The findings demonstrate that construction costs have a notable adverse impact on the performance of construction projects, suggesting that higher expenses are linked to reduced effectiveness in project outcomes. This finding is consistent with the results of studies by Molavi and Barral [23]; and Trach et al. [26]. To explain this finding, it can be said that increasing costs in projects results in dissatisfaction with project quality, project schedules, project costs, reduced productivity, and lack of progress in the projects, ultimately leading to a decline in the performance of projects [32]. Furthermore, as the findings indicate, the characteristics of the design team have an impactful role in reducing construction costs, thereby improving construction project performance. A team operates with shared goals and objectives, enabling members to build strong, collaborative relationships aimed at achieving these goals. Effective teamwork involves individuals working together in a cooperative setting, where knowledge and skills are exchanged to reach common objectives. Research highlights that a primary characteristic of successful teams is their commitment to a clear, collective purpose. Teams play a crucial role in numerous projects, with successful collaboration depending on the synergy among participants. This collaborative spirit fosters a positive environment, encouraging all members to contribute actively to the team's success and effectiveness. Team members must exhibit the flexibility to thrive in collaborative settings where success is attained through cooperation and shared goals, rather than through competition and individual aims [33]. Consequently, the attributes of a construction project design team can be viewed as a vital factor in achieving competitive advantage within the construction industry.

5 PRACTICAL IMPLICATIONS

The results of this research offer substantial insights that can significantly enhance the performance of construction projects. The following key findings highlight practical applications that stakeholders may consider implementing:

The study underscores the importance of design team attributes—specifically competence, skill, and experience—as critical factors in minimizing project costs and improving overall performance. Consequently, it is advisable for organizations to prioritize these characteristics as essential criteria during the selection and recruitment process for design team members. Evidence suggests that robust communication channels within design teams facilitate enhanced collaboration and engagement. It is imperative for organizations to establish conducive platforms for the exchange of ideas and perspectives among team members, as this can lead to improved cooperation and cost efficiency.

The research highlights the value of fostering an environment that motivates team members to propose innovative solutions. By nurturing a culture that encourages the free sharing of ideas, organizations can significantly elevate project performance outcomes. The findings emphasize that successful construction projects hinge on effective teamwork. It is crucial for organizations to cultivate a collaborative culture and enhance teamwork competencies among design team members, thereby reinforcing the collective skill set necessary for project success. The study demonstrates that reducing costs not only improves project performance but also heightens customer satisfaction and ensures timely delivery. Therefore, prioritizing comprehensive cost management strategies throughout all phases of construction projects is essential. Clarifying and establishing shared objectives among design team members can foster synergy and enhance collaborative efforts. By promoting a collective understanding of project aims, team members are more likely to work cohesively toward achieving the overall objectives. By implementing these findings, stakeholders in construction management can significantly improve project outcomes, fostering higher efficiency, satisfaction, and success rates in construction endeavors.

6 CONCLUSION

This study provides strong evidence that the costs associated with construction projects play a crucial role as a mediator between the characteristics of design teams and the overall performance of construction projects. Our results suggest that positive attributes of design teams not only improve project performance but also help in reducing costs, resulting in more favorable project outcomes. By recognizing and nurturing key traits such as collective expertise, effective communication, teamwork, and a unified commitment to objectives, stakeholders can enhance the process of selecting design teams and create an atmosphere that values efficiency and innovation. These traits align well with the project life cycle and significantly boost the likelihood of team success.

Additionally, comprehending the relationship between the qualities of design teams and team dynamics is essential for advancing project management methodologies. Our research highlights the need to cultivate an environment that promotes collaboration and problem-solving, which can lessen the time needed for conflict resolution and improve the quality of ideas exchanged, thereby influencing project schedules and results. However, this study is limited by its focus on project managers from construction projects in Iran, which may restrict the applicability of the findings. Future studies should aim to include a wider range of geographical locations and project settings to gain a comprehensive understanding of the results' relevance. In conclusion, our findings emphasize the importance of investing in the development of design teams as a strategic method to enhance construction project performance. We urge industry professionals and researchers to delve deeper into the intricate characteristics of design teams and their effects on project success, thereby enriching the field and promoting improved practices in construction management globally. In future research, it would be beneficial to include qualitative studies such as Ghorashi et al. [34] and Darvishinia [35] that explore people's experiences and opinions about design teams. Additionally, using data mining such as Bevilacqua et al. [36], deep learning such as Tashakkori et al. [37] and artificial intelligence techniques such as Espahbod et al. [27], Metaheuristics algorithms such as Bahadoran Baghbadorani et al. [38] can help analyze data more effectively and provide deeper insights into team performance and project outcomes.

7 REFERENCES

[1] Yan, J.-K., Zheng, Z., Zhou, Y.-C., Lin, J.-R., Deng, Y.-C. & Lu, X.-Z. (2023). Recent research progress in intelligent construction: A comparison between china and developed countries. *Buildings, 13*(5), 1329. https://doi.org/10.3390/buildings13051329

[2] Roshdieh, N. & Farzad, G. (2024). The effect of fiscal decentralization on foreign direct investment in developing countries: Panel smooth transition regression. *International Research Journal of Economics and Management Studies (IRJEMS), 3*(7), 133-140.

[3] Farzad, G. & Roshdieh, N. (2024). The interplay of destructive work behaviors, organizational citizenship behaviors, and fiscal decentralization: Implications for economic development in developing countries. *International Research Journal of Economics and Management Studies (IRJEMS), 3*(8), 1-8.

[4] Idoko, O. R. & Ifediora, C. O. (2021). Evaluation of procurement systems for project success in real estate construction in Onitsha Anambra State, Nigeria.

[5] Ingle, P. V., Mahesh, G. & Deepak. (2021). Identifying the performance areas affecting the project performance for indian construction projects. *Journal of Engineering, Design and Technology, 19*(1), 1-20. https://doi.org/10.1108/JEDT-01-2020-0027

[6] Yazdi, M. R. T., Mozaffari, M. M., Nazari-Shirkouhi, S. & Asadzadeh, S. M. (2018). Integrated fuzzy DEA-ANFIS to measure the success effect of human resource spirituality. *Cybernetics and Systems, 49*(3), 151-169. https://doi.org/10.1080/01969722.2018.1448221

[7] Nguyen, L. H. & Watanabe, T. (2017). The impact of project organizational culture on the performance of construction projects. *Sustainability, 9*(5), 781. https://doi.org/10.3390/su9050781

[8] Mok, K. Y. Shen, G. Q. & Yang, J. (2015). Stakeholder management studies in mega construction projects: A review and future directions. *International journal of project management, 33*(2), 446-457. https://doi.org/10.1016/j.ijproman.2014.08.007

[9] Evbuomwan, N. F. & Anumba, C. J. (1998). An integrated framework for concurrent life-cycle design and construction. *Advances in engineering software, 29*(7-9), 587-597. https://doi.org/10.1016/S0965-9978(98)00024-6

[10] Azmy, N. (2012). *The role of team effectiveness in construction project teams and project performance*. Iowa State University.

[11] Bevilacqua, C. & Sohrabi, P. (2020). Networking analysis in the urban context. Novel instrument for managing the urban transition. *Urbanistica Informazioni, 12*(Special Issue), 6-10.

[12] Hu, M. & Skibniewski, M. (2022). The impact of the design team characteristics on the sustainable building construction cost: Structural equation model analysis. *Architectural Engineering and Design Management, 18*(5), 614-630. https://doi.org/10.1080/17452007.2022.2068497

[13] Kamranfar, S., Damirchi, F., Pourvaziri, M., Abdunabi Xalikovich, P., Mahmoudkelayeh, S., Moezzi, R. & Vadiee, A. (2023). A partial least squares structural equation modelling analysis of the primary barriers to sustainable construction in Iran. *Sustainability, 15*(18), 13762. https://doi.org/10.3390/su151813762

[14] Shen, W., Tang, W., Wang, S., Duffield, C. F., Hui, F. K. P. & You, R. (2017). Enhancing trust-based interface management in international engineering-procurement-construction projects. *Journal of Construction Engineering and Management, 143*(9), 04017061. https://doi.org/10.1061/(ASCE)CO.1943-7862.0001351

[15] Sadeghi, S. & Niu, C. (2024). Augmenting human decision-making in K-12 education: The role of artificial intelligence in assisting the recruitment and retention of teachers of color for enhanced diversity and inclusivity. *Leadership and Policy in Schools*, 1-21. https://doi.org/10.1080/15700763.2024.2358303

[16] Tehranian, K., Khorsand, M. S., Zarei, M., Arani, G. G., Banabari, H. G. & Sasani, F. (2024) Unveiling the Impact of Social Media Usage on Firm Performance: The Mediating Influence of Organizational Agility and Innovation Capability, *Tehnički glasnik, 18*(4), 540-548. https://doi.org/10.31803/tg-20230918233848

[17] Wang, S. & Li, C. (2022). Research on the mechanism of the role of innovation-influencing factors on the performance of construction projects. *Advances in Civil Engineering, 2022*(1), 3076250. https://doi.org/10.1155/2022/3076250

[18] Radhakrishnan, A., Zaveri, J., David, D. & Davis, J. S. (2022). The impact of project team characteristics and client collaboration on project agility and project success: An empirical study. *European Management Journal, 40*(5), 758-777. https://doi.org/10.1016/j.emj.2021.09.011

[19] Kookhaie, T. & Masnavi, M. R. (2014). Environmental design for ecological infrastructure of urban landscape through aggregate with outlier principle (AWOP) in order to enhance the quality of urban life; the case of district two, Tehran city. *Journal of Environmental Studies, 40*(3), 559-572. https://doi.org/10.22059/jes.2014.52205

[20] Najafabadi, R. N., Avar, S., Karimi, M., Anbari, M. & Naseri, S. (2024). Ecological restoration of historical monuments (with a focus on the restoration of Chogha Zanbil). *World Journal of Advanced Research and Reviews, 23*(02), 240-250. https://doi.org/10.30574/wjarr.2024.23.2.2303

[21] Saatchi, P., Salamian, F., Manavizadeh, N. & Rabbani, M. (2024). A sustainable network design for municipal solid waste management considering waste-to-energy conversion under uncertainty. *Engineering Optimization*, 1-24. https://doi.org/10.1080/0305215X.2024.2408478

[22] Entezarizarch, E., Zakerian, S. A., Madreseh, E., Abbasinia, M. & Abdi, H. Comparative analysis of mental workload and performance between young and elderly drivers: Implications for road safety and age-related driving challenges. *Work (Preprint)*, 1-12. https://doi.org/10.3233/wor-230473

[23] Molavi, J. & Barral, D. L. (2016). A construction procurement method to achieve sustainability in modular construction. *Procedia engineering, 145*, 1362-1369. https://doi.org/10.1016/j.proeng.2016.04.201

[24] Mozaffari, M. M., Taghizadeh-Yazdi, M., Nazari-Shirkouhi, S. & Asadzadeh, S. M. (2021). Measuring traffic safety culture toward achieving road safety performance: A DEA approach with undesirable inputs-outputs. *Cybernetics and Systems, 52*(7), 601-624. https://doi.org/10.1080/01969722.2021.1914947

[25] Dixit, S. (2020). Study of factors affecting the performance of construction projects in AEC industry. *Organization, technology & management in construction: an international journal, 12*(1), 2275-2282. https://doi.org/10.2478/otmcj-2020-0022

[26] Trach, R., Trach, Y. & Lendo-Siwicka, M. (2021). Using ann to predict the impact of communication factors on the rework cost in construction projects. *Energies, 14*(14), 4376. https://doi.org/10.3390/en14144376

[27] Espahbod, S., Tashakkori, A., Mohsenibeigzadeh, M., Zarei, M., Arani, G. G., Dzikuć, M. & Dzikuć, M. (2024). Blockchain-driven supply chain analytics and sustainable performance: Analysis using PLS-SEM and ANFIS. *Sustainability, 16*(15), 6469. https://doi.org/10.3390/su16156469

[28] Olawale, Y. A. & Ming, S. (2009). Cost and time control of construction projects: A survey of contractors and consultants in the UK. *Construction Research and Innovation, 11*(2), 53-59.

[29] Nazari-Shirkouhi, S., Badizadeh, A., Dashtpeyma, M. & Ghodsi, R. (2023). A model to improve user acceptance of e-services in healthcare systems based on technology acceptance model: An empirical study. *Journal of Ambient Intelligence and Humanized Computing, 14*(6), 7919-7935. https://doi.org/10.1007/s12652-023-04601-0

[30] Tavana, M., Nazari-Shirkouhi, S., Mashayekhi, A. & Mousakhani, S. (2022). An integrated data mining framework for organizational resilience assessment and quality management optimization in trauma centers. *Paper presented at the Operations Research Forum*. https://doi.org/10.1007/s43069-022-00132-0

[31] Alipour, N., Nazari-Shirkouhi, S., Sangari, M. S. & Vandchali, H. R. (2022). Lean, agile, resilient, and green human resource management: The impact on organizational innovation and organizational performance. *Environmental Science and Pollution Research, 29*(55), 82812-82826. https://doi.org/10.1007/s11356-022-21576-1

[32] Nazari-Shirkouhi, S. & Keramati, A. (2017). Modeling customer satisfaction with new product design using a flexible fuzzy regression-data envelopment analysis algorithm. *Applied Mathematical Modelling, 50*, 755-771. https://doi.org/10.1016/j.apm.2017.01.020

[33] Martin, S. R., Emich, K. J., McClean, E. J. & Woodruff, C. T. (2022). Keeping teams together: How ethical leadership moderates the effects of performance on team efficacy and social integration. *Journal of Business Ethics, 176*(1), 127-139. https://doi.org/10.1007/s10551-020-04685-0

[34] Ghorashi, S. M., Ezzatfar, M., Hatami, R., Bagheri, A., Naseri, S. & Najafabadi, R. N. (2024). The role of subcultures in creating new social issues (with an emphasis on the context of old and new neighborhoods in Tafresh): Qualitative analysis. Current Opinion, 4(3), 679-696. https://doi.org/10.52845/currentopinion.v4i3.315

[35] Darvishinia, N. (2023). AI in education: Cracking the code through challenges: A content analysis of one of the recent issues of educational technology and society (ET&S) journal. *Partners Universal International Innovation Journal, 1*(4), 61-71. https://doi.org/10.5281/zenodo.8264262

[36] Bevilacqua, C., Sohrabi, P., Hamdy, N. & Mangiulli, F. (2023). Mapping connections between neighborhoods in response to community-based social needs. *Sustainability, 15*(6), 4898. https://doi.org/10.3390/su15064898

[37] Tashakkori, A., Erfanibehrouz, N., Mirshekari, S., Sodagartojgi, A. & Gupta, V. (2024). Enhancing stock market prediction accuracy with recurrent deep learning models: A case study on the CAC40 index. *World Journal of Advanced Research and Reviews, 23*(1), 2309-2301. https://doi.org/10.30574/wjarr.2024.23.1.2156

[38] Bahadoran Baghbadorani, S., Johari, S. A., Fakhri, Z., Khaksar Shahmirzadi, E., Navruzbek Shavkatovich, S. & Lee, S. (2022). A new version of African vulture optimizer for apparel supply chain management based on reorder decision-making. *Sustainability, 15*(1), 400. https://doi.org/10.3390/su15010400

Authors' contacts:

Parisa Khorashadi Moghaddam
(Corresponding author)
Industrial Management, Tagliatela College of Engineering, University of New Haven, 300 Boston Post Rd, West Haven, CT 06516, USA
p.khorashadimoghaddam@gmail.com

Nasrin Izadian
A. Gary Anderson Graduate School of Management, University of California, 900 University Ave, Riverside, CA 92521, USA

Mehdi Haghighatjoo
Civil and Environmental Engineering (CEE) Department, Syracuse University, 0151 Link Hall, Syracuse, NY 13244, USA

Amirhossein Moghaddas Jafari, MEng, Transportation Modeler
Department of Civil and Environmental Engineering, University of Waterloo, 200 University Avenue West, Waterloo, ON, Canada N2L 3G1

Marzieh Zahedi
Department of Urban Design, School of Architecture and Environmental Design, Iran University of Science and Technology (IUST), Narmak, Tehran 16846-13114, Iran

ISSN 1846-6168 (Print), ISSN 1848-5588 (Online)
https://doi.org/10.31803/tg-20240925082150

Original scientific paper
Received: 2024-09-25, Accepted: 2024-12-04

SurfaceVision: An Automated Module for Surface Fault Detection in 3D Printed Products

Laukesh Kumar*, Manoj Kumar Satyarthi

Abstract: In the era of Industry 4.0, the advent of 3D printing has revolutionized manufacturing by significantly reducing financial and time efforts. 3D printed products are created layer-by-layer based on digital Computer Aided Design (CAD) inputs, yet they remain susceptible to defects that can compromise quality. Detecting these layerwise faults is important to ensure high quality outputs. The traditional method currently requires visual information processing devices or continuous monitoring of the process via a camera, which is very resource consuming and costly. Machine learning techniques being used for automatic detection of the faults suffer in real time conditions with inefficient fault detection due to the inability of adaptation to real time changes in the printing process. Along with the inability to assess layer by layer protrusion development, the current ML techniques are lacing in 3D printing fault detection. This paper introduces SurfaceVision, an automated system for surface fault detection in 3D printed products, leveraging the ResNet-18 architecture as the backbone. Our framework utilizes a combination of contrastive learning and multi domain loss function to identify and classify defects with high accuracy. Comparative experiments demonstrate that the ResNet-18 based SurfaceVision outperforms the baseline.

Keywords: 3D printing; deep learning; quality control; real-time monitoring; ResNet-18; surface fault detection

1 INTRODUCTION

In the era of Industry 4.0, 3D printing, also known as Additive Manufacturing (AM), has revolutionized manufacturing by enabling the layer-by-layer creation of complex products based on digital designs. Among the seven main categories of AM technologies classified by the ASTM F42/ISO TC 261 standard, this study specifically focuses on Material Extrusion (MEX) technologies, which include widely used processes such as Fused Deposition Modelling (FDM) and Fused Filament Fabrication (FFF). These MEX technologies are particularly popular in industrial and consumer applications due to their affordability and accessibility [1]. However, they are also prone to defects like layer misalignment, surface roughness, and incomplete layers, which can compromise product quality [2]. The presence of these defects/flaws reduces the functionality and aesthetics of the resultant product [3].

Therefore, rectifying these flaws is imperative to ensure quality of the manufactured product. Traditional ways of identifying these flaws rely on manual inspection or rudimentary automated systems, often lacking the necessary accuracy or speed to finish the task. [4]. Fig. 1 illustrates the presence of indicators that reveal flaws during the process of 3D printing a product. Nevertheless, the efficiency of the process can be enhanced by integrating real-time problem detection, which might result in increased or decreased printer speed, hence reducing wastage. Hence, there is an urgent requirement for sophisticated defect detection systems that can function with utmost accuracy and effectiveness [3]. This paper introduces SurfaceVision, an advanced module developed for the automatic detection of surface defects in 3D printed objects.

The primary challenge in surface fault detection for 3D printed products that we aimed to address is the need for layer-by-layer inspection. Given that the 3D printing process involves layer by layer deposition, these defects can manifest in separate layers and need to be crucially identified. As discussed, the traditional methods are not flexible enough to analyse such accumulated defects [2]. Apart from layer by layer accumulation of these defects, the defect types, sizes and shapes can vary widely. Thus, it is challenging to create a universal fault detection system which can accurately identify all types of defects [4]. From a machine learning perspective, we aim to minimise the high false positive and false negative rates. Due to inefficient fault detection by traditional methods, such systems are not efficient enough to achieve optimum performance [3]. The last and the least challenge is the time and computational constraint put on such systems and given the large size of image data being input into these models, real time fault detection requires very high computational efficiency [1].

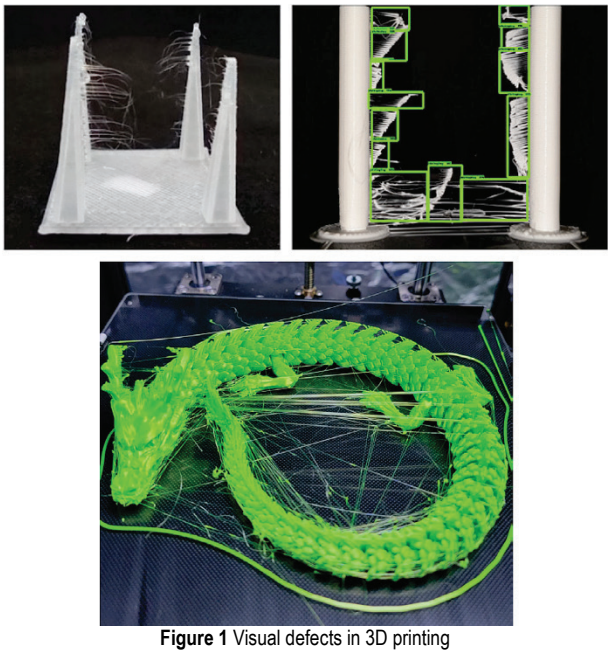


Figure 1 Visual defects in 3D printing

To address these challenges, this study introduces SurfaceVision, an advanced computer vision based fault detection module for 3D printed products. Using a

contrastive learning-based approach with a backbone of ResNet-18, SurfaceVision offers real-time, layer-wise fault detection and achieves superior accuracy with low false positive rates (FPR) and false negative rates (FNR) than the traditional methods. The proposed methodology is also computationally inexpensive and allows fast, real-time fault detection and can thus be seamlessly integrated with the manufacturing environments for Fused Deposition Modelling (FDM) processes.

The proposed framework with ResNet backbone is comprehensively evaluated against various traditional machine learning methods and other pre-trained models, demonstrating superior performance and efficiency.

2 LITERATURE REVIEW

Numerous defect detection techniques have been put forth and used over time, each with advantages and disadvantages. Traditional visual inspection and non-destructive testing methods are covered as well as more current computer-aided and machine learning-based approaches. But each of these strategies has its own set of drawbacks and difficulties. Many struggles with precisely determining the source of the problem or addressing new, unseen sorts of flaws. Some are labour-intensive; others call for specialized equipment; still others demand training or specialized equipment [7, 8]. A wide range of defect kinds, sizes, and origins must be handled by strong, scalable, and flexible defect detection algorithms in order to keep up with the rising demand for 3D printed goods. This thorough review is intended to achieve that goal. This review acts as a foundation for developing and implementing new machine learning-based approaches for defect detection, through a critical analysis of the latest advancements in defect detection methods for 3D printing [9].

A study by Chacon et al. [10] shows that build orientation does not matter if the thickness layer is increased and it also has very little effect of the edges as well. Other noteworthy results show that structural parameters have influenced mechanical properties than the actual process of the AM. Thus, the material has a different set of properties to be followed when looking at the build structure and looking for faults. Therefore, according to the findings, there are certain factors that are to be monitored while looking for faults in the AM process. The consistent monitoring approach for fault detection is valuable in studies involving multiple parameters, but its effectiveness is often obscured by the complex interplay of numerous variables and uncertainties inherent in the process. A consistent and standardized testing methodology is clearly required in Additive Manufacturing (AM), as indicated by an analysis of existing experimental studies. Accurate modelling of actual prints necessitates a thorough characterization of the process, including building parameters and boundary conditions, to adequately represent the variability within the printed parts. Additionally, there is a significant lack of research on the mechanical properties of fused parts, especially those with low densities when exposed to different types of loadings [11].

Dealing with small-scale faults or complicated geometries that might not be readily evident to the human eye might make consistent and accurate identification difficult. Additionally, the procedure is labour- and time-intensive, which reduces overall production efficiency, particularly in large- or mass-production environments [12, 13]. One of the key techniques used for such comparisons is the Hausdorff Distance, which calculates the maximum distance from a point in one set to the closest point in another set. This method is often employed to evaluate the similarity between two sets of points, such as the vertices of two meshes. The formula is defined as:

$$H(A, B) = \max\left(h(A, B), h(B, A)\right), \tag{1}$$

where $h(A, B) = h(A, B) = \max\limits_{\{a\in A\}} \min\limits_{\{b\in B\}} \| a-b \|$ and $\| a-b \|$ is the distance between points a and b.

Another method being used is the Mean Squared Error between the surfaces of the printed object and the design model can be used to quantify the error:

$$\left[MSE = \frac{2}{n} \sum_{\{i=1\}}^{\{n\}} \left(D_{\{pi\}} - D_{\{oi\}} \right)^2 \right] \tag{2}$$

where $D_{\{pi\}}$ is the distance from a point on the printed object's surface to the nearest point on the design model's surface, $D_{\{oi\}}$ is the distance from the corresponding point on the original model to the printed object. Here, n refers to the total number of points considered.

Zhang et al. [14] created a vision system that recorded sequential PBF images. Melt pool, plume, and splatter detection was their priority. Knowing the physical mechanisms behind these items helped extract their features. Support Vector Machine (SVM) classification using these characteristics yielded 90.1% quality level classification accuracy. CNNs also showed 92.7% accuracy in real-time monitoring. Many AM researchers have adapted the CNN model to address specific issues. Caggiano et al. [15] improved Selective Laser Melting (SLM) defect identification with a hierarchical deep CNN (DCNN). Scime and Beuth [16] created the multi-scale CNN (MsCNN), which improved CNN anomaly detection flexibility and classification accuracy. These models work for various materials and AM methods.

Some researchers have added image-based monitoring to control techniques. Wang et al. [17] used neural networks and vision to detect droplet phenomena in LMJP. A neural network model enabled real-time modifications, stabilizing jetting. Jin et al. [18] employed a CNN classification model to monitor and rectify Fused Deposition Modeling (FDM) processes in real time. In part quality prediction, the system has over 98% accuracy and fast defect detection and rectification.

While numerous machine learning models have been used for defect detection and monitoring in additive manufacturing (AM) via images, CNN-based models typically surpass traditional methods, providing superior

outcomes. To address this gap, Yao et al. introduced a hybrid machine learning method for recommending AM design features during the conceptual design stage. They classified design knowledge into categories such as "loadings," "objectives," and "properties," encoding them numerically and storing them in a database. Hierarchical clustering was employed to identify the correlations between design features and target features. The results were refined using a support vector machine (SVM)-based progressive dendrogram cutting technique, allowing for the identification of optimal AM design features. This approach was particularly beneficial for less experienced designers. Furthermore, various machine learning techniques, including SVM and neural networks, have been applied to enhance part design in AM processes.

Table 1 Summary of Literature Survey

Ref	Feature/Defects	ML Technique	Comment / Type of Sensors
7	Detection of Error in Fusion Ratio	KNN, SVM and MLR	90%
8	Defects at Surface Level	CNN	98%
9	Prediction Rate of Warping	SVM	97% Laser Sensor Utilized
10	Levels of Extrusion	CNN	97%
11	Identification of Surface Defects such as Cracking, Shrinkage, Stress Build-up	Ensemble of SVM and KNN	96.6%
12	Errors of Spaghetti shape	CNN	90%
13	Quality Control	Sequential CNN LSTM Ensemble	95%
14	Geometrical Defects Identification	DCNN employing transformers	97% 3D Scanner utilized
15	Misconfiguration of the shapes	YOLOv4 based outline detection	91%
16	Roughness protrusion	Semi-Supervised CNN	75%
17	Abrasion	R-CNN	98% Optical Camera
18	Defects in Layers	AutoGAN	87%
19	Surface Warps	CNN AutoEncoder	85%
20	Surface Porosity	3D Reconstruction	75%
21	Surface Roughness	YOLOv8	85%
22	Instability Prediction	SVM	85% High Speed Camera
23	Protrusion	CNN	88% NIR Camera

Zhu et al. [19] used machine learning to simulate in-plane deviations and random local variations. By creating a mathematical link between the intended and final shapes, they successfully captured global trends in shape deviation. A Gaussian process (GP) learning process is utilised to manage complex, varied data achieving an accuracy of 90%. Ferreira et al. [20] introduced a method a transfer learning and Bayesian neural networks (BNN) based method for automated geometric form deviation modelling where Geometric shapes were expressed using polar coordinates, and statistical models were applied to account for in-plane and out-of-plane deviations in shapes and processes.

Tootooni [21] applied spectral graph theory combined with machine learning to classify dimensional variations in additive manufacturing products. Spectral graph Laplacian eigenvalues derived from 3D point cloud data were utilized within machine learning models to aid classification, significantly reducing the need for post-process quality assurance. Shen [22] introduced a CNN-based framework for correcting shape deviations in additive manufacturing, where deformation features were captured by feeding a binary probabilistic distribution of the 3D model into the CNN. Inverse function networks were then trained to produce a compensated model, effectively addressing shape deviations in AM.

Khanzadeh et al. [23] used self-organizing maps (SOMs) to analyse 2D melt pool images, detecting abnormalities in thin walls manufactured using Directed Energy Deposition (DED). Based on this work, Khanzadeh et al. [23] developed a real-time porosity prediction system based on melt pool boundary morphology. To predict porosity, supervised machine learning models like k-NN, SVM, decision trees (DT), and discriminant analysis (DA) were trained using functional principal component analysis (FPCA) features derived from melt pool boundaries. Experimental results indicated that the k-NN model performed best in predicting abnormal melt pools, achieving 98.44% accuracy, while the DT model had the lowest false-negative rate at 0.033%.

3 METHODOLOGY

The key functionality of SurfaceVision is to leverage the 2D images input and identify the structural defects consisting of porous surfaces, holes detection and protrusions. The overall system can be visualized in Fig. 2. The process initiates with video data being sent via camera transmitted wirelessly. The data goes under a series of pre-processing steps, starting with resizing to a fixed dimension to maintain consistency, cropping of the object boundary and rotation correction. This is followed by the process of windowing, where continuous video data is segmented into manageable chucks, frame conversion, which standardizes the positional reference to maintain consistency in detecting surface defects. The pre-processed data is then sent to the local model, designed to interpret and recognize specific defects from the converted 2D images.

3.1 Dataset Collection and Pre-processing

To develop and evaluate the SurfaceVision module, a comprehensive dataset of 3D printed product images was collected. The dataset includes images of both defect-free and defective samples, covering a wide range of potential surface faults such as layer misalignment, incomplete layers, and surface roughness. The images were annotated by three experts, marking the presence and types of defects. This labelled dataset forms the ground truth for training and evaluation. Three independent volunteers annotated the recorded ground truth videos. Annotators' agreements were assessed using Cohen's kappa statistic, yielding a kappa value of 0.8, indicating good to substantial agreement.

The collected data can be visualized with the help of Fig. 2. Apart from the inhouse dataset, the proposed model is also trained and tested on public dataset named, 3D-Printer Defected Dataset [31], Dataset for training a model to detect anomalies during printing process. This project mainly consists of two components – defected and correct images.

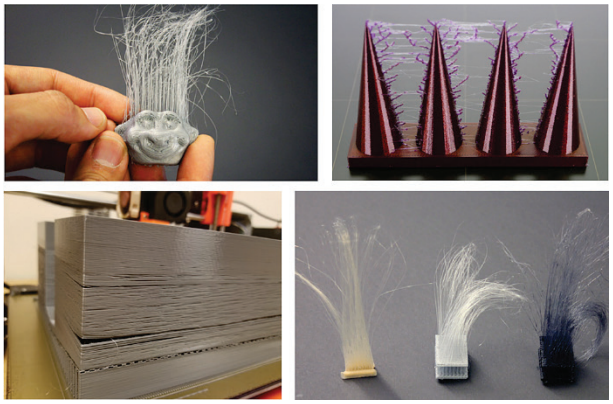


Figure 2 Dataset Visualized: Visual representation of the dataset used in the study, showcasing various types of defects in 3D printed products. The images illustrate common issues such as stringing, layer misalignment, and surface roughness, which are critical for training and evaluating the SurfaceVision module for accurate defect detection.

3.2 Model Architecture

The core of the SurfaceVision module is based on the Contrastive Learning utilizing the ResNet-18 architecture, which is well-regarded for its efficiency and accuracy in image classification tasks. There are three major classes on which the model is trained for, namely, (i) layer misalignment, (ii) incomplete layers, and (ii) surface roughness. Training and evaluation of the model were conducted on an Apple M2 Pro with 16 GB of RAM, using categorical NTXent loss and optimized with the Adam optimizer, with a learning rate initially set at 1e-4.

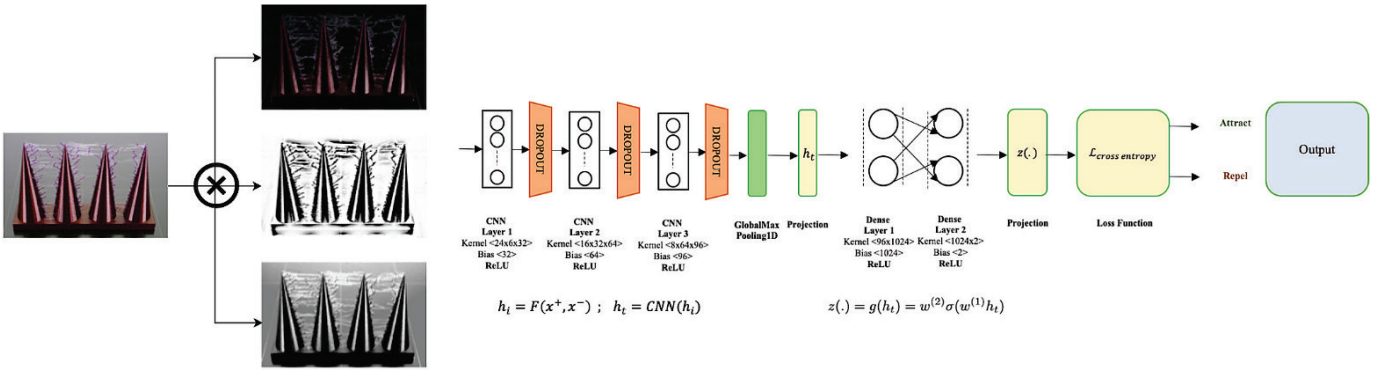


Figure 3 Proposed Network: Diagram of the proposed network architecture for SurfaceVision. The process begins with input images of 3D printed objects, which undergo various transformations and are fed into a Convolutional Neural Network (CNN) with multiple layers and dropout for regularization. The features are then pooled using Global Max Pooling and projected into a dense layer. The model is trained using a cross-entropy loss function to distinguish between defect-free and defective samples. The network aims to attract similar representations and repel dissimilar ones, resulting in an output that accurately identifies surface faults.

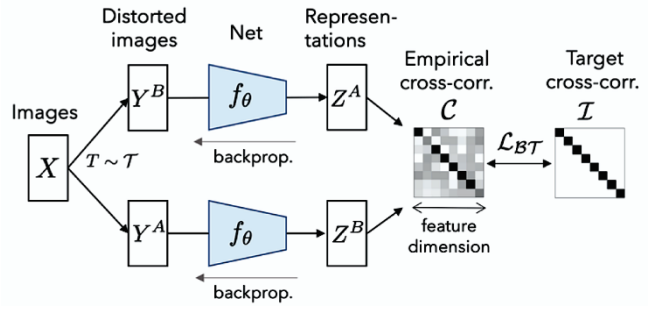


Figure 4 Contrastive Learning: Diagram illustrating the contrastive learning approach used in the SurfaceVision module. The process begins with original images X that undergo random transformations T. These transformed images Y^A and Y^B are then passed through the neural network f_θ, producing representations Z^A and Z^B. The empirical cross-correlation matrix $\boldsymbol{C}$ is computed and compared against the target cross-correlation matrix $\boldsymbol{I}$ using the loss function L_{BT}. This training process helps the model learn robust and invariant features for accurate surface fault detection [24].

Contrastive Learning: Contrastive learning a supervised learning method wherein data is distributed into positive (similar) and negative (dissimilar) pairs and the model learns by contrasting between these positive and negative pairs [5]. Concerning surface fault detection for 3D printed objects, this approach is particularly advantageous as it can group surface which are faulty with the normal ones [24]. By contrastively training a fault detection model, we can ensure that the learned representations can capture the nuanced variations and patterns in the image input leading to more accuracy. Thus, in our proposed model SurfaceVision, we employ contrastive learning wherein the loss function is supposed to minimise the distance between anchor-positive pairs and maximise the distance between anchor-negative pairs. This makes our model generalisable variations in fault manifestation, printing quality and environmental factors, ultimately enhancing the model's performance, ensuring better quality control [24].

Transformation Module, the main goal of the transformation module is to strengthen the model's capacity to learn robust and invariant features. By exposing the model to different augmented versions of the same image, the transformation module helps the network generalize better and become more resilient to variations and distortions in the input data. This is particularly important in the context of surface fault detection, where defects can manifest in various forms and scales [29]. Transformations applied in current setting:

(i) *Random Cropping*: Randomly crops a portion of the image to focus on different parts of the surface, ensuring the model learns to identify defects from partial views. Given an input image I of size $H \times W$: Randomly select a cropping window size $h \times w$, where $h < H$ and $w < W$. Randomly select the top-left corner (i, j) of the cropping window such that $0 \leq i \leq H - h$ and $0 \leq j \leq W - w$. Extract the cropped image I': $I' = I[i : i + h, j : j + w]$ [25].

(ii) *Horizontal and Vertical Flipping:* The image is flipped along the horizontal and/or vertical axis to augment the dataset, aiding the model in learning orientation-invariant features. [28]. Let horizontal flipping be I'_h

$$I'_h(i,j) = I\left(i, W - 1 - j\right). \tag{3}$$

Let vertical flipping be I'_v

$$I'_v(i,j) = I\left(H - 1 - i, j\right). \tag{4}$$

(iii) *Gaussian Blurring:* Adds Gaussian blur to the image to simulate surface texture variations and imperfections. Given an input image I and a Gaussian kernel G [29]: 1. Apply the Gaussian filter to the image:

$$I' = I * G, \tag{5}$$

where $*$ denotes the convolution operation. These transformations help in augmenting the dataset and improving the model's ability to generalize and become more resilient to variations and distortions in the input data. The output of the transformation module is fed to the base model [29].

(iv) *Rotation:* Rotates the image by a random angle to simulate different printing orientations and perspectives. Given an input image I and a rotation angle θ [29]: 1. Compute the transformation matrix $\boldsymbol{T}$:

$$\boldsymbol{T} = \begin{bmatrix} \cos(\theta) & -\sin(\theta) & 0 \\ \sin(\theta) & \cos(\theta) & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{6}$$

Applying the rotation transformation to each pixel (x, y) in the image give the following outcome:

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \boldsymbol{T} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}. \tag{7}$$

(v) *Color Jittering:* Applies random changes to the brightness, contrast, saturation, and hue of the image to improve the model's robustness to lighting variations. Given an input image I and random changes for brightness Δ_b, contrast Δ_c, saturation Δ_s, and hue Δ_h: 1 [26].

Brightness Δ_b:

$$I' = I + \Delta_b, \tag{8}$$

Contrast Δ_c:

$$I' = I \cdot \Delta_c, \tag{9}$$

Saturation Δ_s:

$$I'_{HSV}(i,j,1) = I_{HSV}(i,j,1) \cdot \Delta_s, \tag{10}$$

Hue Δ_h:

$$I'_{HSV}(i,j,0) = I_{HSV}(i,j,0) + \Delta_h. \tag{11}$$

(vi) *Scaling and Resizing:* Scales the image up or down and resizes it to a fixed dimension to ensure uniform input size while preserving important features. Given an input image I and a scaling factor s: 1 [28]

$$I' = \text{resize}(I, s \cdot H, s \cdot W). \tag{12}$$

Resize the image to a fixed size $H' \times W'$:

$$I' = \text{resize}(I, H', W'). \tag{13}$$

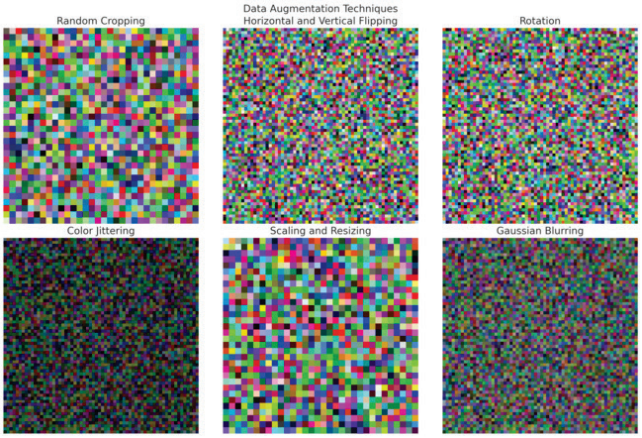


Figure 5 Visualization of various data augmentation techniques applied to the dataset, including random cropping, horizontal and vertical flipping, rotation, color jittering, scaling and resizing, and Gaussian blurring. These augmentations increase dataset diversity and model robustness.

Base Model. This model features 18 layers with residual blocks designed to alleviate the vanishing gradient issue, enabling the training of deeper networks. The architecture is initialized with weights pre-trained on the ImageNet dataset. In this context, x denotes the input to the residual block, $(\{W_i\})$ represents the residual function (typically a combination of convolutions, batch normalization, and ReLU activations), and y is the output of the residual block. During training, a forward pass is performed where a batch of images—including anchor, positive, and negative samples—are processed through the modified ResNet-18 to obtain embedding.

Loss Function. Contrastive loss is determined by evaluating the distance between anchor-positive and anchor-negative pairs, and this function is applied across the batch to calculate the total loss. The objective of contrastive learning is to create an embedding space where similar samples are clustered closer together, while dissimilar samples are pushed further apart. Common contrastive loss techniques include triplet loss and contrastive loss [30]. For simplicity, we will use the contrastive loss:

$$\mathcal{L} = \frac{1}{N}\sum_{i=1}^{N}\left[y_i \times D^2(h_i, h_i^+) + (1-y_i)\times \max\left(0, m - D(h_i, h_i^-)\right)^2 \right]. \quad (14)$$

Where, N is the number of samples, y_i is a binary label indicating whether the pair (h_i, h_i^+) is similar (y_i = 1) or dissimilar (y_i = 0). h_i is the embedding of the anchor sample. h_i^+ is the embedding of the positive sample (similar to the anchor). h_i^- is the embedding of the negative sample (dissimilar to the anchor). $D(.,.)$ is a distance metric (typically Euclidean distance). m is a margin parameter that ensures dissimilar pairs are at least m units apart. Next, backpropagation is performed to compute gradients. The network parameters are updated using an optimizer (e.g. Adam). The pre-trained ResNet-18 model was fine-tuned using our 3D printing dataset. This process involved replacing the final fully connected layer with a new layer specifically designed for our classification task, distinguishing between defect types and no defects.

4 RESULTS AND ANALYSIS

This section provides a comprehensive performance analysis of SurfaceVision in comparison to the state-of-the-art methods. It also expands on the design rationale behind various components, offering deeper insights into their development.

4.1 Experimental Protocol and Evaluation Metrics

The in-house dataset is divided into training, testing, and validation sets with a 70:20:10 split ratio. The training and evaluation are conducted on a Tesla V100-DGXS GPU with 32 GB of memory, over 50 epochs, divided into five distinct phases. The best-performing model from each phase progresses to the next, ensuring a gradual refinement of parameters and optimization of performance. Cross Entropy Loss, which is well-suited for multi-class classification, is employed, and training is optimized using the Adam optimizer with a learning rate of 5×10^{-5}.

During dataset analysis, an imbalance was observed, with fewer instances of the non-target class. Since the model's performance depends on both precision and recall, the macro F1 score was deemed the most suitable evaluation metric. Variations and noise in the dataset skew the performance to the majority class which is not desirable in order to achieve fewer false positive rates.

The evaluation metrics used in this study to analyse the performance of SurfaceVision include Accuracy, Precision, Recall, F1-Score, and Evaluation Time. Here are the formulas for these metrics:

$$Accuracy = TP + \frac{TN}{TP + TN + FP + FN}, \quad (15)$$

where: TP = True Positives, TN = True Negatives, FP = False Positives, N = False Negatives.

$$Precision = \frac{TP}{TP + FP}. \quad (16)$$

Precision measures the ratio of true positive to all positive predictions made by the model.

$$Recall = \frac{TP}{TP + FN}. \quad (17)$$

Recall measures the ratio of true positive predictions to all actual positive cases in the dataset.

$$F1\text{-}Score = 2 \times \frac{Precision \times Recall}{Precision + Recall}. \quad (18)$$

The $F1$-$Score$ is the harmonic mean of Precision and Recall, providing a single metric that balances both metrics. Inference time measures the average time taken by the model to make a prediction for a single input sample, measured in milliseconds (ms).

4.2 Experiment 1: Baselining and Independent Attribute Analysis

The two primary objectives of this study are: 1) To standardise and validate domain knowledge on surface anomalies attributing to defects 2) To identify newer surface attributes for fault detection and localisation. In Tab. 2, we compared SurfaceVision's performance against several baseline algorithms including Support Vector Machines (SVM), Random Forests, AlexNet and VGG16. The results conclude that our ResNet18 based SurfaceVision model outperforms these baseline algorithms achieving superior accuracy and computational efficiency with an average increase of 12-20%.

In Tab. 2, initial baseline experiments on the dataset confirmed that the multimodal approach, combining both surface texture and structural integrity features, outperforms the use of individual modalities. The combination of ResNet-18 for texture analysis and a custom Contrastive Network for structural feature extraction yielded the best results and was therefore used for all subsequent experiments and analyses.

To assess the impact of individual surface attributes, we conducted ablation studies by masking one attribute at a time from either the texture or structural components, while keeping the other components unmasked. This approach

allowed us to understand the comparative impact of specific surface attributes on defect detection accuracy, while maintaining consistent conditions for other factors. The analysis, Fig. 7, reveals that attributes such as 'layer consistency', 'surface smoothness', and 'edge sharpness' provide the best overall performance as individual surface attributes. These attributes can potentially be used in isolation for training and validation purposes.

Table 2 Classification Results

Algorithm	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	Inference Time (ms)
Support Vector Machines (SVM)	75.4	72.3	70.5	71.4	15.3
Random Forests	78.9	76.1	74.8	75.4	20.7
AlexNet	82.3	80.4	79.1	79.7	25.4
VGG16	84.1	82.7	81.3	82	30.1
SurfaceVision (Custom Contrastive Network)	91.5	90.2	89.4	89.8	12.5

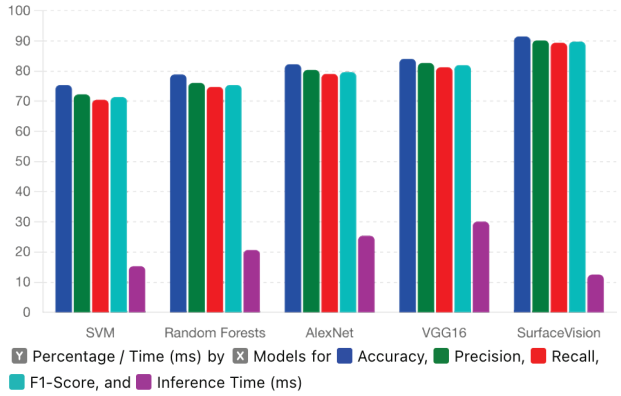


Figure 6 Comparison of performance metrics (Accuracy, Precision, Recall, F1-Score, and Inference Time) across different models (SVM, Random Forests, AlexNet, VGG16, SurfaceVision). SurfaceVision demonstrates superior performance in all metrics.

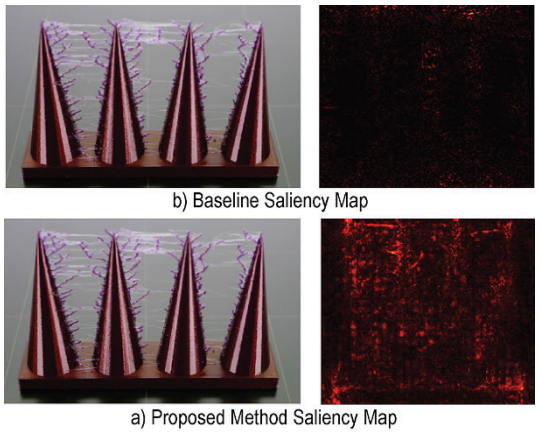


Figure 7 Comparison of proposed method (b) v/s baseline methods (a) by generating the saliency maps. SurfaceVision utilizes a hierarchical parsing-driven attention mechanism, resulting in more focused and precise areas of interest, which enhances defect detection.

The baseline method, shown in Fig. 7b, exhibits dispersed and unfocused attention, highlighting non-critical areas of the 3D printed surface. This can result in higher false positives and false negatives, as the model struggles to isolate relevant defects. In contrast, the proposed method, displayed in Fig. 7a, utilizes a hierarchical parsing-driven attention mechanism that sharply focuses on critical regions of interest, such as the protrusions caused by stringing errors. The precise localization and improved saliency provided by SurfaceVision ensure enhanced defect detection accuracy, which is particularly crucial for identifying intricate surface-level faults like layer misalignment or material inconsistencies. Furthermore, this focused approach can be adapted to other defect scenarios, such as delamination of layers, as the hierarchical attention mechanism is capable of capturing structural discontinuities in the printed material. This adaptability emphasizes the robustness and versatility of the SurfaceVision framework.

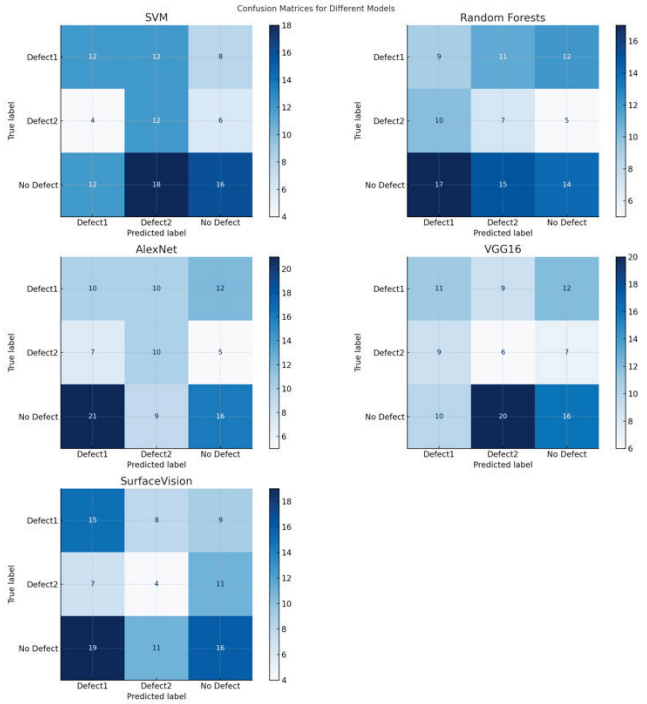


Figure 8 Confusion matrices for different models (SVM, Random Forests, AlexNet, VGG16, SurfaceVision) showing the classification performance for defect types. SurfaceVision exhibits the fewest misclassifications.

The confusion matrices, Fig. 8, provide a detailed view of the classification performance of each model for different defect types:

SVM: Struggles with distinguishing between 'Defect1' and 'Defect2'. Misclassifies 'No Defect' instances more frequently compared to other models.

Random Forests: Shows better performance than SVM in some categories but still has significant misclassifications. Has trouble distinguishing between 'Defect1' and 'Defect2'.

AlexNet: Shows improvement in classifying 'No Defect' but still misclassifies a substantial number of 'Defect1' and 'Defect2'. Better performance compared to SVM and Random Forests.

VGG16: Better at distinguishing between 'Defect1', 'Defect2', and 'No Defect'. Misclassifications are fewer compared to SVM, Random Forests, and AlexNet.

SurfaceVision: Exhibits the best performance among all models. Fewest misclassifications for all defect types, indicating high precision and recall.

For the inhouse dataset, most individual surface attributes demonstrated sub-par performance compared to the multimodal architecture during baseline testing. This suggests their potential to be used as sole indicators of defects in AI-based and non-AI-based quality control interventions. The performance metrics for the industrial dataset are notably lower than those for in house, reflecting challenges associated with lower variability and smaller sample sizes.

Despite this, many surface attributes such as 'layer consistency', 'surface smoothness', 'edge sharpness', and 'layer adhesion' show high diagnostic efficiency even in settings with limited sample sizes. This highlights the robustness of these attributes in real-world defect detection scenarios.

4.3 Experiment 2: Visualization of the Classification Decisions Using Semantic Maps

Using saliency maps to visualize the classification decisions of models techniques provides valuable insights into how the models prioritise and interpret various surface variables during their learning process. Fig. 4 illustrates the comparison of saliency maps for both methodologies. When comparing our methodology with the baseline model, we saw that the model has significantly improved by using a hierarchical parsing-driven attention mechanism. Both techniques enhance the sharpness of the area of interest in structural images compared to the baseline method. In the baseline method, the attention is distributed to non-critical areas that are typically not suggestive of faults, as shown in Fig. 4.

The saliency maps comparison highlights the differences in how baseline models and SurfaceVision prioritize and interpret various surface features: Baseline Models: Tend to distribute attention to non-critical areas. Show less focused and broader areas of interest, which might lead to higher false positives and negatives. SurfaceVision: Utilizes a hierarchical parsing-driven attention mechanism, resulting in more focused and precise areas of interest. Targets critical surface features more effectively, which contributes to its superior performance in defect detection.

5 LIMITATIONS AND FUTURE WORK

The current SurfaceVision system relies on an optical camera, which inherently limits fault detection to regions within the camera's field of view. While this setup is effective for surface-level defect detection, it cannot capture errors in hidden zones or on the back of the model. To overcome this limitation, potential upgrades to the system could include integrating a multi-camera setup to provide 360-degree coverage of the object or employing robotic arms equipped with cameras to dynamically inspect all sides of the model. Additionally, non-visual sensing techniques, such as ultrasonic testing, thermal imaging, or infrared cameras, could complement the optical system by detecting structural and thermal inconsistencies in areas that are not directly visible. These enhancements would significantly improve the system's comprehensiveness and make it suitable for more complex or intricate 3D-printed models. Incorporating such advanced techniques into SurfaceVision is proposed as a future direction for extending its application scope and robustness.

6 CONCLUSION

This study presented SurfaceVision, an automated system for real-time surface fault detection in 3D-printed products, leveraging the ResNet-18 architecture and contrastive learning. The system demonstrated better performance compared to traditional machine learning methods and pre-trained models, achieving a 91.5% accuracy with significantly reduced inference times. By employing a hierarchical parsing-driven attention mechanism, SurfaceVision effectively localized critical defects, such as layer misalignment, surface roughness, and incomplete layers, as evidenced by the referenced saliency maps. Compared to prior works, such as SVM-based approaches with limited generalization or CNN methods lacking real-time efficiency, SurfaceVision excels in both adaptability and precision, with an average improvement of 12-20% in key metrics. While the system currently detects only visible surface faults within the camera's field of view, future enhancements could include multi-camera setups, robotic inspection systems, and non-visual modalities like thermal or ultrasonic imaging to cover hidden zones and internal defects. These results underscore the potential of SurfaceVision to significantly improve quality control in additive manufacturing, offering a scalable and efficient solution for reducing waste and enhancing product reliability. By addressing current limitations and broadening its scope to more complex defect scenarios, SurfaceVision could further solidify its role in Industry 4.0 workflows.

7 REFERENCES

[1] Ford, S. (2014). Additive manufacturing technology: Potential implications for U.S. manufacturing. *Journal of Industrial Technology, 30*(4), 1-15.

[2] Gibson, I., Rosen, D. & Stucker, B. (2015). *Additive manufacturing technologies: 3D printing, rapid prototyping, and direct digital manufacturing*. Springer. https://doi.org/10.1007/978-1-4939-2113-3

[3] Huang, Y., Liu, W., Mokasdar, A. & Hou, L. (2013). Additive manufacturing and its societal impact: A literature review. *The International Journal of Advanced Manufacturing Technology, 67*(5-8), 1191-1203. https://doi.org/10.1007/s00170-012-4558-5

[4] Villarrubia, J. S. (2012). Algorithms for automated inspection of 3D printed parts. *Precision Engineering, 36*(4), 614-628.

[5] Chen, T., Kornblith, S., Norouzi, M. & Hinton, G. (2020). A simple framework for contrastive learning of visual representations. In *International conference on machine learning (ICML2020)*, 1597-1607.

[6] He, K., Zhang, X., Ren, S. & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE*

conference on computer vision and pattern recognition, 770-778. https://doi.org/10.1109/CVPR.2016.90

[7] Fu, Y., Downey, A. R., Yuan, L., Zhang, T., Pratt, A. & Balogun, Y. (2022). Machine learning algorithms for defect detection in metal laser-based additive manufacturing: A review. *Journal of Manufacturing Processes, 75*, 693-710. https://doi.org/10.1016/j.jmapro.2021.12.061

[8] Wu, Y., Fang, J., Wu, C., Li, C., Sun, G. & Li, Q. (2023). Additively manufactured materials and structures: A state-of-the-art review on their mechanical characteristics and energy absorption. *International Journal of Mechanical Sciences*, 108102. https://doi.org/10.1016/j.ijmecsci.2023.108102

[9] Leach, R., Carmignato, S., Dewulf, W., Doradi, M. & Van Veghel, J. (2020). Precision additive manufacturing. *CIRP Annals, 69*(2), 710-733. https://doi.org/10.1201/9780429436543

[10] Ferrando Chacon, J. L. (2015). Fault detection in rotating machinery using acoustic emission. *Doctoral dissertation*, Brunel University London.

[11] Mirzaei, K., Arashpour, M., Asadi, E., Masoumi, H., Bai, Y. & Behnood, A. (2022). 3D point cloud data processing with machine learning for construction and infrastructure applications: A comprehensive review. *Advanced Engineering Informatics, 51*, 101501. https://doi.org/10.1016/j.aei.2021.101501

[12] Parmar, H., Khan, T., Tucci, F., Umer, R. & Carlone, P. (2022). Advanced robotics and additive manufacturing of composites: towards a new era in Industry 4.0. *Materials and manufacturing processes, 37*(5), 483-517. https://doi.org/10.1080/10426914.2020.1866195

[13] Akhavan, J. & Manoochehri, S. (2022, June). Sensory data fusion using machine learning methods for in-situ defect registration in additive manufacturing: a review. In *IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS2022)*, 1-10. https://doi.org/10.1109/IEMTRONICS55184.2022.9795815

[14] Zhang, H., Vallabh, C. K. P. & Zhao, X. (2023). Machine learning enhanced high dynamic range fringe projection profilometry for in-situ layer-wise surface topography measurement during LPBF additive manufacturing. *Precision Engineering, 84*, 1-14. https://doi.org/10.1016/j.precisioneng.2023.06.015

[15] Wegener, K., Spierings, A. B., Teti, R., Caggiano, A., Knüttel, D., & Staub, A. (2021). A conceptual vision for a bio-intelligent manufacturing cell for Selective Laser Melting. *CIRP Journal of Manufacturing Science and Technology, 34*, 61-83. https://doi.org/10.1016/j.cirpj.2020.11.009

[16] Scime, L. & Beuth, J. (2019). Using machine learning to identify in-situ melt pool signatures indicative of flaw formation in a laser powder bed fusion additive manufacturing process. *Additive Manufacturing, 25*, 151-165. https://doi.org/10.1016/j.addma.2018.11.010

[17] Wang, C., Tan, X. P., Tor, S. B. & Lim, C. S. (2020). Machine learning in additive manufacturing: State-of-the-art and perspectives. *Additive Manufacturing, 36*, 101538. https://doi.org/10.1016/j.addma.2020.101538

[18] Li, R., Jin, M. & Paquit, V. C. (2021). Geometrical defect detection for additive manufacturing with machine learning models. *Materials & Design, 206*, 109726. https://doi.org/10.1016/j.matdes.2021.109726

[19] Zhu, Q., Liu, Z. & Yan, J. (2021). Machine learning for metal additive manufacturing: predicting temperature and melt pool fluid dynamics using physics-informed neural networks. *Computational Mechanics, 67*, 619-635. https://doi.org/10.1007/s00466-020-01952-9

[20] Ferreira, R. D. S. B., Sabbaghi, A. & Huang, Q. (2019). Automated geometric shape deviation modeling for additive manufacturing systems via Bayesian neural networks. *IEEE Transactions on Automation Science and Engineering, 17*(2), 584-598. https://doi.org/10.1109/TASE.2019.2936821

[21] Samie Tootooni, M., Dsouza, A., Donovan, R., Rao, P. K., Kong, Z. & Borgesen, P. (2017). Classifying the dimensional variation in additive manufactured parts from laser-scanned three-dimensional point cloud data using machine learning approaches. *Journal of Manufacturing Science and Engineering, 139*(9), 091005. https://doi.org/10.1115/1.4036641

[22] Yuan, S., Shen, F., Chua, C. K. & Zhou, K. (2019). Polymeric composites for powder-based additive manufacturing: Materials and applications. *Progress in Polymer Science, 91*, 141-168. https://doi.org/10.1016/j.progpolymsci.2018.11.001

[23] Francis, J., Khanzadeh, M., Doude, H., Hammond, V. & Bian, L. (2019). Data-driven calibration for infrared camera in additive manufacturing. *International Journal of Rapid Manufacturing, 8*(4), 302-325. https://doi.org/10.1504/IJRAPIDM.2019.102531

[24] He, K., Zhang, X., Ren, S. & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 770-778. https://doi.org/10.1109/CVPR.2016.90

[25] Goodfellow, I., Bengio, Y. & Courville, A. (2016). *Deep learning*. MIT press.

[26] Howard, A. G. (2013). Some improvements on deep convolutional neural network based image classification. *arXiv* preprint arXiv:1312.5402.

[27] Perez, L. & Wang, J. (2017). The effectiveness of data augmentation in image classification using deep learning. *arXiv* preprint arXiv:1712.04621.

[28] Shorten, C. & Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. *Journal of Big Data, 6*(1), 1-48. https://doi.org/10.1186/s40537-019-0197-0

[29] Simard, P. Y., Steinkraus, D. & Platt, J. C. (2003). Best practices for convolutional neural networks applied to visual document analysis. In *Icdar, Vol. 3*, 958-962. https://doi.org/10.1109/ICDAR.2003.1227801

[30] Hadsell, R., Chopra, S. & LeCun, Y. (2006). Dimensionality reduction by learning an invariant mapping. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR2006), Vol. 2*, 1735-1742. https://doi.org/10.1109/CVPR.2006.100

[31] Bo-Kai Ruan. (2023). 3D Printer Defected Dataset. Kaggle. https://www.kaggle.com/datasets/justin900429/3d-printer-defected-dataset

Authors' contacts:

Laukesh Kumar
(Corresponding author)
USICT, Guru Gobind Singh Indraprastha University,
Dwarka, New Delhi, Delhi 110078, India
laukeshbhati@gmail.com

Manoj Kumar Satyarthi
USICT, Guru Gobind Singh Indraprastha University,
Dwarka, New Delhi, Delhi 110078, India

ISSN 1846-6168 (Print), ISSN 1848-5588 (Online)
https://doi.org/10.31803/tg-20241018061352

Original scientific paper
Received: 2024-10-18, Accepted: 2024-10-29

Green Entrepreneurial Orientation and Sustainable Performance: A Moderated Mediation Analysis in Chinese Manufacturing

Wenli Hu, Teetut Tresirichod*

Abstract: This study investigates the mechanisms through which green entrepreneurial orientation (GEO) influences sustainable performance (SuP) in Chinese manufacturing enterprises, focusing on the mediating role of green intellectual capital (GIC) and the moderating effect of environmental uncertainty (EU). Drawing on the Resource-Based View and Natural Resource Base View, we develop and test a moderated mediation model using data from 486 Chinese manufacturing firms. Results from partial least squares structural equation modeling (PLS-SEM) reveal that GEO positively influences SuP both directly and indirectly through GIC. Moreover, EU negatively moderates the relationship between GIC and SuP. These findings contribute to the literature on green entrepreneurship and sustainable performance by elucidating the complex relationships among GEO, GIC, EU, and SuP in the context of Chinese manufacturing. The study offers important implications for theory development and managerial practice in promoting sustainable development in manufacturing industries.

Keywords: Environmental uncertainty; Green Entrepreneurial Orientation; Green Intellectual Capital; SmartPLS; Sustainable Performance

1 INTRODUCTION

The traditional extensive production model, while generating significant short-term economic benefits for enterprises, has also led to severe environmental issues [1]. The negative impacts of economic growth on the environment have gradually increased, contributing to the continuous escalation of global environmental crises [2]. This production approach, characterized by high resource consumption and inefficiency, has caused widespread ecological degradation, including exacerbated pollution, resource depletion, and the imbalance of ecosystems [3, 4]. The conflict between economic efficiency and environmental degradation is intensifying, leading to a public demand for sustainable development. More than 138 countries around the world have proposed "carbon-free" or "carbon-neutral" targets. The Chinese government has additionally enacted various environmental policies and regulations that set higher sustainability requirements for companies and require them to disclose environmental information [5]. Simultaneously, green products are increasingly seen as a consumer advantage and a significant preference for consumers [6], who are willing to buy environmentally friendly products and pay a premium for them. For example, the explosive growth of new energy vehicles in Europe, America and Asia has led to the gradual recognition of "green cars" by consumers [7]. This change has also prompted companies to accelerate the pace of green transformation under the triple pressure of environmental pollution, policy constraints and social demands.

In this context, based on the Resource-Based View (RBV), some scholars argue that entrepreneurial activities can address environmental issues [8]. However, the impact of entrepreneurship on the environment is dual-sided: on one hand, it may result in environmental pollution, ecological degradation, and the depletion of non-renewable resources; on the other hand, it can also drive innovation to solve environmental problems [9]. The RBV posits that a firm's sustained competitive advantage stems from its internally possessed scarce and inimitable resources and capabilities. As a novel topic in the field of entrepreneurship, green entrepreneurial orientation enables firms to protect the environment while pursuing economic benefits, garnering sustained attention from both academia and business managers [10]. Although existing studies have mainly explored the relationship between GEO and green innovation, suggesting that green entrepreneurial orientation has a catalytic effect on green incremental innovation and green radical innovation [11], there are also some studies that confirm the positive impact of green entrepreneurial orientation on firms' financial performance [12], but ignore its key role in driving sustainable performance. Moreover, green intellectual capital, as a core resource of a firm, encompasses the knowledge, skills, and abilities that organizational employees contribute to creating and managing sustainable value, with the aim of integrating environmental protection principles into business operations [13]. Green intellectual capital supports firms in building sustainable development.

However, the existing literature has insufficiently addressed the role of green intellectual capital in implementing green entrepreneurial orientation and enhancing sustainable performance, and there is also a lack of research on environmental uncertainty. Environmental uncertainty refers to technological uncertainty and demand uncertainty [14], and current studies seldom consider the external environmental changes that firms face during the implementation of green entrepreneurial orientation [15]. In highly uncertain environments, firms need to be more flexible and innovative to respond to challenges.

Therefore, this study aims to emphasize how firms can effectively integrate and utilize internal resources to gain sustainable advantages when facing external environmental challenges, based on the Resource-Based View (RBV) theory. It explores the pathways through which green entrepreneurial orientation influences sustainable performance, focusing on the mediating role of green intellectual capital and the moderating effect of environmental uncertainty. By employing Structural Equation Modeling with Partial Least Squares (SEM-PLS) to

analyze questionnaire data from Chinese manufacturing enterprises, the study validates the relationships among the variables. The findings provide empirical support for the theoretical model and offer managerial recommendations for firms on how to respond to environmental uncertainty and enhance sustainable performance during the green transformation process.

2 LITERATURE REVIEW

2.1 Green Entrepreneurial Orientation and Sustainable Performance

The growing number of environmental issues in the world gave rise to the idea of "green entrepreneurship." Many international organizations stress the need of businesses embracing a green business paradigm and call for a worldwide shift to green development [16]. Green business orientation is a strategic approach in which companies proactively adopt various environmental measures, innovations and management practices with environmental protection and sustainable development at the center of their business processes. Academics generally agree that an entrepreneurial mindset improves the success of businesses. Instead of only concentrating on generating short-term profits, entrepreneurs are now also adopting social responsibility and long-term corporate sustainability [17]. Elkington [18] introduced the concept of SuP and argued for corporate development that considers the natural environment, social and economic performance simultaneously. This perspective has been recognized by a wide range of scholars [19] and it is consistent with the triple bottom line principle, which emphasizes that SuP should prioritize social responsibility and environmental preservation in addition to economic rewards [20].

GEO has been shown to make it easier for companies to engage in green activities, enabling them to seize development opportunities and gain new competitive advantages amid complex environmental changes [12]. Previous studies have documented the influence of GEO on various facets of business operations, including green innovation and green management [21]. Ullah and Qaiser Danish [22] found that a GEO can stimulate firms to innovate and adopt environmentally friendly production methods, thereby positively impacting performance. Companies with a GEO have a higher level of environmental performance [23]. Fatoki [20] discovered a strong relationship between GEO and financial, environmental and social performance. Thus, this research posits the following hypotheses.

H1: Green entrepreneurial orientation has a positive impact on sustainable performance.

2.2 Green Entrepreneurial Orientation and Green Intellectual Capital

The concept of GIC was originally introduced by Chen [24] and has since gained wide recognition. Organizational intellectual capital is the combined inventory of intangible assets, information, skills, and connections held by people and organizations inside a company who are involved in environmental preservation or green innovation [25, 26]. Based on the resource-based view (RBV), organizations that implement a particular strategy can accumulate a set of resources aligned with that strategy. Organizations that adopt a green business orientation focus on integrating green technologies into their production processes and strive to minimize resource consumption [27]. This promotes environmental awareness among organizational members, facilitates the accumulation of knowledge, experience and skills related to environmental protection [24], and culminates in the creation of green capital. A company's strategic development and commitment significantly influence employee behavior [19]. Companies actively engage in environmental initiatives and collaborate with various stakeholders such as government agencies, non-governmental organizations (NGOs) and customers to cultivate a green corporate ecosystem. This collaboration enables employees to acquire advanced environmental technologies and management skills while promoting the development of green relational capital. Therefore, the study hypothesizes the following.

H2: Green entrepreneurial orientation has a positive effect on green intellectual capital.

2.3 Green Intellectual Capital and Sustainable Performance from an Intellectual Capital Perspective

Knowledge is the most important strategic resource for any organization. The intellectual capital perspective highlights the central role of intellectual resources, including knowledge, skills and innovation, within organizations, especially in today's knowledge-based economy and innovation-driven era. Within the intellectual capital perspective, sustainability serves as a driving force for the development of future organizations that are able to meet the challenges of sustainable development through knowledge [28]. Intellectual capital plays a crucial role in sustainable development as it provides an effective organizational approach to achieve both economic and social benefits [29]. Cavicchi and Vagnoni [30] believe that intellectual capital forms the basis for sustainable activities in companies. GIC represents a special form of intellectual resource that emphasizes non-economic goals, promotes sustainable thinking and helps companies achieve long-term competitive advantages [24]. In addition to improving environmental performance, GIC facilitates effective resource management in the manufacturing process, which lowers operating costs and boosts the business's overall performance [31]. Therefore, the study hypothesizes the following.

H3: Green intellectual capital has a positive impact on sustainable performance.

2.4 The Mediating Role of Green Intellectual Capital in Green Entrepreneurial Orientation and Sustainable Performance

The RBV theory assumes that resources and capabilities within an organization are critical to achieving competitive advantage. By implementing green strategies, organizations can invest in and accumulate various forms of GIC, including expertise, technical know-how and green brand reputation. The use of GIC to develop new products that meet the environmental protection needs of stakeholders not only

improves a company's environmental image [32], but also strengthens its competitiveness in the field of sustainability and environmental protection [19]. In addition, building GIC contributes internally to the SuP of the company, which brings economic, environmental and social benefits [32]. As a result, the research makes the hypotheses that follow.

H4: Green intellectual capital mediates the relationship between green entrepreneurial orientation and sustainable performance.

2.5 The Moderating Role of Environmental Uncertainty in Green Entrepreneurial Orientation and Sustainable Performance

Green Entrepreneurial Orientation, which is recognized as an important organizational resource, is believed to positively influence the SuP of firms [19]. According to RBV theory, organizations facing technological and demand-side uncertainty tend to rely more on internal resources to compensate for the instability of the external environment. Increasing EU poses a huge challenge for organizations and makes it difficult to accurately predict and respond to future trends [33, 34]. Firms need to consider their own resources and capabilities and adopt proactive approaches to manage environmental performance in order to achieve sustainable development [35]. This can affect the effectiveness of organizational knowledge acquisition, technological innovation and collaboration between organizational members [36], forcing companies to make major investments to acquire additional knowledge and new skills from external sources [34]. In essence, the knowledge and skills inherent in firms' resources may hinder their adaptability and anticipation of changes in the external environment and consequently reduce their efficiency in addressing environmental challenges. This potential reduction in the positive effect of GIC on SuP forms the basis for the following hypothesis put forward in this study.

H5: Environmental uncertainty moderates the relationship between Green intellectual capital and corporate sustainable performance.

Fig. 1 shows the research framework.

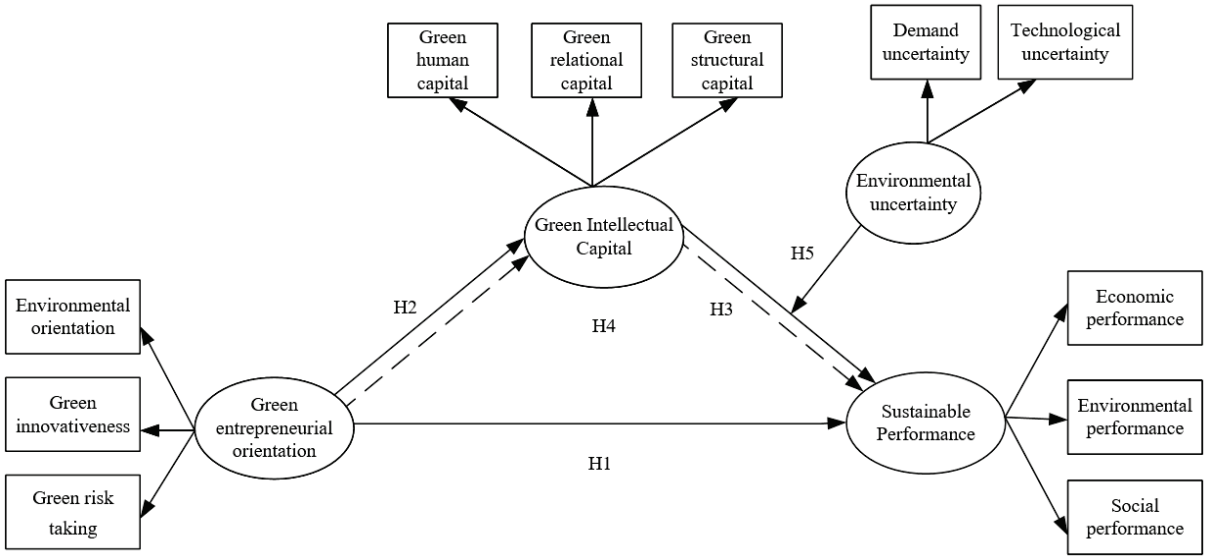


Figure 1 Conceptual Framework

3 MATERIAL AND METHODS

3.1 Measures

The study utilized the Environmental Orientation (EO), Green innovativeness (GI), Green risk taking (GRT) to measure GEO, which has been validated in previous research [37]. This scale comprises 10 items. To measure GIC, the study employed the Green human capital (GHC), Green relational capital (GRC), Green structural capital (GSC) to measure GIC, which has been validated in previous research [38], consisting of 12 items. For SuP, the study used the Economic performance (EcP), Environmental performance (EP), Social performance (SP) to measure SuP, which has been validated in previous research [39], which include 13 items. The study also measured EU using the Demand uncertainty (DU), Technological uncertainty (TU) scales, validated by Wu [14], comprising 6 items. All the reference scales for CEO, GIC, SuP, and EU demonstrated good reliability (with all $CR > 0.7$) and utilized a 5-point Likert scale.

3.2 Population, Sample, and Data Collection

Manufacturing is an important pillar of China's economic development, but also the most resource-consuming industry, so how to balance the development of the manufacturing industry and environmental governance is a difficult problem that needs to be dealt with urgently. In this study, the manufacturing industry in China was selected as the research area, and CEOs or departmental managers and deputy managers of manufacturing enterprises were the research subjects. The study used a questionnaire to obtain data, and data collection lasted two months. After eliminating the samples with the same answers to multiple questions and

too short filling time, 486 valid questionnaires were obtained. Specifically, according to the China Statistical Yearbook 2023, 335 questionnaires were sent to the Light Textile Industry (29%), 424 to the Resource Processing Industry (36%), and 411 to the Machinery and Electronic Manufacturing sector (approximately 35%), totaling 1170 questionnaires. Data collection spanned two months. After excluding samples with multiple identical answers and excessively short completion times, 486 valid questionnaires were obtained. Of these, 323 respondents were male (66.46%) and 163 female (33.54%). In terms of education, 12 respondents had a secondary school degree or less (2.47%), 56 had a junior college degree (11.52%), 322 had a college degree (66.26%), and 96 had a postgraduate degree or higher (19.75%). In terms of age, 45 of the respondents were under 30 years old (9.26%), 138 were between 31 and 40 years old (28.40%), 204 were between 41 and 50 years old (41.98%) and 99 were over 51 years old (20.37%). The sample comprised 79 general managers (16.26%), 342 middle managers (70.37%) and 65 top managers (13.37%). In terms of company size, 77 companies had 51-299 employees (15.84%), 264 had 300-1000 employees (54.32%) and 145 had 1,000 or more employees (29.84%). In addition, 133 companies were in the textile industry (27.37%), 196 in the raw materials processing industry (40.33%) and 157 in the machinery and electronics industry (32.30%).

SPSS was used for data collection, screening, demographic analysis and descriptive analysis. In this study, SmartPLS was used for both measurement model and structural model analysis. Compared to other methods, Partial Least Squares Structural Equation Modeling (PLS-SEM) is particularly well-suited for handling complex multi-structured models and managing intricate causal relationships between variables. Moreover, PLS-SEM does not require the data to strictly adhere to normal distribution assumptions, making it effective in generating reliable results even when sample sizes are limited or data distribution is skewed. Therefore, this study employs PLS-SEM to ensure the robustness of the research model and the validity of the findings.

Additionally, to address common method bias, measures such as reverse-coded items and randomized response order were incorporated into the questionnaire design to mitigate its effects. During the data collection process, ethical norms were followed, all participants gave informed consent, and the data were kept strictly confidential to ensure the legitimacy of the study and the voluntary nature of participation.

4 RESULTS

The model presented in this study is a higher-order model characterized by a reflective-reflective structure and was calculated using the repeated indicators method. Consequently, the analyses of the measurement model and the structural model were performed after the construction of the model using the repeated indicators method (see Fig. 2).

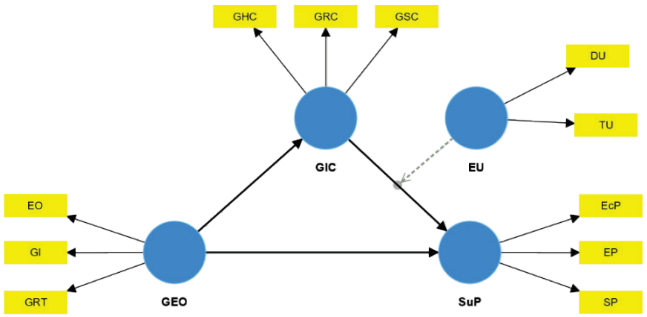


Figure 2 Reduced Form Model

4.1 Assessment of the Measurement Model

Following the methodology proposed by Roldán and Sánchez-Franco [40], we first analyzed the loadings of the indicators (>0.7) [41], Cronbach's alpha (α> 0.7), composite reliability (*CR* > 0.7) [42] and average variance extracted (*AVE* > 0.5) [43, 44]. The present study uses PLS to calculate these metrics and confirms that they exceed the thresholds recommended in the literature (see Tab. 1).

Table 1 Validity and reliability of measurement model

	Item	Loadings	α	*CR*	*AVE*
Green entrepreneurial orientation	Environmental orientation	0.852	0.784	0.793	0.698
	Green innovativeness	0.869			
	Green risk-taking	0.783			
Green intellectual capital	Green human capital	0.838	0.796	0.796	0.710
	Green relational capital	0.844			
	Green structural capital	0.845			
Sustainable performance	Economic performance	0.831	0.794	0.794	0.708
	Environmental performance	0.849			
	Social performance	0.845			
Environmental uncertainty	Demand uncertainty	0.887	0.722	0.723	0.783
	Technological uncertainty	0.882			

Furthermore, we performed a discriminant validity calculation to accurately assess the extent of differentiation and independence between the different constructs. In our analysis, we followed closely the criteria described by Fornell and Larcker [45] and compared the Average Variance Extracted (*AVE*) values of each construct with its squared variance. This comparison ensures that each construct adequately explains its respective metrics while maintaining its exclusivity. The results of these comparisons can be found in Tab. 2.

In addition, we analyzed the values of the external loading metrics for each of the constructs of cross loadings. These values (in bold) were all significantly higher than any of their cross-loading values on the other constructs, in line with the criteria of [46], further confirming the differentiation of the constructs (see Tab. 3).

Table 2 Fornell-Larcker Criterion

	EU	GEO	GIC	SuP
EU	0.885			
GEO	–0.303	0.836		
GIC	–0.301	0.571	0.843	
SuP	–0.324	0.573	0.518	0.842

Table 3 Cross Loading Analysis

	GEO	GIC	SuP	EU	EU × GIC
EO	0.852	0.474	0.448	–0.207	–0.099
GI	0.869	0.540	0.519	–0.267	–0.159
GRT	0.783	0.408	0.465	–0.289	–0.085
GHC	0.482	0.838	0.436	–0.283	–0.109
GRC	0.500	0.845	0.432	–0.251	–0.063
GSC	0.461	0.845	0.441	–0.226	–0.114
EcP	0.471	0.438	0.831	–0.290	–0.308
EP	0.472	0.433	0.849	–0.261	–0.336
SP	0.502	0.436	0.845	–0.267	–0.321
DU	–0.271	–0.233	–0.289	0.887	0.046
TU	–0.266	–0.301	–0.284	0.882	0.027
EU × GIC	–0.140	–0.113	–0.382	0.041	1.000

For a more comprehensive assessment of discriminant validity, we incorporated the Heterotrait-Monotrait Ratio of Correlations (*HTMT* < 0.9) as proposed by Henseler, Ringle and Sarstedt [47] as an evaluation criterion. After completing the PLS calculation, we found that the HTMT for all constructs was below 0.9, which confirmed the discriminant validity of our constructed model. The detailed results can be found in Tab. 4.

Table 4 Heterotrait-Monotrait Ratio (HTMT)

	EU	GEO	GIC	SuP	EU × GIC
EU					
GEO	0.404				
GIC	0.397	0.718			
SuP	0.428	0.724	0.652		
EU × GIC	0.048	0.154	0.127	0.429	

4.2 Structural Modelling Evaluation

(1) Collinearity Issue

In this study, GEO, GIC, SuP and EU are all second-order variables and should ideally show no covariance between the indicators of the individual structures. Multicollinearity is usually assessed using the variance inflation factor (*VIF*). If the *VIF* reaches or exceeds a value of 5 or falls below 0.2, problems with covariance may occur [41]. In our study, the values of *VIF* ranged from 1.453 to 1.820. Hence, the issue of covariance does not adversely affect the path coefficients of the structural model in this study.

(2) Model Fit

The SRMR of this model is 0.066 (<0.08) [48], d_ULS is 0.289 (<0.95), and d_G is 0.170 (<0.95) [49] indicating that the model has a good fit to the set of sample data. Hypothesis testing can be performed.

(3) Structural Model Relationship

Using the bootstrap method offered by SmartPLS, the sample size was set to 5000 and the significance level was set to 0.05. Structural equation modeling was then performed to derive the path relationships (β) between the variables along with the corresponding t-values, p-values and 95% confidence intervals (95% *CI*) (see Tab. 5). The research hypotheses H1, H2, H3, H4 and H5 were then validated (see Fig. 3).

Table 5 Bootstrapping results for structural model evaluation

Hypothesis	Relationship	β	t-value	95% *CI*	Decision
H1	GEO→SuP	0.352	8.254***	[0.268, 0.435]	Supported
H2	GEO→GIC	0.571	19.018***	[0.511, 0.631]	Supported
H3	GIC→SuP	0.243	5.257***	[0.151, 0.335]	Supported
H4	GEO→GIC→SuP	0.139	5.020 ***	[0.086, 0.194]	Supported
H5	EU x GIC→SuP	–0.304	8.643 ***	[-0.372, -0.233]	Supported

Note: * denotes $P < .05$, ** denotes $P < .01$, *** denotes $P < .001$

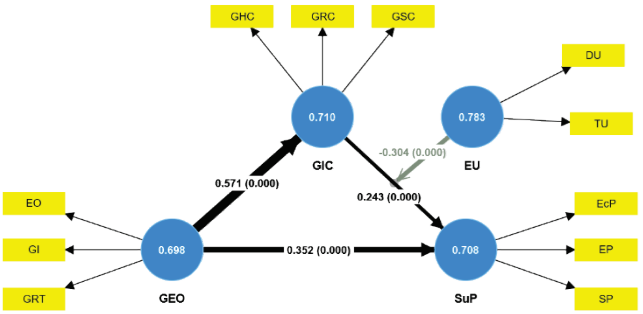


Figure 3 Evaluation of the Structural model (Note: Thicker arrow line means that the effect of independent variable on dependent variable is higher.)

Table 6 Mediation effect test

Hypothesis	Independent variable	Mediating variable	Dependent variable	Direct effect	Indirect effect	Total effect	VAF	Decision
H4	GEO	GIC	SuP	0.352	0.139	0.491	0.283	Partial

Notes: bootstrapping (n = 5000).

In addition, this study investigated the mediating effect of GIC. The mediating effect is only considered valid if the direct effect of the non-mediator is significant and the indirect effect of the mediator is significant. Our results show that GEO exerted a direct effect on SuP and GEO also had a significant indirect effect on SuP through GIC (see Tab. 5), indicating a potential mediating effect. In addition, the variance explained (*VAF*) was estimated to measure the magnitude of the mediating effects. In this study, a *VAF* value of 0.283, ranging from 0.2 to 0.8, suggests that GIC partially mediates the relationship between GEO and SuP (see Tab. 6).

When analyzing the predictive power of the endogenous structure, a key indicator we look at is R^2. The R^2 values for SuP, which are 0.326 and 0.485 respectively, indicate weak predictive power based on the classification criteria established by Hair Jr, Babin and Anderson [42]. Another critical indicator is f^2, which measures the extent of the effect of the explanatory variables on the endogenous variables. The f^2 values in this study ranged from 0.075 to 0.484, which

meets the criteria for judgement. Following Cohen [50] criteria, we find that the effect of GEO on GIC reaches a large effect level. Conversely, GIC shows the smallest effect on SuP and is in the small effect range. To evaluate the predictive accuracy of the model, we calculated the Stone-Geisser test (Q^2), which according to [51] must be greater than 0. Our analysis shows that both GIC (0.227) and SuP (0.334) exceed this threshold (see Table 7). To summarize, despite the R^2 value indicating weak predictive power, the model is reliable to elucidate the relationship between endogenous variables and the predictive accuracy is satisfactory as evidenced by f^2 and Q^2.

5 DISCUSSIONS AND IMPLICATIONS

5.1 Discussion

In response to increasing pressures from environmental concerns, regulatory measures and changing consumer demands, organizations are becoming more motivated to enhance their performance improvement methods by strategically implementing corporate initiatives. This study examines the relationships among green entrepreneurial orientation (GEO), green intellectual capital (GIC), environmental uncertainty (EU), and sustainable performance (SuP) in the context of Chinese manufacturing enterprises. The findings underscore the pivotal role of GEO in driving SuP, as well as the mediating effect of GIC on the relationship between GEO and SuP.

GEO has a positive impact on SuP ($\beta = 0.352, p < 0.001$). We found a positive correlation between GEO and SuP, consistent with recent studies in the business field. For instance, research by Jiang, Chai, Shao and Feng [12] and Fatoki [20], focusing on Chinese enterprises and the South African hotel industry, respectively, confirmed similar phenomena. Sustainable development is closely related to entrepreneurial intentions [52]. Adopting GEO not only helps firms better fulfill their social and environmental responsibilities but also enables them to gain a competitive edge in a fiercely competitive market.

Furthermore, this study shows a positive relationship between GEO and GIC ($\beta = 0.571, p < 0.001$), a finding that differs from previous research. While Wu and Yu [53] postulated a significant positive correlation between GIC and entrepreneurial orientation based on hospital samples and emphasized the constructive effects of relational and structural capital on entrepreneurial orientation, the study conducted by Al-Jinini, Dahiyat and Bontis [54] uncovered the positive influence of intellectual capital on entrepreneurial orientation using Jordanian SMEs as the research cohort. However, despite the intertwining of entrepreneurial orientation and GEO, the latter places a greater emphasis on green business development. This focus not only strengthens market competitiveness but also effectively promotes the accumulation and value enhancement of GIC. In addition, small and medium-sized enterprises (SMEs), constrained by limited resources and facing intense competition in the market, often encounter challenges in pursuing green development before they reach a certain scale. As Chen [24] points out, a deep understanding of the green industry is essential for companies to succeed in the environmentally conscious market. Structuring and utilising corporate resources enables resource accumulation [55]. By adopting a GEO, companies acquire expertise about the green market through practical experience. Consequently, GEO exerts a positive influence on GIC.

Furthermore, this study confirms the positive influence of GIC on SuP ($\beta = 0.243, p < 0.001$). From an intellectual capital perspective, the GIC proves to be a crucial catalyst for the SuP of companies. It facilitates companies to achieve environmental innovation, adaptability and resource efficiency, laying the foundation for long-term success in a changing business landscape. This finding aligns with Yusliza, Yong, Tanveer, Ramayah, Faezah and Muhammad [19], who emphasized the pivotal role of GIC in addressing environmental issues and its significant contribution to sustainable performance. Additionally, Martínez-Falcó, Sánchez-García, Millan-Tudela and Marco-Lajara [31] further confirmed this relationship using a sample of Spanish wineries.

This study highlights the mediating role of GIC between GEO and SuP ($\beta = 0.139, p < 0.001$) and uncovers the inherent link between corporate strategy, intellectual capital and performance. Collaborative networks have an impact in promoting innovation [56]. The adoption of green strategies by companies in conjunction with the advancement of green supply chain management [39] facilitates collaboration and information sharing between different stakeholders. This dissemination of knowledge empowers corporate managers to integrate environmental principles more skillfully and promotes the accumulation of GIC, thereby improving SuP.

Nevertheless, the negative moderating effect of EU on the relationship between GIC and SuP ($\beta = -0.304, p < 0.001$). The application of entrepreneurial methods enables firms to continuously accumulate capital and effectively respond to the ever-changing market environment [57]. Dangelico and Pontrandolfo [58] demonstrated, based on the Natural Resource-Based View, that environmental management capabilities significantly influence both the market competitiveness and performance of enterprises. Due to fluctuations in environmental regulations, uncertainties in technological innovations, and variations in consumer demand for sustainable products, companies face multifaceted challenges in green entrepreneurship [14]. This finding aligns with the resource dependency theory, which posits that organizations facing uncertainties tend to rely more heavily on internal resources to compensate for external instability [36]. These environmental uncertainties not only test a firm's strategic vision and responsiveness but also profoundly impact its green intellectual capital and sustainable performance.

5.2 Theoretical Implication

The RBV emphasizes that firms can achieve competitive advantage by integrating internal and external resources. This study, grounded in the RBV, deepens the understanding of how firms can achieve sustainable development in the context of finite natural resources and environmental pressures. Our research reveals that the implementation of GEO actively fosters the accumulation and development of GIC, providing new insights into the resource accumulation process and extending the application of RBV in green

management. Furthermore, while previous studies have often overlooked the specific pathways through which GEO affects SuP, this study not only confirms the importance of GIC as a mediating variable but also highlights how GEO influences the accumulation of internal green resources, thereby enhancing firms' sustainable performance. This provides a more comprehensive theoretical framework for understanding the relationship between GEO and sustainable development pathways. Finally, the study explores the complex interactions between EU, GEO, GIC, and SuP, revealing how firms navigate environmental uncertainty in the process of green resource accumulation and entrepreneurial orientation to optimize their sustainable performance. This provides a new perspective on how firms can strategically adapt and allocate green resources in uncertain market conditions to achieve green entrepreneurial success and long-term sustainable development.

5.3 Managerial Implication

This study provides significant managerial implications for Chinese manufacturing firms, particularly in resource management and addressing environmental uncertainty. First, firms should position GEO as the core of their strategy, integrating resource and environmental management to address the constraints of finite natural resources and environmental pressures. Managers should prioritize sustainability by embedding it deeply into strategic planning and day-to-day operations to ensure a competitive edge in future markets. Second, firms need to strengthen the accumulation and management of GIC by implementing employee training programs, adopting green technologies and knowledge, and enhancing green innovation capabilities. For Chinese manufacturing enterprises, managers can establish dedicated green R&D departments and collaborate with external research institutions to drive green technology innovation and product-service upgrades, thereby creating a differentiated advantage in green competition. Furthermore, in the face of EU, firms must flexibly adjust their strategies and product designs to adapt to changing conditions. Companies should also establish effective market sensing and response mechanisms, continuously accumulating GIC to promote green technology innovation and product development, thereby driving the green transformation of the supply chain. Managers can leverage both internal R&D and external collaborations to proactively respond to market dynamics and shifts in consumer demands, ensuring long-term sustainable development in a highly competitive landscape. These strategies are not only applicable to the Chinese manufacturing sector but also offer valuable managerial insights for firms in other regions worldwide.

6 CONCLUSION

This study advances our understanding of the relationship between green entrepreneurial orientation and sustainable performance in Chinese manufacturing enterprises. By identifying the mediating role of green intellectual capital and the moderating effect of environmental uncertainty, we provide a more nuanced view of how firms can leverage their green strategies to enhance sustainability outcomes. Our findings highlight the importance of developing green intellectual capital as a key mechanism for translating green entrepreneurial orientation into improved sustainable performance. However, the negative moderating effect of environmental uncertainty underscores the challenges firms face in maintaining performance in volatile environments. These insights offer valuable guidance for managers seeking to implement effective sustainability strategies in the manufacturing sector. Future research should explore these relationships in different industrial and national contexts, consider potential non-linear effects, and investigate additional mediators and moderators to further refine our understanding of the green entrepreneurship-sustainability performance link.

7 REFERENCES

[1] Stern, D. I. (2004). The rise and fall of the environmental Kuznets curve. *World development*, *32*(8), 1419-1439. https://doi.org/10.1016/j.worlddev.2004.03.004

[2] Porcelli, A. M. (2022). After the collapse: Evaluating undone science in the wake of a global environmental crisis. *Geoforum*, *132*, 219-228. https://doi.org/10.1016/j.geoforum.2021.10.011

[3] Lin, R.-J., Tan, K.-H. & Geng, Y. (2013). Market demand, green product innovation, and firm performance: evidence from Vietnam motorcycle industry. *Journal of Cleaner Production*, *40*, 101-107. https://doi.org/10.1016/j.jclepro.2012.01.001

[4] Leonidou, L. C., Christodoulides, P., Kyrgidou, L. P. & Palihawadana, D. (2017). Internal drivers and performance consequences of small firm green business strategy: The moderating role of external forces. *Journal of business ethics*, *140*, 585-606. https://doi.org/10.1007/s10551-015-2670-9

[5] Ahmad, N., Li, H.-Z. & Tian, X.-L. (2019). Increased firm profitability under a nationwide environmental information disclosure program? Evidence from China. *Journal of Cleaner Production*, *230*, 1176-1187. https://doi.org/10.1016/j.jclepro.2019.05.161

[6] Tully, S. M. & Winer, R. S. (2014). The role of the beneficiary in willingness to pay for socially responsible products: A meta-analysis. *Journal of retailing*, *90*(2), 255-274. https://doi.org/10.1016/j.jretai.2014.03.004

[7] Sun, X., Li, Z., Wang, X. & Li, C. (2020). Technology Development of Electric Vehicles: A Review. *Energies*, *13*(1), 90. https://doi.org/10.3390/en13010090

[8] Coase, R. H. (1974). The lighthouse in economics. *The journal of law and economics*, *17*(2), 357-376. https://doi.org/10.1086/466796

[9] Criscuolo, C. & Menon, C. (2015). Environmental policies and risk finance in the green sector: Cross-country evidence. *Energy Policy*, *83*, 38-56. https://doi.org/10.1016/j.enpol.2015.03.023

[10] Gast, J., Gundolf, K. & Cesinger, B. (2017). Doing business in a green way: A systematic review of the ecological sustainability entrepreneurship literature and future research directions. *Journal of Cleaner Production*, *147*, 44-56. https://doi.org/10.1016/j.jclepro.2017.01.065

[11] Guo, Y., Wang, L. & Chen, Y. (2020). Green entrepreneurial orientation and green innovation: The mediating effect of supply chain learning. *SAGE Open*, *10*(1). https://doi.org/10.1177/2158244019898798

[12] Jiang, W., Chai, H., Shao, J. & Feng, T. (2018). Green entrepreneurial orientation for enhancing firm performance: A

dynamic capability perspective. *Journal of Cleaner Production*, *198*, 1311-1323. https://doi.org/10.1016/j.jclepro.2018.07.104

[13] Munawar, S., Yousaf, H. Q., Ahmed, M. & Rehman, S. (2022). Effects of green human resource management on green innovation through green human capital, environmental knowledge, and managerial environmental concern. *Journal of Hospitality and Tourism Management*, *52*, 141-150. https://doi.org/10.1016/j.jhtm.2022.06.009

[14] Wu, G. C. (2013). The influence of green supply chain integration and environmental uncertainty on green innovation in Taiwan's IT industry. *Supply Chain Management: An International Journal*, *18*(5), 539-552. https://doi.org/10.1108/SCM-06-2012-0201

[15] Simangunsong, E., Hendry, L. C. & Stevenson, M. (2012). Supply-chain uncertainty: a review and theoretical foundation for future research. *International journal of production research*, *50*(16), 4493-4523. https://doi.org/10.1080/00207543.2011.613864

[16] Farinelli, F., Bottini, M., Akkoyunlu, S. & Aerni, P. (2011). Green entrepreneurship: the missing link towards a greener economy. *Atdf Journal*, *8*(3/4), 42-48.

[17] İyigün, N. Ö. (2015). What could entrepreneurship do for sustainable development? A corporate social responsibility-based approach. *Procedia-Social and Behavioral Sciences*, *195*, 1226-1231. https://doi.org/10.1016/j.sbspro.2015.06.253

[18] Elkington, J. (1998). Partnerships from cannibals with forks: The triple bottom line of 21st-century business. *Environmental quality management*, *8*(1), 37-51. https://doi.org/10.1002/tqem.3310080106

[19] Yusliza, M. Y., Yong, J. Y., Tanveer, M. I., Ramayah, T., Faezah, J. N. & Muhammad, Z. (2020). A structural model of the impact of green intellectual capital on sustainable performance. *Journal of Cleaner Production*, *249*, 119334. https://doi.org/10.1016/j.jclepro.2019.119334

[20] Fatoki, O. (2019). Green entrepreneurial orientation and firm performance in South Africa. *Entrepreneurship and Sustainability Issues*, *7*(1), 247. https://doi.org/10.9770/jesi.2019.7.1(19)

[21] Li, X., Yang, J., Liu, H. & Zhuang, X. (2021). Entrepreneurial orientation and green management in an emerging economy: The moderating effects of social legitimacy and ownership type. *Journal of Cleaner Production*, *316*, 128293. https://doi.org/10.1016/j.jclepro.2021.128293

[22] Ullah, S. & Qaiser Danish, R. (2020). The impact of green entrepreneurial orientation on firm performance through green innovation: The moderating role of strategic green marketing orientation. *European Online Journal of Natural and Social Sciences*, *9*(2), 306-317.

[23] Makhloufi, L., Laghouag, A. A., Meirun, T. & Belaid, F. (2022). Impact of green entrepreneurship orientation on environmental performance: The natural resource-based view and environmental policy perspective. *Business Strategy and the Environment*, *31*(1), 425-444. https://doi.org/10.1002/bse.2902

[24] Chen, Y.-S. (2008). The positive effect of green intellectual capital on competitive advantages of firms. *Journal of business ethics*, *77*, 271-286. https://doi.org/10.1007/s10551-006-9349-1

[25] Asiaei, K., O'Connor, N. G., Barani, O. & Joshi, M. (2023). Green intellectual capital and ambidextrous green innovation: The impact on environmental performance. *Business Strategy and the Environment*, *32*(1), 369-386. https://doi.org/10.1002/bse.3136

[26] Marco-Lajara, B., Zaragoza-Sáez, P., Martínez-Falcó, J. & Sánchez-García, E. (2022). Green intellectual capital in the Spanish wine industry. In *Innovative Economic, Social, and Environmental Practices for Progressing Future Sustainability* (pp. 102-120). IGI Global. https://doi.org/10.4018/978-1-7998-9590-9.ch006

[27] Triguero, A., Moreno-Mondéjar, L. & Davia, M. A. (2013). Drivers of different types of eco-innovation in European SMEs. *Ecological economics*, *92*, 25-33. https://doi.org/10.1016/j.ecolecon.2013.04.009

[28] López-Gamero, M. D., Zaragoza-Sáez, P., Claver-Cortés, E. & Molina-Azorín, J. F. (2011). Sustainable development and intangibles: building sustainable intellectual capital. *Business Strategy and the Environment*, *20*(1), 18-37. https://doi.org/10.1002/bse.666

[29] Secundo, G., Ndou, V., Del Vecchio, P. & De Pascale, G. (2020). Sustainable development, intellectual capital and technology policies: A structured literature review and future research agenda. *Technological Forecasting and Social Change*, *153*, 119917. https://doi.org/10.1016/j.techfore.2020.119917

[30] Cavicchi, C. & Vagnoni, E. (2017). Does intellectual capital promote the shift of healthcare organizations towards sustainable development? Evidence from Italy. *Journal of Cleaner Production*, *153*, 275-286. https://doi.org/10.1016/j.jclepro.2017.03.175

[31] Martínez-Falcó, J., Sánchez-García, E., Millan-Tudela, L. A. & Marco-Lajara, B. (2023). The role of green agriculture and green supply chain management in the green intellectual capital–Sustainable performance relationship: A structural equation modeling analysis applied to the Spanish wine industry. *Agriculture*, *13*(2), 425. https://doi.org/10.3390/agriculture13020425

[32] Khan, A., Hussain, S. & Sampene, A. K. (2023). Investing in green intellectual capital to enhance green corporate image under the Influence of green innovation climate: A Case of Chinese Entrepreneurial SMEs. *Journal of Cleaner Production*, *418*, 138177. https://doi.org/10.1016/j.jclepro.2023.138177

[33] Firdausy, A. N., Bastian, E. & Fitriyani, F. (2022). Interaction of Environmental Uncertainty, Characteristics of Management Accounting Information Systems and Decentralization on Managerial Performance. *Journal of Applied Business, Taxation and Economics Research*, *2*(2), 169-187. https://doi.org/10.54408/jabter.v2i2.143

[34] Haarhaus, T. & Liening, A. (2020). Building dynamic capabilities to cope with environmental uncertainty: The role of strategic foresight. *Technological Forecasting and Social Change*, *155*, 120033. https://doi.org/10.1016/j.techfore.2020.120033

[35] Ramanathan, R., He, Q., Black, A., Ghobadian, A. & Gallear, D. (2017). Environmental regulations, innovation and firm performance: A revisit of the Porter hypothesis. *Journal of Cleaner Production*, *155*, 79-92. https://doi.org/10.1016/j.jclepro.2016.08.116

[36] Helfat, C. E. & Peteraf, M. A. (2015). Managerial cognitive capabilities and the microfoundations of dynamic capabilities. *Strategic management journal*, *36*(6), 831-850. https://doi.org/10.1002/smj.2247

[37] Xia, H. (2019). Relationship between green entrepreneurial orientation, positive market orientation and innovation performance of technology start-ups. *Sci. Technol. Manage. Res*, *16*, 264-274.

[38] Yusoff, Y. M., Omar, M. K., Zaman, M. D. K. & Samad, S. (2019). Do all elements of green intellectual capital contribute toward business sustainability? Evidence from the Malaysian context using the Partial Least Squares method. *Journal of Cleaner Production*, *234*, 626-637. https://doi.org/10.1016/j.jclepro.2019.06.153

[39] Habib, M. A., Bao, Y. & Ilmudeen, A. (2020). The impact of green entrepreneurial orientation, market orientation and green supply chain management practices on sustainable firm performance. *Cogent Business & Management*, *7*(1), 1743616. https://doi.org/10.1080/23311975.2020.1743616

[40] Roldán, J. L. & Sánchez-Franco, M. J. (2012). Variance-based structural equation modeling: Guidelines for using partial least squares in information systems research. In *Research methodologies, innovations and philosophies in software systems engineering and information systems* (pp. 193-221). IGI global. https://doi.org/10.4018/978-1-4666-0179-6.ch010

[41] Hair, J. F., Ringle, C. M. & Sarstedt, M. (2011). PLS-SEM: Indeed a silver bullet. *Journal of Marketing theory and Practice*, *19*(2), 139-152. https://doi.org/10.2753/MTP1069-6679190202

[42] Hair Jr, J. F., Babin, B. J. & Anderson, R. E. (2010). A global p-erspect-ivie. *Kennesaw: Kennesaw State University*.

[43] Hair, J. F., Sarstedt, M., Ringle, C. M. & Mena, J. A. (2012). An assessment of the use of partial least squares structural equation modeling in marketing research. *Journal of the academy of marketing science*, *40*, 414-433. https://doi.org/10.1007/s11747-011-0261-6

[44] Hulland, J. (1999). Use of partial least squares (PLS) in strategic management research: A review of four recent studies. *Strategic management journal*, *20*(2), 195-204. https://doi.org/10.1002/(SICI)1097-0266(199902)20:2%3C195::AID-SMJ13%3E3.0.CO;2-7

[45] Fornell, C. & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of marketing research*, *18*(1), 39-50. https://doi.org/10.1177/002224378101800104

[46] Chin, W. W. (1998). The partial least squares approach to structural equation modeling. *Modern methods for business research*, *295*(2), 295-336.

[47] Henseler, J., Ringle, C. M. & Sarstedt, M. (2015). A new criterion for assessing discriminant validity in variance-based structural equation modeling. *Journal of the academy of marketing science*, *43*, 115-135. https://doi.org/10.1007/s11747-014-0403-8

[48] Hu, L.-t. & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological methods*, *3*(4), 424. https://doi.org/10.1037/1082-989X.3.4.424

[49] Dijkstra, T. K. & Henseler, J. (2015). Consistent and asymptotically normal PLS estimators for linear structural equations. *Computational statistics & data analysis*, *81*, 10-23. https://doi.org/10.1016/j.csda.2014.07.008

[50] Cohen, J. (1988). Set correlation and contingency tables. *Applied psychological measurement*, *12*(4), 425-434. https://doi.org/10.1177/014662168801200410

[51] Stone, M. (1974). Cross-validatory choice and assessment of statistical predictions. *Journal of the royal statistical society: Series B (Methodological)*, *36*(2), 111-133. https://doi.org/10.1111/j.2517-6161.1974.tb00994.x

[52] Kuckertz, A. & Wagner, M. (2010). The influence of sustainability orientation on entrepreneurial intentions—Investigating the role of business experience. *Journal of business Venturing, 25*(5), 524-539. https://doi.org/10.1016/j.jbusvent.2009.09.001

[53] Wu, W. & Yu, L. (2023). How Does Environmental Corporate Social Responsibility Affect Technological Innovation? The Role of Green Entrepreneurial Orientation and Green Intellectual Capital. *Journal of the Knowledge Economy*, 1-32. https://doi.org/10.1007/s13132-023-01132-z

[54] Al-Jinini, D. K., Dahiyat, S. E. & Bontis, N. (2019). Intellectual capital, entrepreneurial orientation, and technical innovation in small and medium-sized enterprises. *Knowledge and Process Management*, *26*(2), 69-85. https://doi.org/10.1002/kpm.1593

[55] Wong, C. Y., Wong, C. W. & Boon-Itt, S. (2015). Integrating environmental management into supply chains: a systematic literature review and theoretical framework. *International Journal of Physical Distribution & Logistics Management*, *45*(1/2), 43-68. https://doi.org/10.1108/IJPDLM-05-2013-0110

[56] Liang, X. & Liu, A. M. (2018). The evolution of government sponsored collaboration network and its impact on innovation: A bibliometric analysis in the Chinese solar PV sector. *Research Policy*, *47*(7), 1295-1308. https://doi.org/10.1016/j.respol.2018.04.012

[57] Schaltegger, S. & Wagner, M. (2011). Sustainable entrepreneurship and sustainability innovation: categories and interactions. *Business Strategy and the Environment*, *20*(4), 222-237. https://doi.org/10.1002/bse.682

[58] Dangelico, R. M. & Pontrandolfo, P. (2015). Being 'green and competitive': The impact of environmental actions and collaborations on firm performance. *Business Strategy and the Environment*, *24*(6), 413-430. https://doi.org/10.1002/bse.1828

Authors' contacts:

Wenli Hu, PhD
Faculty of Business Administration,
Rajamangala University of Technology Thanyaburi,
Pathum Thani 12110, Thailand
wenli_h@mail.rmutt.ac.th

Teetut Tresirichod, D.B.A, Assistant Professor
(Corresponding author)
Faculty of Business Administration,
Rajamangala University of Technology, Thanyaburi
Pathum Thani 12110, Thailand
teetut_t@rmutt.ac.th

ISSN 1846-6168 (Print), ISSN 1848-5588 (Online)
https://doi.org/10.31803/tg-20241106104357

Original scientific paper
Received: 2024-11-06, Accepted: 2024-11-25

Extending VUML with Behavioral Modeling: A UML-Based Approach for Multi-View Object Specification

Chaimae Ouali-Alami*, Abdelali El Bdouri, Younes Lakhrissi

Abstract: The VUML (View-based UML) profile makes it possible to create unified, multiview models for complex systems. Nevertheless, the work already done focuses mostly on structural issues and does not fully cover behavioral concerns. With an emphasis on multi-view objects, this study adds capabilities to the VUML profile to represent system behavior. We suggest modifying UML state machines to express view object behaviors and how they combine to form global multi-view object behaviors. Our method allows for the independent construction of view behaviors, followed by a methodical composition process. The outcomes demonstrate how our approach improves VUML's capacity to represent intricate, multi-view system behaviors while preserving concern separation. The work being done here helps to close the gap that exists between structural and behavioral modeling in multi-view systems, which has implications for developing complex software systems using model-driven development.

Keywords: behavioural; fusion; model composition; modeling; UML; VUML

1 INTRODUCTION

In today's ever-changing technological landscape, the development of complex IT systems [1], such as embedded systems or large-scale enterprise software, presents significant challenges. These systems must integrate a multitude of structural components and dynamic behaviors, making it difficult to create global models that capture all aspects simultaneously. Instead, the system's size and complexity are decreased by breaking the application down into many component models. A compositional phase is then initiated to create the application's final release.

The problem of behavioral specification in the context of the VUML profile framework is discussed in this study. The main goal is to describe the distinct behavior of multi-view objects, each of which is made up of view objects that contain data relevant to an actor. This involves addressing two critical aspects: specifying the behaviors of view objects and composing these behaviors to form the global behavior of multi-view objects [2].

Complexity stems from the separate development of views in VUML and their subsequent composition to create multi-view object behaviors [3]. The proposed approach aims to strike a balance by allowing maximum freedom in view of development while providing means to facilitate fusion.

The View-based UML (VUML) profile has been developed to meet the need for a unified modeling approach that considers multiple perspectives within complex systems. Although VUML excels in structural modeling, including the decomposition of systems into manageable views, it has significant limitations regarding behavioral modeling, for the specification and integration of multi-view object behaviors.

This paper aims to close a significant hole in the VUML framework: the effective specification and integration of multi-view object behaviors. The central question of this research is: how can VUML be improved to better support behavioral modeling, thus enabling a more comprehensive representation of complex systems? Our goal is to develop a method that allows the independent development of visual behaviors while ensuring their harmonious composition into a global behavior.

To address this issue, we suggest reusing UML mechanisms [4] for specifying behavior and communication between UML objects [5]. The method makes use of common UML methods to control how VUML view objects behave and how views communicate with one another. This involves using state machines that interact with one another through method calls or signal exchanges to indicate how objects should behave. The method is based on a distinct description, after which the base and view objects' state machines are coordinated.

Our approach utilizes existing UML mechanisms; state machines, to define and coordinate the behavior of the visible objects within the VUML framework. In doing so, we enable developers to independently design the behaviors of each view, which can then be systematically compiled into a global behavior model. This method not only preserves modularity and separation of concerns [6]; but also, significantly reduces the complexity associated with integrating behavioral models into complex systems. The effectiveness of this approach is demonstrated through a detailed case study of the management of an automotive repair agency, showing tangible improvements in the system's behavioral modeling capabilities.

The modeling of complex systems has seen considerable growth, with frameworks like SysML and UML-RT being widely used. SysML, an extension of UML (Unified Modeling Language), is specifically designed for modeling complex engineering systems, such as those encountered in automotive, aerospace, or cyber-physical systems. Although SysML is effective in capturing requirements and structuring systems, it shows its limitations in managing dynamic behaviors and synchronizing different views. In particular, SysML may lack flexibility when behaviors need to be modeled through evolving scenarios.

Moreover, UML-RT, a variant of UML adapted for real-time systems, excels in contexts requiring rapid responses, such as telecommunications, medical devices, or embedded

systems. However, UML-RT presents difficulties in synchronizing states and transitions between its capsules, especially for complex multi-view systems.

The extended VUML approach proposed in this article stands out for its ability to model modularly and integrate multi-view behaviors coherently. This method offers greater flexibility and reduces synchronization conflicts, thus overcoming the limitations encountered by SysML and UML-RT in modeling complex systems.

Behavioral modeling allows for better management of complex systems by clearly defining the dynamic interactions between components. This improves flexibility by allowing different parts to evolve independently, while ensuring their smooth integration into the overall system. By anticipating and representing possible behaviors, this modeling facilitates the predictability of the system's responses to various scenarios, thereby reducing the risks of errors and incompatibilities. It also allows for simulating and testing behaviors before implementation, increasing the robustness and responsiveness of the system.

for example, in the field of Cyber-Physical Systems; In applications such as smart electrical grids, the absence of behavioral modeling can prevent the correct simulation of interactions between the grid and IoT devices. This could lead to errors in load management or risks of overloads, thereby disrupting the overall system.

The following is how the document develops: A few definitions and concepts are presented in Section 2. Viewpoint modeling based on the VUML technique is briefly explained in Section 3. Our method will be used in the implementation given in Section 4. We examine our case study's current general structural and behavioral analysis in this part.

2 STRUCTURE OF A MULTI-VIEW CLASS

A multi-view class typically consists of a base class and a set of view classes. The base class includes structural and behavioral properties that are shared among all views. On the other hand, each view class represents a specific viewpoint or concern and contains attributes and behaviors specific to that view [7]. The views are extensions of the base class, allowing for the segregation of concerns and fine-grained access rights management at the class level. This structure enables the creation of a comprehensive and shareable model accessible from different viewpoints in a multi-view modeling approach.

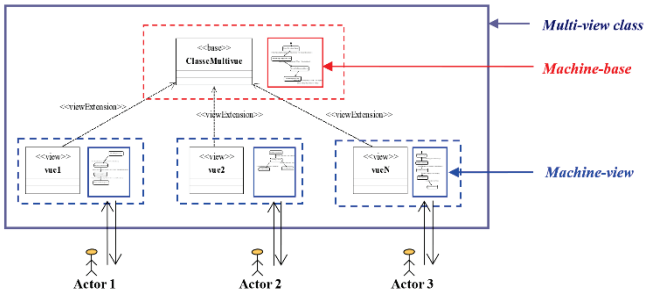


Figure 1 Structure of a multi-view class

The stereotyped data classes, labeled "base" and "view," portray the static aspects of the system [8, 9]. Conversely, the dynamic behavior is illustrated by the state machines, denoted as "machine-base" and "machine-view," which are linked to these classes (Fig. 1).

In the structure and behavior of a multi-view class, the terms "machine base" and "machine view" typically refer to state machines associated with the "base" and "view" classes [10, 11], respectively. Here's a breakdown of their roles:

- **Machine Base:**
  - Role: The "machine base" represents the behavior of the base class within a multi-view class.
  - Functionality: It specifies the dynamic, transitions, and state changes that are specific to the base class [12].
  - Focus: The "machine base" is concerned with the behavior that is common to all views and actors associated with the base class.
  - Scope: It captures the core behavior shared among different perspectives or viewpoints.
- **Machine View:**
  - Role: The "machine view" is responsible for representing the behavior specific to a particular view or actor within the multi-view class.
  - Functionality: It details the dynamic aspects, transitions, and state changes that are unique to the view or actor associated with a specific perspective [12].
  - Focus: The "machine view" concentrates on the behavior that differentiates one view from another within the multi-view class.
  - Scope: It captures the specialized behavior tailored to the requirements of a particular viewpoint.

In summary, the "machine base" encompasses behavior shared across all views, providing a foundation for common functionality, while the "machine view" accounts for variations in behavior specific to individual perspectives or actors within the multi-view class.

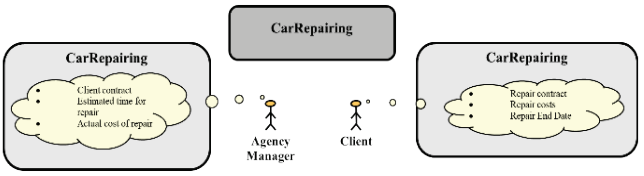


Figure 2 Multi-View machine - Example of the Car Repairing

For example, the car's "CarRepairing" state (in the diagram's center) in Fig. 2 depicts a multi-view state, with varied interpretations based on the actor type. A mechanic's perspective centers around faults, repairs, required tools, and spare parts. On the other hand, a workshop manager views the repair from a logistics standpoint, emphasizing track assignments, equipment reservations, and spare parts allocation. For a client, technical repair details are less relevant, and their interest lies in the repair contract specifics, associated costs, and the repair completion date. Meanwhile, the agency manager's concerns revolve around the financial aspects of the repair, encompassing actual costs, estimated completion times, and contract arrangements with the client.

3 VUML APPROACH

Existing work on VUML (View Unified Modeling Language) primarily focuses on extending classical UML to enable modeling from different perspectives, particularly in complex systems. However, these approaches have certain limitations, such as the difficulty in maintaining consistency among the multiple views, the lack of a standardized framework for integrating behaviors, and increased complexity when implementing models at a large scale. In parallel, other behavioral modeling approaches in UML have been developed, particularly using state diagrams, sequence diagrams, and activity diagrams to represent the interactions and transitions of the system. Extensions like UML-RT or SysML have also emerged to meet the needs of real-time systems and complex systems. However, these methods are often limited by their ability to effectively manage concurrent and distributed behaviors, especially in a multi-perspective context. Another major challenge lies in composing behaviors from multiple viewpoints, which raises issues of coherence and integration of perspectives. It is difficult to ensure correct synchronization of behaviors arising from different views while avoiding conflicts or redundancies. The management of model granularity and the coordination of various levels of abstraction further complicate the task, raising issues of scalability and traceability of the modeled systems.

The VUML (View-Based Unified Modeling Language) approach is a modeling methodology that emphasizes the creation and management of multiple views to capture different aspects of a complex system. It is an extension of the Unified Modeling Language (UML) tailored to address the challenges of modeling large and intricate systems. Here are key aspects of the VUML approach:

- View-Based Modeling: VUML employs a view-based modeling strategy where different views of a system are created to represent various perspectives and concerns. Each view focuses on specific aspects of the system relevant to a particular stakeholder or aspect of the system.
- Multiview Classes: In VUML, the notion of multiview classes is introduced. A multiview class consists of a base class (shared by all actors) and multiple view classes (extensions of the base class), each specific to a particular viewpoint. This allows for fine-grained access rights management and segregation of concerns at the class level.
- Separation of Concerns: VUML emphasizes the separation of concerns in system modeling. By creating distinct views for different stakeholders, concerns such as requirements, design, and behavior can be addressed independently in each view.
- Horizontal and Vertical Modeling: The approach combines both horizontal and vertical modeling.
- Horizontal modeling involves creating models at each level of abstraction, while vertical modeling aligns with the principles of Model-Driven Architecture (MDA) [13, 14].
- Event Observation: VUML introduces the concept of event observation as a first-class object interaction method. Events and probes are formalized to handle interactions between system components. Probes can be used to observe events and work with event data.
- Formalization of Concepts: VUML formalizes various concepts, including the definition of multiview classes, the use of probes, and the creation of a VUML Probe Profile to support event observation.
- Decentralized Design Process: The design process in VUML consists of decentralized phases, including the identification of actor desires, the creation of various PIM (Platform-Independent Model) models, and a composition operation to combine independently created design models into a global VUML design model [15].
- Tool Support: The approach may include tool support to facilitate the modeling process, ensuring consistency and manageability of models and code.

Overall, the VUML approach provides a structured and systematic way to address the complexities of modeling large and diverse systems by employing multiple views and emphasizing the separation of concerns.

In simpler terms, the following is a description of the main ideas of VUML:

- Actor: symbolizes a logical or human being interacting with the system.
- Point of view: represents the viewpoint of an actor on the entire structure or a portion of it. A single point of view is linked to every actor.
- View: A modeling entity, primarily static, that results from applying a point of view to a specific entity (like a class) or, in a broader sense, to the entire system.

The View-based UML (VUML) profile has been developed to meet the need to model complex systems by allowing a unified view from different perspectives. Previous research on VUML has mainly explored its application in structural modeling, where it excels in breaking down systems into separate and manageable views. However, although VUML offers powerful tools for structural modeling, it has notable limitations with regard to behavioral modeling. In particular, the specification and composition of multi-view object behaviors remain major challenges. Existing work has often failed to deal in depth with how individual behaviors of views can be coherently integrated to form a global behavior. This gap limits the effectiveness of VUML in capturing complex system dynamics, thus justifying the need to develop additional mechanisms to improve its ability to model complex behaviors in a multi-view framework.

4 METHODOLOGY

This methodology enhances VUML by introducing "machine-base" and "machine-view" state machines to capture the behaviors of the actors. It unfolds in three phases: identifying the needs of the stakeholders, creating PIM

models corresponding to different viewpoints, and composing these models to form a unified design. The overall behavioral analysis and viewpoints ensure consistency among the perspectives of the different stakeholders. Finally, the merging of behaviors ensures a harmonious integration of interactions and workflows, creating a coherent system.

Three primary stages of development are included in the VUML approach:

Identification of Actor Desires (Global Analysis):

- The initial step focuses on recognizing the desires of the actors.
- The primary objective is to formulate a requirements model, typically represented as a UML use case diagram.

Creation of Various PIM Models:

- The second decentralized stage involves generating multiple PIM (Platform-Independent Model) models, each corresponding to a distinct viewpoint.
- These models encompass various UML diagrams such as class diagrams, state machines, and sequence diagrams.
- The output of this phase is a set of UML models that cater to the needs of different actors.

Composition Operation:

- The final stage is a composition operation, which entails combining independently created design models to formulate a comprehensive VUML design model.
- This phase is critical for achieving an integrated and holistic representation of the system, as the independently developed models are combined to create a unified design.

We propose a three-step approach: (1) global behavioral analysis, (2) view-specific behavioral analysis, and (3) behavioral composition [16] (Tab. 1).

Table 1 Methodology of Behavioral Analysis and Design

Phase	Sub-steps	Objectives	Tools and representations
Global Behavioral Analysis	- Identify the needs of the stakeholders (analysis of the desires of the stakeholders) - Create a needs model in the form of a use case diagram - Specify the overall behavior using state machines	- Understand the overall behavior of the system - Ensure a coherent basis for specific viewpoints	- UML use case diagrams - UML state machines compliant with the UML Omega standard
Analysis/Design by Viewpoint	- Generate PIM models for each viewpoint - Include UML diagrams: classes, state machines, sequences - Refine the behaviors specific to each actor (design to each viewpoint)	- Capture the specific behaviors of each actor - Reflect the particular needs in the form of separate models	- UML diagrams (classes, sequences, state machines) - UML Omega for complex transitions
Merging/Synchronization of Behaviors	- Identify and resolve conflicts between viewpoints - Align the states and transitions between the actors - Create a global model integrating all viewpoints	- Integrate the behavioral models of the actors into a coherent system - Resolve conflicts and maintain overall coherence	- Workflow alignment diagrams - IFx tool for formal verification - State transition synchronization

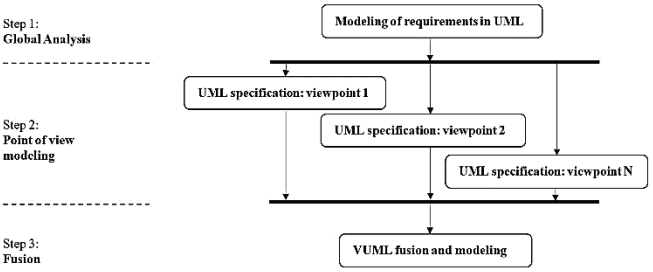


Figure 3 Development phases of VUML Global behavioral analysis

In this subsection, we describe how the VUML approach was modified to include the behavioral components of the technique.

4.1 Global Behavioral Analysis

Global behavioral analysis in VUML refers to the examination and understanding of the overall behavioral aspects of a system represented in the VUML modeling approach [17]. It involves analyzing the interactions, dependencies, and dynamic behaviors of various components within the system to gain a comprehensive view of how the system functions as a whole. This analysis helps in identifying potential issues, ensuring consistency across different viewpoints, and refining the system's behavior to meet the specified requirements. The global behavioral analysis is an integral part of the VUML approach to ensure a thorough understanding and effective representation of both structural and behavioral aspects in the modeling process.

4.2 Behavioral Analysis/Design by Point of View

Behavioral analysis/design by point of view in VUML (View-Based Unified Modeling Language) involves focusing on the specific behaviors and interactions of system elements from the perspective of individual actors or stakeholders. The VUML approach recognizes that different actors have distinct viewpoints, and each viewpoint is associated with a particular set of concerns and behaviors [18].

Here's an overview of Behavioral Analysis/Design by Point of View in VUML:

- Actor-Centric Approach: VUML emphasizes an actor-centric modeling approach where each actor has a unique point of view. This point of view defines how an actor perceives and interacts with the system.

- Actor-Centric Approach: VUML emphasizes an actor-centric modeling approach where each actor (human or logical entity interacting with the system) has a unique point of view. This point of view defines how an actor perceives and interacts with the system.
- Viewpoints: A viewpoint in VUML corresponds to the modeling of a specific aspect of the system from the perspective of an actor. It includes both structural and behavioral elements, providing a holistic representation of the system for a particular actor.
- Behavioral Analysis: The behavioral analysis involves capturing the dynamic aspects of the system, such as interactions, state transitions, and workflows, based on the concerns of a specific actor. This analysis ensures that the system behaves following the expectations and requirements of that actor.
- Designing for Specific Concerns: Behavioral design within a specific viewpoint considers the unique concerns and requirements of the associated actor. It includes the definition of states, transitions, events, and operations that are relevant to that actor's perspective.
- Consistency Across Viewpoints: While each actor's viewpoint is unique, VUML also emphasizes the importance of maintaining consistency across different viewpoints. This involves ensuring that the behaviors specified for each actor align cohesively to create a coherent and functional overall system.
- Iterative Process: Behavioral analysis and design in VUML are iterative processes. As the understanding of each actor's needs evolves, the corresponding viewpoint can be refined, and the overall system behavior can be adjusted to accommodate changes.
- By adopting a behavioral analysis and design approach by point of view, VUML enables a more focused and actor-specific representation of system behavior. This helps in creating models that accurately reflect the intended functionality and interactions from the perspective of different stakeholders in the system.

By adopting a behavioral analysis and design approach by point of view, VUML enables a more focused and actor-specific representation of system behavior [19]. This helps in creating models that accurately reflect the intended functionality and interactions from the perspective of different stakeholders in the system.

4.3 Behavioral Fusion/Synchronisation

The Composition of behaviors from multiple perspectives poses significant challenges in the modeling of complex systems [20]. Each view often represents a specific perspective, capturing aspects of the system, such as actors' interactions or internal processes. However, integrating these disparate behaviors into a coherent global behavior is far from trivial. Key challenges include managing conflicts between behaviors, synchronizing states between different views, and preserving overall system coherence. In addition, interview interactions can lead to unexpected emerging behaviors, making it difficult to predict and verify overall behavior. These challenges underline the importance of developing robust behavioral composition mechanisms that effectively coordinate the dynamics between different views while the integrity of the overall model [21].

Behavioral Fusion/Synchronization in VUML (View-Based Unified Modeling Language) refers to the process of combining and harmonizing the behavioral aspects of multiple viewpoints or actors within a system. In VUML, each actor has a unique viewpoint that captures its specific concerns and behaviors. Behavioral fusion/synchronization is crucial for ensuring that the overall system functions cohesively when considering the perspectives of different actors.

Here are key aspects of behavioral fusion/ synchronization in VUML:

- Integration of Viewpoints: VUML recognizes that a complete system involves the collaboration of multiple actors, each with its viewpoint. Behavioral fusion involves integrating the behavioral models associated with each actor's viewpoint into a unified representation of system behavior.
- Coherence and Consistency: The goal of behavioral fusion/synchronization is to achieve coherence and consistency across different viewpoints. It ensures that the behaviors specified for each actor complement each other and collectively contribute to the intended system functionality.
- State and Event Alignment: Fusion involves aligning the states and events defined in different viewpoints. This ensures that the system transitions seamlessly between states based on the interactions and events triggered by different actors.
- Workflow Alignment: Fusion addresses the alignment of workflows and sequences of activities across different actors. It ensures that the overall system workflow is coherent and aligns with the expectations of all stakeholders.
- Conflict Resolution: In cases where conflicts arise between behaviors specified in different viewpoints, synchronization aims to resolve these conflicts. This may involve negotiation, compromise, or the introduction of additional mechanisms to reconcile conflicting requirements.
- Iterative Process: Like other aspects of VUML, behavioral fusion/synchronization is an iterative process. As the understanding of system requirements and actor viewpoints evolves, the behavioral models may need to be adjusted and synchronized to reflect the most up-to-date specifications.
- Fusion at Different Levels: Behavioral fusion can occur at different levels, ranging from the synchronization of high-level system behaviors to the alignment of detailed state transitions and interactions between specific elements in the system.

By incorporating behavioral fusion/synchronization into the VUML approach, the modeling process aims to create a comprehensive and integrated representation of system behavior. This ensures that the system, as perceived from the

diverse viewpoints of different actors, functions harmoniously and meets the overall objectives and requirements of the stakeholders.

As mentioned before, the state machines that are suggested to illustrate how views behave comply with the UML standard [22, 23]. While UML provides diverse concepts for specifying behavior, their semantics often remain insufficient or vague. The drive to make UML a universal modeling standard capable of analyzing/designing any domain has resulted in ambiguous and nonspecific semantics. The Omega UML profile serves as the basis for the syntax we use to represent transitions in the created state machines. With a collection of techniques tailored to communication and execution characteristics, this profile offers a comparatively extensive semantics.

Utilizing the UML Omega profile and the IFx tool is fully justified in the context of the behavioral modeling of complex systems, particularly those requiring rigorous and formal verification. The UML Omega profile is an extension of UML designed to model real-time, distributed, or critical systems, providing mechanisms to specify complex behaviors with formal rigor. It stands out for its ability to capture temporal, concurrent, and synchronous aspects, making it particularly suitable for modeling systems where timing accuracy and concurrency management are essential. This not only allows for a better representation of the systems in question but also integrates formal verification techniques in the early design phases, thereby limiting costly errors during the development stage.

The IFx tool, for its part, is used to perform formal verification of models specified in UML Omega. It allows for the simulation, verification, and analysis of the properties of complex systems through rigorous formal semantics. By utilizing techniques such as model checking, IFx helps to detect potential errors, such as unwanted race conditions, violations of timing constraints, or deadlock situations, even before the deployment phase.

Thus, in this context, the UML Omega profile ensures accurate and formal modeling of complex behaviors, while IFx provides the necessary tools for validating these models, ensuring the robustness and reliability of the modeled systems.

5 RESULTS – CASE STUDY

In this section, we demonstrate the application of the afore aforementioned to our case study.

The results of applying the proposed approach to our case study reveal several significant advancements. The global structural analysis has made it possible to model the needs of the various stakeholders by organizing them into functional units represented by use cases. (e.g., car management). The global behavioral analysis identified reactive multi-view classes and detailed their lifecycle (e.g., the states of the Car class, from diagnosis to repair). Then, the behavioral design phase from the perspective of each actor (e.g., Client, Agency Manager) refined these states. Finally, the behavioral composition has allowed for the integration of these multi-view behaviors while ensuring coherence and a smooth transition between states. Although this approach has improved clarity and precision in modeling complex behaviors, it has introduced increased complexity in coordinating signals between the states of different machines.

5.1 Global Structural Analysis

Global analysis is the first step in the process, which acts as a central location for requirements modeling. Finding the demands of different actors and grouping them into functional units—represented as use cases—is the aim. An overview of the system's functions is shown in Fig. 4, which includes (1) agency monitoring, (2) people management, (3) material management, (4) financial management, and (5) automobile management. Every feature is broken down into smaller, more manageable functional parts that show the roles and responsibilities of each actor. The client, agency manager, workshop manager, and maintainer are the particular actors that are the subject of our investigation.

To provide clarification, we will use the "cars management" feature as an example (Fig. 4). The administration of the agency's vehicles includes duties including registration, supervision of each vehicle's expertise and repairs, and final verification testing. In the "save" use case, the customer gives details regarding their vehicle, works with the agency manager to carry out expert and repair contracts, and verifies the repair by giving their automobile one more functioning test. While the workshop manager is responsible in supervising the automobiles' maintenance chores, maintainers are involved in technical aspects such as expertise, repair, and testing.

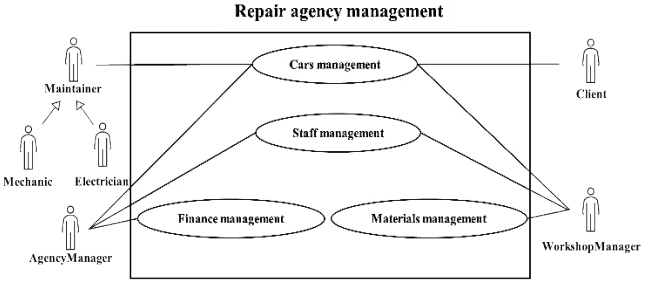


Figure 4 Repair Agency Management" use case

5.2 Global Behavioral Analysis

In this section, we outline the steps involved in the global behavioral analysis phase.

- In this phase, we **identify reactive multi-view classes** based on the comprehensive structural analysis of the case study. The detailed examination of use cases and user requirements, through tools like sequence and activity diagrams, aids in pinpointing classes with potential multi-view and reactive behavior. The following classes were found to be multi-view in our case study: Car, Expertise, Breakdown, Repair, and Contract. For simplicity, we attribute reactive behavior specifically to the Car class, while categorizing others as data classes (static). It is crucial to remember that throughout the full analysis by point of view in the next

phase of the method, this list of multi-view classes and their classification as static or reactive may change.

- In this step, we **identify potential states for each reactive multi-view class** by considering their entire life cycle. Taking the Car class as an example, the life cycle involves several stages: When the vehicle breaks down, the owner contracts with the agency and gives them the details they need to register the vehicle. The first step in the agency's repair procedure is to bring the vehicle to the garage, where mechanical and electrical specialists will inspect it for problems. A repair contract is created based on the findings of the experts, and the identified flaws are fixed by the maintainers. Testing is the last phase to make sure the vehicle operates as intendedThe several probable states of an automobile in the garage are included in this notional life cycle.

The initial state, OutOfOrders, signifies the car's breakdown condition, marking the beginning of the maintenance journey. This state is crucial as it sets the context for subsequent processes. Following this, the InGarageRouting state illustrates the logistics involved in transporting the vehicle from its location of failure to the garage, emphasizing the importance of efficient routing in minimizing downtime.

Once the car is in the garage, it enters the InExpertiseProcedure phase, where expert assessments are conducted to detect faults. This stage is critical, as it lays the groundwork for informed decision-making regarding the necessary repairs. The subsequent InRepairProcedure state is multifaceted, comprising two essential components: the negotiation of repair contracts based on expert evaluations and the technical repair itself. This duality highlights the complex interplay between administrative and technical aspects of automotive repair.

Finally, the process culminates in the InTest state, where the vehicle undergoes rigorous testing to confirm its operational readiness before exiting the garage. The Exit state signifies the conclusion of the car's lifecycle within the garage, encapsulating the entire process from breakdown to resolution.

The base-state machine of the Car class is shown in Fig. 5. This device provides a logical temporal sequence of an object's states, encapsulating its life cycle inside the system. In the second step of the process, the designers then integrate this machine into the lexicon for shared and reusable reasons.

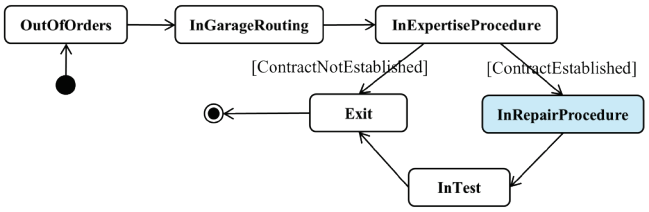


Figure 5 "machine-base" of Car Class

5.3 Analysis/Design of Behavior by Point of View

The behavioral specification phase is started based on a particular viewpoint when the structural model for that viewpoint has been built (see section 4.2). The "machine-base" states found in the first phase are analyzed from a single point of view. A sub-machine may be produced from each basic state, which can then be refined and specialized to meet the needs of the particular point of view. We choose to use our example to restrict the illustration:

- One part of the "machine-base." We'll focus on the InRepairProcedure stage, when the vehicle goes through a sequence of steps to return to its typical operational condition.
- The client and the agency manager are the two actors to whom this is given.

Client Perspective: We extract the part of the SM_Client_Car state machine associated with the InRepairProcedure state (Fig. 6) from the class diagram created for the Client Perspective (see section 4.2), which focuses on the repair and contract negotiation process.

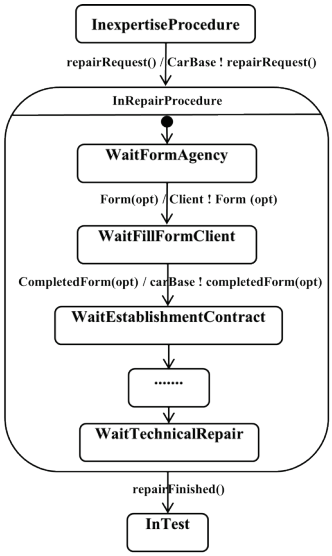


Figure 6 Enhancement of the InRepairProcedure State for the Client Viewpoint

The Repair() signal, which the client actor requests, initiates the change in the SM_Client_Car view from the ExpertiseProcedure state to the RepairProcedure state. However, the analysis/design by point of view is carried out in a decentralized fashion throughout this process stage. Therefore, without worrying about which objects will get the signal, the SM_Client_Car state machine transfers this signal to the base machine. When one actor waits for a signal from another entity, the same logic is applicable. From this perspective, the developer in these situations presumes that the signal comes from the "machine-base." The form(opt) signal, which starts the changeover from the WaitFormAgency state to the WaitFillFormClient state in Fig. 7, serves as an example of this.

As seen in Fig. 7, the Agency Manager's viewpoint entails improving the InRepairProcedure state. The same idea of centralizing communications inside the machine base is used in this improvement. The signalRepair() is received by the SM_AgencyManager_Car from the "machine-base," and it is then sent on to the agency manager actor. The SM_AgencyManager_Car then creates a form with a list of

the failures that need to be repaired along with the choices for fixing each one. After preparing the form, the agency manager sends it to the SM_AgencyManager_Car, which, upon receiving the form(opt) signal, initiates a transition to the WaitReturnFormCompleted state.

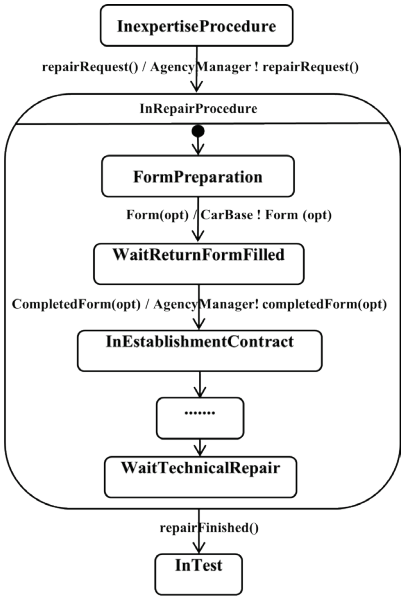


Figure 7 Enhancement of the InRepairProcedure state for the AgencyManager Viewpoint

This could be relevant in situations where a complex system needs to be understood, developed, or maintained by different teams or individuals with specialized concerns. The design might involve composing behaviors from the viewpoint of each stakeholder or participant in the development process.

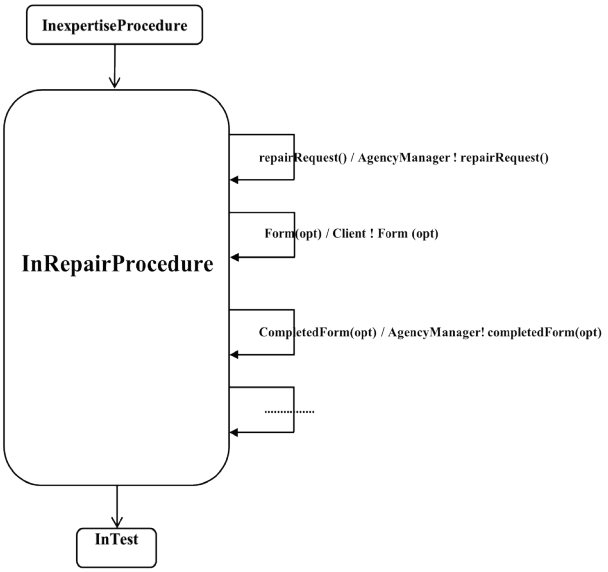


Figure 8 Including Synchronization Messages for the InRepairProcedure State (Extract) in the Machine Base.

5.4 Behavioral Composition

The term "behavioral composition model by point of view" isn't a standard phrase, and its meaning might depend on the context in which it's used. However, I can offer an interpretation based on common concepts in software engineering.

In the context of software design or system architecture, "behavioral composition by point of view" could refer to a method where the composition of behaviors is guided or structured based on different perspectives or viewpoints. Here's a breakdown:

Behavioral Composition: This generally involves combining the behaviors of individual components, objects, or modules to create a larger system behavior. It's about how different pieces of functionality come together to form a coherent and functional whole.

Point of View: In the context of software engineering, a "point of view" often refers to a particular perspective or set of concerns related to a system. Different stakeholders or participants in the development process may have distinct viewpoints, such as end-users, developers, system administrators, etc.

Bringing these concepts together, "behavioral composition model by point of view" could imply a design or modeling approach where the composition of behaviors is structured or organized according to different perspectives or viewpoints. Each viewpoint might focus on specific aspects of the system's behavior, and the behavioral composition is done in a way that aligns with these different perspectives.

Compared to existing VUML approaches, our method has made it possible to model more accurately the complex behaviors of multi-view objects in the way state transitions were managed between different views.

The proposed approach has shown significant advantages in terms of clarity and consistency in multi-view behavior modeling. However, it also has limitations, including increased complexity in the coordination of signals between the different states of the machine.

Compared to existing VUML approaches, the proposed method shows significant improvements in terms of accurately modeling the complex behaviors of multi-view objects. While traditional VUML approaches primarily focused on the separate structural and behavioral modeling of viewpoints, our method allows for a more nuanced management of state transitions between different views, thereby providing better clarity and coherence in the composition of behaviors. However, this increased precision also introduces an additional complexity, particularly in coordinating signals between the different states of the machines, which can make the process more challenging to manage than in traditional VUML approaches.

Tab. 2 compares the proposed VUML method to other behavioral modeling frameworks, such as standard UML and SysML. It highlights the improvements brought by VUML in terms of viewpoint management, behavior fusion, and formal verification.

This work has important implications for software engineering and model-driven development (MDD). By integrating a multi-view approach for modeling complex behaviors, it promotes better management of complexity in software systems. This method allows for a more coherent

and detailed representation of the interactions between the different actors, which enhances the modularity and reusability of the models. Furthermore, by facilitating the synchronization of states and behaviors in multi-view systems, this approach contributes to the creation of more robust and adaptable software, a crucial asset for large-scale projects where multiple perspectives need to be harmonized.

Table 2 VUML in comparison to other frameworks

Criteria	VUML (Proposed)	Standard UML	Other (e.g., SysML, BPMN)
Points of View	Explicit integration, distinct actors	Generic approach	Limited to structural aspects
Merging Behaviors	Synchronization, conflict resolution	Weak support	Underdeveloped fusion mechanisms
Formal Verification	Support with IFx	Not supported	Limited to external tools
Complex Systems	Optimized for critical systems	Less suited to complex needs	Limited support for rich interactions
Clarity and Precision	Reduction of ambiguities	Possible ambiguities	Variable depending on the context

6 CONCLUSION

This article makes a notable contribution to the VUML approach by proposing a systematic method for the specification and composition of multi-view object behaviors. This advance improves complex systems modeling by effectively integrating behavioral aspects while preserving the separation of concerns. The implications of this method for VUML and behavioral modeling open new perspectives in the field for the development of complex software systems [24]. For future research, it will be crucial to address the challenges of scalability concerning large-scale system management and to explore solutions to further automate behavioral composition.

This method allows a clear separation of behavior development by point of view while providing effective mechanisms for their composition into coherent behaviors of multi-view objects.

This research expands the capabilities of VUML by introducing rigorous mechanisms for behavioral modeling, offering a new path for managing complex systems with multiple perspectives.

As previously said, the original method only uses UML principles. Finding two different kinds of specialized machinery linked to either a "base" class or a "view" class for a given multi-view class is the main result of this method. According to the UML2.0 specification, these state machines are used to represent the dynamic behavior displayed by instances of the multi-view class under consideration.

A "machine-base" is intended to create coherence and coordination among the "machines-view" and express the collective behavior of the actors involved, whereas a "machine-view" depicts the life cycle of an object-view. A multi-view machine is a combination of a "machine-base" and its dependent "machines-view".

By supporting autonomous development in the second phase, this strategy tackles the problem. The "machine-base" idea was introduced in order to ease this, as the second phase required to be directed by a pre-established model. During the global modeling phase, the "machine-base" is created to express the aggregate behavior of the actors. When creating behaviors from other angles, this behavior is thought of as a model to follow.

During the decentralized modeling stage, the machines-view and object-view are defined independently, according on the "machine-base" state. The integration of incomplete behaviors from various perspectives is then achieved by adding signal exchanges in the fusion phase. The coordination of machines with the object-view and object-based states is made possible by these interactions.

However, as it grows, this UML-based method has drawbacks, especially in two crucial areas:

- Behavior Specification Challenges: Interconnected Views: Difficulties arise in ensuring the independence of view development due to intricate connections between them. This interdependence poses a risk during scaling, making it challenging to implement the approach without modifying the development of other views to gather necessary information.
  Early Identification: It becomes challenging when multi-view classes and their "machine-bases" must be identified early in the global analysis phase. Developing a state machine that encompasses an object's life cycle necessitates a detailed analysis of multi-view object use cases, which is problematic for big systems.
- Behavior Composition Challenges: Complexity of Integration: Many adjustments and changes may be required at the base machine and vision machine levels when integrating independently created vision machines. Maintaining the coherence of the entire system throughout the composing process when scaling necessitates considerable work.

This framework could transform software engineering by improving the modeling of complex behaviors, particularly in critical and collaborative systems. For example, case studies in the fields of industrial automation or intelligent transportation systems could validate the effectiveness of the method for synchronizing multi-view behaviors and resolving conflicts. These validations would demonstrate how this approach ensures consistency, reduces errors, and optimizes design in environments with high interconnectivity and complexity.

In the future, work should focus on developing tools to automate certain parts of the composition process, as well as exploring techniques to better manage complexity in large-scale multi-view behavioral models. Furthermore, it would be relevant to study the integration of this approach with other model-driven development methodologies in order to enhance its effectiveness and applicability and could include applications in demanding fields where the management of multi-view behaviors is crucial. Furthermore, additional research on the optimization of synchronization mechanisms and adaptation to large-scale systems would provide increased robustness to the method.

7 REFERENCES

[1] El Hamlaoui, M., Ebersold, S., Bennani, S., Anwar, A., Dkaki, T., Nassar, M. & Coulette, B. (2021). A Model-Driven

Approach to align Heterogeneous Models of a Complex System. *The Journal of Object Technology, 20*(2), 1-24. https://doi.org/10.5381/jot.2021.20.2.a2

[2] Kriouile, A. (1995). VBOOM, une méthode orientée objet d'analyse et de conception par points de vue. *Unpublished doctoral dissertation*, University Mohammed V, Rabat, Morocco. (in French)

[3] Anwar, A. (2009). Formalisation par une approche IDM de la composition de modèles dans le profil VUML (Doctoral dissertation, *Thèse de doctorat*, Université de Toulouse). (in French)

[4] Ober, I., Graf, S. & Ober, I. (2006). Validating timed UML models by simulation and verification. *International Journal on Software Tools for Technology Transfer, 8*(2), 128-145. https://doi.org/10.1007/s10009-005-0205-x

[5] Ouali-Alami, C., El Bdouri, A., Elmarzouki, N., & Lakhrissi, Y. (2022). View-based Modelling: Behaviour Specification based on UML Concept. https://doi.org/10.5220/0010730100003101

[6] Horman, Y. & Kaminka, G. A. (2004). Improving Sequence Learning for Modeling Other Agents. *Proceedings of the AAMAS 2004 Workshop on Learning and Evolution in Agent-Based Systems*.

[7] Nassar, M. (2005). Analyse/conception par points de vue: le profil VUML (*Doctoral dissertation*). (in French)

[8] Lakhrissi, Y., Coulette, B., Ober, I., Nassar, M. & Kriouile, A. (2008). Démarche VUML Statique et Dynamique. (in French)

[9] El Marzouki, N., Lakhrissi, Y., Nikiforova, O., El Mohajir, M. & Gusarovs, K. (2017). Behavioral and Structural Model Composition Techniques: State of Art and Research Directions. *Transactions on Computers, WSEAS*, 16, 39-50.

[10] Dávid, I. (2016). A multi-paradigm modeling foundation for collaborative multi-view model/system development. In *Proceedings of the ACM Student Research Competition at MODELS 2016*, October 3-4, 2016, St. Malo, France/Gray, Jeff [edit.]; et al.

[11] Kühne, T. (2022). Multi-dimensional multi-level modeling. *Software and Systems Modeling, 21*(2), 543-559. https://doi.org/10.1007/s10270-021-00951-5

[12] Hannousse, A. (2019). Dealing with crosscutting and dynamic features in component software using aspect-orientation: requirements and experiences. *IET Software, 13*(5), 434-446. https://doi.org/10.1049/iet-sen.2018.5324

[13] Essebaa, I. & Chantit, S. (2017). QVT transformation rules to get PIM model from CIM model. In *Europe and MENA Cooperation Advances in Information and Communication Technologies*, 195-207. Springer, Cham. https://doi.org/10.1007/978-3-319-46568-5_20

[14] Kim, W. Y., Son, H. S., Park, Y. B., Park, B. H., Carlson, C. R. & Kim, R. Y. C. (2008). *Automatic MDA (model driven architecture) transformations for heterogeneous embedded systems*.

[15] Anwar, A., Dkaki, T., Ebersold, S., Coulette, B. & Nassar, M. (2011, April). A formal approach to model composition applied to VUML. In *the 16th IEEE International Conference on Engineering of Complex Computer Systems*, 188-197. https://doi.org/10.1109/ICECCS.2011.26

[16] El Marzouki, N., Lakhrissi, Y., Nikiforova, O. & El Mohajir, M. (2017, April). The application of an automatic model composition prototype on the-Two hemisphere model driven approach. In *IEEE International Conference on Wireless Technologies, Embedded and Intelligent Systems (WITS2017)*, 1-6. https://doi.org/10.1109/WITS.2017.7934673

[17] Wong, P. Y., Bubel, R., de Boer, F. S., Gómez-Zamalloa, M., De Gouw, S., Hähnle, R., ... & Sindhu, M. A. (2015). Testing abstract behavioral specifications. *International Journal on Software Tools for Technology Transfer, 17*(1), 107-119. https://doi.org/10.1007/s10009-014-0301-x

[18] Asteasuain, F. & Braberman, V. (2017). Declaratively building behavior by means of scenario clauses. *Requirements Engineering, 22*(2), 239-274. https://doi.org/10.1007/s00766-015-0242-2

[19] El Marzouki, N., Ņikiforova, O., Lakhrissi, Y. & El Mohajir, M. (2016). Enhancing conflict resolution mechanism for automatic model composition. *Applied Computer Systems, 19*(1), 44-52. https://doi.org/10.1515/acss-2016-0006

[20] Nikiforova, O., El Marzouki, N., Gusarovs, K., Vangheluwe, H., Bures, T., Al-Ali, R., ... & Leon, F. (2017). The two-hemisphere modelling approach to the composition of cyber-physical systems. In *Proceedings of the 12th International Conference on Software Technologies*, 286-293. https://doi.org/10.5220/0006424902860293

[21] El Marzouki, N., Nikiforova, O., Lakhrissi, Y. & El Mohajir, M. (2017). Toward a generic metamodel for model composition using transformation. *Procedia Computer Science, 104*, 564-571. https://doi.org/10.1016/j.procs.2017.01.173

[22] Lakhrissi, Y., Ober, I., Coulette, B., Nassar, M. & Kriouile, A. (2007, July). Vers la notion de machine à états multivue dans le profil VUML. In *Workshop WOTIC, 5*(07), 2007-06. (in French)

[23] Lakhrissi, Y., Anwar, A., Nassar, M. & Kriouile, A. (2008). Composition des machines à états par point de vue dans VUML. In *Workshop JIMD, 3*(07), 2008-05. (in French)

[24] Ouali-Alami, C., El Bdouri, A., Elmarzouki, N., Korchi, A. & Lakhrissi, Y. (2022). Approaches to Design Complex Software Systems. https://doi.org/10.5220/0010740200003101

Authors' contacts:

Chaimae Ouali-Alami, PhD student
(Corresponding author)
SIGER Laboratory,
Sidi Mohamed Ben Abdellah University,
36000 Fez, Morocco
E-mail: chaimae.oualialami@usmba.ac.ma

Abdelali El Bdouri, PhD student
SIGER Laboratory,
Sidi Mohamed Ben Abdellah University,
36000 Fez, Morocco
E-mail: abdelali.elbdouri@usmba.ac.ma

Younes Lakhrissi, Professor (Full)
SIGER Laboratory,
Sidi Mohamed Ben Abdellah University,
36000 Fez, Morocco
E-mail: younes.lakhrissi@usmba.ac.ma

ISSN 1846-6168 (Print), ISSN 1848-5588 (Online)
https://doi.org/10.31803/tg-20241213154237

Original scientific paper
Received: 2024-12-13, Accepted: 2024-12-30

Understanding Vietnamese Tourists' Re-Browse Intentions for Smart Tourism Apps: An Extended TAM Approach

Thai-Ngoc Pham

Abstract: This study examines Vietnamese young tourists' perceptions of smart tourism applications, focusing on their influence on flow experiences and re-browse intentions. Using an extended Technology Acceptance Model (TAM), the research surveyed 455 respondents, analyzing how app dimensions—user interface design, informativeness, interactivity, personalization, and accessibility—affect perceived ease of use, usefulness, enjoyment, and security. The findings reveal that perceived enjoyment and security significantly enhance e-flow experiences, which in turn foster re-browse intentions. Practical implications include strategies for app developers to optimize design features and personalization to improve user engagement and loyalty.

Keywords: e-flow experience; re-browse intention; smart app dimensions; Vietnamese young tourists

1 INTRODUCTION

Thanks to the breakthrough development of mobile devices and information technologies, smart tourism is now in its era [1]. While various technologies enhance smart tourism, information and communication technologies (ICTs) are essential [2], and smart tourism technologies (SSTs) are indeed transforming tourist behaviors and providing exceptional tools for tourism providers to gain competitive advantages [3], [4]. Enhancing the tourist experience is challenging, but the achievement is worth it because a better tourist experience will result in favorable responses [3]. Regardless of previous results on SSTs, scholars believe that the findings could not be generalized because of differences among research contexts [5]. Therefore, Vietnamese tourists will be targeted in this study with the expectation that better understandings on how SSTs drive on tourists' e-experiences and their intentions will provide strategic design guidelines for practitioners operating in this market [4].

Vietnam's online travel ranks the top five in the Asia-Pacific area behind Indonesia, Thailand, and Singapore. VITA's 2023 figures indicated that the volume of individuals seeking tourist information online surged by nearly 32-fold compared to 2022. Statista (2023) [6] reports that 54 percent of Vietnamese respondents indicated they have utilized an online travel agency, and among online booking users in Vietnam, 40 percent reported traveling at least three times year. Furthermore, youthful travelers constitute the majority. This figure indicates both the segment's familiarity with e-commerce and its appeal to the hotel and tourist sectors [6]. Nguyen and colleagues (2021) [7] emphasize that younger demographics exhibit enthusiasm, interests, and a strong readiness to participate in travel activities. Vietnamese travel patterns have seen substantial changes over the past decade due to the digital revolution of the tourist sector. Nonetheless, research on smart tourism in Vietnam is inadequate, particularly at the individual level [8]. Nevertheless, research on smart tourism in Vietnam remains limited, particularly at the individual level, which constrains efforts to enhance tourists' e-experiences and foster the development of behavioral intention [8].

Azis and colleagues (2020) [9] assert that current research on self-service technologies (SSTs) emphasizes on the reasons for tourists' use of SSTs and methods to promote this adoption, rather than elucidating the impact of SSTs on electronic experiences and fostering re-browsing intentions. For example, the role of various dimensions of SSTs, such as personalization and interactivity, in shaping young tourists' e-experiences and influencing their intentions to re-browse remains underexplored [10]. Morover, digital specialists have noted that a substantial segment of the market has progressed beyond the introduction phase, and scholars advocate for research that considers post-behaviors in pursuit of more sustainable growth [10]. Additionally, e-experience is a novel concept that warrants thorough examination, as the dynamics of the virtual world differ significantly from those of the traditional world [11]. What constitutes an experience in real life is distinctly different from what defines an e-experience [11]. Consumer researchers believe that post-behaviors must be examined within a comprehensive model to emphasize the impact of STTs and e-experiences on specific behaviors [12]. This study aims to investigate the subsequent research questions:

RQ1: In what manner will the dimensions of STTs affect e-flow experiences?

RQ2: In what manner will the dimensions of STT affect re-browsing intention?

To address the aforementioned questions, this study will utilize a quantitative research design and implement a survey targeting Vietnamese e-tourists. A research model will be developed to hypothesize and assess the impact of features of smart tourism applications on users' e-flow experiences and their intention to re-browse. The anticipated findings aim to provide valuable insights for both scholars and practitioners within the context of Vietnam's tourism sector, offering guidance on how the functionalities of smart tourism apps can be optimized to enhance user experiences and facilitate re-browsing intention.

2 LITERATURE REVIEW

2.1 Theoretical Definitions

Smart tourism has emerged as a prominent concept in academic literature in the late 2020s, rapidly being adopted in practice due to its potential to enhance travel experiences and optimize operational and resource management [13]. Mehraliyev and colleagues (2020) [1] define smart tourism as a transformation in tourism management that involves the participation of all stakeholders within a cohesive ecosystem. While information and communication technology (ICT) is recognized as a pivotal element in the advancement of smart tourism, it is not the only factor, as technological innovations continue to evolve [14]. Cutting-edge technologies such as the Internet of Things (IoT), cloud computing, artificial intelligence (AI), augmented reality (AR), and virtual reality (VR), along with advanced devices and robust networks, are enhancing the attributes of tourism and driving it towards greater intelligence [4]. Furthermore, Gretzel and colleagues (2015) [15] categorize smart tourism into three components: smart destinations, smart business ecosystems, and smart experiences, all of which rely on data collection, analysis, and sharing.

2.1.1 Smart Tourism Technologies

Smart tourism technologies (STTs) mention all forms of smart tourism applications and tourism-related information sources such as online travel agents (OTAs), websites, personal blogs, social networking sites, and other smart tourism apps that continuously incorporate new functions as decision support systems, history records, VR/AR, facial recognition, and integrated payment methods, continuously adding and co-creating values for tourists, destination marketing organizations (DMOs), and tourism operators [3]. STTs in user-based platforms manifest six attributes, including: user interface design, accessibility, informativeness, interactivity, and personalization [13]. In this study, mobile apps will be examined as target forms of STTs.

User interface design refers to the finest visualization of a display that users can see [16]. The attribute manifests itself in an online platform as the general graphical presentation presented to the users, including design, structures, animation, layout, etc. [17]. In a basic level, interface design requires to be apparent, easy to look at, read, and find, while in a more advanced level, interface design requires to be professional, fashionable, artist, and eye-catching [18]. Users interact with smart tourism apps through front-ends, and their designs will probably more or less influence users's performing actions [19]. Based on empirical results, several design techniques have been applied to enhance the effectiveness of interface design, such as presentation optimization, semantic conversion, zooming functions, and focus and context tactics [20].

Informativeness refers to the extent of quality, credibility, and accuracy provided by an information source [21]. The attribute applied in a tourism-related platform demonstrates the volume of sufficient information satisfies users' needs about tourism [4, 22]. To save time and effort, users prefer using general apps to private ones because they often contain more information, such as accommodations, tourist sites, foods, and restaurants, rather than about only one sector [5]. Keeping up with the development of mobile technology, smart tourism apps are sources of updated, frequent, and authentic information generated and shared by users as truly experienced tourists [13]. Information plays a crucial role for tourists in evaluating and making decisions, and thus it contributes to the quality of a smart tourism app [2].

Interactivity refers to the capability that allows multiple relationships among related stakeholders [21]. The attribute in a tourism platform will enhance the two-way communication of all involved parties [22]. For example, users can easily connect with other users sharing about tourism experiences, or users can connect with app operators in assisting with technical issues, or even users can work with hotel owners in booking their accommodations. Smart tourism apps are operating as social networks where all users can connect, share, and perform supported functions through their account registrations [4]. Interativity encourages e-tourists to be actively involved and engaged in tourism behaviors [13].

Personalization refers to the degree to which individual attention is employed to a specific product or service [22]. The attribute expresses itself in a tourism platform as the ability to collect personal data and is able to offer suitable and appropriate recommendations [4]. For instance, based on provided information and behavioral preferences, the app may suggest fitting options, which in turn helps users save their time and effort in searching. Personalization is an enhanced function that is embedded in user platforms to increase flow experience and other users' action [9]. Personalization facilitates e-tourists to finalize their selections since they offer relevant information and offer 'best' alternatives' as users' demands [13].

Accessibility refers to the capability of having access to an information source [21]. The attribute applied in a tourism-related platform exhibits the ease that users can connect to [4, 22]. Often, many technologies will be implemented in an app since they will help enhance users' experience [1], and accessibility will also account for the quality of reachability of all STTs in that app. Users can get more information in the situation of high accessibility and will make decisions faster [5]. Accessibility plays as a facilitating condition for a smart app to be feasible to the e-tourists [13]. Accessibility is considered a prerequisite for other attributes to exert their influences on the user's intention [2].

2.1.2 Customer Perception

Perceived ease of use: Davis (1989) [23] initially characterized perceived ease of use as the extent to which an individual believes that utilizing a certain system would be effortless. In other words, perceived ease of use refers to the customer's perception about how difficult it is to use a system [24, 25]. Perceived ease of use is widely empirically supported as an antecedent explaining the adoption of new technology in customers [26-28]. In the context of smart tourism apps, perceived ease of use is the user's perception of the ease of performing functions on smart apps [24].

Perceived usefulness: Davis (1989) [23] defined perceived usefulness as the extent to which a person thinks that employing a certain system would efficiently enhance his/her job performance. In other words, perceived usefulness refers to the customer's perception about how much usefulness they gain when using a system [24, 29]. Perceived usefulness is empirically supported as a key factor in understanding why a person prefers to accept a new technology [26-28]. In the context of smart tourism apps, perceived usefulness is the user's perception of the utility of employing smart apps in their consumption behaviors related to tourism [24].

Perceived enjoyment: Davis (1992) [23] defined perceived enjoyment as the extent to which a person thinks that employing a certain system would be fun and enjoyable beyond its main functions. In other words, perceived enjoyment refers to the customer's perception about the pleasure they feel when using a system [30, 31]. Perceived enjoyment is empirically found as a factor in keeping a user accepting a new technology [28]. In the context of smart tourism apps, perceived enjoyment is the user's perception of the hedonism when employing smart apps in their search for tourism information [31].

Perceived security: No and Kim (2015) [22] defined perceived security as the extent of the overall safety of a specific system for all involved users. In other words, perceived security exhibits the degree to protect personal information and data circulating in the system [4, 32]. Security is the latest attribute required in developing and maintaining apps due to the cyber risks that appeared recently, and in turn, perceived security is increasingly required by online users toward current systems [33]. Security must be guaranteed in the context of smart platforms because they often involve multiple parties, and when risks occur, they are often spread with more intense influence [4]. In the context of smart tourism apps, perceived security is the user's perception of the safety when using the apps for their needs [13].

2.1.3 E-Flow Experience

E-experience is a key concept in understanding a user's behaviors in an online environment [34]. Flow experience is defined as a total sensation that one feels when they focus on a specific object with all involvement [35]. Flow experience is described as the merging of states that a person feels, such as awareness, attention, self-consciousness, and control [36]. Flow experience often occurs during network navigation searching, and that is when the concept of e-experience has emerged. When flow is experienced, nothing else seems to matter, and that keeps users in that environment [35, 37]. E-flow experience is a critical construct in the online environment and is considered an evaluative criterion for the success of an e-program. When users experience e-flow, they tend to spend more time and engage more deeply with the program [35]. Research has shown that e-flow experience is closely linked to expect virtual behaviors, such as e-satisfaction and e-loyalty, across various online platforms [34]. Liu and colleagues (2016) argue that e-flow experience should be examined within the context of e-commerce, as it is a key determinant of user engagement [37].

2.1.4 Re-Browsing Intention

Re-browsing intention is a behavioral concept in which it refers to the willingness to browse the searching from the same source [38]. Re-browsing intention is a reliable metric used to determine users' e-satisfaction because it shows they are satisfied with the browsing and thus likely to continue to browse for the next time [39]. Re-browsing intention is also considered a requirement in nurturing user loyalty because the repeated browsing is favorably associated with intense satisfaction. In the context of smart apps, re-browsing is highly valued as it contributes to both increased app traffic and improved business performance [40]. Re-browsing intention has garnered significant attention from app operators, as understanding the factors that drive this behavior can help identify strategic determinants to encourage returned users [41]. The stronger the intention to re-browse a smart app, the higher the probability that current customers will engage in desired behaviors such as making bookings or purchasing items [39].

2.2 Underlying Theory

The objective of this study is to navigate the re-browse intentions of users on smart tourism apps through app dimensions and users perceptions. Therefore, the TAM model has been adapted to establish the framework among factors [42]. The TAM is widely employed and validated to understand technological acceptance across information systems [43]. This study has employed the extended TAM to determine factors influencing the repurchase intention. Because the smart apps have already been adopted by young travelers and the information systems have been continuously upgraded, it is logical to utilize the extented TAM to propose hypotheses among external factors and intention behaviors [42]. Thus, smart app features will be hypothesized to influence users' perceptions, which will facilitate behavioral intentions

2.3 Research Hypotheses

2.3.1 The Relationship between Smart App Dimensions and Perceived Ease of Use

It is believed that app dimensions will significantly influence perceived ease of use toward that app, and this conclusion has also been empirically supported in various studies [5]. First, Azis and colleagues (2020) [9] highlight that the design of an app's dimensions will determine the user's satisfaction through the perception of ease in using. The result is explained by the reason that the app should be designed in a way that the target user can use it easily and find effortless in performing its functions [5]. Verkijika and De Wet (2019) [44] point out that users get frustrated with a system that they find difficult to use, and that will drive user dissatisfaction. Second, informativeness is believed to have a positive influence on perceived ease of use, as the sufficiency and quality of relevant information enhance the ease of using the smart app [47]. Third, interactivity is expected to increase perceived ease of use, as users can engage with app supporters and other users, facilitating their interaction with the app [49]. Fourth, personalization is proposed to improve

perceived ease of use, as greater customization for target users tends to make the app feel easier to use [50]. Finally, accessibility is thought to contribute to the perception of ease of use, as users who can easily access a smart app are more likely to find it effortless to perform the app's functions [51]. While few studies have examined the impact of smart app dimensions on perceived ease of use, previous empirical research has established a positive relationship between app attributes—such as user interface design [45, 46], informativeness [47, 48], interactivity [49], personalization [50], and accessibility [51]—and perceived ease of use. Thus, the study proposes:

- H1-a: User interface design positively influence on perceived ease of use.
- H1-b: Informativeness positively influence on perceived ease of use.
- H1-c: Interactivity positively influence on perceived ease of use.
- H1-d: Personalization positively influence on perceived ease of use.
- H1-e: Accessibility positively influence on perceived ease of use.

2.3.2 The Relationship between Smart App Dimensions and Perceived Usefulness

Findings have reported that app dimensions will prominently influence perceived usefulness toward that app, and they are explained by the fact that app dimensions are designed in a way that helps users perform app functions [5]. To achieve the most optimal perceived usefulness, app operators should consider the familiarity and skillfulness of the target user to design the app [50]. Baker-Eveleth and Stone (2020) [52] have argued that any failure in app design would more or less drive down the perceived usefulness in target users because the users perform app functions through its dimensions. First, interface design enables users to navigate app functions quickly and efficiently, thereby enhancing their perception of the app's usefulness [46]. Second, informativeness provides users with the necessary information while using smart apps, which in turn increases their assessment of the app's utility [53]. Third, interactivity facilitates communication among users, thereby supporting their use of the app [54]. Fourth, a higher level of personalization ensures that the app's functionalities align with users' specific needs, thereby increasing perceived usefulness. Finally, accessibility is believed to strengthen perceived usefulness, as easy access to the app allows users to fully utilize its features, leading them to perceive it as more useful [56]. Previous empirical studies have found the positive relationship between app dimensions and perceived usefulness across research contexts: user interface design [46], informativeness [53], interactivity [54, 55], personalization [50], and accessibility [56]. Thus, the study proposes that:

- H2-a: User interface design positively influence on perceived usefulness.
- H2-b: Informativeness positively influence on perceived usefulness.
- H2-c: Interactivity positively influence on perceived usefulness.
- H2-d: Personalization positively influence on perceived usefulness.
- H2-e: Accessibility positively influence on perceived usefulness.

2.3.3 The Relationship between Smart App Dimensions and Perceived Enjoyment

Findings have pointed out that the dimensions of smart app could influence the users' perceived enjoyment, and these relationships have been extensively enhanced in the context of a virtual environment to increase the adoption level [57]. For interface design, both scholars and practitioners agree that the design of a smart app can significantly influence the perceived fun and happiness users experience while interacting with it [58]. For instance, app designers suggest that a carefully chosen color scheme can positively impact users' moods [58]. Regarding informativeness, the availability of relevant information can enhance users' perceived enjoyment, particularly when it aligns with their hedonic values. For example, the inclusion of entertaining content might increase users' enjoyment [58]. In terms of interactivity, it is believed that users experience greater happiness when they can engage with others. For example, app operators suggest that features such as chat functions can boost users' happiness by allowing them to share information and experiences [59]. Concerning personalization, customization ensures that the app aligns with users' preferences, thereby enhancing their enjoyment [60]. For instance, offering users access to their favorite topics can make them feel more cheerful. Finally, with regard to accessibility, users are more likely to experience delight when the app is easily accessible, as it allows them to quickly obtain what they need without delays [61]. Previous studies have found results on the positive relationship among each dimension of smart app dimensions on perceived enjoyment across virtual applications, such as user interface design, informativeness [58], interactivity [59], personalization [60], and accessibility [61]. Thus, the study proposes that:

- H3-a: User interface design positively influence on perceived enjoyment.
- H3-b: Informativeness positively influence on perceived enjoyment.
- H3-c: Interactivity positively influence on perceived enjoyment.
- H3-d: Personalization positively influence on perceived enjoyment.
- H3-e: Accessibility positively influence on perceived enjoyment.

2.3.4 The Relationship between Smart App Dimensions and Perceived Security

Scholars highlight that smart app dimensions will also positively contribute to the perceived security toward that app [62]. Perceived security is getting the users' attention and becoming a requirement of users employing any online system [63]. Turner-McGrievy and colleagues (2017) [64] demonstrate that the design of an app will help users trust its safety. Since perceived security is considered an evaluation of risk-free, an app should be operated in a way that users are

not concerned about employing it [65]. In terms of informativeness, app operators believe that high-quality and up-to-date information strengthens the perception of security, as users associate these attributes with official management and authoritative oversight [68]. Regarding interactivity, users are more likely to feel trust when an app allows communication with others, as it reassures them that they have companions with whom they can share risks [69]. In terms of personalization, apps can be tailored to meet the security preferences of target users [70]. For example, younger users may require more advanced safety features, while older users may prefer simpler, more reassuring security modes [9]. Concerning accessibility, users are more likely to feel secure in an app when they can easily access it [71]. In practice, users have been shown to be skeptical if they are unable to access an app on the first try, often leading them to forgo subsequent attempts to access the app [5]. Previous studies have found the favorable link between app dimensions and perceived usefulness in various online function systems: user interface design [66, 67], informativeness [68], interactivity [69], personalization [70], and accessibility [71]. Thus, the study proposes that:

- H4-a: User interface design positively influence on perceived security.
- H4-b: Accessibility positively influence on perceived security.
- H4-c: Informativeness positively influence on perceived security.
- H4-d: Interactivity positively influence on perceived security.
- H4-e: Personalization positively influence on perceived security.

2.3.5 The Relationship between Perceived Ease of Use, E-Flow Experience, and Re-Browsing Intention

Lu and colleagues (2022) [72] have found that perceived ease of use positively contributes to the flow experience since the ease of using a system would keep the experience flow continously. Users will be interrupted if they find difficulty in using an app, and it will unfavorably drive down the flow experience. Ashraf, Thongpapanl, and Spyropoulou (2016) [73] argue that flow experience is crucial in enhancing other related behaviors such as visiting intention, positive word-of-mouth, or destination loyalty [74, 75]. Previous studies have supported the positive relationship between perceived ease of use and e-flow experience as well as repurchase intention across app settings [72, 75-77]. Thus, this study proposes:

- H5-a: Perceived ease of use positively influence on e-flow experience.
- H5-b: Perceived ease of use positively influence on re-browsing intention.

2.3.6 The Relationship between Perceived Usefulness E-Flow Experience, and Re-Browsing Intention

Kong and Wang (2021) [78] highlight the positive association between perceived usefulness and flow experience in their study on visual programming. Scholars support the relationship by explaining that users have deeper flow experience in an app when they find the system useful and help them perform tasks better and more effectively [79], [80]. App operators are advised to enhance the extent of usefulness that a target user perceives in enhancing the e-flow that the user experiences [72]. In addition, scholars recommend that perceived usefulness will result in post-behavioral intentions as the fact that users will continuous use apps that they evaluate useful [74]. Previous empirical studies have found the favorable relationship between perceived usefulness and flow experience as well as repurchase intention [75, 78, 81]. Thus, this study proposes:

- H6-a: Perceived usefulness positively influence on e-flow experience.
- H6-b: Perceived usefulness positively influence on re-browsing intention.

2.3.7 The Relationship between Perceived Enjoyment E-Flow Experience, and Re-Browsing Intention

Zhao and Khan (2022) [82] found that perceived enjoyment was positively associated with flow experience among students in using online English learning platforms. It is explained by the reality that when users feel enjoyable, they will be addicted and tend to spend more time [83]. Leung (2016) [84] suggested that mobile applications should be designed in a way that can create happiness for users while using them so that it will increase the adoption level as well as revisit intention. Previous empirical studies have supported the similar pattern across online systems [85-87]. Thus, this study proposes that:

- H7-a: Perceived enjoyment positively influence on e-flow experience.
- H7-b: Perceived enjoyment positively influence on re-browsing intention.

2.3.8 The Relationship between Perceived Security E-Flow Experience, and Re-Browsing Intention

Since 2005, Wu and Chang (2005) [88] recognized the role of perceived security to flow experience through the influence of trust toward an information system. From then, various authors have found the direct relationship between perceived security and flow experience as well as repurchase intention as the increasing awareness of cyber risks along with the popularity of mobile applications [65, 89, 90]. The link is explained by the fact that online users perceive security as a requirement to guarantee their safety when working in virtual environments [65]. Chen and colleagues (2017) [91] indicated that a higher perception of risks would reduce the flow experience as well as international tourist satisfaction toward revisiting Myanmar. Thus, this study proposes that:

- H8-a: Perceived security positively influence on e-flow experience.
- H8-b: Perceived security positively influence on re-browsing intention.

2.3.9 The Relationship between Perceived Enjoyment E-Flow Experience, and Re-Browsing Intention

Chen and colleagues (2017) [91] highlight the vitality of flow experience and repurchase intention in their study

among international tourists toward Myanmar. Flow experience is an overall sensation that one feels with total involvement, and when it works, it will increase people's favorable responses toward the system that creates that flow [92]. In an online game context, intense e-flow experience will lead to addiction and continuous playing [93]. Previous authors believe that the positive association between e-flow experience and re-browsing intention is similar across other virtual platforms and has been supported through empirical studies [91, 92, 94]. Thus, this study proposes that:

- H9: E-Flow experience positively influence on browsing intention.

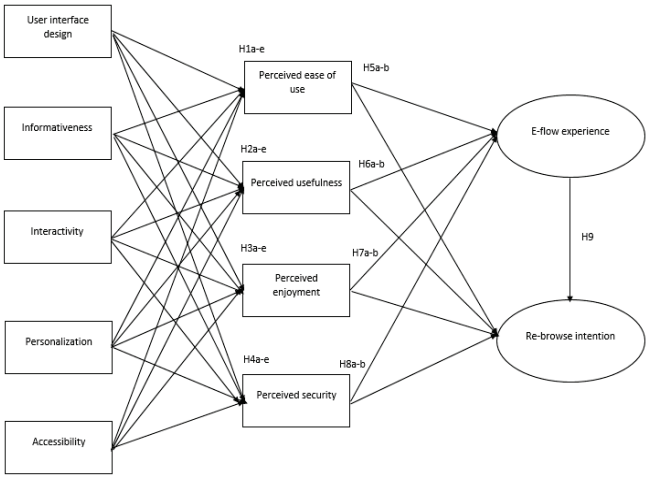


Figure 1 The research model

3 RESEARCH METHODOLOGY

3.1 Research Design

The research objectives are to identify and determine the influence of smart app dimensions on customers' perceptions and their behaviors in terms of e-flow experience and re-browsing intention toward that app, so quantitative research methodology has been employed. Research hypotheses have been developed after an extensive literature review and an adapted research model. Quantitative data has been collected and analyzed to test research hypotheses. Discussions and implications will be generated through quantitative results.

3.2 Research Method

3.2.1 Data Collection

A survey has been conducted to collect quantitative data. Target participants are young tourists (age from 16-34) who have used smart apps for their tourism searching, planning, and purchasing. Some smart tourism apps are named for participants' references: Online Travel Agents (e.g., Traveloka, Booking.com, Mytour, etc.) or destination websites (e.g., Vietnamtourism, etc.). Participants are approached through social media (Facebook, Zalo, e-mail) by online questionnaires.

The questionnaire is designed with two filtered questions and two main sections. The filtered questions are required participants to answer, including: (1) Have you ever used any smart tourism apps on mobile devices before? and (2) Please kindly name smart tourism apps that you have used. Only valid answers will be accepted for further analysis. The two main sections are: (1) evaluative items and (2) demographic information. The first section includes factors in the research model, and they have been adapted from previous reflective measurement scales. Independent factors include: (1) user interface design adapted from Sonderegger and Sauer (2010) [18]; (2) informativeness, interactivity, personalization, and accessibility adapted from [22]. Dependent factors include (1) perceived ease of use adapted from [25]; (2) perceived usefulness adapted from [29]; (3) perceived enjoyment adapted from [30]; (4) perceived security adapted from [32]; (5) e-flow experience adapted from [37]; and (6) re-browsing intention adapted from [38]. The second section includes demographic information for descriptive statistics such as gender, age, marital status, income, and occupation.

Table 1 Constructs measurement items

Constructs	Items	Authors
User interface design (DE)	The design is visually engaging.	Sonderegger and Sauer (2010) [18]
	The design is aesthetically pleasing.	
	The design has a intuitive interface.	
	The design is easy to use.	
Informativeness (IF)	During my travels, tourism apps offer valuable information about my travel destinations	No and Kim (2015) [22]
	During my travels, tourism apps are helpful for assessing both the destination(s) and the overall trip.	
	During my travels, tourism apps enable me to complete my trip with the detailed information provided.	
	During my travels, tourism apps allow me to successfully complete my trip by providing detailed and essential information.	
Interactivity (RAC)	During my travels, I can access a wide range of questions and answers from other travelers on tourism apps.	No and Kim (2015) [22]
	During my travels, the tourism apps I use are highly responsive to my needs.	
	During my travels, the tourism apps I use are highly interactive.	
	During my travels, sharing tourism-related content on apps is easy and straightforward.	
Personalization (PER)	During my travels, tourism apps provide me with personalized information tailored to my needs.	No and Kim (2015) [22]
	During my travels, tourism apps offer clear and easy-to-follow paths and links.	
	During my travels, I can interact with tourism apps to receive personalized information.	
	During my travels, the tourism information provided by the apps effectively meets my needs.	
Accessibility (ACC)	During my travels, I can use tourism apps anytime and anywhere.	No and Kim (2015) [22]
	During my travels, I can easily use tourism apps.	
	During my travels, I can easily find tourism apps.	
	During my travels, I can search without a complicated sign-up process at tourism apps.	

Table 1 Constructs measurement items (continuation)

Constructs	Items	Authors
Perceived ease of use (PE)	Learning to use the smart tourism app would be very easy.	Van der Heijden (2004) [25]
	I find it easy to use the smart tourism app.	
	It is easy to become skillful in using the smart tourism app.	
Perceived usefulness (PU)	Using the smart tourism app increases my learning/life/social interaction performance.	Yang (2017) [25]
	Smart tourism app is useful for learning/life/social networking.	
	Using the smart tourism app enhances my effectiveness in learning/life/social networking.	
	Using the smart tourism app provides me with information that would lead to better learning/life/ social networking.	
Perceived enjoyment (PJ)	I have fun interacting with the smart tourism app.	Kim, Chan, and Gupta (2007) [30]
	Using the smart tourism app provides me with a lot of enjoyment.	
	I enjoy using the smart tourism app.	
	Using the smart tourism app does not bores me.	
Perceived security (PS)	This smart app is a secure site through which to send sensitive information.	Salisbury and colleagues (2001) [32]
	I would feel totally safe providing sensitive information about myself through this app.	
	Overall, this app is a safe place to transmit sensitive information.	
E-flow experience (FL)	It is fun to use smartphone.	Liu and colleagues (2016) [37]
	Using smartphone is interesting.	
	When using smartphone, I feel the excitement of exploring.	
	I am absorbed when using smartphone.	
Re-browsing intention (RI)	I am going to re-browse smart tourism app in next trip as well.	Ajjan and Hartshorne (2008) [38]
	I think I will re-browse smart tourism app in the next trip as well.	
	I plan to re-browse smart tourism app in the next trip as well.	

The questionnaire is first developed in English and back-translated to Vietnamese to ensure compatibility between two versions of the language. A pilot test is conducted with a sample of 15 participants to minimize any errors or misunderstandings that might occur. Convenience sampling will be employed to reach an optimal number of target participants within the constraints of time and available resources. While this sampling method is subject to various biases, it is considered suitable for achieving the study's objectives [95]. The limitations of the study will be acknowledged, with suggestions provided for future research.

3.2.2 Data Analysis

Descriptive statistics will be analyzed for demographic information. Before the research hypotheses have been qualified, constructs will be tested in terms of reliability and validity, and model fit will be ensured. Reliability in terms of internal consistency (>0.8) and composite (>0.7) will be checked, while validity in terms of convergent (AVE>0.5) and discriminant ($MSV > MSA$) will be checked by SPSS and AMOS.

Model fit will be ensured and followed as suggested by [95]. Research hypotheses are supported at a p-value lower than 0.05.

4 DATA ANALYSIS AND RESULTS

4.1 Descriptive Statistics

Over five months, 518 questionnaires were collected but only 455 were qualified to continue for further analysis. Among the 455 participants, 41.1% were female. The age groups of 16-19 and 20-24 collectively comprised 68.1% of the sample. Additionally, 86.2% of participants reported an income of less than 20 million VND. Tab. 2 provides the profile of the sample.

Table 2 E-tourist respondent profiles

Respondent characteristics	No. of respondents	%
Gender		
Male	268	58.9
Female	187	41.1
Age		
16-19	112	24.6
20-24	198	43.5
25-29	86	18.9
30-34	59	13.0
Income		
< 10 million VND	176	38.7
10 – 20 million VND	216	47.5
> 20 million VND	63	13.8
Marital status		
Single	288	63.3
Married	167	36.7

4.2 Construct Measurement

Before the research model has been tested, constructs were examined to ensure their reliability and validity. In terms of reliability, internal consistency and composite reliability were conducted through Cronbach's alpha and CR. The results indicated that all constructs pass the threshold of internal consistency (Cronbach's alpha > 0.7) and composite reliability (CR > 0.6). Then, constructs were examined for their validity in terms of convergence and discriminant. Tab. 3 provides the validity of constructs.

The results indicated that all constructs were qualified with the threshold of convergent validity (AVE > 0.5) and discriminant validity ($MSV < AVE$). In addition, the model fit was ensured: (1) $C_{min}/df = 3.537$; (2) $CFI = 0.949$; (3) $PCFI = 0.829$; (4) $RMSEA = 0.075$. After the construct measurement, the structural equation modeling was analyzed to test the research hypotheses.

4.3 Structural Equation Modelling

The research of the model fit was also qualified: (1) $C_{min}/df = 3.606$; (2) $CFI = 0.947$; (3) $PCFI = 0.847$; (4)

RMSEA = 0.076. Tab. 4 provides the standardized regression weight of each research hypotheses.

The results pointed out that smart app dimensions are determined by users' evaluations toward those app functions. Specially, user interface design and informativeness positively influence perceived ease of use (β = 0.486 and β = 0.162, $p < 0.05$), supporting for H1-a and H1-b. User interface design, informativeness, personalization, and accessibility positive facilitate perceived usefulness (β = 0.13, β = 0.105, β = 0.304, and β = 0.126, $p < 0.05$), supporting for H2-a, H2-b, H2-d, and H2-e.

Interactivity and personalization positively manipulates perceived enjoyment (β = 0.481, and β = 0.109, $p < 0.05$), supporting for H3-c and H3-d. Personalization and accessibility are found to positively affect perceived sercurity (β = 0.248 and β = 0.194, $p < 0.05$), supporting for H4-d and H4-e.

Table 3 The constructs validity

	AVE	*MSV*	*ASV*	*RIa*	*DEa*	*IFa*	*RACa*	*PERa*	*ACCa*	*PEa*	*PUa*	*PJa*	*PSa*	*FLa*
RIa	0.927	0.403	0.128	0.963										
DEa	0.827	0.213	0.058	0.170	0.910									
IFa	0.942	0.205	0.046	0.189	0.042	0.971								
RACa	0.942	0.272	0.072	0.364	0.098	0.099	0.971							
PERa	0.940	0.205	0.113	0.405	0.281	0.453	0.347	0.970						
ACCa	0.979	0.124	0.034	0.063	0.352	0.051	0.154	0.177	0.989					
PEa	0.973	0.213	0.036	0.149	0.462	0.153	-0.018	0.136	0.130	0.986				
PUa	0.932	0.201	0.100	0.448	0.271	0.262	0.231	0.439	0.244	0.153	0.965			
PJa	0.899	0.272	0.107	0.466	0.068	0.163	0.522	0.299	0.094	0.011	0.385	0.948		
PSa	0.974	0.077	0.043	0.265	0.183	0.105	0.084	0.277	0.250	0.169	0.190	0.208	0.987	
FLa	0.841	0.403	0.123	0.635	0.103	0.279	0.302	0.396	0.082	0.118	0.384	0.511	0.250	0.917

Table 4 The path analysis

Hypotheses	Relationship	Path coefficient	p-value	Conclusion
H1-a	**DE → PE**	**0.486**	***	**Supported**
H1-b	**IF → PE**	**0.162**	***	**Supported**
H1-c	RAC → PE	-0.061	0.179	Unsupported
H1-d	PER → PE	-0.047	0.372	Unsupported
H1-e	ACC → PE	-0.032	0.482	Unsupported
H2-a	**DE → PU**	**0.13**	**0.005**	**Supported**
H2-b	**IF → PU**	**0.105**	**0.026**	**Supported**
H2-c	RAC → PU	0.086	0.054	Unsupported
H2-d	**PER → PU**	**0.304**	***	**Supported**
H2-e	**ACC → PU**	**0.126**	**0.005**	**Supported**
H3-a	DE → PJ	-0.012	0.797	Unsupported
H3-b	IF → PJ	0.069	0.134	Unsupported
H3-c	**RAC → PJ**	**0.481**	***	**Supported**
H3-d	**PER → PJ**	**0.109**	**0.033**	**Supported**
H3-e	ACC → PJ	0.000	0.996	Unsupported
H4-a	DE → PS	0.052	0.302	Unsupported
H4-b	IF → PS	-0.015	0.765	Unsupported
H4-c	RAC → PS	-0.034	0.484	Unsupported
H4-d	**PER → PS**	**0.248**	***	**Supported**
H4-e	**ACC → PS**	**0.194**	***	**Supported**
H5-a	PE → FL	0.064	0.125	Unsupported
H5-b	PE → RI	0.052	0.158	Unsupported
H6-a	**PU → FL**	**0.203**	***	**Supported**
H6-b	**PU → RI**	**0.202**	***	**Supported**
H7-a	**PJ → FL**	**0.423**	***	**Supported**
H7-b	**PJ → RI**	**0.145**	***	**Supported**
H8-a	**PS → FL**	**0.122**	**0.004**	**Supported**
H8-b	**PS → RI**	**0.077**	**0.038**	**Supported**
H9	**FL → RI**	**0.464**	***	**Supported**

The results indicated the relationship between user evaluation of a specific smart app and flow experience on users' intention to re-browse it. Perceived usefulness, perceived enjoyment and perceived security are supported to significantly drive on both e-flow experience and re-browse behavior. Perceived usefulness poses its influences on e-flow experience at β = 0.203 ($p < 0.05$), supporting for H6-a, and β = 0.202 ($p < 0.05$), supporting for H6-b. On the other hand, perceived enjoyment positively associates with e-flow experience at a stronger level (β = 0.423, $p < 0.05$), supporting for H7-a than with re-browse intention (β = 0.145, $p < 0.05$), supporting for H7-b. Additionally, perceived security favorably influence on both e-flow experience and re-browse intention but at smaller levels (β = 0.122, and β = 0.077, $p < 0.05$, respectively), supporting for H8-a and H8-b.

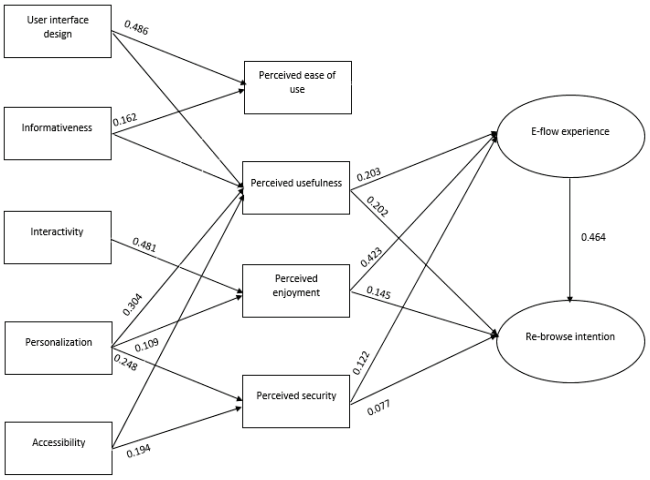


Figure 2 Structural model testing results

Finally, the results prove the positive link between e-flow experience and re-browse intention (β = 0.464, $p < 0.05$), supporting for H9. It is supported that perceived usefulness, perceived enjoyment, perceived security and flow experience play mediating roles in bridging the influence of smart app features and user re-browse intention.

The results do not support a statistically significant relationship between: (1) interactivity, personalization, accessibility, and perceived ease of use; (2) interactivity and perceived usefulness; (3) user interface design, informativeness, accessibility, and perceived enjoyment; (4) user interface design, informativeness, interactivity, and perceived security; and (5) perceived ease of use, e-flow experience, and re-browsing intention.

The results reveal discrepancies when compared to the proposed hypotheses. None of the dimensions significantly influence users' perceptions of ease of use, usefulness, enjoyment, or security, and none of the users' perceptions facilitate e-flow experience or re-browsing intention. First, perceived ease of use is not affected by interactivity, personalization, or accessibility, which was different with [49], [50], and [51]. This suggests that Vietnamese users do not find a smart tourism app easier to use simply because they can communicate with others or access the app quickly. Additionally, high customization of an app does not improve users' evaluations of ease of use. Second, perceived usefulness is not influenced by interactivity, meaning that the ability to interact with others within the app does not enhance its perceived usefulness, and was dissimilar with [54], and [55]. Third, perceived enjoyment is not driven by user interface design, informativeness, or accessibility, and was unlike with [58], and [61]. This indicates that Vietnamese users do not experience greater enjoyment simply because the app is well-designed, informative, or easily accessible. Fourth, perceived security is not determined by user interface design, informativeness, or interactivity, meaning that users' trust is not increased when the design is more polished, the information is more qualified, or the app allows for communication with other users. The findings were different with [67], [68], and [69]. Moreover, this study finds no support for a positive relationship between perceived ease of use and both e-flow experience and re-browsing intention. This results was constrast with Vahdat and colleagues (2021) [99] Chen and colleagues (2023) [100] when perceived ease of use was the strongest determinant of users' behaviors. It suggests that users' e-flow experiences are not enhanced, nor are their intentions to re-browse facilitated, by simply perceiving the app as easy to use, and the results were divergent with the study of [75]. Fig. 2 provides the result of path analysis.

5 DISCUSSIONS AND IMPLICATIONS

5.1 Theoretical and Methodological Implications

There are several theoretical implications drawn from this study. First, the study provides empirical validity for the TAM model in the context of smart tourism apps among Vietnamese. The adaptation of the TAM helps develop a research model that illuminates the relationships among smart app features in terms of (1) user interface design, (2) informativeness, (3) accessibility, (4) interactivity, and (5) personalization and users perceptions in terms of (1) perceived ease of use, (2) perceived usefulness, (3) perceived enjoyment, and (4) perceived security and e-flow experience and re-browse intention. As the TAM model provides a conceptual framework to determine external variables that contribute to perceived ease of use and perceived usefulness, which will later facilitate users behaviors, the study indicates that smart app dimensions, together with perceived enjoyment and perceived security, play as external variables in the TAM model, while e-flow experience and re-browse intention play as behavioral intention in the model [96], [97], [98], [100]. The study provides statistical results supporting the extension of the TAM model by incorporating SST dimensions and customer perceptions as determinants of customer behavior [98, 100]. Additionally, the findings emphasize e-flow experience in predicting the re-browsing intention of young Vietnamese tourists toward a smart tourism app. Mininal previous studies have recognized the role of e-flow experience in understanding users' behaviors in virtual environments.

Second, this study provides an explanation of how the SST dimensions of a smart tourism app influence users' perceptions of ease of use, usefulness, enjoyment, and security. The statistical results find that perceived ease of use is directly influenced by user interface design and informativeness, and informativeness poses a more significant impact. The findings are compatible with the study of [19] supporting that interface design contributes to users perceived ease of use. This study reveals that perceived enjoyment is directly driven by interactivity and personalization, while perceived security is both driven by personalization and accessibility. The findings reflect the same pattern with [59]. In addition, the study also demonstrates that perceived usefulness is facilitated by user interface design, informativeness, personalization and accessibility. The results align with the study by [46], offering valuable insights for practitioners on how tourism apps should be designed to enhance user evaluations. Furthermore, it contributes to the literature on SSTs by conceptualizing them as a multi-dimensional construct consisting of five dimensions.

Third, e-flow experience is positively sharpened by perceived usefulness, perceived enjoyment and perceived security. The findings are similar to the study in the same research context [78, 87, 99]. Since hedonism is regarded as the core value of a smart tourism app, perceived usefulness and enjoyment are found to be key factors driving users' experience. Furthermore, in the context of virtual environment, perceived security is required as it assures the personal safety. Finally, the study specifies that re-browse intention is manipulated by perceived usefulness, perceived enjoyment, perceived security and e-flow experience. It is highlighted the role of users evaluation of a smart tourism app in terms of its usefulness, enjoyment, and security because they will both enhance users e-flow experience and encourage users' re-browse intention [100]. Unlike previous studies in this vein, the study focuses on identifying the antecedents of re-browse intention rather than browsing intention. This is significant because the era of mobile applications is already determined by the development phase, whereas most users are familiar with smart apps. Thus, it is logical for scholars to continuously spend effort on post-behavioral intentions for more intense and long-term strategies. Furthermore, this also helps explain why in this study, perceived ease-of-use does not pose significantly impact on users' experience as well as re-browse intention.

Methodically, the study employs quantitative research, and it is considered an appropriate research design to achieve its study's objectives. Quantitative data is collected through a survey with structurally designed questionnaires and is statistically analyzed by SPSS and AMOS to qualify construct measurements and research hypotheses. The conclusions are drawn by statistical results at p-value < 0.05. All statistical results are confirmed with threshold values suggested by prominent scholars. The results are in line with

prior studies and reflect the theoretical relationship among proposed constructs.

5.2 Managerial Implications

There are several managerial implications highlighted from these study findings. For the first implication, it is once again affirmed the important role of a smart app design since its features play a significant foundation for enhancing users' e-flow experience and re-browse intention. Because five dimensions of a smart tourism app are identified as determinants of users perceived ease of use, usefulness, enjoyment, and security, smart app operators and digital managers in Vietnam tourism are recommended to prioritize the quality of their smart app dimensions in terms of interface design, informativeness, interactivity, personalization, and accessibility. However, these features should be designed and prioritized based on users' preferences, as the findings indicate that users' process and experience based on smart app dimensions. With respect to the results, the study recommends some strategic designs for a smart tourism app to be highly experienced by users in the context of Vietnam e-tourism commerce channels.

First, operators should simplify app design and enrich provided information because they will increase users' perceived ease of use and usefulness. More importantly, informativeness plays a stronger impact as the explanation that the fundamental intention of a smart tourism app is to provide related tourism information and the quality of the information will definitely add value to the app performance. Thus, tourism experts have elucidated that online users prefer to find information in a single source rather than multiple ones, and thus an app should offer diversified related information. The poor information will deteriorate users' processing and experiencing, which later will dim the re-browse intention.

Secondly, a smart app will be perceived as enjoyable if it offers a high level of interactivity as well as personalization among Vietnamese users. Interactivity allows users to interact with other parties, such as other users, app operators, or customer services. The two-way communication helps an app to be more enjoyable because users can exchange information and share what they have actually experienced. Since the core value of a tour is hedonism, and in turn, the interactivity is one way users can share that happiness with others. Subsequently, app operators nowadays consider interactivity a must-have feature rather than an optional one as before. Direct and open communication between tourism providers and users, as well as other users, may boost trust in online business transactions in the e-commerce process. Personalization also improves the perception of enjoyment since the customization of a smart app ensures the fitting with users. The high level of personalization in terms of contents and images will enhance the perceived enjoyment among users.

Thirdly, the study indicated that personalization and accessibility enhanced users' perceived security toward a smart tourism app. Unlike other previous studies, personalization is often negatively associated with perceived security as the reality that the higher degree of customization requires more revealed personal information. However, the study found the difference when personalization is favorably associated with users' perceived security. It reflects the fact that the newly designed app is even customized but still guarantees users' personal information. Regardless of tailored information provided for each individual, personalization is to enhance the user's experience but not to be used for other purposes. Fourthly, accessibility is another dimension that also increases the users' perceptions of security. It is suggested that app operators improve the accessibility because the low accessibility will cause users suspicious and prevent them from continuing to sign up due to security concerns.

For the second implication, the study accentuates the important role of users' perceptions in deciding their behaviors directly and indirectly through e-flow experience. Perceived usefulness, perceived enjoyment, and perceived security remain critical factors in promoting e-flow experience and researching intention toward a smart tourism app among Vietnamese users. The study recommends app operators should focus on increasing users perceptions of usefulness, enjoyment and security as ways to improve e-flow experience and encourage re-browsing intention. For strengthening perceived usefulness and enjoyment, this study also identifies app operators who put effort into personalization and interactivity dimensions when designing app features because two dimensions pose the most significant influences.

For the third implication, digital managers should prioritize e-flow experience in measuring online channels since it reflects the total involvement of users experience when browsing and it will facilitate future behavioral intentions. Previous studies have demonstrated that e-flow experience will result in engaged behaviors such as citizenship behaviors, loyalty, and addiction, and in this study, re-browsing intention is found to be driven by e-flow experience. Re-browsing intention is an expected behavior that app operators put effort into promoting since it not only exhibits users' satisfaction with prior browsing but also brings on promising benefits. Experts show that smart apps have already passed the adoption level among young users, and subsequently academics are required to focus more on rebrowsing post-adoption behaviors. This study recommends that in order to improve the users experience and encourage future behavior, tourism providers should implement smart tourism technologies, as the results state that users perceptions are influenced by smart app dimensions, and the findings also highlight that users perceptions will enhance e-flow experience and facilitate re-browsing intention.

5.3 Limitations and Future Research Directions

Regardless of implications drawing from empirical findings, this study has also suffered from several limitations. The first limitation lies in the research model. Employing the extended TAM model, the study only considered smart app features, perceived enjoyment, and perceived security as external variables, while there were various other factors. In the same vein, research was determined at behavioral intention while others were excluded. Further studies can continue to take into consideration other factors that have not been researched in this study to provide a more

comprehensive picture about the relationships among related factors in this field. For example, exploring the relationship between e-flow experience, e-satisfaction, and e-word of mouth could provide valuable insights for developing more strategic app designs. The second limitation lies in the sampling method. This study employed the convenience sampling method to collect data. This non-probability method has its limitations. Future studies should improve the quality of data by employing methods that help reduce the weaknesses caused by this method. The last limitation lies in the sample. Although demographic information was collected, this study did not take into account the variance among demographic criteria. Future studies should consider the differences among online users because it is believed that digital behaviors are widely segmented. Future studies can compare the similarities and dissimilarities among groups to better highlight their behaviors in a virtual environment. For example, urban tourists may perceive the app differently from tourists in other regions of the country, or female tourists may experience e-flow differently than male tourists, leading to variations in re-browsing intentions.

6 CONCLUSION

This study demonstrates the significant impact of smart app dimensions, particularly personalization and interactivity, on young Vietnamese tourists' perceptions and re-browse intentions. By applying an extended TAM framework, the research highlights the importance of perceived enjoyment and security in fostering e-flow experiences. These findings offer valuable insights for app developers aiming to enhance user engagement and loyalty. Future research should explore diverse demographic groups, integrate additional app dimensions, and assess long-term behavioral outcomes to further validate and extend these findings.

7 REFERENCES

[1] Mehraliyev, F., Chan, I. C. C., Choi, Y., Koseoglu, M. A. & Law, R. (2020). A state-of-the-art review of smart tourism research. *Journal of Travel & Tourism Marketing*, *37*(1), 78-91. https://doi.org/10.1080/10548408.2020.1712309

[2] Ng, K. S. P., Wong, J. W. C., Xie, D. & Zhu, J. (2023). From the attributes of smart tourism technologies to loyalty and WOM via user satisfaction: the moderating role of switching costs. *Kybernetes*, *52*(8), 2868-2885. https://doi.org/10.1108/k-09-2021-0840

[3] Um, T. & Chung, N. (2021). Does smart tourism technology matter? Lessons from three smart tourism cities in South Korea. *Asia Pacific Journal of Tourism Research*, *26*(4), 396-414. https://doi.org/10.1080/10941665.2019.1595691

[4] Zhang, Y., Sotiriadis, M. & Shen, S. (2022). Investigating the impact of smart tourism technologies on tourists' experiences. *Sustainability*, *14*(5), 3048. https://doi.org/10.3390/su14053048

[5] Lee, H., Lee, J., Chung, N. & Koo, C. (2018). Tourists' happiness: are there smart tourism technology effects? *Asia Pacific Journal of Tourism Research*, *23*(5), 486-501. https://doi.org/10.1080/10941665.2018.1468344

[6] Statista (2023). Retrieved from: https://www.statista.com/statistics/1211507/vietnam-number-of-domestic-tourist-arrivals/

[7] Nguyen, V. H., Truong, T. X. D., Pham, H. T., Tran, D. T. & Nguyen, P. H. (2021). Travel intention to visit tourism destinations: A perspective of generation Z in Vietnam. *The Journal of Asian Finance, Economics and Business*, *8*(2), 1043-1053.

[8] Hien, H. N. & Trang, P. H. (2024). Decoding smart tech's influence on tourist experience quality. *Asian Journal of Business Research Volume*, *14*(1). https://doi.org/10.14707/ajbr.230167

[9] Azis, N., Amin, M., Chan, S. & Aprilia, C. (2020). How smart tourism technologies affect tourist destination loyalty. *Journal of hospitality and tourism technology*, *11*(4), 603-625. https://doi.org/10.1108/jhtt-01-2020-0005

[10] Elgarhy, S. D., Abdel Rahieem, W. M. A. N. & Abdulmawla, M. (2024). Influences of gamification on repurchase intention and intrinsic motivations in Egyptian hotels and travel agencies: The mediating role of customer engagement. *Journal of Quality Assurance in Hospitality & Tourism*, *25*(2), 397-419. https://doi.org/10.1080/1528008x.2023.2194705

[11] Roy, S. K., Singh, G., Hope, M., Nguyen, B. & Harrigan, P. (2022). The rise of smart consumers: role of smart servicescape and smart consumer experience co-creation. In *The Role of Smart Technologies in Decision Making* (pp. 114-147). Routledge. https://doi.org/10.4324/9781003307105-6

[12] Hamzah, Z. L., Waqas, M., Rahman, R. B. & Ghazali, E. (2023). Consumer repurchase intention of smart wearable devices: an extension of the technology integration model. *International Journal of Technology Marketing*, *17*(1), 78-98. https://doi.org/10.1504/ijtmkt.2023.127329

[13] Ionescu, A. M. & Sârbu, F. A. (2024). Exploring the Impact of Smart Technologies on the Tourism Industry. *Sustainability*, *16*(8), 3318. https://doi.org/10.3390/su16083318

[14] Shen, S., Sotiriadis, M. & Zhang, Y. (2020). The influence of smart technologies on customer journey in tourist attractions within the smart tourism management framework. *Sustainability*, *12*(10), 4157. https://doi.org/10.3390/su12104157

[15] Gretzel, U., Sigala, M., Xiang, Z. & Koo, C. (2015). Smart tourism: foundations and developments. *Electronic markets*, *25*, 179-188. https://doi.org/10.1007/s12525-015-0196-8

[16] Bhandari, U., Neben, T., Chang, K. & Chua, W. Y. (2017). Effects of interface design factors on affective responses and quality evaluations in mobile applications. *Computers in Human Behavior*, *72*, 525-534. https://doi.org/10.1016/j.chb.2017.02.044

[17] Lee, Y. E. & Benbasat, I. (2004). A framework for the study of customer interface design for mobile commerce. *International Journal of Electronic Commerce*, *8*(3), 79-102. https://doi.org/10.1080/10864415.2004.11044299

[18] Sonderegger, A. & Sauer, J. (2010). The influence of design aesthetics in usability testing: Effects on user performance and perceived usability. *Applied ergonomics*, *41*(3), 403-410. https://doi.org/10.1016/j.apergo.2009.09.002

[19] Punchoojit, L. & Hongwarittorrn, N. (2017). Usability studies on mobile user interface design patterns: a systematic literature review. *Advances in Human-Computer Interaction*, *2017*(1), 6787504. https://doi.org/10.1155/2017/6787504

[20] Yu, N. & Kong, J. (2016). User experience with web browsing on small screens: Experimental investigations of mobile-page interface design and homepage design for news websites. *Information Sciences*, *330*, 427-443. https://doi.org/10.1016/j.ins.2015.06.004

[21] Huang, C. D., Goo, J., Nam, K. & Yoo, C. W. (2017). Smart tourism technologies in travel planning: The role of exploration and exploitation. *Information & Management*, *54*(6), 757-770. https://doi.org/10.1016/j.im.2016.11.010

[22] No, E. & Kim, J. K. (2015). Comparing the attributes of online tourism information sources. *Computers in human behavior*, *50*, 564-575. https://doi.org/10.1016/j.chb.2015.02.063

[23] Davis, F. D., Bagozzi, R. P. & Warshaw, P. R. (1989). Technology acceptance model. *J Manag Sci*, *35*(8), 982-1003. https://doi.org/10.1287/mnsc.35.8.982

[24] Li, T. & Zhu, Z. (2023). A meta-analysis of online travel booking services adoption: the moderating effects of national or regional economic development level and culture differences. *Aslib Journal of Information Management*, *75*(6), 1209-1234. https://doi.org/10.1108/ajim-04-2022-0198

[25] Van der Heijden, H. (2004). User acceptance of hedonic information systems. *MIS quarterly*, 695-704. https://doi.org/10.2307/25148660

[26] Hubert, M., Blut, M., Brock, C., Zhang, R. W., Koch, V. & Riedl, R. (2019). The influence of acceptance and adoption drivers on smart home usage. *European journal of marketing*, *53*(6), 1073-1098. https://doi.org/10.1108/ejm-12-2016-0794

[27] Wang, W. (2019). The influence of perceived technological congruence of smartphone application and air travel experience on consumers' attitudes toward price change and adoption. *Journal of Hospitality and Tourism Technology*, *10*(2), 122-135. https://doi.org/10.1108/jhtt-01-2018-0004

[28] Chang, Y. W. & Chen, J. (2021). What motivates customers to shop in smart shops? The impacts of smart technology and technology readiness. *Journal of Retailing and Consumer Services*, *58*, 102325. https://doi.org/10.1016/j.jretconser.2020.102325

[29] Yang, F. X. (2017). Effects of restaurant satisfaction and knowledge sharing motivation on eWOM intentions: the moderating role of technology acceptance factors. *Journal of Hospitality & Tourism Research*, *41*(1), 93-127. https://doi.org/10.1177/1096348013515918

[30] Kim, H. W., Chan, H. C. & Gupta, S. (2007). Value-based adoption of mobile internet: an empirical investigation. *Decision support systems*, *43*(1), 111-126. https://doi.org/10.1016/j.dss.2005.05.009

[31] Chung, N., Tyan, I. & Han, H. (2017). Enhancing the smart tourism experience through geotag. *Information Systems Frontiers*, *19*, 731-742. https://doi.org/10.1007/s10796-016-9710-6

[32] Salisbury, W. D., Pearson, R. A., Pearson, A. W. & Miller, D. W. (2001). Perceived security and World Wide Web purchase intention. *Industrial Management & Data Systems*, *101*(4), 165-177. https://doi.org/10.1108/02635570110390071

[33] Pai, C. K., Liu, Y., Kang, S. & Dai, A. (2020). The role of perceived smart tourism technology experience for tourist satisfaction, happiness and revisit intention. *Sustainability*, *12*(16), 6592. https://doi.org/10.3390/su12166592

[34] Liu, Y. & Sun, X. (2024). Tourism e-commerce live streaming: The effects of live streamer authenticity on purchase intention. *Tourism Review*, *79*(5), 1147-1165. https://doi.org/10.1108/tr-04-2023-0245

[35] Liu, X., Zhang, L. & Chen, Q. (2022). The effects of tourism e-commerce live streaming features on consumer purchase intention: The mediating roles of flow experience and trust. *Frontiers in Psychology*, *13*, 995129. https://doi.org/10.3389/fpsyg.2022.995129

[36] Xie, C., Yu, J., Huang, S. S. & Zhang, J. (2022). Tourism e-commerce live streaming: Identifying and testing a value-based marketing framework from the live streamer perspective. *Tourism management*, *91*, 104513. https://doi.org/10.1016/j.tourman.2022.104513

[37] Liu, H., Chu, H., Huang, Q. & Chen, X. (2016). Enhancing the flow experience of consumers in China through interpersonal interaction in social commerce. *Computers in Human Behavior*, *58*, 306-314. https://doi.org/10.1016/j.chb.2016.01.012

[38] Ajjan, H. & Hartshorne, R. (2008). Investigating faculty decisions to adopt Web 2.0 technologies: Theory and empirical tests. *The internet and higher education*, *11*(2), 71-80. https://doi.org/10.1016/j.iheduc.2008.05.002

[39] Ho, M. H. W. & Chung, H. F. (2020). Customer engagement, customer equity and repurchase intention in mobile apps. *Journal of business research*, *121*, 13-21. https://doi.org/10.1016/j.jbusres.2020.07.046

[40] Cha, S. S. & Lee, S. H. (2021). The effects of user experience factors on satisfaction and repurchase intention at online food market. *The Journal of Industrial Distribution & Business*, *12*(4), 7-13.

[41] Tresna, P. W., Suryanto & Haidysirly, H. (2019). Effect of perceived value on repurchase intention in e-commerce: a study on the Female Daily mobile application. *International Journal of Trade and Global Markets*, *12*(3-4), 373-380. https://doi.org/10.1504/ijtgm.2019.101560

[42] Xia, M., Zhang, Y. & Zhang, C. (2018). A TAM-based approach to explore the effect of online experience on destination image: A smartphone user's perspective. *Journal of destination marketing & management*, *8*, 259-270. https://doi.org/10.1016/j.jdmm.2017.05.002

[43] Lin, S. Y., Juan, P. J. & Lin, S. W. (2020). A tam framework to evaluate the effect of smartphone application on tourism information search behavior of foreign independent travelers. *Sustainability*, *12*(22), 9366. https://doi.org/10.3390/su12229366

[44] Verkijika, S. F. & De Wet, L. (2019). Understanding word-of-mouth (WOM) intentions of mobile app users: The role of simplicity and emotions during the first interaction. *Telematics and Informatics*, *41*, 218-228. https://doi.org/10.1016/j.tele.2019.05.003

[45] Cho, V., Cheng, T. E. & Lai, W. J. (2009). The role of perceived user-interface design in continued usage intention of self-paced e-learning tools. *Computers & Education*, *53*(2), 216-227. https://doi.org/10.1016/j.compedu.2009.01.014

[46] Arifah, I. D. C. & Juniarti, R. P. (2021, May). Interface aesthetic, perceived value, perceived ease of use, and perceived usefulness on purchase intention of smartwatch consumers. In *International Conference on Business and Engineering Management (ICONBEM 2021)* (pp. 25-33). Atlantis Press. https://doi.org/10.2991/aebmr.k.210522.005

[47] Wu, R., Wang, G. & Yan, L. (2020). The effects of online store informativeness and entertainment on consumers' approach behaviors: Empirical evidence from China. *Asia Pacific Journal of Marketing and Logistics*, *32*(6), 1327-1342. https://doi.org/10.1108/apjml-03-2019-0182

[48] Parikh, A., Patel, J. D. & Jaiswal, A. K. (2021). Managing job applications online: integrating website informativeness and compatibility in theory of planned behaviour and technology acceptance model. *Decision*, *48*, 97-113. https://doi.org/10.1007/s40622-020-00266-2

[49] Gupta, P., Prashar, S., Vijay, T. S. & Parsad, C. (2021). Examining the influence of antecedents of continuous intention to use an informational app: the role of perceived usefulness and perceived ease of use. *International Journal of Business Information Systems*, *36*(2), 270-287. https://doi.org/10.1504/ijbis.2021.112829

[50] Choi, B., Kwon, O. & Shin, B. (2017). Location-based system: Comparative effects of personalization vs ease of use. *Telematics and Informatics*, *34*(1), 91-102. https://doi.org/10.1016/j.tele.2016.04.011

[51] Shen, C. C. & Chiou, J. S. (2010). The impact of perceived ease of use on Internet service adoption: The moderating effects of temporal distance and perceived risk. *Computers in human behavior*, *26*(1), 42-50. https://doi.org/10.1016/j.chb.2009.07.003

[52] Baker-Eveleth, L. & Stone, R. W. (2020). User's perceptions of perceived usefulness, satisfaction, and intentions of mobile application. *International Journal of Mobile Communications*, *18*(1), 1-18. https://doi.org/10.1504/ijmc.2020.10025192

[53] Wang, E. S. T. (2016). The moderating role of consumer characteristics in the relationship between website quality and perceived usefulness. *International Journal of Retail & Distribution Management*, *44*(6), 627-639. https://doi.org/10.1108/ijrdm-03-2015-0049

[54] Kirk, C. P., Chiagouris, L. & Gopalakrishna, P. (2012). Some people just want to read: The roles of age, interactivity, and perceived usefulness of print in the consumption of digital information products. *Journal of retailing and consumer services*, *19*(1), 168-178. https://doi.org/10.1016/j.jretconser.2011.11.006

[55] Abdullah, D., Jayaraman, K., Shariff, D. N., Bahari, K. A. & Nor, N. M. (2017). The effects of perceived interactivity, perceived ease of use and perceived usefulness on online hotel booking intention: A conceptual framework. *International Academic Research Journal of Social Science*, *3*(1), 16-23. https://doi.org/10.7456/1080sse/102

[56] Chandra, S., Shirish, A. & Srivastava, S. C. (2020). Theorizing technological spatial intrusion for ICT enabled employee innovation: The mediating role of perceived usefulness. *Technological Forecasting and Social Change*, *161*, 120320. https://doi.org/10.1016/j.techfore.2020.120320

[57] Yang, Y. & Koenigstorfer, J. (2021). Determinants of fitness app usage and moderating impacts of education-, motivation-, and gamification-related app features on physical activity intentions: cross-sectional survey study. *Journal of medical Internet research*, *23*(7), e26063. https://doi.org/10.2196/26063

[58] Won, D., Chiu, W. & Byun, H. (2023). Factors influencing consumer use of a sport-branded app: The technology acceptance model integrating app quality and perceived enjoyment. *Asia Pacific Journal of Marketing and Logistics*, *35*(5), 1112-1133. https://doi.org/10.1108/apjml-09-2021-0709

[59] Arghashi, V. & Yuksel, C. A. (2022). Interactivity, Inspiration, and Perceived Usefulness! How retailers' AR-apps improve consumer engagement through flow. *Journal of Retailing and Consumer Services*, *64*, 102756. https://doi.org/10.1016/j.jretconser.2021.102756

[60] Liu, H., Shao, M., Liu, X. & Zhao, L. (2021). Exploring the influential factors on readers' continuance intentions of E-Book APPs: personalization, usefulness, playfulness, and satisfaction. *Frontiers in Psychology*, *12*, 640110. https://doi.org/10.3389/fpsyg.2021.640110

[61] Sauer, J., Sonderegger, A. & Schmutz, S. (2020). Usability, user experience and accessibility: towards an integrative model. *Ergonomics*, *63*(10), 1207-1220. https://doi.org/10.1080/00140139.2020.1774080

[62] Hubert, M., Blut, M., Brock, C., Backhaus, C. & Eberhardt, T. (2017). Acceptance of smartphone-based mobile shopping: Mobile benefits, customer characteristics, perceived risks, and the impact of application context. *Psychology & Marketing*, *34*(2), 175-194. https://doi.org/10.1002/mar.20982

[63] Treen, E., Pitt, L., Bredican, J. & Farshid, M. (2017). App service: How do consumers perceive the quality of financial service apps on smart devices? *Journal of Financial Services Marketing*, *22*, 119-125. https://doi.org/10.1057/s41264-017-0029-2

[64] Turner-McGrievy, G. M., Hales, S. B., Schoffman, D. E., Valafar, H., Brazendale, K., Weaver, R. G., ... & McGrievy, M. J. (2017). Choosing between responsive-design websites versus mobile apps for your mobile behavioral intervention: presenting four case studies. *Translational behavioral medicine*, *7*(2), 224-232. https://doi.org/10.1007/s13142-016-0448-y

[65] Wu, L., Chiu, M. L. & Chen, K. W. (2020). Defining the determinants of online impulse buying through a shopping process of integrating perceived risk, expectation-confirmation model, and flow theory issues. *International Journal of Information Management*, *52*, 102099. https://doi.org/10.1016/j.ijinfomgt.2020.102099

[66] Kamoun, F. & Halaweh, M. (2012). User interface design and e-commerce security perception: An empirical study. *International Journal of E-Business Research (IJEBR)*, *8*(2), 15-32. https://doi.org/10.4018/jebr.2012040102

[67] Zhang, J., Luximon, Y. & Song, Y. (2019). The role of consumers' perceived security, perceived control, interface design features, and conscientiousness in continuous use of mobile payment services. *Sustainability*, *11*(23), 6843. https://doi.org/10.3390/su11236843

[68] Mao, E. & Zhang, J. (2017). What affects users to click on display ads on social media? The roles of message values, involvement, and security. *Journal of Information Privacy and Security*, *13*(2), 84-96. https://doi.org/10.1080/15536548.2017.1322434

[69] Shankar, A. (2021). Impact of mobile banking application interactivity on consumer engagement: An Experiment-based investigation. *Journal of Global Information Management (JGIM)*, *30*(5), 1-18. https://doi.org/10.4018/jgim.290368

[70] Alalwan, A. A., Baabdullah, A. M., Rana, N. P., Dwivedi, Y. K. & Kizgin, H. (2019). Examining the influence of mobile store features on user e-satisfaction: extending UTAUT2 with personalization, responsiveness, and perceived security and privacy. In *Digital Transformation for a Sustainable Society in the 21st Century: 18th IFIP WG 6.11 Conference on e-Business, e-Services, and e-Society, I3E 2019, Trondheim, Norway, September 18–20, 2019, Proceedings 18* (pp. 50-61). Springer International Publishing. https://doi.org/10.1007/978-3-030-29374-1_5

[71] Friman, M., Lättman, K. & Olsson, L. E. (2020). Public transport quality, safety, and perceived accessibility. *Sustainability*, *12*(9), 3563. https://doi.org/10.3390/su12093563

[72] Lu, A., Deng, R., Huang, Y., Song, T., Shen, Y., Fan, Z. & Zhang, J. (2022). The roles of mobile app perceived usefulness and perceived ease of use in app-based Chinese and English learning flow and satisfaction. *Education and Information Technologies*, *27*(7), 10349-10370. https://doi.org/10.1007/s10639-022-11036-1

[73] Ashraf, A. R., Thongpapanl, N. T. & Spyropoulou, S. (2016). The connection and disconnection between e-commerce businesses and their customers: Exploring the role of engagement, perceived usefulness, and perceived ease-of-use. *Electronic Commerce Research and Applications*, *20*, 69-86. https://doi.org/10.1016/j.elerap.2016.10.001

[74] Matute, J., Polo-Redondo, Y. & Utrillas, A. (2016). The influence of EWOM characteristics on online repurchase intention: Mediating roles of trust and perceived usefulness. *Online Information Review*, *40*(7), 1090-1110. https://doi.org/10.1108/oir-11-2015-0373

[75] Yeo, S. F., Tan, C. L., Teo, S. L. & Tan, K. H. (2021). The role of food apps servitization on repurchase intention: A study of FoodPanda. *International Journal of Production Economics*, *234*, 108063. https://doi.org/10.1016/j.ijpe.2021.108063

[76] Zhou, T. (2012). Examining mobile banking user adoption from the perspectives of trust and flow experience. *Information Technology and Management*, *13*, 27-37. https://doi.org/10.1007/s10799-011-0111-8

[77] Rodrigues, L. F., Oliveira, A. & Costa, C. J. (2016). Does ease-of-use contributes to the perception of enjoyment? A case of gamification in e-banking. *Computers in Human Behavior*, *61*, 114-126. https://doi.org/10.1016/j.chb.2016.03.015

[78] Kong, S. C. & Wang, Y. Q. (2021). The influence of parental support and perceived usefulness on students' learning motivation and flow experience in visual programming: Investigation from a parent perspective. *British Journal of Educational Technology*, *52*(4), 1749-1770. https://doi.org/10.1111/bjet.13071

[79] Aboelmaged, M. G. (2018). Predicting the success of Twitter in healthcare: A synthesis of perceived quality, usefulness and flow experience by healthcare professionals. *Online Information Review*, *42*(6), 898-922. https://doi.org/10.1108/oir-01-2017-0018

[80] Wang, X., Lu, A., Lin, T., Liu, S., Song, T., Huang, X. & Jiang, L. (2022). Perceived usefulness predicts second language learners' continuance intention toward language learning applications: a serial multiple mediation model of integrative motivation and flow. *Education and Information Technologies*, 1-17. https://doi.org/10.1007/s10639-021-10822-7

[81] Gloria, T. V. & Achyar, A. (2018). Effects of externalities and flow on perceived usefulness, satisfaction, and loyalty in mobile instant messaging promotion. *ASEAN Marketing Journal*, *8*(2), 2.

[82] Zhao, H. & Khan, A. (2022). The students' flow experience with the continuous intention of using online English platforms. *Frontiers in Psychology*, *12*, 807084. https://doi.org/10.3389/fpsyg.2021.807084

[83] Mahfouz, A. Y., Joonas, K. & Opara, E. U. (2020). An overview of and factor analytic approach to flow theory in online contexts. *Technology in Society*, *61*, 101228. https://doi.org/10.1016/j.techsoc.2020.101228

[84] Leung, R., Hastings, J. F., Keefe, R. H., Brownstein-Evans, C., Chan, K. T. & Mullick, R. (2016). Building mobile apps for underrepresented mental health care consumers: A grounded theory approach. *Social work in mental health*, *14*(6), 625-636. https://doi.org/10.1080/15332985.2015.1130010

[85] Bartzik, M., Bentrup, A., Hill, S., Bley, M., von Hirschhausen, E., Krause, G., ... & Peifer, C. (2021). Care for joy: Evaluation of a humor intervention and its effects on stress, flow experience, work enjoyment, and meaningfulness of work. *Frontiers in public health*, *9*, 667821. https://doi.org/10.3389/fpubh.2021.667821

[86] Abuhamdeh, S. (2021). On the relationship between flow and enjoyment. *Advances in flow research*, 155-169. https://doi.org/10.1007/978-3-030-53468-4_6

[87] Oliveira, W., Hamari, J., Joaquim, S., Toda, A. M., Palomino, P. T., Vassileva, J. & Isotani, S. (2022). The effects of personalized gamification on students' flow experience, motivation, and enjoyment. *Smart Learning Environments*, *9*(1), 16. https://doi.org/10.1186/s40561-022-00194-x

[88] Wu, J. J. & Chang, Y. S. (2005). Towards understanding members' interactivity, trust, and flow in online travel community. *Industrial Management & Data Systems*, *105*(7), 937-954. https://doi.org/10.1108/02635570510616120

[89] Aggarwal, A. & Rahul, M. (2018). The effect of perceived security on consumer purchase intensions in electronic commerce. *International Journal of Public Sector Performance Management*, *4*(1), 1-20. https://doi.org/10.1504/ijpspm.2018.088691

[90] Kumar, A., Adlakaha, A. & Mukherjee, K. (2018). The effect of perceived security and grievance redressal on continuance intention to use M-wallets in a developing country. *International Journal of Bank Marketing*, *36*(7), 1170-1189. https://doi.org/10.1108/ijbm-04-2017-0077

[91] Chen, J. V., Htaik, S., Hiele, T. M., & Chen, C. (2017). Investigating international tourists' intention to revisit Myanmar based on need gratification, flow experience and perceived risk. *Journal of Quality Assurance in Hospitality & Tourism*, *18*(1), 25-44. https://doi.org/10.1080/1528008x.2015.1133367

[92] Prodanova, J., San-Martín, S. & Jiménez, N. (2020). Achieving customers' repurchase intention through stimuli and site attachment. *Journal of Organizational Computing and Electronic Commerce*, *30*(3), 187-208 https://doi.org/10.1080/10919392.2020.1739395

[93] Chang, C. C. (2013). Examining users′ intention to continue using social network games: A flow experience perspective. *Telematics and Informatics*, *30*(4), 311-321. https://doi.org/10.1016/j.tele.2012.10.006

[94] Cho, Y. K. (2015). Creating customer repurchase intention in Internet retailing: The effects of multiple service events and product type. *Journal of retailing and consumer services*, *22*, 213-222. https://doi.org/10.1016/j.jretconser.2014.11.002

[95] Hu, L. T. & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural equation modeling: a multidisciplinary journal*, *6*(1), 1-55. https://doi.org/10.1080/10705519909540118

[96] Huang, J. H., Lin, Y. R. & Chuang, S. T. (2007). Elucidating user behavior of mobile learning: A perspective of the extended technology acceptance model. *The electronic library*, *25*(5), 585-598. https://doi.org/10.1108/02640470710829569

[97] Rafique, H., Almagrabi, A. O., Shamim, A., Anwar, F. & Bashir, A. K. (2020). Investigating the acceptance of mobile library applications with an extended technology acceptance model (TAM). *Computers & Education*, *145*, 103732. https://doi.org/10.1016/j.compedu.2019.103732

[98] Lin, S. Y., Juan, P. J. & Lin, S. W. (2020). A tam framework to evaluate the effect of smartphone application on tourism information search behavior of foreign independent travelers. *Sustainability*, *12*(22), 9366. https://doi.org/10.3390/su12229366

[99] Vahdat, A., Alizadeh, A., Quach, S. & Hamelin, N. (2021). Would you like to shop via mobile app technology? The technology acceptance model, social factors and purchase intention. *Australasian Marketing Journal*, *29*(2), 187-197. https://doi.org/10.1016/j.ausmj.2020.01.002

[100] Chen, C. C., Liu, C. C., Chiu, T. H., Lee, Y. W. & Wu, K. C. (2023). Role of perceived ease of use for augmented reality app designed to help children navigate smart libraries. *International Journal of Human–Computer Interaction, 39*(13), 2606-2623. https://doi.org/10.1080/10447318.2022.2082017

Author's contacts:

Thai-Ngoc Pham, PhD
Faculty of Economics – Law, Dong Thap University,
783 Pham Huu Lau, Ward 6, Cao Lanh city, Dong Thap province, Vietnam
ptngoc@dthu.edu.vn

ISSN 1846-6168 (Print), ISSN 1848-5588 (Online)
https://doi.org/10.31803/tg-20241216050348

Original scientific paper
Received: 2024-12-16, Accepted: 2025-01-13

Enhancing Inventory Simulation Models for Retail: Addressing Design Flaws Using Arena Software

Daniel Lukito*, Andi Kusdiana, Fergyanto E. Gunawan, Rida Zuraida

Abstract: This study aims to propose an improvement to the simulation model of inventory management system in retail stores developed in the reference paper. This paper analyses the process flow of the original simulation model and also conducted an interview with a retail store owner. This study uses the customer arrival and customer demand distribution patterns and most of other parameter data from the reference paper. Both the original and the enhanced models are optimized for the lowest demand lost. Arena simulation software was used to perform the simulations for both models. The results of both models were then compared. When analysing the original model a design flaw was found. In the event where a customer come and demand a quantity that is more than the current inventory level, the original model will only report a stockout without reordering the product. If this event happened while there is no undelivered order to refill the stock, then all the following demand would go to the same route. Hence, from there on, no order will be fulfilled. In the interview, the store owner also shares about a part of the original model which is not a common practice, such as not offering the remaining stock to the customer who needs more than the current inventory level. The new simulation model was then developed. This study addresses critical design flaws in an inventory simulation model for retail stores. By leveraging insights from field interviews and utilizing Arena software, an enhanced model was developed, achieving a 26 % reduction in lost customers and a 35 % improvement in demand fulfilment. The findings provide actionable insights for optimizing inventory management systems in retail environments.

Keywords: arena simulation; improvement; inventory management; retail; simulation model

1 INTRODUCTION

Inventory management systems (IMS) are central to organizational success because they control flow and costs of inventory and customer reactions. When well implemented IMS should ensure minimum levels of stock out and over stocking while it timely delivers the products. Research shows that strong IMS practices result into improved financial performance and supply chain competitive advantage through reducing inventory related costs and decisions based on customer needs [1, 2]. In fact, it is established that organizations that incorporate these advanced inventory tools receive great benefits in terms of resource and profitability [3].

The retail industry defines a huge number of companies and is considered one of the most promising and highly committed to the world market. The nature of the retail business varies because there are different product categories and distribution types, thus facing problems in the store when demand characteristics change and when the market constantly evolves [4]. Intense focus on retailing has been evident in the past few years and any organisation involved in the retail business requires effective IMS to manage the massive volumes of products and customer expectations. The emergence of this industry underscores the necessity to develop new methods for dealing with this industry's inventory issues, with the aim of achieving performance stability in a constantly changing environment [5].

Another theme common to both IMS research and practice is the application of simulation methods to describe and predict the behaviours of systems and objects. By doing so, it is possible to bench test the changes with a number of realistic scenarios in inventory policies while avoiding the impact of the changes on actual business environment; thus providing ideas on the possible benefits – and drawbacks – of certain changes [6]. Simulation henceforth leads an important role as a decision support tool in inventory management. Simulation technologies are continually used to mitigate the unpredictability and variability in the demand of inventory management systems in retail stores, to perfectly optimize its stock levels and, in particular, to overcome threats of supply chain interruptions. Discrete event simulation (DES) has been frequently applied in analysing inventory processes with a focus on such issues as ordering policies and safety stock, which can be convenient for retailers if different demand patterns are considered [7]. In this way, all dynamic aspects of the system are reflected in DES, and knowing how it is possible to optimise inventory policies in order to decrease the frequency of stockouts and, at the same time, eliminate overstock conditions, the efficiency of the system is increased [8]. Moussavi et al. [9] show that discrete-event simulation can be used in managing uncertainty especially in demand and supply chain systems.

A model cannot represent the real thing perfectly [10]. Inventory models should be expanded incrementally by including the concepts of sustainable and efficient and is typical of the tendency towards more intricate, data-based approaches [11]. When researchers develop simulation models to address these complexities, design flaws may arise due to oversimplifications or inadequate representation of real-world dynamics. For instance, many models fail to capture the interplay between multiple variables, such as pricing strategies, demand elasticity, and lead time variability, resulting in limited applicability to practical scenarios [12]. Furthermore, the lack of accurate or comprehensive data can lead to flawed assumptions, reducing the reliability of simulation outcomes [13]. In some cases, the models are not updated to reflect evolving market conditions or emerging technologies, which can render their recommendations obsolete [14].

Simulation models presented in academic literature have a vast implication in theoretical and management practice. Thus, they need to be verified and validated against the current customer behaviour. However, research on evaluating and improving a simulation model, especially in retail store's inventory management system, is still scarce.

Therefore, the current study aims to build upon the simulation model of IMS outlined by [15] by pointing to its design flaw as well as areas of improvement. Consequently, this paper seeks to increase the understanding and improvement of the simulation model through enabling practical solutions that correspond to the research question: In what way can the simulation model be enhanced? It is believed that this endeavour will be of successful interest to the academia as well as the retail practitioners by offering fresh insights to existing gaps in literature.

2 LITERATURE REVIEW

2.1 Inventory Management System

Inventory systems are a must to have in any business as they ensure appropriative stocking and storage, minimizing on over stocking, or running out of stock. With the help of digital technology, the visibility of supply chain is extended. It will also certainly help the company to track better and improve the time of delivery which of course leads to improved levels of customer satisfaction [16]. Concerning the realisation of the contemporary inventory systems, perhaps due to the boost in the methods of data processing in decision-making processes, inventory performance has improved in terms of precision and reaction time in supply chain networks [17]. In particular, a strategic management of inventory impacts cost control and subsequent customer satisfaction, which are elements critical to diverse industries.

The ever-changing nature of markets has resulted to updates in the use of more appropriate technology of real-time inventory and metrics. Examples include IoT connected inventories through which a business can easily gather fresh knowledge and modify its stock turnover frequency to meet the changing needs while at the same time reducing holding costs [18]. This innovation is most useful in organizations that have inconsistent demand characteristics including the retail and e-commerce businesses.

Research indicates that the deployment of big data, coupled with predictive analytics and machine learning models, complements the traditional inventory management procedures that are actually a part of inventory management and increase the efficiency of demand forecasting of new inventory [19]. These models are useful where the ecosystem structure is larger, and the consumer wants could differ greatly.

Fig. 1 was generated using VOSviewer based on a bibliometric data analysis on the keywords "inventory management system" within Scopus database. The search results in 368 documents. The colour yellow shows the more recent topics. Besides those related with new technology, the new emerging topics of interest within the scope of inventory management system include sustainability, reinforcement learning, and supply chain resilience.

2.2 Inventory Management in Retail Stores

There is always the emphasis to keep high variety of SKUs, fast turnover of inventory and store inventory location with the customer in mind. It is important to understand that retail is not like manufacturing or even the healthcare sectors, where promotions are rare, seasonality is not a key factor, and the use of physical stores along with online platforms is not a requirement [20]. These factors make the inventory systems require special solutions different from the ones used in other industries.

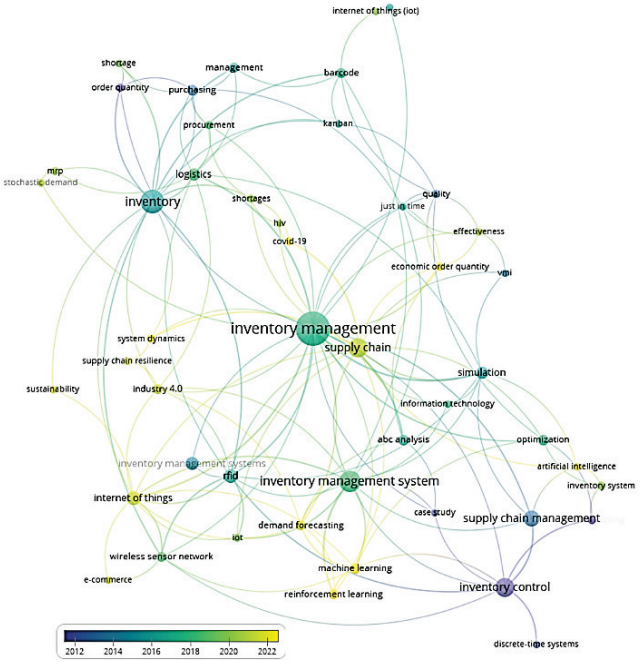


Figure 1 Overlay visualization of keywords co-occurrence

Another difference entails the aspect of quantity and quality of data they manage with regards to the velocities processed. Most retail inventory management depends on POS data and real-time consumer behaviour analysis so that correct restocking techniques are developed. For instance, some retailers such as the grocery stores may work hard to ensure that their perishable stocks are frequently replaced whereas other retail businesses such as the apparel retail stores have to focus on issues to do with high return rates and fashion trends [21]. Another example is different brand loyalty level depends on the product category of the brand being advertised. Where previously brand equity is very strong, current customer can easily seek substitute products for some brands. They would not want to spend time going to another shop in order to purchase what they regard as their own brand [22]. Such behaviour might lead to rising of the number of factors influencing the complexity of the inventory control policy.

The emergence of omnichannel only serves to complicate retail inventory management even more at this stage. Physical and inventory management needs to be in harmony with the digital platform, that requires a strong infrastructure that supports same day updates and cross-channel supply chain execution [23]. Such a high level of integration is not particularly essential in industries like manufacturing as inventory movements are comparatively more efficient and thus more easily predictable.

2.3 Research on Simulation in Retail Inventory Management

Inventory management simulation models are used substantially in retail to forecast the consequences of the

corresponding decisions. DES can model and predict different and unstable characteristics of retail operations such as demand and supply changes, and therefore DES especially discrete event simulation is capable of modelling retail operations accurately. These models come in handy because they help retailers to conduct experiments with actual and virtual operations, while not necessarily affecting actual retail stores business.

In the recent past, the use of simulation in the handling of uncertainties has been a focus in many papers. For instance, stochastic models with DES can improve understanding of how to mitigate such risks as the demand shocks, or supply chain disruptions [24]. It is crucial to increase knowledge regarding technologies that facilitate overall business strategy [25]. This capability is especially useful for the retailers dealing with high variability in consumers' behaviour.

The literature also supports the importance of simulation for sustainable retail inventory management. Simulation-based models can evaluate how different approaches toward reducing waste and optimizing energy consumption match company inventory policies with environmentally friendly objectives [26].

2.4 Simulation Software for Inventory Management

Simulation software such as Arena, AnyLogic, and Simio are commonly used in inventory management research and practice. Arena, a leading software for DES, is lauded for its user-friendly interface and extensive library of functions, making it ideal for modelling complex retail environments [15]. However, it requires expertise for customization, which can be a limitation for small businesses.

AnyLogic offers flexibility with its ability to combine DES, system dynamics, and agent-based modelling. This hybrid approach is beneficial for analysing interconnected systems, such as inventory and customer service [27]. However, its advanced features may lead to steeper learning curves and higher costs.

Simio is noted for its graphical modelling interface, which simplifies the creation of simulations. While it is accessible for beginners, its limited scalability for large-scale models can be a drawback [28]. Each software has unique strengths, and the choice often depends on the complexity and specific requirements of the inventory system. Simulation software such as Arena, AnyLogic, and Simio are commonly used in inventory management research and practice. Arena, a leading software for DES, is lauded for its user-friendly interface and extensive library of functions, making it ideal for modelling complex retail environments [15]. However, it requires expertise for customization, which can be a limitation for small businesses.

AnyLogic offers flexibility with its ability to combine DES, system dynamics, and agent-based modelling. This hybrid approach is beneficial for analysing interconnected systems, such as inventory and customer service [27]. However, its advanced features may lead to steeper learning curves and higher costs.

Much recent research on inventory management discusses about simulating and modelling the use of technology in inventory management. Simulation models was developed to compare the use of RFID and barcode in manufactures [29]. Positioning technology in the packaging of agricultural products also uses passive ultra-high frequency RFID [30]. Neuroevolution Reinforcement Learning (NERL) framework is created to minimize total system cost in the context of Multi-Echelon Inventory Optimization with Delivery Options and Uncertain Discount (MEIO-DO-UD) [31]. A simulation model was developed for the use of Unmanned Aerial Vehicle (UAV). It simulates the energy consumption difference of UAV under different the flight conditions [32]. Model-based deep reinforcement learning is developed simulating the offline and online environment conditions to optimize inventory control of a new retail product where there the historical data is insufficient [33].

2.5 Arena Simulation Software for Inventory Management

DES employing the Arena has been widely used in inventory management research because of its accuracy and flexibility. The software allows order replenishment cycles and stock out to be effectively modelled, decision makers being able to assess situations in order to come up with better strategies [34].

The key strength of Arena is the capacity to model variation nature of demand and lead time usually encountered in retail environment. Therefore, Arena models can capture real-life random characteristics by integrating stochastic features [35]. This feature makes it a good tool for use by researchers who have an interest in inventory management with an intention of improving the strategies to be used.

The last-mentioned focus on improving the usability of Arena through integration with machine learning algorithms with the possibility of a dynamic adjustment and the continuing improvement of simulation models [36]. These advancements take the application beyond merely an inventory tracking tool.

3 METHODOLOGY

The main objective of this study is to develop a better simulation model for inventory management in retail store. This study uses an existing simulation model reported in a literature as the base model [15]. The authors of the literature used as the reference studied on how the retail store responds to customers' demand with regard to the level of inventory. It includes the procedures of fulfilling customers' demand and the reorder or restock procedures.

The product studied in the reference paper is sugar. Fig. 2 represents the simulation model of inventory management system in the reference paper. It is drawn slightly differently from the one presented in the reference paper to avoid copyright issues. However, Fig. 2 shows the process flow of the simulation model exactly as it is in the reference paper.

- The process starts with customer arrival to place an order (demand). In the arena simulation model, the distribution patterns form the reference paper were used. The distribution pattern for customer arrival is –0.001 + Gamma(0.697, 0.529) while the distribution pattern for customer demand is –0.001 + Expo(51.4). These

distribution data are collected from the retail store being studied.

- After the customer place an order, the sugar inventory was checked. If the stock is smaller than the demand, then the shopkeeper will report a stockout and end the process.
- If the sugar stock is larger than the amount requested by the customer, then the transaction is completed.
- After the transaction, the shopkeeper will check whether the remaining stock is lower than the reorder point.
- If the remaining stock is higher than the reorder point, then the process ends.
- However, if the remaining stock is lower than the reorder point, the shopkeeper will place an order to the wholesaler. After a certain lead time, the order arrives at the store and the shopkeeper updates the inventory level.

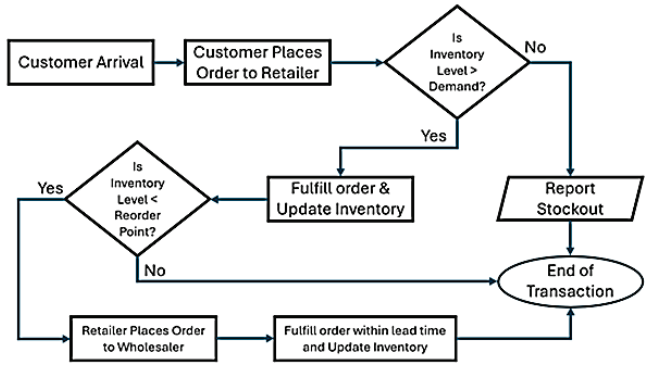


Figure 2 The simulation model of the reference paper [15]

The customer arrival pattern and demand pattern from the reference study are also used in this study. However, there are still some unknown data from the reference paper that make it impossible to re-create the exact results of the simulation model. For example, the initial value of inventory level and the delay time before reordering. This study is more concerned with simulation model improvement rather than the results. Therefore, showing an improvement in the inventory management system using the new proposed simulation model when compared to the reference model previously developed.

This study was conducted following a few steps. First, the authors studied the simulation model in the reference paper. Second, the authors interviewed a retail store owner or shopkeeper regarding the inventory management system there. Third, the new simulation model was developed. Fourth, confirm the new model to the interviewee. Fifth, develop both simulation models using Arena 16.20.03. Sixth, adjusted the parameter values to optimize each model. Seventh, compared the results.

4 RESULTS AND DISCUSSIONS

The Arena Simulation model was then developed. Fig. 3 shows the Arena Simulation model for the original model from the reference paper. It represents the same workflow as the original model in Fig. 2. There are some data needed to be placed into the variables within the model. Aside from the previously mentioned Customer arrival and demand patterns in the methodology section, the reorder point (288) and the reorder amount (667) also follow the reference paper. Both numbers come from the calculation of re-order point and economic order quantity (EOQ) performed by the authors that developed the original model in the reference paper. However, there are some data and information that are not mentioned in the reference article. For example, data such as the delay period from detecting the inventory level to reorder the stock and the initial values of the inventory level are unknown. Therefore, the simulation in this study cannot replicate the original model perfectly.

The simulation is run for 30 replications. Each replication uses 365 days without warm up period. In the reference paper, it is mentioned that the store is open from 08:00 to 20:00. Therefore, the number of hours per day in the simulation is set to 12. The delay for starting the reorder process was unknown from the reference article. In this study, it is set at a constant of 1 hour. The initial inventory level value was also previously unknown. The original model was developed to minimize the demand lost. Therefore, in this study, the authors tested the number manually until the demand lost is minimal. It is found that 50 is the initial value that creates the optimum result.

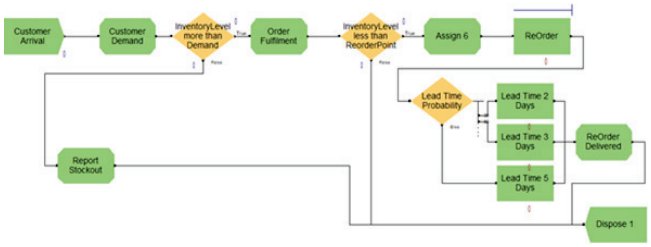


Figure 3 Arena Simulation model for the original model

The reference paper mentioned that the lead time (probability) for reorder delivery is 2 days (0.3), 3 days (0.5), and 5 days (0.2). Since it is not exactly a triangular distribution, in the Arena Simulation model, it is modelled as shown in Fig. 4. The entity passes through a probability choice first and then the lead time delay processes are placed according to the probability outcome.

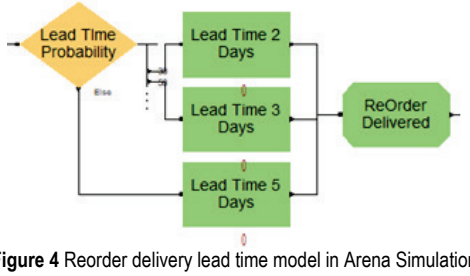


Figure 4 Reorder delivery lead time model in Arena Simulation

The authors simulated the original model of the reference paper and found a design flaw. It is when the shopkeeper checks whether the shop has enough stock to fulfil customer's demand and found out that the stock is not enough, the shopkeeper only report a stockout before ending the process. In the simulation model, the shopkeeper does not place an order to the wholesaler in the event of a stock out. It does not seem like a usual practice in a retail shop. In the

event of a stockout is reported and there is no previous order that is undelivered, the stock in the store will remain lacking. As the consequence, when the simulation software is run, the rest of customer demands will always go straight to stockout report, because the inventory is never replenished.

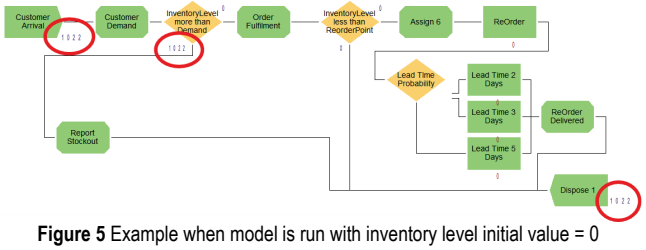


Figure 5 Example when model is run with inventory level initial value = 0

This design flaw will be directly observable when the initial value of the inventory level is set to zero. Fig. 5 shows the results when the simulation of the original model is run with the inventory level initial value is set to 0. It shows that from 1022 customers that come to the store (the left red circle), not one customer's demand is fulfilled. All 1022 is reported as stockout (the middle red circle). However, the flaw might go unnoticed if there is always undelivered order when the stock put was reported within the time limit of the simulation.

Before developing the new and enhanced model, the authors managed to interview a retail store owner. This person is a co-founder of "Toko 27" a retail shop in Semarang City. During the interview, when the authors explained the model used in the reference paper, the store owner gave important feedback. It is highly unusual for a store to not sell anything at all if the store does not have the requested amount. For example, if a customer came and wanted to buy 5 kg of sugar while there were only 3 kg in stock, it was not likely that the customer would walk away emptyhandedly. The shopkeeper usually asks if the customer wants to purchase the remaining 3 kg or not. About 80 % of the time, the customer will purchase the remaining stock, instead of leaving the store with nothing.

The new simulation model was developed taking the design flaw and offering the remaining stock into consideration. Fig. 6 shows the new simulation model. The changes made from the original model are as the following:

- It addresses the need to reorder when stockout is reported (the second decision point → report stockout → places order).
- It also addresses the process of offering the customer the remaining stock when the inventory level is not enough to fulfil the original demand (the third decision point).
- In the case of the customer accepts the offer, the sales will be processed, and the inventory will be updated. In the case of the customer declines the offer, the shopkeeper will report a stockout and then make an order to the wholesaler.
- Based on the interview, the simulation will have 80 % of the customers accept the offer.

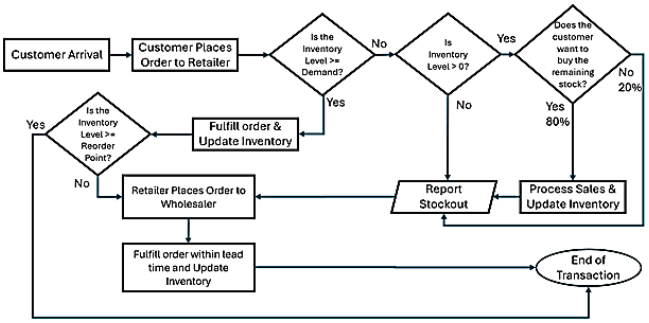


Figure 6 The new and improved simulation model

Fig. 7 shows the Arena simulation model or the new and improved model in this study. It tells the same flow as the new model in Fig. 6. All of the parameter values used in the simulation variables of the new model are the same as the ones used in the original model, including the distribution patterns, initial value of inventory level, re-order point, re-order amount, and time delays. The simulation was also run as much as 30 replications, each uses 365 days limit.

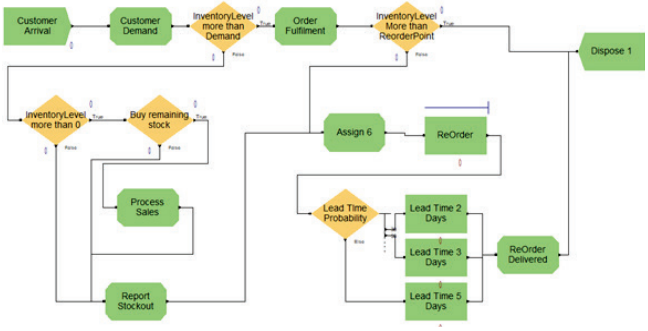


Figure 7 Arena Simulation model of the new and improved model

The next step is performing a comparison analysis. Tab, 1 shows several results of the two models as 30 replications were run in the arena simulation. The complete statistical data are shown in Tab. 2. The study shows an improvement in the service level by 1.25 % from a 26 % lower number of lost customers. It means that the new model really can reduce the number of customers that go empty-handed. The potential loss of sales is also lower with the new model by 35 %. It is caused by the additional activities of offering the customer the remaining stock instead of just sending them away and look for the good in another store. The new inventory management model maximizes the revenue of the store and arguably reduce the level of customer dissatisfaction.

Table 1 Results comparison of the two models

Parameters	Original Model	New Model	Improvement
Service Level	0.9608	0.9728	1.25 %
Lost Customer	39	29	26 %
Lost Demand	2,908.00	1,898.00	35 %
Total Number of Re-orders	79	74	6 %
Total Inventory Cost (INR)	21,11,880.50	19,78,280.50	6.33 %

Using the new model, the number of times which the store needs to re-order is also lowered by 6 %, thus reducing the annual ordering cost by INR 200 and unit cost by INR

1,33,400. Finally, the total inventory cost is also lower by 6.33 % with the new model. The total inventory cost is calculated using the values from the reference paper. The authors of the reference paper use Indian Rupee (INR) to calculate cots and price. Ordering Cost (Co) = INR 40/order, Holding Cost (Cc) = INR 3/unit and the Unit Cost (Cu) = INR 40.

The Total Annual Inventory cost is calculated using the formula $((D/Q)\cdot Co + (Q/2)\cdot Cc + D\cdot Cu)$ where D represent Demand and Q represent Volume per order (667). (D/Q) is the same as the total numbers of reorders that is shown in Tab. 1. The overall results provide evidence that the new simulation model contributes to raising the inventory management performance. The overall saving from lower cost of inventory is INR 1,33,600. Additionally, the customer satisfaction will increase due to the increase in service level and lower number of customers that leave the store empty handedly. An increase of satisfaction might deliver a positive effect to future sales from an increase of customer loyalty and providing word of mouth referral which will attract new customers.

Table 2 Statistical Data from 30 Replications of Model Simulation using ARENA Software

Repli-cation No.	Count Customer Arrival		Count Customer Lost				Count Reorder		Lost Demand	
				Enhanced Model						
	Original Model	Enhanced Model	Original Model	No Stock Lost	Refuse buy Remaining Stock	Total	Original Model	Enhanced Model	Original Model	Enhanced Model
1	919	968	12	19	0	19	77	68	1,136.92	1,483.48
2	960	982	36	39	1	40	79	75	2,562.31	2,734.54
3	1,015	950	32	24	1	25	82	64	2,903.82	1,675.61
4	977	952	36	31	2	33	71	64	2,440.21	1,820.33
5	972	964	19	27	1	28	72	68	1,482.70	1,705.35
6	941	926	30	37	1	38	70	61	1,853.57	1,728.58
7	1,004	1058	52	26	3	29	79	72	4,719.74	1,922.73
8	1,018	953	40	12	3	15	99	75	3,202.61	2,151.68
9	976	970	40	7	2	9	76	68	2,344.18	857.91
10	967	993	41	37	1	38	91	88	3,181.99	3,150.06
11	991	992	39	43	0	43	82	81	2,936.73	2,678.73
12	1,008	959	46	28	1	29	76	65	3,595.02	1,835.42
13	1,034	994	55	19	0	19	77	79	4,196.38	1,345.74
14	1,034	1073	46	46	1	47	90	84	3,715.95	2,857.12
15	1,055	1000	30	42	1	43	76	92	2,086.42	2,377.52
16	1,048	1066	54	29	1	30	84	79	3,637.29	2,176.20
17	993	1014	33	14	3	17	81	74	2,889.05	1,281.14
18	1,011	1025	37	27	1	28	80	79	2,748.64	1,822.82
19	993	1020	45	25	0	25	74	72	3,921.50	2,335.41
20	985	1043	39	28	2	30	77	86	3,174.85	1,799.37
21	967	977	34	28	0	28	88	73	2,733.42	1,699.38
22	1,007	969	47	32	0	32	82	80	3,403.68	1,682.52
23	1,006	976	38	28	0	28	80	77	2,915.09	1,589.50
24	968	1058	50	24	5	29	73	72	3,956.55	2,422.75
25	934	928	32	24	0	24	66	58	2,458.79	1,287.10
26	1,131	1053	55	26	0	26	82	83	3,758.67	1,295.62
27	954	987	58	27	2	29	73	66	3,798.90	2,669.33
28	1,036	983	27	20	0	20	81	83	1,795.27	1,425.24
29	930	960	43	24	2	26	75	65	2,634.04	1,208.69
30	986	967	21	27	2	29	76	68	1,066.61	1,910.58
Rounded Average	994	992	39	27	1	29	79	74	2,908.00	1,898.00

While the results of the new model for retail store inventory management system seem promising, the study is limited to a single dataset. The original research studied a grocery type retail stores and this study uses the data from the original research. The inventory management system model might be different for other types of retail shops. For example, it is quite rare for a customer to purchase more than one microwave in an electronic retail store. In a beauty retail shop, it is also rare that a customer purchase more than one lipstick of the same colour. Besides, in a beauty retail shop, it is easier for the seller to persuade the customer to pick another colour of lipstick the requested colour is not available. Retail stores that sell musical instruments, sport related products, or even pet shops have their own specific activities that might need a different approach in designing their inventory management system.

However, with the critical points are offering the remaining stock and not overlooking the moments to replenish the stock, there are also some types of retail stores that operate similarly with grocery stores. For example, PET food and sand for cat litters in a pet supplies store, raw materials in restaurants or cafes, office furniture retail stores, and many more. These types of stores will benefit from using the new model developed in this study.

The model in this study still has some weaknesses. First, Arena simulation software assumes the customer arrival and demand patterns as a continuous distribution pattern, while in the reality they are not. People do not visit a grocery store

and ask for 0.36 kg of sugar like they do not ask for 1.22 Liter of cooking oil. Second, for many types of goods, the distributor set a minimum order or sell the good in a package with a certain quantity. For example, a box of 12 packs of 1 kg of sugar. In this case, the reorder can only be made in multiples of 12 kg. Third, it is also unusual for a retail shop to order instantly whenever the stock is lower than the re-order amount. Unless the supplier can deliver quickly or urgently required, it make more sense if the order to supplier is made daily after closing the store. Due to these weaknesses and maybe many more, although the simulation results can offer a valuable insight for decision making in inventory management, the store manager's intuitions might also have an important role in inventory management, especially in a non-normal condition of the market.

5 CONCLUSIONS

This study enhances an inventory simulation model for retail stores, addressing a significant design flaw that hindered efficient stock replenishment. The improved model reduces lost customers by 26% and lost demand by 35%, demonstrating significant efficiency gains. These findings underscore the importance of iterative model improvements and highlight the value of simulation tools like Arena in addressing real-world challenges. This study also contributes to the literature by showing the need to review the models presented in published articles and to continuously improve the existing model.

The authors of this study acknowledge that the original inventory management model for retail stores developed by the authors of the article referred to has already contributed a lot in improving the existing system in the retail store being studied. Therefore, the improvement of the financial performance of the inventory management system from the new model may seems too little. However, the common goods in retail stores usually have a low profit margin. Having the annual inventory cost cut by 6.33% is quite significant for the retail store profitability. Therefore, the new model is beneficial for the inventory management practitioners.

The inventory management system in one company is most likely different from another. Business and management practitioners can always try to improve the existing model to get more desired results for a certain company. Future research should explore the model's application to diverse retail contexts, integrate additional factors such as customer behaviour and seasonal trends, and leverage advanced simulation technologies to further enhance inventory systems.

Acknowledgement

The authors are grateful to the shop-owner for the interview and for giving many valuable insights. Due to privacy reasons, the name of the store mentioned in this study for the interview is not the real name. Other data and information from the store owner used in this paper are as it is mentioned in the interview.

Data Availability

All the data and information needed to reproduce this study are already mentioned in this work.

Authors' Contribution

Daniel Lukito: Conceptualization, Methodology, Writing - Original Draft, Formal analysis. Andi Kusdiana: Validation, Writing - Review & Editing. Fergyanto E. Gunawan: Supervision, Review. Rida Zuraida: Supervision, Review.

6 REFERENCES

[1] Becerra, P., Mula, J. & Sanchis, R. (2022). Sustainable inventory management in supply chains: Trends and further research. *Sustainability, 14*(5), p. 2613. https://doi.org/10.3390/su14052613

[2] Fu, Q., Abdul Rahman, A. A., Jiang, H., Abbas, J. & Comite, U. (2022). Sustainable supply chain and business performance: The impact of strategy, network design, information systems, and organizational structure. *Sustainability, 14*(3), p. 1080. https://doi.org/10.3390/su14031080

[3] Heniey, H., Gebrehiwot, K., Desta, T. & Gebrehiwot, L. (2022). Optimization of Inventory Controlling System Using Integrated Seasonal Forecasting and Integer Programming. *Journal of Optimization in Industrial Engineering*, 1(1), p. 57.

[4] Tian, X., Wang, H. & Erjiang, E. (2021). Forecasting intermittent demand for inventory management by retailers: A new approach. *Journal of Retailing and Consumer Services*, 62, p. 102662. https://doi.org/10.1016/j.jretconser.2021.102662

[5] Isharyani, M. E., Sopha, B. M., Wibisono, M. A. & Tjahjono, B. (2024). Retail technology adaptation in traditional retailers: A technology-to-performance chain perspective. *Journal of Open Innovation: Technology, Market, and Complexity*, 10(1), p. 100204. https://doi.org/10.1016/j.joitmc.2023.100204

[6] Liu, S., Wang, J., Wang, R., Zhang, Y., Song, Y. & Xing, L. (2024). Data-driven dynamic pricing and inventory management of an omni-channel retailer in an uncertain demand environment. *Expert Syst Appl, 244*, p. 122948. https://doi.org/10.1016/j.eswa.2023.122948

[7] Hammoud, S., As'ad, R., Ben-Daya, M. & Hariga, M. (2024). Simulation-Based Optimization of Logistics Decisions under Horizontal Collaboration Following the Can-Order Policy. *Operations and Supply Chain Management: An International Journal, 17*(2), 220-235. https://doi.org/10.31387/oscm0570424

[8] Buisman, M. E., Haijema, R. & Bloemhof-Ruwaard, J. M. (2019). Discounting and dynamic shelf life to reduce fresh food waste at retailers. *Int J Prod Econ, 209*, 274-284. https://doi.org/10.1016/j.ijpe.2017.07.016

[9] Vidalakis, C., Tookey, J. E. & Sommerville, J. (2013). Demand uncertainty in construction supply chains: a discrete event simulation study. *Journal of the Operational Research Society*, 64(8), 1194-1204. https://doi.org/10.1057/jors.2012.156

[10] Moussavi, S.-E., Sahin, E. & Riane, F. (2024). A discrete event simulation model assessing the impact of using new packaging in an agri-food supply chain. *International Journal of Systems Science: Operations & Logistics, 11*(1), p. 2305816. https://doi.org/10.1080/23302674.2024.2305816

[11] M. Santos, J. A., Martins, M. S. E., Pinto, R. M. & Vieira, S. M. (2024). Towards Sustainable Inventory Management: A

Many-Objective Approach to Stock Optimization in Multi-Storage Supply Chains. *Algorithms, 17*(6), p. 271. https://doi.org/10.3390/a17060271

[12] Márquez, A. C., Del Castillo, A. C. & Fernández, J. F. G. (2020). Integrating artificial intelligent techniques and continuous time simulation modelling. Practical predictive analytics for energy efficiency and failure detection. *Comput Ind, 115*, p. 103164. https://doi.org/10.1016/j.compind.2019.103164

[13] Sridhar, P., Vishnu, C. R. & Sridharan, R. (2021). Simulation of inventory management systems in retail stores: A case study. *Mater Today Proc, 47*, 5130-5134. https://doi.org/10.1016/j.matpr.2021.05.314

[14] Sharma, A. & Khanna, P. (2020). Relevance of adopting emerging technologies in outbound supply chain: new paradigm for cement industry. *Operations and Supply Chain Management: An International Journal, 13*(2), 210-221. https://doi.org/10.31387/OSCM0410263

[15] Dutta, A., Yadav, R. K. & Muthupandi, G. (2023). The Convergence of Modern Digital Solutions with Online Retail Offers a Comprehensive Examination from Seller and Buyer Standpoints. In *IEEE International Conference on Recent Advances in Science and Engineering Technology (ICRASET2023)*, 1-6. https://doi.org/10.1109/ICRASET59632.2023.10419945

[16] İzmirli, D., Ekren, B. Y., Kumar, V. & Pongsakornrungsilp, S. (2021). Omni-chanel network design towards circular economy under inventory share policies. *Sustainability, 13*(5), p. 2875. https://doi.org/10.3390/su13052875

[17] Ghodake, S. P., Malkar, V. R., Santosh, K., Jabasheela, L., Abdufattokhov, S. & Gopi, A. (2024). Enhancing Supply Chain Management Efficiency: A Data-Driven Approach using Predictive Analytics and Machine Learning Algorithms. *International Journal of Advanced Computer Science & Applications, 15*(4). https://doi.org/10.14569/ijacsa.2024.0150469

[18] Fildes, R., Ma, S. & Kolassa, S. (2022). Retail forecasting: Research and practice. *Int J Forecast*, 38(4), 1283-1318. https://doi.org/10.1016/j.ijforecast.2019.06.004

[19] Maheshwari, P., Kamble, S., Pundir, A., Belhadi, A., Ndubisi, N. O. & Tiwari, S. (2021). Internet of things for perishable inventory management systems: an application and managerial insights for micro, small and medium enterprises. *Ann Oper Res*, pp. 1-29. https://doi.org/10.1007/s10479-021-04277-9

[20] Shah, N., Chaudhari, U. & Jani, M. (2019). Inventory control policies for substitutable deteriorating items under quadratic demand. *Operations and Supply Chain Management: An International Journal*, 12(1), 42-48. https://doi.org/10.31387/oscm0360221

[21] Zhou, Y.-W., Zhang, X., Zhong, Y., Cao, B. & Cheng, T. C. E. (2021). Dynamic pricing and cross-channel fulfillment for omnichannel retailing industry: An approximation policy and implications. *Transp Res E Logist Transp Rev, 156*, p. 102524. https://doi.org/10.1016/j.ejor.2020.01.011

[22] Burgos, D. & Ivanov, D. (2021). Food retail supply chain resilience and the COVID-19 pandemic: A digital twin-based impact analysis and improvement directions. *Transp Res E Logist Transp Rev, 152*, p. 102412. https://doi.org/10.1016/j.tre.2021.102412

[23] Lukito, D., Suharnomo, & Perdhana, M. S. (2023). Investigating the Relationship of Change Leadership, Knowledge Acquisition, and Firm Performance in Digital Transformation Context. *Calitatea, 24*(194), 286-295. https://doi.org/10.47750/QAS/24.194.32

[24] Abbasi, S., Taghizade, K. & Noorzai, E. (2020). BIM-based combination of takt time and discrete event simulation for implementing just in time in construction scheduling under constraints. *J Constr Eng Manag, 146*(12), p. 04020143. https://doi.org/10.1061/(ASCE)CO.1943-7862.0001940

[25] Gu, Y. & Kunc, M. (2020). Using hybrid modelling to simulate and analyse strategies. *Journal of Modelling in Management, 15*(2), 459-490. https://doi.org/10.1108/JM2-02-2019-0039

[26] Koulouris, A., Misailidis, N. & Petrides, D. (2021). Applications of process and digital twin models for production simulation and scheduling in the manufacturing of food ingredients and products. *Food and Bioproducts Processing, 126*, 317-333. https://doi.org/10.1016/j.fbp.2021.01.016

[27] Shinde, A., Haghnevis, M., Janssen, M. A., Runger, G. C. & Janakiram, M. (2023). Scenario analysis of technology products with an agent-based simulation and data mining framework. *International Journal of Innovation and Technology Management, 10*(05), p. 1340019. https://doi.org/10.1142/S0219877013400191

[28] Bajjou, M. S. & Chafi, A. (2021). Lean construction and simulation for performance improvement: a case study of reinforcement process. *International Journal of Productivity and Performance Management, 70*(2), 459-487. https://doi.org/10.1108/IJPPM-06-2019-0309

[29] Greasley, A. & Edwards, J. S. (2021). Enhancing discrete-event simulation with big data analytics: A review. *Journal of the Operational Research Society, 72*(2), 247-267. https://doi.org/10.1080/01605682.2019.1678406

Authors' contacts:

Daniel Lukito
(Corresponding Author)
Industrial Engineering Department, BINUS Graduate Program,
Master of Industrial Engineering, Bina Nusantara University,
Jl. K. H. Syahdan No. 9, Kemanggisan, Palmerah, Jakarta 11480, Indonesia
daniel.lukito@binus.ac.id

Andi Kusdiana
Industrial Engineering Department, BINUS Graduate Program,
Master of Industrial Engineering, Bina Nusantara University,
Jl. K. H. Syahdan No. 9, Kemanggisan, Palmerah, Jakarta 11480, Indonesia

Fergyanto E. Gunawan
Industrial Engineering Department, Bina Nusantara University,
Jl. K. H. Syahdan No. 9, Kemanggisan, Palmerah, Jakarta 11480, Indonesia

Rida Zuraida
Industrial Engineering Department, Bina Nusantara University,
Jl. K. H. Syahdan No. 9, Kemanggisan, Palmerah, Jakarta 11480, Indonesia

ISSN 1846-6168 (Print), ISSN 1848-5588 (Online)
https://doi.org/10.31803/tg-20250106193820

Original scientific paper
Received: 2025-01-06, Accepted: 2025-01-08

Porosity Assessment of Melt Electrospun PLA

Emilija Zdraveva

Abstract: Melt electrospun materials are by far less explored than solution based ones, due to the challenges that come from both temperature and electrical voltage simultaneously. Within this respect, the fibers produced by melt electrospinning are much thicker than 1 micron. This paper focuses on the design of polylactic acid (PLA) microfilament materials melt spun on Spraybase, with previously generated 2D models in three different geometries. The PLA materials have difference in terms of filament densities (distance between the filaments of 0.5, 1 and 2 mm) and filament positions (vertical, vertical/horizontal, diagonal or at the angle less than 90°). The porosity of the materials is determined using microscope images of the samples and by analyzing their pore sizes and total void volume in ImageJ. The porosity is an important parameter that affects the thermal comfort evaluation of materials and it is dependent upon their e.g. thickness, filament diameter, filament distribution or geometry and pore sizes.

Keywords: electrospun; melt; microfilaments; PLA; porosity; 2D models

1 INTRODUCTION

1.1 Melt Electrospinning Direct Writing

Melt electrospinning is the process of microfibers (up to 500 microns) formation with the aid of high voltage power supply that stretches a polymer melt. The production of nanofibers via melt electrospinning is more challenging than in solution electrospinning due to the absence of a solvent that suppresses the bending instabilities and jet thinning [1]. However fine fibers can be formed with careful control of the temperature that is the viscosity and solidification of the filaments, the electrical field, as well as by the increase of the electrical conductivity through targeted additives [2, 3]. The polymer parameters that will also affect the final fiber diameter are not to be omitted. These include: 1) electrical conductivity, 2) molecular weight, 3) tacticity and 4) extensional viscosity [4]. As reported, electrical conductivity is the dominant factor that affects the jet stability that is, a semi-conductive polymer will be ideal for a stable jet formation. Jet instability is present when electrospinning highly conducting materials or the jet can't be initiated due to low stretching forces in case of insulating materials [4, 5]. Stronger stretching forces were generated due to the increased jet electrical charges when sodium oleate was added to the polymer melt [6]. The jet stability also depends upon the polymer molecular weight which is related to its viscosity, modulus and tensile strength. Lower molecular weight will also result in thinner fibers [4, 7]. It is also suggested that more oriented molecular structure that is higher tacticity polymer melts will produce finer fibers [4, 8]. The polymer melt extensional viscosity is directly related to the device set temperature or the polymer melting point. In the usual non-isothermal electrospinning conditions the fiber diameter is restricted of further thinning when the jet flies far away from the spinneret and solidifies due to temperature gradual reduction [9].

The need to control the electrospun structure has emerged a process that combines melt electrospinning and 3D printing. Melt electrospinning direct writing is the process where the melt polymer jet is deposited through an automated translating collection system in order to form highly organized fibrous architecture. These polymer melt jets are drawn into continuous lines according to previously designed geometry and can be laid in the third direction to form a 3D structure [10]. It was reported that the excellent control of the microfibers layout and interspacing can result in the production of aligned, crimped and random fibrous materials as well. Initially the application field was biomedicine for the development of 3D tissue scaffolds [11], but these materials have huge potential due to the controlled porosity and fibers/pores sizes.

In this study melt electrospun materials with three different target geometries were prepared through direct writing electrospinning. The same were evaluated in terms of total porosity and pore size as the dominant factors that will affect i.e. materials thermally insulating performance or in case of chemically protective clothing where the breathability is directly related to the wearer comfort.

2 EXPERIMENTAL PART

2.1 Materials and Methods

The polymer used to produce the melt electrospun materials was polylactic acid (PLA), Luminy L175 with density of 1.24 g/cm^3, purchased from Total Corbion. The electrospun materials were produced on a melt-electrospinning direct writing device Spraybase, Avectas Ltd., Ireland.

The processing parameters were as follows: 1) electrical voltage of 11 kV, 2) temperature of 200 °C, 3) head nozzle to collector distance of 1.2 cm and 4) air pressure of 11 bars. Three groups of materials were produced depending on the filament density and filament position. In the first group, the filaments were positioned vertically with several perpendicular reinforcements, in the second group the filaments were positioned vertically and horizontally at the angle of 90º, while in group three there were additional positions at an angle smaller than 90º, Fig. 1. The filament density concerns the distance among each filament and it was set to 0.5, 1 and 2 mm, except for group three where materials with only 1 and 2 mm filaments distances were produced. Prior melt-electrospinning, 2D models representing the three

groups of materials' geometry were designed using the device software, SEL generator.

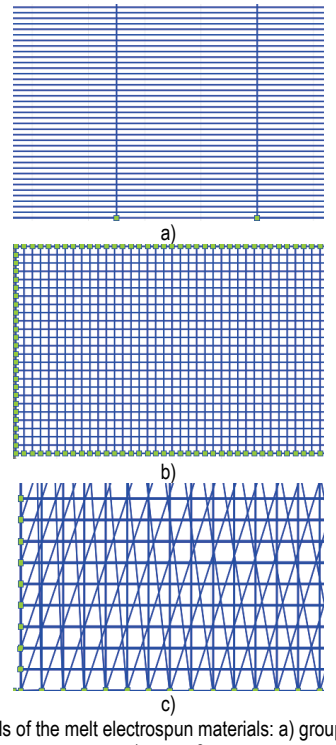


Figure 1 2D models of the melt electrospun materials: a) group 1, b) group 2 and c) group 3.

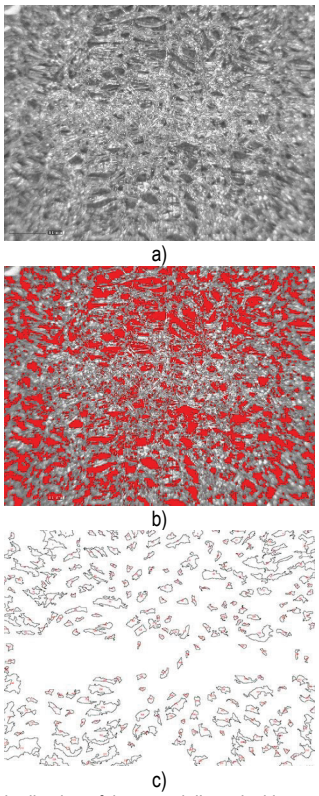


Figure 2 Threshold adjusting of the material's optical image: a) 8-bit image, b) threshold adjusting and c) marked pores to be measured; Sample group 1, filament density of 0.5 mm.

2.2 Evaluation of Melt Electrospun Materials' Porosity

Melt electrospun materials' porosity was evaluated in ImageJ software by analyzing their images taken with DinoLite optical microscopy. The procedure to evaluate the porosity includes several steps: 1) 8-bit image selection, 2) image scale setting, 3) threshold adjustment – red pores setting, 4) cutting off pores connections or filling holes if any, 5) analyze particles, 6) automatic mean pore area measurement and porosity calculation by using the total pores area and the total image area. The samples were analyzed in triplicates. Fig. 2 shows the threshold setting of one of the samples taken images.

3 RESULTS AND DISCUSSION

Fig. 3 shows the optical images of the three groups of electrospun materials. Generally, all samples have curly filaments or are forming continuous loops despite the straight lines of the 2D models. This suggests that during electrospinning the combination of the two dominant parameters that is the high temperature and the high electrical voltage resulted in the melt jet instability, thus the formation of the loops. The regular position (in straight lines) of the looped filaments is easily detected in the samples of group 1 and 2, that is samples with the filament density of 1 and 2 mm, Figs. 3a2, 3a3, 3b2 and 3b3, as well as of 0.5 mm, but only in case of group 2, Fig. 3b1. The regularity of the filament positioning is greater vertically than in the perpendicular direction, where certain deformations of the loops can be observed. Figs. 3b3, 3a1, 3c1 and 3c2, show almost random positioning of the microfilaments, which suggests that the lowest filament density of 0.5 mm, Fig. 3a1, as well as the diagonal positioning of the filaments, Fig. 3c1 and c2, affect their overlaying and loop extensions or deformations.

Fig. 4 shows the total porosity vs. mean pore area of the melt electrospun materials. The given results are in accordance with the images of the samples in Fig. 3. The total calculated porosity of the electrospun materials in group 1 is increasing with the increase of the filament density. This is expected as the increasing distance between the filaments increases the void volume thus, materials' total pore area. The significant increase (the max observed as well) of the total electrospun materials porosity is evident in case of sample 3 - group 1, that is of (64 ± 3) %. This material has the highest pore area of (394 ± 57) μm^2 of all groups as well. The same increasing trend for both porosity and pore area is present in group 2 and 3 with the increase of the two parameters for 22 % (118 μm^2) and 17 % (55 μm^2), respectively.

When comparing the three groups of materials an opposite trend can be seen thus the increasing filament density reduces the porosity/pore area and it is most significant again in case of the 2 mm density. This is due to the deposition of the filaments on top of each other in case of group 2 and 3 when the same are laid perpendicularly and diagonally which eventually results in the closure of surface pores.

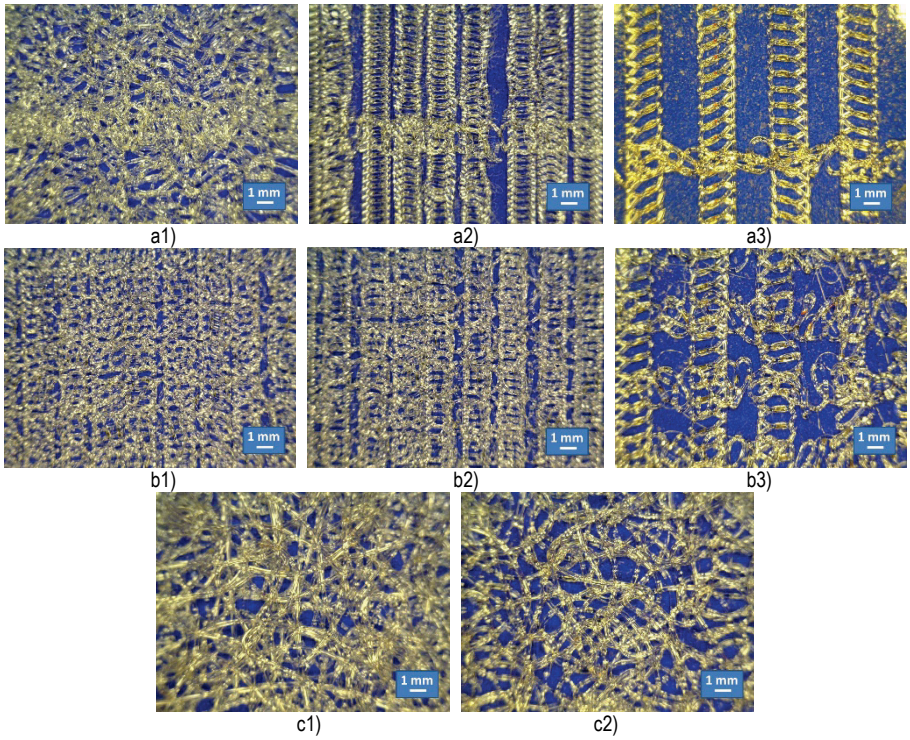


Figure 3 Optical microscopy images of the three groups of samples: a1) group 1 – 0.5 mm, a2) group 1 – 1 mm, a3), group 1 – 2 mm, b1) group 2 – 0.5 mm, b2) group 2 – 1 mm, b3), group 2 – 2 mm, c1) group 3 – 1 mm and c2) group 3 – 2 mm.

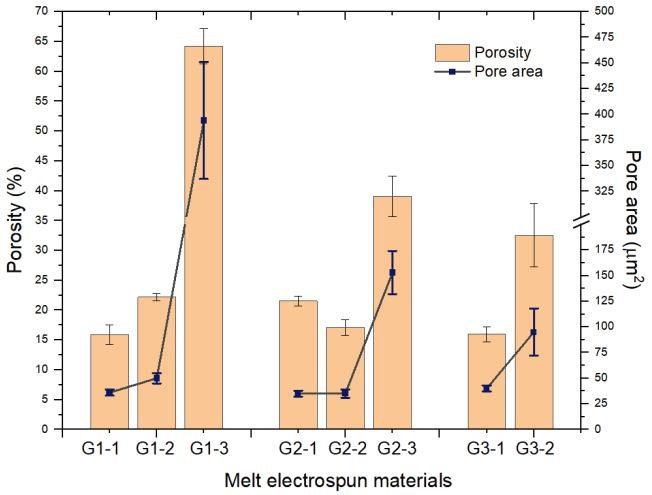


Figure 4 Total porosity vs. mean pore area of the melt electrospun materials.

4 CONCLUSION

In this paper, three groups of materials were prepared via melt electrospinning direct writing having differences in their structure in terms of filament density (filament distances of 0.5, 1 and 2 mm) and position (horizontal, horizontal/vertical and horizontal/vertical/diagonal). The materials' total porosity and mean pore area were evaluated using ImageJ and their optical microscopy images. It was revealed that the higher the filament density was, the greater the total porosity and mean pore area were observed. Significant increase of these parameters was present in case of the samples of all groups with the filament distances of 2 mm, which is due to the greater void area formed during melt electrospinning. The deposition of the second and third layer in case of the materials of group 2 and 3 resulted in the reduction of the two parameters with the less significant effect in case of the 1 mm filament density. By optimizing the porous structure, one can design i.e. breathable and thermally protective clothing.

5 REFERENCES

[1] Mitchell, S. & Sanders, J. (2006). A unique device for controlled electrospinning. *Journal of Biomedical Materials Research Part A, 78*(1), 110-120. https://doi.org/10.1002/jbm.a.30673

[2] Domingos, M., Chiellini, F., Gloria, A., Ambrosio, L., Bartolo, P. & Chiellini, E. (2012). Effect of process parameters on the morphological and mechanical properties of 3D bioextruded poly (ε-caprolactone) scaffolds. *Rapid Prototyping Journal, 18*(1), 56-67. https://doi.org/10.1108/13552541211193502

[3] Nayak, R., Kyratzis, I. L., Truong, Y. B., Padhye, R. & Arnold, L. (2012). Melt-electrospinning of polypropylene with conductive additives. Journal of Material Science 47:6387-6396. https://doi.org/10.1007/s10853-012-6563-3

[4] Brown, T. D., Dalton, P. D. & Hutmacher, D. W. (2016). Melt electrospinning today: an oppportune time for an emerging polymer process. *Progress in Polymer Science, 56*,116-166. https://doi.org/10.1016/j.progpolymsci.2016.01.001

[5] Lyons, J. & Ko, F. (2005). Melt electrospinning of polymers: A review. *Polymer News, 30*(6), 170-178. https://doi.org/10.1080/00323910500458666

[6] Nayak, R., Padhye, R., Kyratzis, I. L., Bach Truong, Y. & Arnold, L. (2013). Effect of viscosity and electrical conductivity on the morphology and fiber diameter in melt electrospinning of polypropylene. *Textile Research Journal, 83*(6), 606-617. https://doi.org/10.1177/0040517512458347

[7] Chung, C. I. (2010). *Extrusion of polymers: theory and practice*. 2nd ed. Hanser Verlag, Munich.

Lyons, J., Li, C. & Ko, F. (2004). Melt-electrospinning part I: processing parameters and geometric properties. *Polymer 45*(22), 7597-603. https://doi.org/10.1016/j.polymer.2004.08.071

[8] Zhmayev, E., Zhou, H. & Joo, Y. L. (2008). Modeling of non-isothermal polymer jets in melt electrospinning. *Journal of Non-newtonian Fluid Mechanics, 153*(2-3), 95-108. https://doi.org/10.1016/j.jnnfm.2007.11.011

[9] Brown, T. D., Dalton, P. D. & Hutmache, D. W. (2012). Direct Writing By Way of Melt Electrospinning. *Advanced Materials, 23*(47), 5651-5657. https://doi.org/10.1002/adma.201103482

[10] Gwiazda, M., Kumar, S., Świeszkowski, W., Ivanovski, S. & Vaquette, C. (2020). The effect of melt electrospun writing fiber orientation onto cellular organization and mechanical properties for application in Anterior Cruciate Ligament tissue engineering. *Journal of the Mechanical Behavior of Biomedical Materials, 104*, 103631. https://doi.org/10.1016/j.jmbbm.2020.103631

Author's contacts:

Emilija Zdraveva, Assistant Professor, dr. sc.
University of Zagreb, Faculty of Textile Technology,
Prilaz baruna Filipovića 28a, 10000 Zagreb, Croatia
emilija.zdraveva@ttf.unizg.hr

ISSN 1846-6168 (Print), ISSN 1848-5588 (Online)
https://doi.org/10.31803/tg-20231220161257

Preliminary communication
Received: 2023-12-20, Accepted: 2024-03-15

Advancing Image Forensic: Detecting Facial Manipulations via Meso_KNN

Hazem Issa, Bayan Zarnahji*

Abstract: Advancements in the fields of computer vision and deep learning have enabled the creation of highly realistic images, especially in generating human faces with an unprecedented level of realism. However, the misuse of these capabilities, such as in creating malicious content, has made image manipulation one of the most significant challenges in our daily lives. Therefore, it has become essential to develop innovative methodologies to distinguish between genuine and computer-generated multimedia, which continuously improves in terms of quality and realism. As a result, an effective model has been developed using deep learning techniques, relying on the deep neural network known as Meso Net and the K-nearest neighbors algorithm. This model, referred to as Meso_KNN, is presented in this paper. What distinguishes this model is its focus on important features in facial images that represent vital characteristics of facial manipulation. Additionally, it harnesses the capabilities of K-nearest neighbors for classification, achieving outstanding efficiency in detecting various types of facial manipulation. The model has been tested on a diverse set of facial images collected in the HFF dataset and has achieved an accuracy rate of up to 100 %. It stands as one of the current leading results in this field.

Keywords: deep learning; face image manipulation; machine learning; Meso Net

1 INTRODUCTION

Since the advent of digital visual media, there has always been a need to manipulate them for various purposes. Initially, manipulating media required expertise and consumed a significant amount of time and effort. With the advancement in computer technology and image processing software, modifying images, especially faces, has become much easier, enabling manipulations that were not possible in the past. Nowadays, anyone can create a face that looks realistic but is not from the real world, even without any prior experience in this field, thanks to the advancements in deep learning techniques and their capacity to produce high-resolution images using different versions of Variational AutoEncoders (VAEs) [1] and Generative Adversarial Networks (GAN) [2]. Despite being used in various applications such as gaming and filmmaking, their use in fabricating fake news, spreading it, and manipulating public opinion has become a tangible and significant threat to media information integrity. This could lead to harmful consequences. Deepfake, known for its ability to create and manipulate facial appearance (features, identity, and expressions) through deep methods, can be classified into the following 4 categories: 1) Entire Face Synthesis; 2) Identity Swap; 3) Expression Swap; 4) Attribute Manipulation [3]. The different manipulations of Deepfake present varying levels of risk. Entire Face Synthesis creates high-quality fake images of entirely imaginary individuals [4]. Identity swap involves replacing one person's face with another's; both techniques have the ability to change crucial personal information [5]. Manipulating facial features involves editing aspects like hair and skin colour, gender, age, and accessories, translating from image to image. Expression swap refers to creating images with specific facial expressions or replacing one person's facial expressions with another person's [6]. Techniques like identity swap replace the face of one person with another's, making actors appear in videos they never participated in; the face is replaced by exploiting comprehensive and adaptive facial information, with FaceSwap being one of the most famous applications. Some methods focus on whole-face synthesis, like PGGAN [7] and StyleGAN [8], aiming to produce and control highly detailed images up to 1024 × 1024, making it challenging to distinguish between real and fake. Expression Swap techniques propose generative models to create fake facial images without leaving any tangible traces, such as Glow [9] and GANination [10], using the effect of realistic images. Face2Face, relying on computer graphics (CG), animates facial expressions for the target video with facial expressions from the source face [11]. As for manipulating facial features based on GAN, applications like StarGAN [12] and AttGAN [13] can enhance feature editing. As shown in Fig. 1, in some examples of facial images, any user can now produce realistic synthetic faces that are challenging for humans to evaluate as real or fake. The impact that intentionally altered and maliciously used facial images have on people has made facial image manipulation detection an important problem in the field of Image Forensics.

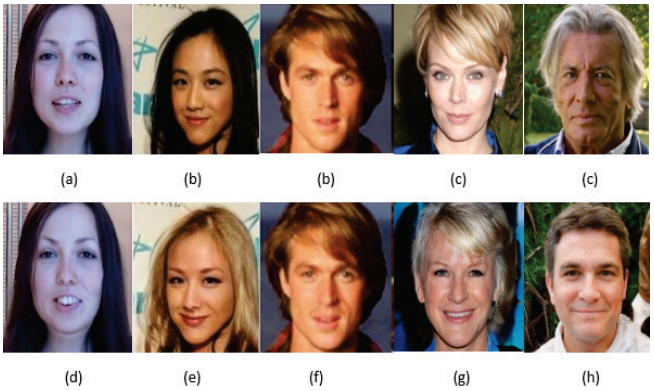


Figure 1 Example faces HFFD: Real images/frames from a) FaceForensics, b) CelebA, and c) CelebA-HQ datasets; Fake images generated by d) Face2Face, e) StarGAN, f) Glow, g) PGGAN and h) StyleGAN, respectively.

The purpose of this paper is to design an effective methodology for detecting various facial image manipulation techniques based on the Meso Net neural network and the K_nearest_neighbours' algorithm (Meso_KNN).

The Meso-4 model is an integrated model based on deep neural networks that perform well in classifying facial

images to detect facial manipulation [14]. Its main feature is the extraction of essential facial features for manipulation without any human intervention. It has a simple structure and is scalable and modifiable. The Meso model involves input layers, hidden layers, and an output layer. It is combined with K_nearest_neighbors in the output layer to improve classification accuracy. KNN is a fast and easy-to-use algorithm used for binary and multiclass classification and is usually employed to enhance system performance. Thus, the proposed hybrid model, Meso_KNN, consists of the Meso model and the KNN classifier connected in the Meso model's output layer. The following is a summary of the proposed model's contributions:

1) Proposing a composite model like Meso_KNN for detecting various facial image manipulation techniques based on Meso Net and the KNN classifier.
2) Achieving good results using the KNN classifier.
3) Comparing the results with other classification models on the same dataset, demonstrating that the final model performs highly in accuracy.

The paper's remaining sections are arranged as follows: presents related studies in Section 2, introduces the Meso_KNN model in Section 3, in Section 4 provides experimental results, conclusions, and future work in Section 5.

2 RELATED WORKS

Previous studies in the field of face manipulation detection have varied in their approaches, ranging from traditional techniques to modern artificial intelligence-based methods. These methods include the analysis of digital images and the use of neural networks to achieve more precise discrimination between real and manipulated images. In this paragraph, we will discuss some studies focusing on the techniques employed in face manipulation detection. Neural networks have been widely used to detect DeepFake manipulations. Li et al. (2020) utilised X-rays to identify the boundaries of manipulated faces; however, this method struggled to detect random noise and showed decreased performance with low-resolution images [15]. Dang et al. (2020) worked on identifying manipulated facial regions by estimating the attention map of the image [16]. They successfully detected visual manipulations; however, estimating the attention map in an unsupervised manner posed challenges. Models based on trainable convolutional neural networks (CNN) have demonstrated their ability to classify and recognise images, differentiating between manipulated and real images [17]. Transfer learning techniques from deep models like ResNet50 and VGG16 were integrated with CNN to enhance the model's performance [18]. In order to extract features related to manipulation, Afchar et al. (2018) built a model named Meso Net that included convolutional layer components that were optimised [14]. This model was characterised by its flexibility, allowing it to be modified and integrated with other networks. Guarnera et al. (2020) utilised the Expectation-Maximization (EM) algorithm to extract local features representing manipulated image effects [19]. SVM, LDA, and KNN were among the basic classifiers that effectively classified these features. These studies mainly focused on the development of models and their ability to utilise a diverse set of images for training and improving the model's performance. Therefore, we present the Meso_KNN approach, based on Meso Net and the KNN algorithm, to improve the performance of detecting different face manipulation techniques under a variety of complex and varied conditions, since the most successful methods heavily rely on CNN models.

3 PROPOSED APPROACH

Our methodology for detecting various face manipulation techniques involves analysing data based on the Meso Net and Meso_KNN models.

3.1 Meso Net Framework

Techniques that rely on analysing data at the micro, meso, and macroscopic levels are vital tools in the process of detecting facial manipulation. These techniques allow researchers to examine minute aspects and structural changes that can occur at the cellular and tissue levels. Microscopic-level analysis is used to study subtle changes on the skin surface and cells, focusing on pixel-level details to identify any alterations, while analysis at the Macroscopic-level is used to explore the biological and molecular details of facial structure manipulations. On the other hand, mesoscopic-level analysis combines the ability to detect fine changes and structural analysis simultaneously. It can identify manipulations that occur at both cellular and molecular levels accurately and comprehensively. The Meso-4 model is an efficient and integrated deep learning (DL) model designed for detecting face video forgeries. It was proposed in 2018, and the term "Meso" refers to the mesoscopic level of analysis, meaning it operates at an intermediate level for feature extraction between pixel-level and high-level semantic analysis. The goal of using Meso Net is to efficiently detect manipulation, making it suitable for real-time applications and scenarios that require processing large amounts of data using relatively few layers and parameters for high model performance. Meso Net consists of four consecutive convolutional layers, each followed by a Max Pooling layer and a batch normalization layer, this is followed by two fully connected layers [14]. Each convolutional layer extracts features, also known as feature maps. These feature maps are utilised for pixel value prediction as follows in Eq. (1):

$$y_i = \sum_{i=1}^{n} (x_i * w_{i,j} + b_i), \tag{1}$$

where b_i represents the bias term for the j^{th} convolution kernel, y_i represents the feature map output by the convolutional layer, and $x_i * w_{i,j}$ represents the convolution between the ith channel of the input image x and the ith channel of the j^{th} convolution kernel in the convolution operation. The feature maps are then passed through the ReLU activation function, as given in Eq. (2):

$$r(y) = \max(y, 0). \tag{2}$$

Where it changes the negative pixel values in the feature maps to zero, providing linear activation for neurons, neurons are not activated simultaneously. In order to extract maximum values from the feature map for critical data feature analysis, the corrected feature maps are also passed through a Max Pooling layer. Finally, a Batch Normalisation layer is employed to speed up convergence and lessen network overfitting.

At last, the learned features are forwarded to the classification unit, consisting of two fully connected layers. In the first layer, the linking between deep features is learned, containing 16 neural cells and a dropout layer to reduce sample dropout and enhance network robustness. In this study, we developed the Meso Net model, adding a Leaky ReLU layer [20] after the first fully connected layer. It is used in model training to speed up the training process, being efficient, and safeguarding the system from dying out due to the ReLU problem given in Eq. (3) below:

$$r(y) = 1(y < 0)(\alpha y) + 1(y \geq 0)(y). \tag{3}$$

Where α is a relatively small constant. The last fully connected layer's neurons contain an activation function that activates the required neural cells; the softmax activation function, which is used in multi-class output layer classification problems [21], was used in this research, as given in Eq. (4) below:

$$P = \frac{e^{zi}}{\sum_{j=1}^{n} e^{zj}}. \tag{4}$$

Where e^{zi} is the exponential function to measure the input ray, e^{zj} is the exponential function to measure the output ray, and P is the probability value. The output of the FC layer is responsible for predicting the final classification, where the output is [1 × 1 × N], where N indicates the number of categories. In the iterative training process, the error function is employed to minimize the loss between the true labels of the data and the network output during the training stage to make them converge. Categorical cross-entropy loss, which is a standard cost function for classification problems, aiming to reduce the gap between the expected and actual distribution to increase accuracy, is selected as the primary cost function in this study [22]. Meso adjusts the weights repeatedly to achieve the best results, as given in Eq. (5):

$$E = -\sum_{i=1}^{n} y_i * \log(P_i). \tag{5}$$

Where y_i represents the actual distribution for sample i, P_i represents the expected distribution given by the model for sample i and E represents the loss function. In binary classification P = 2, this indicates a real face image and a manipulated face image, respectively. In multi-class classification, P denotes the number of classes that better align with each type of image manipulation, with each class representing a method of face manipulation.

Given that the fully connected layer has all its neurons connected, passing input values and their weights is necessary for model training and classification. Training occurs through an iterative process involving forward passes of fed data and backpropagation. We randomly initialize the parameters of convolutional layers, then update the convolutional filters' weights and fully connected layers at each iteration of backpropagation. In this research, we employed Adaptive moment estimation (Adam) optimizer [23] to improve this loss value. Adam is a dynamic learning rate optimization technique that incorporates both scaling and momentum. It proves to be efficient and less memory demanding when dealing with intricate tasks involving extensive data or parameters. The formula for updating weights in Adam, represented as w_t for parameter w at time step t, as given in Eq. (6):

$$w_t = w_{t-1} - \eta \frac{\hat{m}_t}{\sqrt{\hat{v}_t} + \in}. \tag{6}$$

Where $\hat{m}_t$ represents the first moment average of gradients at step t, $\hat{v}_t$ represents the second moment average of squared gradients (uncentered variance), $\in$ represents a small constant to avoid division by zero and numerical issues, and η is the step size for optimizing the cross-entropy loss value.

3.2 Meso_KNN Framework

Illustrated in Fig. 2, the Meso_KNN framework primarily comprises two primary parts: the Meso part and the KNN part. The details of each of these parts are explained below:

3.2.1 Meso Net

The Meso part as illustrated in Fig. 2 is the main component of the model. Its primary objective is to extract crucial features from the dataset. With the same components mentioned earlier, except for isolating the last layer after the weights have been tuned for layers and replacing it with a KNN classifier for classification.

3.2.2 *k*-Nearest Neighbors' Algorithm

It is a simple and effective algorithm in machine learning, primarily used for classification problems. This algorithm utilizes the entire dataset as the training set instead of splitting it into training and testing sets because KNN algorithm is working on a pre-graded data separation, is based on the concept of nearest neighbors, where new points are classified based on the classes of their neighboring points. This significantly enhances classification efficiency and training speed.

Advantages:

- No need for prior training.
- Capability to handle non-linear data.
- Detection of outliers.

The algorithm is based on the principle of calculating the distance between items and all other items, then selecting a certain number of these items (k) as neighbors. These items are the closest ones to the pixel units in the dataset,

considering the calculated distance [24]. Euclidean distance is commonly used for distance calculation, as given in Eq. (7):

$$d(p,q)=\sqrt{\sum_{i=1}^{n}(q_i-p_i)^2}. \quad (7)$$

Where n represents the total number of samples in the dataset, and p, q represents the distance between the unknown test sample and the known training sample. This function is used to calculate the similarity and dissimilarity between samples. Therefore, both real and manipulated face images match the output.

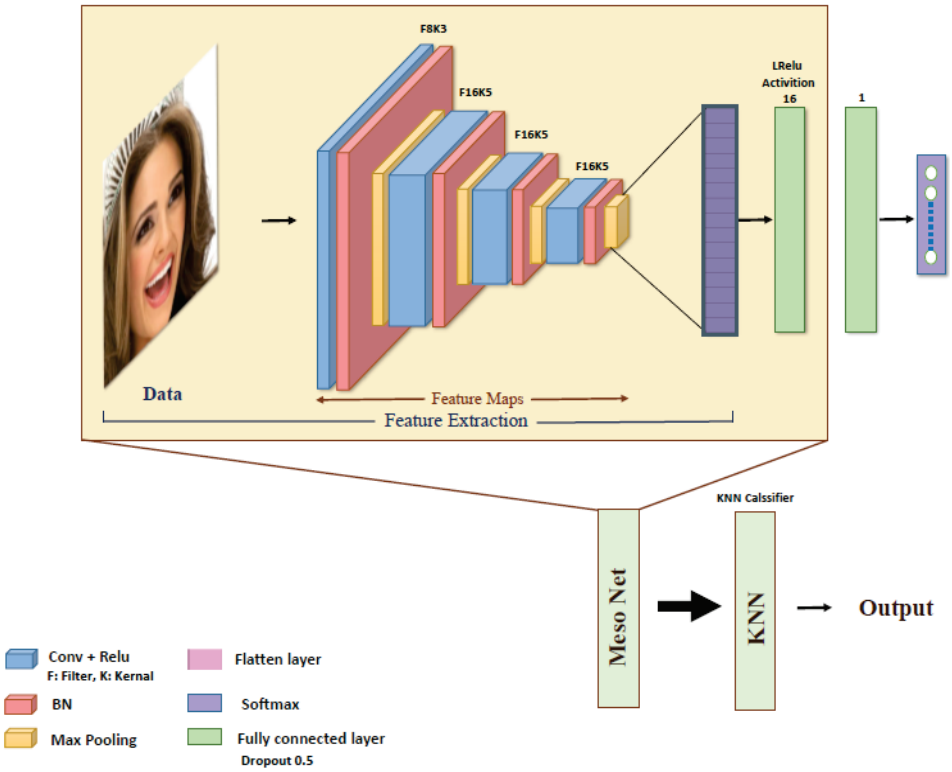


Figure 2 The Meso_KNN Proposed framework.

4 EXPERIMENTAL RESULTS

4.1 Experimental Settings

In this section, we will analyse the experimental results, evaluate the performance of our proposed model Meso_KNN and outperforms some other techniques.

4.1.1 Datasets

For the mentioned experiments, the Hybrid Fake Face Dataset (HFF) [25] dataset was used. It's a varied dataset for fake faces containing 8 types of face images. For real face images, it includes three types of images from three different open datasets. These types include low-resolution images from the CelebA dataset [26], high-resolution images from the CelebA-HQ dataset [7], and face video frames from the FaceForensics dataset [27], respectively. Thus, real face images are simulated under internet scenarios as realistically as possible. For fake face images, it includes PGGAN [7] and StyleGAN [8] for identity manipulation; it should be noted that both PGGAN and StyleGAN can generate no existing face images at a spatial resolution of 1024 × 1024; Face2Face [11] and Glow [9] for facial expression manipulation; and StarGAN [12] for transferring facial features, such as hair color and gender by multidomain image-to-image translation producing fake facial images. The HFF dataset is actually a large fake face dataset that consists of over 155K face images.

4.1.2 Implementation Details

The batch size was set to 32, and all face images in the dataset were resized to 64 × 64 for training the model. The detection accuracy on the test set was recorded after 100 epochs for model training.

4.1.3 Evaluation Criterion

The performance of these models is evaluated using Accuracy, Precision, Recall, and the F1-Score as the evaluation criteria. Accuracy in this case simply refers to how close the values of the model's predictions are to the actual (true vs. false) results. In other words, the number of times the model was able to accurately predict outcomes out of all the predictions it made. Eq. (8) shows the general formula used to calculate the prediction, where *TPR* represents the true predictions and *TOPR* represents the total predictions made by the model.

$$acc=\frac{TPR}{TOPR}. \quad (8)$$

On the other side, Precision (P) provides information about how consistent the obtained results are, even though they are not close to the actual values utilized by the target labelling. Eq. (9) illustrates the proportion of the identifications that were successfully made with respect to the actual ones. In Eq. (9), TP represents the number of true positives, and FP represents the number of false positives.

$$P = \frac{TP}{TP + FP}. \tag{9}$$

Recall (R) is the ratio of true positives correctly identified by the model to the total actual positives. Eq. (10) illustrates this ratio, where TP represents the number of true positives, and FN represents the number of false negatives. Recall intuitively captures the classifier's ability to find all positive samples [28].

$$R = \frac{TP}{TP + FN}. \tag{10}$$

The *F1-Score* works by taking both Precision and Recall into account, providing a balanced measure of a model's ability to predict both true positive and true negative instances. *F1-Score* can be interpreted as a harmonic mean of Precision and Recall. For the process of distinguishing between deep fake and real images, the *F1-Score* is the most suitable evaluation metric as both positive and negative classes are significant, and the relative contribution of Precision and Recall in *F1-Score* is better than equal weight. Eq. (11) illustrates how to calculate the *F1-Score* [28].

$$F1\text{-}Score = \frac{2(P \times R)}{P + R}. \tag{11}$$

4.2 Detection of Multiple Forgeries of Facial Image Manipulation

In our experiments, using the proposed model made it possible to simultaneously detect several facial image manipulations. The dataset contained different types of facial images that were randomly divided into two sub-groups for training (80%) and testing (20%), respectively.

In these experiments, the number of training images was approximately 124K, including five types of fake images and three types of real images with varying resolutions. For Meso_KNN model, we applied Adam optimizer for parameter estimation and the suggested default values for the moments, where the parameters $\beta_1 = 0.9$ and $\beta_2 = 0.999$ were used in Meso Net training. In addition, the learning rate was set to 0.001 in order to reach the best categorical cross-entropy loss value [29] for the Meso Net. The results shown are for a Max Epoch value of 100, displaying the training progress of the Meso Net according to the illustrated graphs Figs. 3 and 4. The first graph shows the accuracy values, while the second one illustrates the loss values during the training period.

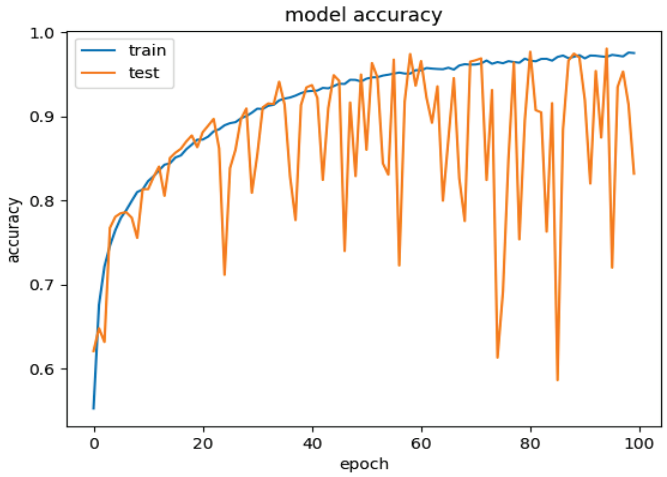


Figure 3 The variations in the accuracy of both the training and test samples during the training period of the Meso Net.

At the beginning of training, it's observed that the accuracy values start low but progressively increase at the end of the training period. Initially, the accuracy values for the test samples exceed those of the training samples, eventually converging towards the end of the training. In this study, the accuracy is reported at 98% for the training samples and 82% for the test samples for the Meso Net.

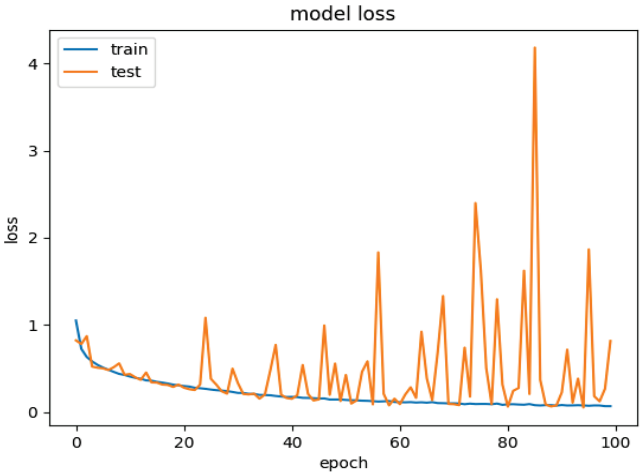


Figure 4 The variations in the loss of both training and test samples during the training period of the Meso Net.

At the beginning of training, it's observed that the loss values are relatively high, gradually diminishing at the end of the training period. Initially, the loss values of test samples are lower than those of training samples, gradually decreasing and fluctuating until they stabilize towards the end of training. In this study, the training samples achieve a loss value of 0.133, while the test samples achieve a loss value of 0.750 for the Meso Net.

A neighborhood size of k=10 and the Euclidean distance metric were adopted for KNN classifier in the Meso_KNN model. To evaluate the model's performance and its ability to separate and distinguish between categories, the following metrics were used: Precision, Recall, F1-score, and Accuracy.

Table 1 The rate of multiple classification identification achieved by the Meso Net model on HFFD calculated from the provided performance equations.

The class	Precision	Recall	F1-score
CelebA	52%	100%	68%
CelebA-HQ	96%	61%	75%
Youtube-Frame	90%	100%	94%
Glow	98%	31%	47%
StarGAN	96%	100%	98%
PGGAN	91%	71%	80%
StyleGAN	100%	95%	98%
Face2Face	100%	88%	93%

In Tabs. 1 and 2, the Precision and Recall rates for the proposed Meso_KNN model indicate that integrating KNN with the Meso architecture instead of the final fully connected classification layer had a positive impact on the model's behaviour. We evaluated the performance of the Meso Net and Meso_KNN models by creating confusion matrices in Tabs. 3 and 4, respectively. Their respective detection accuracies were 82 % and 100 %.

Table 2 The rate of multiple classification identification achieved by the Meso_KNN model on HFFD calculated from the provided performance equations.

The class	Precision	Recall	F1-score
CelebA	100%	100%	100%
CelebA-HQ	100%	100%	100%
Youtube -Frame	100%	100%	100%
Glow	100%	100%	100%
StarGAN	100%	100%	100%
PGGAN	100%	100%	100%
StyleGAN	100%	100%	100%
Face2Face	100%	100%	100%

Table 3 Confusion matrix to identify different types of manipulations using Meso Net. The asterisks "*" represent the value 0 %.

		Predicted class							
		CelebA	CelebA-HQ	Youtube -Frame	Glow	StarGAN	PGGAN	StyleGAN	Face2Face
The class	CelebA	82.18%	10.18%	0.08%	6.42%	*	0.08%	1.06%	*
	CelebA-HQ	1.85%	97.15%	*	*	0.8%	0.2%	*	*
	Youtube -Frame	0.2 %	*	99.96%	*	0.2 %	*	*	*
	Glow	72.66%	3.74%	*	23.12%	0.44%	0.04%	*	*
	StarGAN	0.1%	0.08%	0.2%	0.02%	99.76%	%0.02	*	*
	PGGAN	1.85%	96.75%	*	*	0.85%	0.4%	0.05%	0.1%
	StyleGAN	5.45%	7.85%	0.65%	*	0.05%	*	85.85%	0.15%
	Face2Face	*	*	32%	*	0.3%	*	*	67.7%

Table 4 Confusion matrix to identify different types of manipulations using Meso_KNN. The asterisks "*" represent the value 0 %.

		Predicted class							
		CelebA	CelebA-HQ	Youtube -Frame	Glow	StarGAN	PGGAN	StyleGAN	Face2Face
The class	CelebA	100 %	*	*	*	*	*	*	*
	CelebA-HQ	*	100 %	*	*	*	*	*	*
	Youtube -Frame	*	*	100 %	*	*	*	*	*
	Glow	*	*	*	100 %	*	*	*	*
	StarGAN	*	*	*	*	100 %	*	*	*
	PGGAN	*	*	*	*	*	100 %	*	*
	StyleGAN	*	*	*	*	*	*	100 %	*
	Face2Face	*	*	*	*	*	*	*	100 %

In Tab. 3, we can observe that false detection rates for Glow and CelebA are high. These two types of images share common characteristics, making it challenging for the model to distinguish, especially when resized to 64x64. Similarly, for PGGAN and CelebA-HQ, false detection rates are high for the same reason. On the other hand, the KNN classifier in the Meso_KNN model helped identify common features for every type, proving its effectiveness in identifying face manipulation.

4.3 Comparison of Performance

In this section, we compare the performance of the proposed model to that of other models by presenting experimental results on the given dataset.

Meso-4 [14]: It mainly exploits microscopic features of facial images to detect facial manipulation.

AMTEN [25]: AMTEN is used to extract facial manipulation artifacts for face manipulation detection.

Capsule [30]: It improves capsule networks to enable them to detect various types of counterfeits, such as the reuse of old data or computer-generated images and videos.

GRnet [31]: A guided residuals network utilises both spatial and residual information to better detect manipulated images.

Meso_KNN: The Meso_KNN model utilises both Meso data for analysis and feature extraction. Additionally, this model incorporates KNN classifier to enhance its efficiency and performance.

Table 5 Performance comparison for forensic models in multiple classification of different types of manipulation.

Methods	Accuracy, %
Meso-4	82
AMTEN	98.52
Capsule	96.75
GRnet	99.96
Meso_KNN (our)	100

Therefore, Tab. 5 illustrates a performance comparison between Meso_KNN and some recent works. We can notice that the proposed model shows higher accuracy compared to other models, achieving superior and top performance among current research efforts.

5 CONCLUSION AND FUTURE WORK

In this work, we proposed a novel methodology Meso_KNN for detecting manipulated images created by GANs. Meso_KNN is a robust framework based on the deep neural network Meso and the KNN classifier. Our proposed methodology is effective in image analysis and extracting important features for classifying fake and real images. Meso Net has proven its effectiveness in detecting images. Our model achieves a higher level of accuracy in detecting manipulated and real images in the HFF dataset. The results of our experiments have proven that the proposed method outperforms other techniques in terms of performance. Despite the high performance of the model on the HFF dataset, it is necessary to test the proposed model on other datasets. Although the system has been tested on RGB images, fake images using other color channels should be studied in the future, and this was not possible due to the limited resources available to us. This exploration is necessary to determine the process inputs for facial fake recognition.

6 REFERENCES

[1] Kingma, D. P. & Welling, M. (2013). Auto-encoding variational Bayes. *arXiv*. https://arxiv.org/abs/1312.6114

[2] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A. & Bengio, Y. (2014). Generative adversarial nets. In *Advances in Neural Information Processing Systems*, 2672-2680.

[3] Tolosana, R., Vera-Rodriguez, R., Fierrez, J., Morales, A. & Ortega-Garcia, J. (2020). Deepfakes and beyond: A survey of face manipulation and fake detection. *Information Fusion*, *64*, 131-148. https://doi.org/10.1016/j.inffus.2020.06.014

[4] Arjovsky, M., Chintala, S. & Bottou, L. (2017). Wasserstein GAN. *arXiv*. https://arxiv.org/abs/1701.07875

[5] Korshunova, I., Shi, W., Dambre, J. & Theis, L. (2017). Fast face-swap using convolutional neural networks. In *Proceedings of the IEEE International Conference on Computer Vision*, 3677-3685. https://doi.org/10.1109/ICCV.2017.397

[6] Liu, M., Ding, Y., Xia, M., Liu, X., Ding, E., Zuo, W. & Wen, S. (2019). Stgan: A unified selective transfer network for arbitrary image attribute editing. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 3673-3682. https://doi.org/10.1109/CVPR.2019.00379

[7] Karras, T., Aila, T., Laine, S. & Lehtinen, J. (2017). Progressive growing of gans for improved quality, stability, and variation. *arXiv* preprint *arXiv*: 1710.10196.

[8] Karras, T., Laine, S. & Aila, T. (2019). A style-based generator architecture for generative adversarial networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 4401-4410. https://doi.org/10.1109/CVPR.2019.00453

[9] Kingma, D. P. & Dhariwal, P. (2018). Glow: Generative flow with invertible 1x1 convolutions. In *Advances in Neural Information Processing Systems*, 10215-10224.

[10] Pumarola, A., Agudo, A., Martinez, A. M., Sanfeliu, A. & Moreno-Noguer, F. (2018). Ganimation: Anatomically-aware facial animation from a single image. In *Proceedings of the European conference on computer vision (ECCV)*, 818-833. https://doi.org/10.1007/978-3-030-01249-6_50

[11] Thies, J., Zollhofer, M., Stamminger, M., Theobalt, C. & Nießner, M. (2016). Face2face: Real-time Face Capture and Reenactment of RGB Videos. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 387-2395. https://doi.org/10.48550/arXiv.2007.14808

[12] Choi, Y., Choi, M., Kim, M., Ha, J. W., Kim, S. & Choo, J. (2018). Stargan: Unified generative adversarial networks for multi-domain image-to-image translation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 8789-879.

[13] He, Z., Zuo, W., Kan, M., Shan, S. & Chen, X. (2019). AttGAN: Facial attribute editing by only changing what you want. *IEEE transactions on image processing, 28*(11), 5464-5478. https://doi.org/10.1109/TIP.2019.2916751

[14] Afchar, D., Nozick, V., Yamagishi, J. & Echizen, I. (2018). Mesonet: a compact facial video forgery detection network. *In IEEE international workshop on information forensics and security (WIFS2018),* 1-7. https://doi.org/10.1109/WIFS.2018.8630761

[15] Li, L., Bao, J., Zhang, T., Yang, H., Chen, D., Wen, F. & Guo, B. (2020). Face x-ray for more general face forgery detection. *In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 5001-5010. https://doi.org/10.1109/CVPR42600.2020.00505

[16] Dang, H., Liu, F., Stehouwer, J., Liu, X. & Jain, A. K. (2020). On the detection of digital face manipulation. *In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 5781-5790. https://doi.org/10.1109/CVPR42600.2020.00582

[17] Tariq, S., Lee, S., Kim, H., Shin, Y. & Woo, S. S. (2019). Gan is a friend or foe? A framework to detect various fake face images. *In Proceedings of the 34th ACM/SIGAPP Symposium on Applied Computing*, 1296-1303. https://doi.org/10.1145/3297280.3297410

[18] Sharma, J., Sharma, S., Kumar, V., Hussein, H. S. & Alshazly, H. (2022). Deepfakes Classification of Faces Using Convolutional Neural Networks. *Traitement du Signal*, 39(3). https://doi.org/10.18280/ts.390330

[19] Guarnera, L., Giudice, O. & Battiato, S. (2020). Deepfake detection by analyzing convolutional traces. *In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, 666-667. https://doi.org/10.1109/CVPRW50498.2020.00341

[20] Dubey, A. K. & Jain, V. (2019). Comparative Study of Convolution Neural Network's Relu and Leaky-Relu Activation Functions. In: Mishra, S., Sood, Y. & Tomar, A. (eds) Applications of Computing, Automation and Wireless Systems in Electrical Engineering. *Lecture Notes in Electrical Engineering, 553*. Springer, Singapore. https://doi.org/10.1007/978-981-13-6772-4_76

[21] Bishop Christopher, M. (2006). Pattern recognition and machine learning. *Information science and statistics*, New York: Springer.

[22] Gordon-Rodriguez, E., Loaiza-Ganem, G., Pleiss, G. & Cunningham, J. P. (2020). Uses and abuses of the cross-entropy loss: Case studies in modern deep learning. https://doi.org/10.48550/arXiv.2011.05231

[23] Kingma, D. P. & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv* preprint *arXiv*: 1412.6980.

[24] Sutton, O. (2012). Introduction to *k* nearest neighbor classification and condensed nearest neighbor data reduction. *University lectures*, University of Leicester, 1.

[25] Guo, Z., Yang, G., Chen, J. & Sun, X. (2021). Fake face detection via adaptive manipulation traces extraction network. *Computer Vision and Image Understanding*, 204, 103170. https://doi.org/10.1016/j.cviu.2021.103170

[26] Liu, Z., Luo, P., Wang, X. & Tang, X. (2015). Deep learning face attributes in the wild. In *Proceedings of the IEEE International Conference on Computer Vision*. https://doi.org/10.1109/ICCV.2015.425

[27] Rössler, A., Cozzolino, D., Verdoliva, L., Riess, C., Thies, J. & Nießner, M. (2018). Faceforensics: A large-scale video dataset for forgery detection in human faces. *arXiv* preprint *arXiv*: 1803.09179.

[28] Ferreira, A., Nowroozi, E. & Barni, M. (2021). VIPPrint: A large scale dataset of printed and scanned images for synthetic face images detection and source linking. *arXiv* preprint *arXiv*: 2102.06792.

[29] Dang, M. (2022). Efficient vision-based face manipulation identification framework based on deep learning. *Electronics*, *11*(22), 3773. https://doi.org/10.3390/electronics11223773

[30] Nguyen, H. H., Yamagishi, J. & Echizen, I. (2019). Capsule-forensics: Using Capsule Networks to Detect Forged Images and Videos. *The IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP2019)*, Brighton, UK, 2307-2311. https://doi.org/10.1109/ICASSP.2019.8682602

[31] Guo, Z., Yang, G., Chen, J. & Sun, X. (2023). Exposing Deepfake Face Forgeries with Guided Residuals. *IEEE Transactions on Multimedia*, *25*, 8458-8470. https://doi.org/10.1109/TMM.2023.3237169

Authors' contacts:

Hazem Issa, PhD
Department of Computer Engineering,
College of Electrical and Electronic Engineering, University of Aleppo,
Aleppo, Syria
hazemisaa17@gmail.com

Bayan Zarnahji, Master's degree student
Department of Computer Engineering,
College of Electrical and Electronic Engineering, University of Aleppo,
Aleppo, Syria
bayanzernadjei@gmail.com

ISSN 1846-6168 (Print), ISSN 1848-5588 (Online)
https://doi.org/10.31803/tg-20240111081254

Preliminary communication
Received: 2024-01-11, Accepted: 2024-05-18

Survey Based on Edge Structured File Systems in Edge Computing

Ravula Rajesh*, Ripon Patgiri

Abstract: Edge Structured File systems (ESFs) are distributed file systems designed to provide efficient and reliable storage solutions in edge computing environments. By employing distributed and decentralized designs, these file systems address specific challenges such as limited resources, inconsistent connectivity, and fluctuating network conditions, resulting in quicker access times, reduced latency, and enhanced resilience. ESFs adapt to dynamic edge settings through flexible data placement tactics, network congestion detection and resolution, and seamless integration with cloud-based storage systems. These techniques enable data portability and, when necessary, the outsourcing of computation-intensive operations to the cloud. Overall, edge-computing ecosystems rely on ESFs to deliver optimal performance, resilience, and data availability. The article discusses several studies on edge-structured file systems, highlighting their features and limitations of previous works. Furthermore, it identifies requirements and discusses research challenges in edge computing, laying the groundwork for future advancements in this rapidly evolving field. By providing insights into the state-of-the-art technologies, features, and limitations of ESFs, as well as their broader implications for edge computing, this article aims to offer valuable guidance to researchers, practitioners, and stakeholders interested in harnessing the full potential of edge computing technologies.

Keywords: cloud computing; decentralized designs; distributed file systems; edge computing; edge devices; edge environment; edge structured file system; stakeholders

1 INTRODUCTION

Performing computation at the edge, which refers to a local resource, is the foundation of the idea of edge computing. In edge computing, the nearby resource might be a computer or network resource that is situated among the client & the cloud data centers. A few of the choices provided by Edge Computing are described in [1] and [2]. Essentially, Edge Computing, Cloudlet Computing, and Fog Computing all operate under the same concepts.

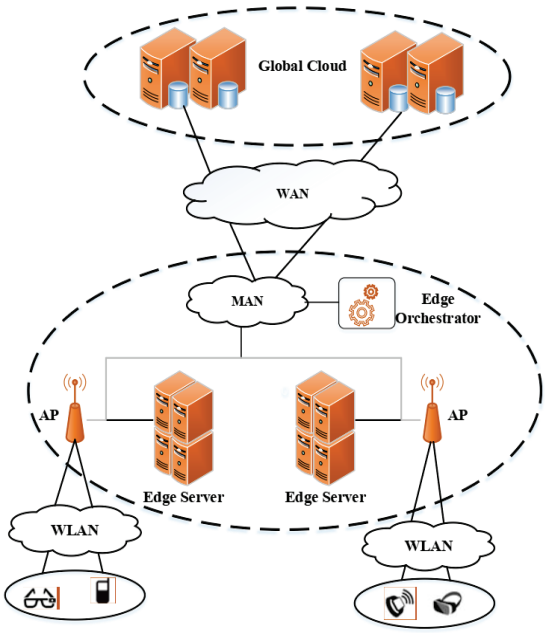


Figure 1 General Edge Computing architecture

Mobile Cloud Computing (MCC) [3, 4], Fog Computing [5], and other computing methods are all included in edge computing. To acquire a service from a nearby server located in the user's geographic area is the common purpose of several computer technologies. Overall, time-sensitive applications or situations where cloud connectivity is unavailable clearly benefit from this approach. Implementing a program or a portion of an application on the edge brings considerable benefits to mobile devices as well. To begin with, by utilizing an edge server, mobile devices may accomplish complicated processes while utilizing less battery power. Since clients do not face WAN (wide area network) delay in Edge Computing, it also has less transmission latency than cloud computing. Activities can be sent to the global cloud by mobile gadgets in two-tier architectures using the WAN connection provided by the corresponding access point. Since only the two-tier with Edge Offload (EO) model allows work performed on the initial layer to be offloaded to any edge server located in different buildings, the two-tier with EO design offers a substantial advantage. It is intended for the edge servers and orchestrators to be connected to an identical network. A general Edge Computing design is shown in Fig. 1.

An edge structure file system (ESFs) is a form of file system developed for edge computing, in which information is processed closer to the source rather than being sent to a centralized server. ESFs distinguishes itself from typical file systems by designing it for low power, resource-constrained devices encountered in edge computing environments. To reduce the quantity of storage and processing power necessary, ESFs often employs improved information structures and algorithms. Furthermore, to increase performance and security in edge computing environments, ESFs frequently adds technologies such as data compression and encryption. Overall, the purpose of a ESFs is to create a lightweight, efficient, and secure file system capable of meeting the particular requirements of edge computing settings [6]. Edge-structured storage systems cater to edge computing demands by organizing data at the edge, enabling efficient storage, retrieval, and evaluation. These systems integrate edge-optimized approaches like distributed file systems, data replication, caching, metadata management, and data placement schemes for efficient resource utilization. Fig. 2 shows the overview of edge structured file systems.

High-speed connections are used in the edge computing surroundings to connect neighboring edge servers in an area [7], creating an edge server network that doubles as an edge storage system (ESS) [8, 9, 10]. The single-point failure along with bottleneck issues with the edge-cloud structure are addressed by ESS [11]. Studies are working toward different optimization goals by caching data replicas on edge servers, including minimum data retrieval latency [12], maximum cache hit ratio [13, 14], maximum caching advantages [15, 16], and maximum caching capacity [17, 18], have recently been drawn to new issues raised by ESS.

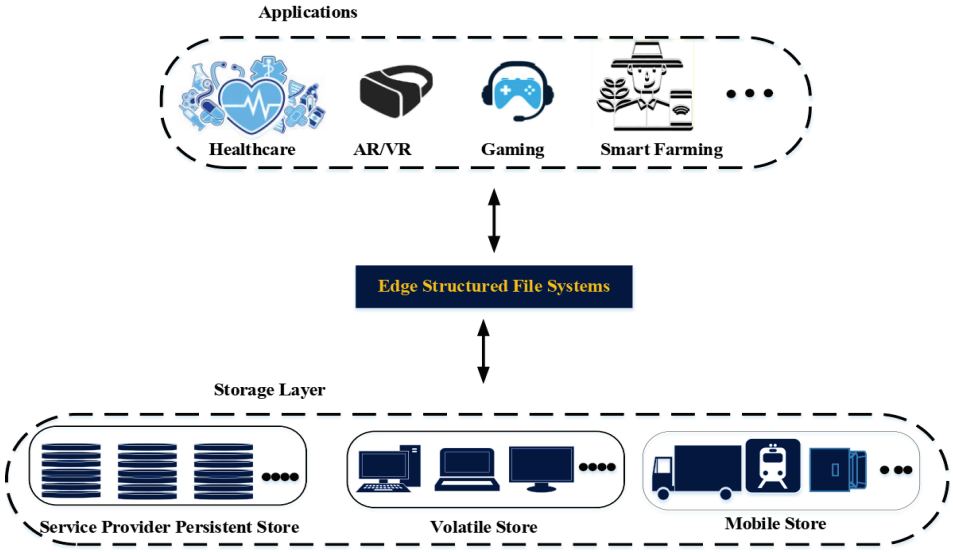


Figure 2 Overview of Edge Structured File Systems

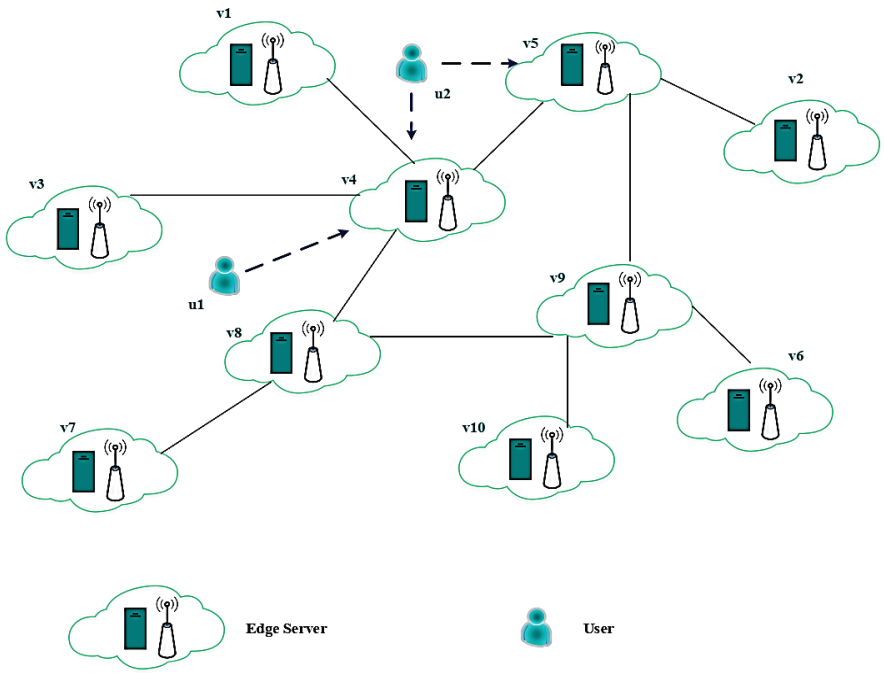


Figure 3 Example of edge storage system

Fig. 3 depicts an ESS made up of 10 networked edge servers that serve users in a specified region collectively. A basic replica-based distribution method for every of the 10 edge servers. The edge servers that protect consumers in the edge computing architecture are accessible to the consumer. Based on specific studies, the client's distance from the edge server affects data rate but not latency. As an outcome, this research considers the latencies among clients and edge servers.

This study aims to evaluate various edge-enabled storage solutions by examining cloudlet-based architectures, decentralized storage systems, and software-based edge computing platforms. It also investigates the integration of network resources into offloading decisions and the impact of network conditions on edge computing outcomes. The study also assesses the feasibility and effectiveness of utilizing existing distributed database platforms for edge infrastructure, focusing on attributes like data mobility,

scalability, and accessibility. The analysis provides insights into designing edge storage solutions that meet diverse edge computing scenarios, considering factors like energy efficiency and reliability. These studies provide a comprehensive understanding of challenges and opportunities in edge storage. The objective and contribution of this survey paper is as follows:

- The research aims to tackle offloading decisions in edge computing frameworks, focusing on cloudlet-based architectures with a centralized multi-tiered tree structure, by developing a formulation considering computing and network resources' impact on edge computing outcomes.
- The study explores edge storage, focusing on adapting distributed database platforms, using S3 file systems in fog environments, and proposing efficient frameworks. The research also discusses decentralized data storage solutions on different edge nodes with unique operating systems and file systems.
- The research explores the integration of block-chain technology with edge storage to improve security and data integrity.

The organization of this article is as listed below: The survey study's concept is described in Section 2, and the contents of the current research papers are covered in Section 3. A summary of this study and its future prospects are given in Section 4, and the paper is wrapped up in Section 5.

2 THE STRUCTURE OF THIS SYSTEMATIC LITERATURE REVIEW'S PROCEDURE

This part describes the approach used to conduct a comprehensive of the literature for edge structured file systems. It sheds insight on several areas of the review process, including details on preferred article databases, the research method or conceptual structure, criteria for adding or omitting research papers, and the distribution of articles utilized in the research.

2.1 Search Strategy

Six digital libraries—IEEExplorer, ACMDigital Library, Science Direct, Web of Science, SCOPUS, and Google Scholar—were used to find the relevant research. The most widely utilized and well-liked digital libraries for edge-structured file systems. In order to find pertinent research based on study topics like "Edge computing," "cloud computing," "Edge environment," and "edge devices," a specific database is searched using a search query string. For the research, related terms and alternatives to important words such as ‘distributed file systems’, ‘edge server’, ‘computing environments’ are evaluated. The numerous ways for achieving the objectives are also given.

2.2 Research Questions

The research questions are essential in selecting the search strategy, data extraction, and analysis in the context of a systematic literature review. The following have been chosen as the research questions for this investigation:

RQ1. Which type of edge structured file systems used in the existing literature paper?

RQ2. What are the various techniques used?

RQ3. What are the various future suggestive techniques used for edge computing?

2.3 Criteria for Article Selection Process

The inclusion and exclusion standards have been established after thorough consideration of the quality of literature already in existence. With the help of an expert advisor, the writers reviewed the publications and decided which ones to include and exclude. The following standards evaluated the papers.

Inclusion Criteria: Depending on routing protocol categorization, we issued the papers between 2014 and 2022, and they were completely searchable.

Exclusion Criteria: Papers that fail to satisfy the aforementioned inclusion requirements.

2.4 Quality Assessment

The articles acquired and reviewed by qualified professionals connected to reputable journals and conferences. No specific quality requirements were required.

2.5 Data Extraction

The qualities have been identified from existing literature and published literature reviews. The characteristics taken into account in the extraction are as follows:

- Publication year
- Methods discussed
- Edge structured file systems.

3 LITERATURE SURVEY

Routaib et al. [19] recommend a cloudlet-based framework with a centralized multi-tiered tree architecture. It might be challenging to choose where to offload in a situation like this since an edge-computing framework may consist of numerous edge servers. Previously, only computing resources were included while framing the offloading issue. Yet since network resources may utilize local edge computational units, the formulation must account for both kinds of resources.

Betzler et al. [20] discovered that network resources have a substantial impact on outcomes in edge circumstances. For example, Chen et al.'s [21] optimization of offloading choices in a multi-tier scenario where mobile clients offload faraway cloud and compute access points. The researchers employ a heuristic approach that considers both resource and interaction assessments.

Saguna et al. [22] introduced Sagunas Open-RAN, a software-based Edge Computing solution, in 2016. It is a platform that enables third-party Multi-access Edge Computing (MEC) applications via virtualized environments. DNS caching, retail-targeted video marketing, safe cities, and smart security were among the application cases for the project.

The majority of study is being done on edge storage via utilizing existing distributed database platforms and adapting it to the edge infrastructure. The utilization of an S3 file system in a fog environment is the main focus of the writers Confais et al. [23]. The following list of attributes of an edge storage framework: Data should be moved or copied from one location to another, matching client locations, and should be accessible in splitting circumstances and network confinement. It should also have high scalability to handle a large number of sites, customers, and products.

Table 1 Papers with their specifications

Authors and Reference no	Technique Used	Storage Systems	Limitations/ Future work	Published Journal
Routaib et al. [19]	Centralized multi-tiered tree architecture	Cloudlet-based storage systems	Further studies will concentrate on simulating various applications and assessing workload migration between clouds or cloudlets.	International Conference on Intelligent Systems: Theories and Applications, IEEE
Betzler et al. [20]	wireless SDN-based architecture	Heuristic approach for optimizing offloading choices in a multi-tier scenario.	Experiments on real constrained devices demonstrate successful implementation and execution of these policies.	European conference on networks and communications, IEEE
Chen et al. [21]	Heuristic approach considering both resource and interaction assessments	Multi-tier storage systems	In future, Integrate edge computing in mobile cloud architecture to optimize offloading, resource allocation, and energy efficiency by utilizing network resources near mobile devices.	International Conference on Acoustics, Speech and Signal Processing, IEEE
Saguna et al. [22]	Software-based Edge Computing solution enabling third-party MEC applications via virtualized environments	Edge computing storage systems	Explore optimal edge-server placement strategies considering user density, mobility patterns, and network traffic demands. Develop algorithms for optimal deployment locations and explore mobility management techniques for seamless connectivity.	White paper
Confais et al. [23]	Adaptation of existing distributed database platforms to edge infrastructure	S3 file systems	RozoFS reduces local reading times by 34%, but more realistic edge-Fog networking issues remain.	Transactions on Large-Scale Data-and Knowledge-Centered Systems XXXIII
Gheorghe et al. [24]	Wireless communication for file transfer, shared data with continuous Cloud backup	Distributed and decentralized storage applications	Performance is improved overwall, and experimental evaluation covers proposed attack scenarios.	Designing privacy enhancing technologies: international workshop on design issues in anonymity and unobservability Berkeley, Springer Berlin Heidelberg.
Freenet et al. [25]	Epidemic method for building decentralized storage networks	Decentralized storage systems	Address the challenge of data availability and redundancy in a distributed system. Explore replication strategies, data erasure coding, and redundancy schemes to ensure high availability of data while maintaining anonymity. Investigate mechanisms to incentivize users to contribute storage resources and maintain data replication.	Resource Management for Big Data Platforms. Springer
Wilkinson et al. [26]	Block-chain system for recording item positions, periodic integrity and availability checks	Block-chain storage systems	Develop strategies to ensure fault tolerance and resilience in the Storj network. Investigate techniques for data replication, redundancy, and erasure coding to protect against data loss and ensure high availability.	
Cohn et al. [27]	Integration of edge storage with decentralized block-chain technologies	Decentralized block-chain storage systems	Improve the scalability and performance of ADEPT systems. Develop resource allocation and energy management strategies for ADEPT systems.	Patent and trademark office
Shreshth et al. [28]	Prototype system called APEX, pre-optimized weights for low adapting time	APEX file system	The research may be expanded to include secured file-systems, wear leveling optimization, and new non-volatile memories like RRAMs for power-efficient edge devices.	International Conference on Cloud Computing Technology and Science IEEE
Jin et al. [29]	EC-EDP-O solution using integer programming and EC-EDP-V solution for addressing processing complexity	EC based storage systems and distributed storage systems	Research on reliability of data and storage expenses in EC-based edge storage will be made possible by experimental findings that show effective, cost-efficient methods.	IEEE Transactions on Services Computing
Sonbol et al. [30]	EdgeKV decentralized storage system with data replication and location-transparent architecture	Decentralized storage systems	EdgeKV can be employed with alternative replication mechanisms in the future to give weaker types of consistency for faster response times.	Journal of Parallel and Distributed Computing, Elsevier publication

A distributed, decentralized data storage solution was suggested by Gheorghe et al. [24] that will operate on a variety of edge nodes, every having a unique operating system, design, and file system. The edge nodes will use a wireless communication environment to transfer files from multiple sensors and host sharded data with continuous Cloud backup.

Freenet [25] presents an epidemic method for building decentralized storage networks that support data publication, replication, and retrieval. Without a centralized index or broadcast searches, the files regularly copied into nodes near users and discarded from nodes with low interest. Freenet allows customers to share spare disc space in a cooperative distributed network. In terms of safety, the platform ensures anonymity for the consumers who enter information and for customers who retrieve data.

Requests for an item that exists locally cannot be forwarded to the Cloud in the Edge context. Thus, Wilkinson et al. [26] employed the block-chain system to record the item positions. Because the block-chain saves all transactions that are duplicated on every node, network clients might utilize it to locate every item kept on the edge. The approach presents a challenging method built around a periodic check of a file's integrity and availability.

In Autonomous Decentralised Peer-to-Peer Telemetry (ADEPT) by Cohn et al. [27], proposed IBM and Samsung Electronics provide a proof of concept combining edge storage with decentralized Block-chain technologies. Three open-source protocols serve as the foundation of ADEPT: BitTorrent for distributed sharing of files, Telehash for message-based interactions, and Ethereum for device coordination. Ethereum-based behaviors include authentication with proximity-based norms of interaction, for instance.

Shreshth et al. [28] propose a lightweight, adaptable, portable, and efficient file allocation mechanism that is adaptable, reliable, and independent of storage design, as well as a set of pre-optimized weights that call for only a small change of hyper-parameters based on utilization resulting in low adapting time for new circumstances. In addition, they develop a prototype system of files called APEX and show how it may be used to maliciously overwrite or delete video surveillance records in real life.

Cloud computing may now store information on network edge servers, opening up new options and difficulties. Traditional replica-based edge storage techniques might be expensive. Thus, in order to ascertain the most efficient method for positioning coded data blocks, Hai Jin et al. [29] investigate the application of erasure codes for low-cost storage of data at the edge. The researchers propose an ideal solution called EC-EDP-O that utilizes integer programming and additional approximation approach named EC-EDP-V to address the high processing complexity of large-scale situations.

According to the diverse and dispersed nature of the edge along with its limited resources, it is challenging to create an efficient edge-enabled storage platform. Several distributed applications are frequently meant for the cloud. EdgeKV is a decentralized storage system designed for the network edge that Sonbol et al. [30] propose. Through data replication with solid consistency guarantees, EdgeKV offers speedy and trustworthy storage. EdgeKV's location-transparent and interface-based architecture enable it to expand with a variety of edge node platforms. At last, they look at how employing edge resources with EdgeKV instead of centralized cloud results in better energy efficiency.

Various studies explore different aspects of Edge Computing, particularly focusing on the challenges and solutions related to edge storage and decentralized systems. One study proposes a cloudlet-based structure with a centralized multi-tiered tree design, highlighting the importance of considering both computing and network resources in offloading decisions. Other research investigates the impact of network resources on edge computing outcomes and presents software-based solutions for Edge Computing, such as Saguna's Open-RAN platform. Additionally, there's a focus on edge storage solutions, including the adaptation of existing distributed database platforms, the use of decentralized storage applications, and epidemic methods for building decentralized storage networks. These solutions aim to address issues like data locality, scalability, and data accessibility in partitioning situations.

Furthermore, some studies explore the integration of edge storage with block-chain technologies for enhanced data integrity and availability. Proposed systems like ADEPT combine edge storage with decentralized block-chain technologies for telemetry applications. Moreover, researchers propose novel file allocation systems optimized for post-deletion recovery and efficient utilization of resources. They also investigate the application of erasure codes for cost-effective data storage at the edge. Finally, there's a proposal for EdgeKV, a decentralized storage system designed for the network edge, offering quick and reliable storage with strong consistency guarantees and increased energy efficiency compared to centralized cloud solutions.

4 METHODS AND EXPERIMENTAL SETUP DETAILS

4.1 Optimization of Edge Offloading and Resource Considerations

The framework aims to use a cloudlet-based, centralized multi-tiered tree architecture, following Routaib et al.'s [19] recommendations. It also considers computing and network resources, considering edge server challenges. A heuristic approach is used to optimize offloading choices based on resource and interaction assessments, inspired by Chen et al.'s work [21].

4.2 Development and Optimization of Edge Storage Frameworks

The study explores the use of existing distributed database platforms for edge storage, focusing on their adaptation to edge infrastructure. It also evaluates the use of an S3 file system in a fog environment, focusing on data movement, scalability, and accessibility [23]. The study also explores the implementation of a distributed, decentralized data storage solution, considering various edge nodes with diverse operating systems and file systems [24]. Finally, it explores Freenet's epidemic method for building decentralized storage networks, ensuring anonymity and safety for users.

4.3 Integration of Block-chain with Edge Storage

The study aims to implement Wilkinson et al.'s [26] block-chain system for item position recording, enabling periodic file integrity and availability checks. It also evaluates IBM and Samsung Electronics' proof of concept in ADEPT [27], combining edge storage with decentralized block-chain technologies using BitTorrent, Telehash, and Ethereum protocols.

4.4 Lightweight and Adaptable File Allocation Mechanism

The proposed mechanism aims to implement Shreshth et al.'s lightweight, adaptable, and efficient file allocation mechanism [28], while the APEX prototype system will be developed and evaluated, showcasing its usage and potential vulnerabilities in real-life scenarios.

4.5 Low-Cost Storage with Erasure Codes and Energy Efficiency with EdgeKV

The study explores the use of erasure codes for low-cost edge storage using Hai Jin et al.'s methodology [29]. It also investigates the implementation of EdgeKV, a decentralized storage system for network edge, and evaluates its performance in terms of data replication, consistency guarantees, and energy efficiency compared to centralized cloud storage [30]. The study also assesses the energy efficiency implications of EdgeKV with edge resources.

4.6 Experimental Setup Details

Aspect	Details
Focus Area	Edge Offloading, Edge Storage, Block-chain Integration
Methodologies	Evaluate heuristic approaches for offloading decisions, implement distributed storage solutions, and integrate block-chain with edge technologies.
Instruments	Edge servers, network simulation tools, block-chain platforms.
Materials	Cloudlet-based frameworks, Open-RAN, distributed database platforms, Freenet, block-chain systems.
Procedures	Offloading optimization, storage framework implementation, block-chain integration.
Measurements	Performance metrics for offloading decisions, scalability, anonymity, and safety measures for storage solutions.
Variables	Computing resources, network resources, file systems, block-chain transactions.
Challenges Addressed	Efficient offloading, decentralized storage, security, and integrity in edge environments.

5 DISCUSSION AND FUTURE STUDIES

The literature presents a comprehensive overview of diverse approaches and challenges in edge computing, particularly focusing on offloading decisions, edge storage, and the integration of block-chain technologies. Notably, the recommendations for a cloudlet-based framework, network-aware offloading optimization, and the introduction of Sagunas Open-RAN demonstrate the evolving strategies to address the complexities of edge computing environments. The emphasis on edge storage solutions, such as the S3 file system, decentralized data storage, and the application of erasure codes, highlights the ongoing efforts to optimize data management at the edge. Block--chain integration, as seen in Wilkinson et al.'s and ADEPT, addresses issues related to item position tracking and file integrity. Shreshth et al.'s [28] proposed file allocation mechanism, APEX, while showcasing adaptability, also raises concerns about security vulnerabilities. Furthermore, EdgeKV's decentralized storage system provides insights into efficient and energy-effective storage at the network edge. Future work should focus on refining security measures for decentralized storage, addressing potential vulnerabilities in lightweight file allocation mechanisms, and exploring the broader implications of edge computing, especially concerning the integration of emerging technologies like block-chain. Additionally, efforts should be directed towards standardization and interoperability to foster a more seamless integration of diverse edge computing solutions.

Take home messages from this survey:

Consideration of Network Resources: It is important to consider both network and computational resources while creating edge computing frameworks, as noted by Routaib et al. [19] and Betzler et al. [20].

Diverse Applications of Edge Computing: Saguna et al. [22] showcased the variety and potential impact of edge computing by demonstrating a range of applications such as DNS caching, video marketing, safe cities, and smart security.

Focus on Edge Storage: Numerous academics are experimenting with distributed database platforms, proposing decentralized storage systems such as the S3 file system in fog environments, and investigating edge storage solutions (Confais et al. [23], Gheorghe et al. [24]).

Decentralized Approaches: With a focus on data privacy, redundancy, and integrity, Freenet and Wilkinson et al. support Block-chain-based solutions and decentralized storage networks.

Integration of Edge Storage with Block-chain: Block-chain technologies and edge storage can be integrated for improved security and decentralization, as shown by projects like Cohn et al.'s ADEPT [27].

Efficiency and Adaptability: The goal of Shreshth et al.'s [28] lightweight and flexible file allocation algorithms is to achieve dependability and efficiency in a variety of storage scenarios.

Optimization for Edge Storage: In order to overcome the difficulties brought on by scarce resources and sophisticated processing, Hai Jin et al. [29] investigate the use of erasure codes for inexpensive storage at the edge.

Edge-specific Storage Solutions: Sonbol et al.'s [30] EdgeKV decentralized storage system is designed with the network edge in mind, with a focus on energy efficiency, speed, and dependability.

6 CONCLUSION

In conclusion, this survey paper has thoroughly explored the current landscape of edge computing, with a particular focus on edge storage solutions. Through an exhaustive

review of the literature, several key findings have emerged. Firstly, the study identifies various approaches to edge storage, ranging from cloudlet-based frameworks to decentralized storage systems like EdgeKV, addressing the challenges of managing data at the network edge. Secondly, the importance of considering both computational and network resources in offloading decisions has been underscored, as evidenced by optimization techniques proposed by several researchers. Additionally, the integration of Block-chain technology, exemplified by projects like ADEPT, holds promise for enhancing the security and efficiency of edge storage solutions. Lastly, the need for lightweight and adaptable file allocation mechanisms, coupled with the exploration of erasure codes for low-cost storage, underscores ongoing efforts to optimize edge storage platforms for diverse and resource-constrained environments. These key findings provide valuable insights into the current state of edge storage research and offer directions for future exploration in this rapidly evolving field. The take-home messages from this survey highlight the importance of considering various factors, such as architectural design, resource management strategies, and security measures, in the development and implementation of edge storage systems. Furthermore, the significance of the study lies in its contribution to a broader scientific consensus on edge storage by synthesizing diverse research findings and fostering a more comprehensive understanding of the topic. By offering practical implications for decision-making and guiding future research and innovation, this survey equips scholars and industry practitioners with the knowledge needed to address the challenges and opportunities in edge computing effectively. Overall, this collaborative endeavor aims to provide scholars and industry practitioners with a comprehensive understanding of current trends and challenges in edge computing, paving the way for informed advancements in this dynamic and rapidly evolving field. Moving forward, future research endeavors should prioritize strategies for seamlessly integrating edge and cloud resources to optimize storage, computation, and data management across distributed environments. It is imperative to further investigate novel approaches aimed at enhancing security and privacy in edge storage systems, particularly considering the unique challenges posed by decentralized and distributed architectures. Additionally, the development of edge-native storage solutions tailored to the specific requirements and constraints of edge computing environments is paramount, focusing on scalability, reliability, and energy efficiency. As a collaborative endeavor, this survey aims to equip scholars and industry practitioners with a comprehensive understanding of current trends and challenges in edge computing, paving the way for informed advancements in this dynamic and rapidly evolving field.

7 REFERENCES

[1] Shi, W., Cao, J., Zhang, Q., Li, Y. & Xu, L. (2016). Edge computing: Vision and challenges. *IEEE internet of things journal, 3*(5), 637-646. https://doi.org/10.1109/JIOT.2016.2579198

[2] Jararweh, Y., Doulat, A., AlQudah, O., Ahmed, E., Al-Ayyoub, M. & Benkhelifa, E. (2016). The future of mobile cloud computing: integrating cloudlets and mobile edge computing. *The 23rd IEEE International conference on telecommunications (ICT2016)*, 1-5. https://doi.org/10.1109/ICT.2016.7500486

[3] Dinh, H. T., Lee, C., Niyato, D. & Wang, P. (2013). A survey of mobile cloud computing: architecture, applications, and approaches. *Wireless communications and mobile computing, 13*(18), 1587-1611. https://doi.org/10.1002/wcm.1203

[4] Bonomi, F., Milito, R., Zhu, J. & Addepalli, S. (2012). Fog computing and its role in the internet of things. *Proceedings of the first edition of the MCC workshop on Mobile cloud computing*, 13-16. https://doi.org/10.1145/2342509.2342513

[5] Satyanarayanan, M., Bahl, P., Caceres, R. & Davies, N. (2009). The case for vm-based cloudlets in mobile computing. *IEEE Pervasive Computing, 8*(4), 14-23. https://doi.org/10.1109/MPRV.2009.82

[6] Ovesen, A. B., Nordmo, T. A. S., Johansen, H. D., Riegler, M. A., Halvorsen, P. & Johansen, D. (2021). File System Support for Privacy-Preserving Analysis and Forensics in Low-Bandwidth Edge Environments. *Information, 12*(10), 430. https://doi.org/10.3390/info12100430

[7] Li, Y., Wang, X., Gan, X., Jin, H., Fu, L. & Wang, X. (2019). Learning-aided computation offloading for trusted collaborative mobile edge computing. *IEEE Transactions on Mobile Computing*, *19*(12), 2833-2849. https://doi.org/10.1109/TMC.2019.2934103

[8] Ceselli, A., Premoli, M. & Secci, S. (2017). Mobile edge cloud network design optimization. *IEEE/ACM Transactions on Networking*, *25*(3), 1818-1831. https://doi.org/10.1109/TNET.2017.2652850

[9] Xia, X., Chen, F., He, Q., Grundy, J. C., Abdelrazek, M. & Jin, H. (2020). Cost-effective app data distribution in edge computing. *IEEE Transactions on Parallel and Distributed Systems*, *32*(1), 31-44. https://doi.org/10.1109/TPDS.2020.3010521

[10] Luo, R., Jin, H., He, Q., Wu, S., Zeng, Z. & Xia, X. (2021). Graph-based data deduplication in mobile edge computing environment. *Service-Oriented Computing: 19th International Conference, ICSOC 2021, Virtual Event, November 22–25, 2021, Proceedings 19*. Springer International Publishing, 499-515. https://doi.org/10.1007/978-3-030-91431-8_31

[11] Zhang, F., Ge, J., Wong, C., Li, C., Chen, X., Zhang, S., ... & Chang, V. (2019). Online learning offloading framework for heterogeneous mobile edge computing system. *Journal of Parallel and Distributed Computing*, *128*, 167-183. https://doi.org/10.1016/j.jpdc.2019.02.003

[12] Xie, J., Qian, C., Guo, D., Li, X., Shi, S. & Chen, H. (2019). Efficient data placement and retrieval services in edge computing. *The 39th IEEE International Conference on Distributed Computing Systems (ICDCS2019)*. 1029-1039. https://doi.org/10.1109/ICDCS.2019.00106

[13] Tran, T. X. & Pompili, D. (2018). Adaptive bitrate video caching and processing in mobile-edge computing networks. *IEEE Transactions on Mobile Computing*, *18*(9), 1965-1978. https://doi.org/10.1109/TMC.2018.2871147

[14] Chen, X., He, L., Xu, S., Hu, S., Li, Q. & Liu, G. (2019). Hit ratio driven mobile edge caching scheme for video on demand services. *The IEEE International Conference on Multimedia and Expo (ICME2019)*, 1702-1707. https://doi.org/10.1109/ICME.2019.00293

[15] Ren, C., Lyu, X., Ni, W., Tian, H. & Liu, R. P. (2019). Profitable cooperative region for distributed online edge

caching. *IEEE Transactions on Communications*, *67*(7), 4696-4708. https://doi.org/10.1109/TCOMM.2019.2908574

[16] Cao, X., Zhang, J. & Poor, H. V. (2018). An optimal auction mechanism for mobile edge caching. *The 38th IEEE International Conference on Distributed Computing Systems (ICDCS2018)*, 388-399. https://doi.org/10.1109/ICDCS.2018.00046

[17] Lyu, F., Ren, J., Cheng, N., Yang, P., Li, M., Zhang, Y. & Shen, X. S. (2020). LEAD: Large-scale edge cache deployment based on spatio-temporal WiFi traffic statistics. *IEEE Transactions on Mobile Computing*, *20*(8), 2607-2623. https://doi.org/10.1109/TMC.2020.2984261

[18] Wu, H., Lyu, F., Zhou, C., Chen, J., Wang, L. & Shen, X. (2020). Optimal UAV caching and trajectory in aerial-assisted vehicular networks: A learning-based approach. *IEEE Journal on Selected Areas in Communications*, *38*(12), 2783-2797. https://doi.org/10.1109/JSAC.2020.3005469

[19] Routaib, H., Badidi, E., Elmachkour, M., Sabir, E. & ElKoutbi, M. (2014). Modeling and evaluating a cloudlet-based architecture for mobile cloud computing. *The 9th IEEE International Conference on Intelligent Systems: Theories and Applications (SITA-14)*, 1-7. https://doi.org/10.1109/SITA.2014.6847290

[20] Betzler, A., Quer, F., Camps-Mur, D., Demirkol, I., & Garcia-Villegas, E. (2016). On the benefits of wireless SDN in networks of constrained edge devices. *The IEEE European conference on networks and communications (EUCNC2016)*, 37-41. https://doi.org/10.1109/EuCNC.2016.7561000

[21] Chen, M. H., Dong, M. & Liang, B. (2016). Joint offloading decision and resource allocation for mobile cloud with computing access point. *The IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP2016)*, 3516-3520. https://doi.org/10.1109/ICASSP.2016.7472331

[22] Saguna and Intel. (2016). Using mobile edge computing to improve mobile network performance and profitability. White paper.

[23] Confais, B., Lebre, A. & Parrein, B. (2017). Performance analysis of object store systems in a fog and edge computing infrastructure. *Transactions on Large-Scale Data-and Knowledge-Centered Systems, XXXIII*, 40-79. https://doi.org/10.1007/978-3-662-55696-2_2

[24] Gheorghe, A. G., Crecana, C. C., Negru, C., Pop, F. & Dobre, C. (2019). Decentralized storage system for edge computing. *The 18th IEEE International Symposium on Parallel and Distributed Computing (ISPDC2019)*, 41-49. https://doi.org/10.1109/ISPDC.2019.00009

[25] Clarke, I., Sandberg, O., Wiley, B. & Hong, T. W. (2001). Freenet: A distributed anonymous information storage and retrieval system. In: Federrath, H. (eds) *Designing Privacy Enhancing Technologies. Lecture Notes in Computer Science, vol 2009*. Springer, Berlin, Heidelberg. https://doi.org/10.1007/3-540-44702-4_4

[26] Wilkinson, S., Boshevski, T., Brandoff, J. & Buterin, V. (2014). Storj a peer-to-peer cloud storage network.

[27] Cohn, J. M., Finn, P. G., Nair, S. P., Panikkar, S. B., & Pureswaran, V. S. (2019). U.S. Patent No. 10,257,270. Washington, DC: U.S. Patent and Trademark Office.

[28] Tuli, S., Tuli, S., Jain, U. & Buyya, R. (2019). APEX: Adaptive Ext4 File System for Enhanced Data Recoverability in Edge Devices. *The IEEE International Conference on Cloud Computing Technology and Science (CloudCom2019)*, 191-198. https://doi.org/10.1109/CloudCom.2019.00037

[29] Jin, H., Luo, R., He, Q., Wu, S., Zeng, Z. & Xia, X. (2022). Cost-effective data placement in edge storage systems with erasure code. *IEEE Transactions on Services Computing*, *16*(2), 1039-1050. https://doi.org/10.1109/TSC.2022.3152849

[30] Sonbol, K., Özkasap, Ö., Al-Oqily, I. & Aloqaily, M. (2020). EdgeKV: Decentralized, scalable, and consistent storage for the edge. *Journal of Parallel and Distributed Computing, 144*, 28-40. https://doi.org/10.1016/j.jpdc.2020.05.009

Authors' contacts:

Ravula Rajesh, Dr.
(Corresponding author)
National Institute of Technology Silchar,
SH 39, Masimpur, Silchar, Assam 788010, India
ravula21_rs@cse.nits.ac.in

Ripon Patgiri, Assistant Professor Dr.
Department of Computer Science and Engineering
National Institute of Technology Silchar,
SH 39, Masimpur, Silchar, Assam 788010, India
ripon@cse.nits.ac.in
patgiri@ieee.org

ISSN 1846-6168 (Print), ISSN 1848-5588 (Online)
https://doi.org/10.31803/tg-20240117203325

Preliminary communication
Received: 2024-01-17, Accepted: 2024-02-29

Public Relations of Complementary Micro-enterprises in the Maintenance of Construction Facilities

Katarina Štavlić*, Krešimir Lacković, Igor Štavlić

Abstract: The subject of this paper is the possibilities of developing relations with the public of micro-enterprises from the activity of building maintenance. Given that micro and small businesses do not have large financial opportunities for marketing activities, it is logical to join a cluster in order to reduce costs, and for the purpose of developing strategic activities through public relations. In this paper, a hypothesis is put forward about the association into a cluster of complementary construction micro-enterprises that deal with the maintenance of construction facilities for the sake of public relations efficiency. The development of IT technologies has created the conditions for public relations to be organized online, which increases efficiency, but also reduces costs. In addition, information technologies enable the development of a number of other types of cooperation through the development of activities, market research, exchange of experience and other things of common interest. A model of public relations management was established in order to prove the hypothesis and objectives of this paper.

Keywords: cluster; communication; information; micro enterprise; public relations

1 INTRODUCTION

Construction is an economic branch that, with its dynamics, can stimulate the overall development of any economy. In addition, the increase in construction activities is one of the signs of positive developments in the economy and society as a whole. That is why the subject of research on this topic is exactly construction, that is, the maintenance of construction facilities. Various construction and complementary activities usually participate in this process (Examples: construction works, electrical and other energy and information technologies (IT) installations and finishing works in construction). In addition, the maintenance of construction facilities involves mostly smaller companies, but exceptional efficiency is required. The fact that smaller companies have a number of advantages, but also one fundamental problem, can be particularly emphasized. Namely, every small company has to invest funds in development as well as marketing activities, where compared to larger companies, they have fewer financial opportunities.

In this paper, some suggestions to the problem of public relations in micro and small enterprises in construction are proposed by joining a cluster. In this way, all participants retain their autonomy, and the subject of association is common interests. In this sense, market research, determination of the market segment and the position of the cluster can be highlighted. In this literature review research, the special goal of the association is joint activity on public relations. The main reason for this is the fact that public relations is not only a marketing, but also a long-term strategic activity of everyone, even small companies in maintenance activities. That is why complementary types of micro-enterprises should be determined and the goals and process of public relations activities should be defined.

In the modern conditions of technical and technological development, especially informatics, there is a new opportunity to perform certain activities of common interest, but with minimal costs. In this sense, several complementary smaller companies in the field should join together in a cluster and then use information technologies (IT) to define the goals, process and functioning model. This is exactly the main goal of this paper, with the hypothesis that companies, due to complementarity, are deprived of mutual competition, and informatics makes it possible to conduct relations with the public online, thus with minimal costs. The activity process is carried out using a functioning model that, in addition to public relations, also enables other marketing activities, especially the development of products or services and companies. The content analysis method was used in the paper in order to determine the complementarity of companies and essential elements of the cluster. In addition, the activity of public relations that is adapted to smaller companies was investigated. Finally, the development of IT enables the setting up of process and functioning models, which is particularly emphasized in the discussion.

The dynamics and development of construction, as well as information technology, guarantees the successful continuation of research on this topic.

2 PARTICIPANTS IN BUILDING MAINTENANCE

Building maintenance is described as a set of permanent activities that contain technical and associated administrative activities in order to ensure the level of easement within the available (satisfactory) limits of the building as a whole and all its parts during the entire life of use. [1] In addition, maintenance implies ensuring the reliable function of a residential building as well as other buildings, regardless of purpose. [2] According to the Construction Act strictly formally, maintenance is defined (Article 3, paragraph 1) as follows: "Building maintenance is the performance of construction and other works on an existing building in order to preserve the basic requirements for the building during its life, which it does not change the compliance of the building with the location conditions according to which it was built. [3] In addition to the Construction Act, other by-laws have

determined the method of financing or covering costs for certain types of works, such as the following [4]:

- regular maintenance and improvement of common parts and devices of the building
- urgent repairs of common parts of the building and devices
- replacement of existing and installation of new common parts and devices of the building.

The object of this topic is a cluster of companies that deal with the maintenance of construction facilities. Clusters can be defined as a modern model of business networking, by forming cooperation chains based on production cooperation, transport and service provision, in which the development of small and medium-sized enterprises is initiated in a flexible way. [5] In this case, it is an example where the cluster consists of four basic types of construction works. The carrier of the activity is a construction company that combines basic types of work such as: masonry, reinforcement, roofing, etc. The rest refers to all mechanical installations (water supply, sewage, other energy sources and air conditioning). The third group are companies that perform all types of electrical and IT installations. Other participants in the maintenance cluster are finishing works (installation of floors and ceilings, painting, etc.). In addition, the maintenance of construction facilities usually involves works of lesser value than the construction of new facilities. [6] That is why maintenance works are usually efficiently carried out by smaller companies known as micro-enterprises. These are economic entities, natural and legal persons that employ less than 10 workers on average per year, generate a total annual income of up to EUR 2,000,000.00, or have total assets of up to EUR 2,000,000.00. [7] In addition, small and micro enterprises have certain advantages such as: "greater business flexibility, universality of employees, greater level of innovation, greater networking opportunities, etc." [8]

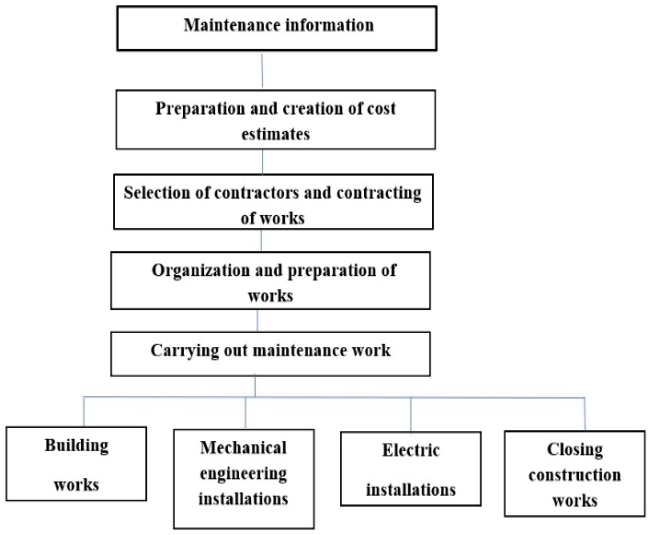


Figure 1 The process of building maintenance [6]

Fig. 1 shows a representation of the common maintenance procedures where a company specializing in building maintenance receives information or an order for intervention due to the occurrence of damage or a defect in the building or the need for regular maintenance. Every well-organized company must carry out certain activities for the preparation and execution of maintenance work. In addition, it should be borne in mind that these are usually jobs that need to be completed as soon as possible, so in addition to quality, price and deadline, efficiency is also an important factor. Process efficiency can be achieved if several complementary construction micro-enterprises join together in a cluster. In terms of organization, the participants in the cluster choose the person responsible for the information and communication process who leads the maintenance process according to the type and quantity of individual works. By agreement between the members of the cluster, other jobs can be combined, especially commercial or marketing jobs and public relations. In addition to marketing, public relations activity is becoming more and more strategic, both for large and small business entities and groups. That is why the starting point of consideration is precisely the joint organization of relations with the public in order to achieve maximum efficiency and minimum costs.

3 PUBLIC RELATIONS AS AN IMPORTANT STRATEGIC ACTIVITY

Public relations, often abbreviated as PR, is an important strategic, communication and marketing tool that takes on an increasingly important strategic role in the business operations of both larger and smaller economic entities. There are numerous definitions of the concept of public relations, from which the activities and functions of public relations derive. One of the basic definitions is that "public relations is a strategic communication process that builds mutually beneficial relationships between organizations and their publics." [9] Public means all stakeholders in the environment, which can be: media, market, local self-government, investors, suppliers, non-profit organizations, financial institutions, employees, and others. It is necessary to communicate effectively with the public, and this requires leadership competencies. That is why the emphasis is on the managerial function of public relations, whereby public relations "is a management function that classically focuses on long-term patterns of interaction between an organization and all of its various publics" [9]

From the aspect of small entrepreneurship, public relations represent communication between a business entity and its public with the aim of achieving mutual understanding and achieving common interests, ensuring the realization of long-term goals of creating a good image. For more successful operations and achieving a competitive advantage, strategic planning and measurement of achieved goals by monitoring defined indicators of business success are necessary. Success can be measured through non-financial and qualitative indicators, and in this sense, the function of public relations is reflected in strategic planning that brings the company into balance with the needs of its public. In doing so, it is necessary to choose the appropriate and most functional communication channels and tools for clearly communicating set business goals with the public, as well as for implementing feedback into strategic decisions.

There are two types of communication channels. They can be personal and indirect [10] and it is important to adapt the communication channels to the content of the communication, the stakeholders and the goal of the communication. In business practice, integrated marketing and communication tools are used, aligned with the manager's managerial and marketing competencies. The above is more acceptable for larger and more financially stable organizations, while for smaller companies, especially for micro-companies, such models are not financially acceptable. In support of the strategic application of public relations in micro and small companies is the development of IT technologies and the information revolution, which enables faster, more frequent, targeted, more accessible and differentiated communication with the public. Modern means of communication, availability of information, greater education of the public and openness of the market are also in favor [11].

The mentioned functions and activities of public relations and communication with the environment can be extremely useful for micro-enterprises. But their problem is the lack of financial and professional opportunities for organizing such activities. That is why the solution to this problem lies in association, and in this case, it is exactly a cluster. In this way, even small or micro companies, especially with the help of IT, can successfully organize public relations. Bearing in mind that public relations, according to the definitions, is not only a marketing but also a very important strategic managerial activity and small companies, especially in maintenance activities, should pay special attention to it. Given the great possibilities offered by information technology today, joint public relations activities can be organized very efficiently and with minimal costs.

4 A PUBLIC RELATIONS MODEL OF A CLUSTER OF CONSTRUCTION COMPANIES FOR MAINTENANCE IN ONLINE CONDITIONS

By establishing a cluster of reflection of construction facilities, conditions were created for effective cooperation, but also for other development and commercial activities of common interest. In maintenance activities, it is extremely important to have contact with the environment, because the maintenance of construction facilities is essential for the functioning of the living space of the population and the infrastructure. In addition, joining a cluster creates greater opportunities for business development and covering various commercial costs. Given the increasing importance of public relations, these costs can be more easily financed by pooling. The development of IT, especially the Internet, enables public relations activities to be carried out in a very efficient manner with minimal involvement of all resources.

Fig. 2 shows an example of the cooperation of participants in the cluster for the maintenance of construction facilities with the aim of effective relations with the public. In this example, all four basic types of construction work are included, with the inclusion of construction supervision. This is important because supervision enables the detection of various problems during the execution of works. Construction supervision is an independent professional activity that is not a participant in the cluster and ensures quality and efficient execution of works. It is also a source of information on the basis of which actions are taken to increase the quality of future works. The activity system should usually be managed by the construction company as the holder of all other maintenance work.

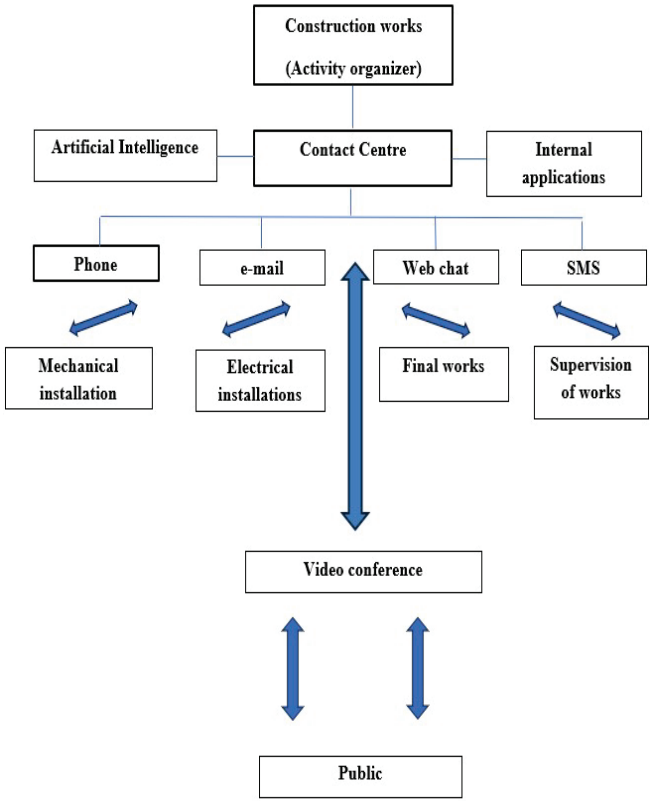


Figure 2 Public relations model of a cluster of construction companies for maintenance in online conditions

An effective public relations activity can be enabled by a combined system shown in Fig. 2. It can be seen that the participants in the public relations activity are all types of construction and subcontracting works, including supervision works with the formal management of the construction company. This is made possible with the help of IT, and in this case, it is CRM (Customer Relationship Management). [12] The basic configuration of this model consists of: a contact centre, internal applications and artificial intelligence, all of which can function virtually. The contact centre is a hardware-software module with which the activity leader can communicate with other members of the cluster, as well as all together via video conferences with the public. This complex communication is made possible by internal applications as special software, and special support can be provided by artificial intelligence. This software support is a novelty in modern IT systems. An intelligent system is considered to be any system that shows adaptive behaviour, learns based on experience, uses large amounts of knowledge, shows properties of awareness, communicates with humans in natural language and speech, but allows mistakes and ambiguities in communication, etc. [13]. Apart

from that, and in a number of other definitions, there is an opinion that artificial intelligence includes the following IT systems that:

- think rationally,
- behave rationally,
- have all the appearance of intelligence (rational or human),
- with their internal functions, they try to be in harmony with a human being, that is, a rational being,
- have adaptability, where work ability is improved by learning from experience and independence and the ability to perform tasks in complex environments without someone's help and constant guidance [14].

In the last 10 years, artificial intelligence has brought enormous changes to the world market and the operations of business entities. The changes are reflected in faster analysis of user data, in personalized advertising, automated communication with the larger public, faster forecasting of trends and generation of content for the needs of more successful business. The ability of artificial intelligence to collect, analyse and interpret data, in addition to its other features, enables its contribution in improving public relations in micro-enterprises. Artificial intelligence is a memory that contains various information from concrete practice and the latest theoretical knowledge, and the management part can enable quick and efficient management of public relations activities. Some of the activities of applying the public relations strategy in the described cluster can go in the direction of: educating the public about the importance of maintenance; sharing stories about successful projects; displaying positive reviews from satisfied clients, and socially responsible activities and care for the local environment and community. This is important because maintenance activities also need to monitor the development of their activities in order to be able to effectively inform and understand clients and the rest of the public.

5 DISCUSSION

Research on this topic shows, first of all, the topicality of the research subject of this paper, which is public relations at construction maintenance companies. In particular, the association in a cluster can be highlighted due to their complementarity, i.e. synergy where their activities complement each other. This creates the conditions for other forms of cooperation, especially regarding the exchange of experience, innovations and joint marketing activities. In this case, special attention is focused on public relations. Definitions and other information point to the fact that public relations is not only a marketing but an extremely important strategic activity. This is particularly evident in the maintenance of construction facilities. Namely, for these activities, all the mentioned participants, who in this case represent the public to this cluster, are very important. Some, such as clients, are interesting as users of services, while others are carriers of information, so that participants in the cluster are up-to-date on all events related to direct business or the image of the cluster. In addition, public relations can be more effective with the use of information technology. The application of CRM as software for virtual communication ensures maximum efficiency with minimum costs. Everything can take place through joint video conferencing enabled by internal applications in the contact centre. In this case, artificial intelligence deserves special attention as additional software support for public relations activities. From the very definition and other features of artificial intelligence, new possibilities for increasing the quality of relations with the public can be seen. The current level of software enables quick access to a large amount of information, but also a certain level of creativity. This opens the possibility for the moderator of the activity to have a more varied and useful approach to communication. Lalić et al. [15] wrote about the importance of association and cooperative business. In the results of their research on the influence of mutual cooperation on the level of innovation, they indicated that manufacturing companies that cooperate with other organizations in product development and research and development activities are significantly more innovative, in terms of innovative products, from non-cooperating companies. A higher level of innovation, according to some authors, leads to a more successful business. So that association with the goal of joint relations with the public has the potential to lead to business improvement. These are all facts that arise from consideration of this contemporary topic and contribute to the realization of the research goal and hypothesis.

6 CONCLUSIONS

Several conclusions can be drawn from the previous considerations. First of all, it can be said that this topic deserves special attention due to its modernity and the needs of micro-enterprises. Research and consideration of the literature and proposed solutions is aimed at small or micro enterprises that deal with the maintenance of construction facilities. This activity is extremely important and useful both for the economy, infrastructure and citizens. The paper describes the process of building maintenance and public relations model of a cluster of construction companies for maintenance in online conditions. In addition, for micro-enterprises, a complementary association into a cluster is suggested for greater efficiency and quality, but also for lower costs. Participants of the proposed cluster can cooperate in the exchange of various technical, development and commercial information. On this occasion, the starting point of cooperation is a joint relationship with the public. The reason for this is the fact that public relations is a useful marketing tool, but in modern conditions an increasingly important strategic activity of every company. Given that everything takes place online, the effectiveness of this activity can be increased with the support of IT and the application of new IT systems, especially artificial intelligence. In this sense, a model of public relations of microenterprises in the activity of building maintenance with the support of hardware-software support, known as CRM, is proposed. This solution contains a contact centre for help, which, with the support of internal applications, allows the

manager to contact other participants of the cluster and together with representatives of the public. Special support can be provided by artificial intelligence, especially in terms of greater speed, quality and variety of information with the aim of better organizational solutions. The above conclusions point to the justification of research on this topic and the achievement of the goal of this paper. Proposals for future research are based on the market needs of micro-enterprises in the field of construction, and go in the direction of detecting key factors for improving communication within and outside these enterprises, and factors that influence the application of new technologies in business. By detecting the aforementioned factors, the possibility of faster and targeted directing of financial means and other resources to improve relations with the public of all stakeholders of that system would be opened.

The modernity and usefulness of this topic, as well as the technical and special IT development, are sufficient reason for continuing the research.

7 REFERENCES

[1] Marenjak, S. (2008). *Održavanje zgrada – skripta*. Osijek: Građevinski fakultet Osijek. (in Croatian)

[2] Cerić, A. & Katavić, M. (2000). Upravljanje održavanjem. *Građevinar, 53*/2000, HSGI, Zagreb, 83-89. (in Croatian)

[3] Construction Act, Zakon o gradnji, Narodne novine br. 153/13, 20/17, 39/19, 125/19. (in Croatian)

[4] Uredba o održavanju zgrada, NN 64/97.

[5] Izsak, K., Ketels, C., Meier zu Koecker, G. & Laemmer-Gamp, T. (2016). *Smart Guide Cluster Policy*, European Commission. https://data.europa.eu/doi/10.2873/729624

[6] Lacković, Z. (2017). *Lean-građevinarstvo*. Osijek: Alberta

[7] Narodne novine br. 29/02, 63/07, 53/12, 56/13, 121/16, čl. 3, See https://www.zakon.hr/z/527/Zakon-o-poticanju-razvoja-malog-gospodarstva (in Croatian)

[8] Štavlić, K. (2018). *Mikro poduzetništvo – resursi, potencijali i uspješnost*. Požega: Veleučilište u Požegi (in Croatian)

[9] Smith, D. R. (2021). *Strategic Planning for Public Relations*. 6th Edition. New York and London: Routledge.

[10] Tomaš, A. & Topić Stipić, D. (2022). Convergence in public relations. *South Eastern European Journal of Communication, 4*(2), 63-73. https://doi.org/10.47960/2712-0457.2022.2.4.63

[11] Vasilj, D. & Korda, K. I. (2022). Uloga i aktivnost funkcije odnosa s javnošću u menadžmentu poduzeća. *South Eastern European Journal of Communication, 4*(2), 51-62. (in Croatian) https://doi.org/10.47960/2712-0457.2022.2.4.51

[12] Panian, Ž. (2003). *Odnosi s klijentima u e-Poslovanju*. Zagreb: Sinergija. (in Croatian)

[13] Prister, V. (2019). Umjetna inteligencija. *Media, culture and public relations, 10* (1), 67-72. (in Croatian) https://doi.org/10.32914/mcpr.10.1.7

[14] Russell, S. & Norvig, P. (2011). *Artificial Intelligence: A Modern Approach*. Third Edition. New Jersey: Prentice Hall. *Artificial Intelligence, 175*(5-6), 935-937. https://doi.org/10.1016/j.artint.2011.01.005

[15] Lalic B., Tasic, N., Marjanovic, U., Delic, M., Cvetkovic, N. (2017). Inter-Organizational Collaboration for Innovation in Manufacturing Firms. *Vallis Aurea,* 3(2), 55-66. https://doi.org/10.2507/IJVA.3.2.5.38

Authors' contacts:

Katarina Štavlić, PhD, Assistant Professor
(Corresponding author)
Faculty of Tourism and Rural Development Požega,
J. J. Strossmayer University of Osijek,
Vukovarska 17, 34000 Požega, Croatia
kstavlic@ftrr.hr

Krešimir Lacković, PhD, Assistant Professor
University North,
Trg dr. Žarka Dolinara 1, 48000 Koprivnica, Croatia
klackovic@unin.hr

Igor Štavlić, Master of Science
IGOR, Lifelong Learning Centre in Nature,
Marindvor, Marindvor 39, 34000 Požega, Croatia
igorstavlic@gmail.com

ISSN 1846-6168 (Print), ISSN 1848-5588 (Online)
Ohttps://doi.org/10.31803/tg-20240208164505

Preliminary communication
Received: 2024-02-08, Accepted: 2024-04-15

Decision-Making Model for Reliable Electronic Component Manufacturing in a Blockchain Environment

Sung Yong An, Guy Ngayo, Seng-Phil Hong*

Abstract: This study focuses on improving the efficiency of 5G ceramic antenna manufacturing through the application of blockchain and smart contract technologies. Using the Analytic Hierarchy Process (AHP), the relative importance of three key factors was assessed: 'Security', defined as the protection against unauthorized access and data integrity; 'Standards & Regulations', which refers to compliance with industry norms and legislative requirements; and 'Efficiency', highlighting process optimization and resource management. The findings indicate that 'Efficiency' is the most critical factor, with a weight of 0.545, followed by 'Security' (0.318) and 'Standards & Regulations' (0.137). Furthermore, the sub-factor analysis identifies 'Unit Gain', related to productivity improvements, and 'Integrity', concerning data veracity, as significant, with global weights of 0.309 and 0.111, respectively. These insights emphasize the critical role of chip efficiency in the manufacturing process. The study proposes an AHP-based decision model for effectively integrating blockchain technology in the 5G sector, offering strategies for optimizing antenna manufacturing. This approach not only delineates critical process elements but also promotes the advancement of blockchain applications in 5G technology, potentially catalysing substantial technological innovation.

Keywords: Analytic Hierarchy Process (AHP); blockchain; decision-making; Consistency Rate; smart contract; 5G ceramic antenna

1 INTRODUCTION

In the era of the Fourth Industrial Revolution, blockchain and smart contracts have emerged as transformative technologies, playing a crucial role in enhancing the reliability and security of enterprise manufacturing and supply chain management. Their implementation is particularly effective in addressing the logistical and security challenges posed by global business operations. This is especially pertinent given the recent surge in information leakage incidents, underscoring the need for the transparency, data integrity, and security that blockchain technology affords. Smart contract technology, in particular, is receiving increasing attention for its role in mitigating these concerns [1-6]. Parallel to these developments, 5G communication technology stands at the forefront of this industrial revolution. The advent of 5G ceramic antennas, developed as an advancement over traditional PCB-based antennas [7], marks a significant leap in communication efficiency. Recent initiatives are exploring the integration of blockchain and smart contracts to bolster the security and efficiency of these 5G ceramic antennas [8].

The progression of 5G technology is catalyzing innovations across various sectors, paving the way for new services and products. In the realm of 5G ceramic antenna manufacturing, there is a growing recognition of the need for enhanced security, compliance with standards and regulations, and operational efficiency. Adherence to regulatory standards enhances the credibility of antenna manufacturers, aligning them with international benchmarks. Furthermore, operational efficiency, a cornerstone in antenna production, necessitates a focus on streamlined processes, optimal resource allocation, and cost minimization [9, 10].

This research is dedicated to improving both the efficiency and reliability of the 5G ceramic antenna manufacturing process through the strategic application of the Analytic Hierarchy Process (AHP). Given the complex nature of antenna production, which is susceptible to defects that can necessitate expensive corrective measures, our approach integrates blockchain technology for meticulous tracking and smart contracts for streamlined issue resolution. AHP will be used to systematically prioritize aspects such as security, standards and regulations, and efficiency—focusing on key areas that significantly impact cost-effectiveness and performance.

Despite the advances in blockchain and smart contract applications within the 5G technology space, there remains a notable gap in systematic decision-making frameworks that leverage these technologies for optimizing manufacturing processes. Previous studies have predominantly focused on the theoretical potentials of blockchain in enhancing security and operational transparency but have not sufficiently explored its integration into practical manufacturing workflows, particularly in the context of ceramic antennas for 5G networks. This study seeks to fill this gap by developing an AHP-based decision model that not only prioritizes critical manufacturing factors but also integrates blockchain technology to enhance the efficiency and reliability of 5G ceramic antenna production.

2 LITERATURE REVIEW

2.1 Analytic Hierarchy Process (AHP)

The Analytic Hierarchy Process (AHP), developed by Professor T. L. Saaty in the 1970s, revolutionized decision-making in complex scenarios [11-16]. Initially designed to address decision-making challenges in the U.S. State Department, particularly in arms control and disarmament, AHP has since broadened its application, significantly impacting public sector investment projects. Its core aim is to manage subjective decision-making elements like intuition, emotion, and perception, traditionally challenging to quantify [17, 18].

AHP operates on the principle of pairwise comparisons to evaluate relative importance, making it particularly effective in complex scenarios involving multiple criteria or diverse stakeholders [19, 20]. This methodology provides a

structured approach to decision-making, enhancing the clarity and robustness of choices made in intricate situations.

In public administration and policy analysis, AHP has been instrumental in evaluating infrastructure projects, including dams, airports, subways, and road construction [21, 22]. Its systematic process involves organizing decision-making factors hierarchically into main and sub-factors and determining their importance through pairwise comparisons. This process, supported by ratio scales, allows decision-makers to calculate weights and prioritize factors effectively, thereby simplifying complex decision-making processes [23, 24].

AHP is particularly valuable in situations where mathematical quantification of problems is challenging, such as in research and development (R&D) projects. It leverages expert judgment to assign weights to factors that are difficult to quantify objectively, aiding in the formulation of diverse and sometimes conflicting evaluation criteria [25, 26]. The methodology is underpinned by three foundational principles: building a hierarchical structure, determining relative importance through subjective pairwise comparisons, and maintaining logical consistency. Hierarchies help break down complex problems into simpler components, making them easier to analyse and solve. Pairwise comparisons involve subjective judgments to assess the relative significance of each element in the hierarchy. Logical consistency is assured using a consistency index, ensuring reliable and consistent decision-making [27].

2.2 Blockchain and Smart Contracts

2.2.1 Blockchain Technology: The Theoretical Foundation

Blockchain technology, a cornerstone of modern digital transactions, gained prominence with the introduction of Bitcoin by Satoshi Nakamoto [28]. Characterized by enhanced security and transparency, blockchain overcomes the limitations of traditional centralized systems through its use of distributed ledgers and cryptographic techniques. The technology’s core is a distributed database, relying on a shared ledger that ensures transaction records are simultaneously distributed and validated across the network, thereby heightening security and reliability [29, 30].

The primary unit of blockchain is the "block," which contains multiple transaction records linked to previous blocks via cryptographic hashes. Each new block incorporates data from its predecessor, forming a chain and ensuring data immutability. Altering recorded data would require changes across all subsequent blocks, a feature that solidifies blockchain's security. Its applications span various sectors, from finance, where it is used in payments and smart contract-based financial products, to supply chain management, healthcare, intellectual property, and even voting systems [31-33].

In manufacturing, blockchain's decentralized nature promotes trust across production stages, addressing concerns like product quality and cost-efficiency. Unlike centralized systems, blockchain’s distributed architecture supports transparent, direct peer-to-peer transactions, with all network participants verifying transactions. This structure guarantees data immutability and resistance to tampering, further bolstered by distributed computing that allows nodes to concurrently store and verify transactions, enhancing data security [29, 34]. Blockchain's potential extends to car sharing, medical in-formation exchange, and other industries, promising innovation and development. Its secure platform positions it as a catalyst for cross-industry advancements.

Incorporating AHP in blockchain, organizations can systematically evaluate and prioritize projects. By setting criteria and sub-criteria, decision-makers can objectively compare projects, considering aspects like security, cost, and expected profitability. AHP's methodology quantifies the relative importance of each aspect, aiding in informed project selection [35-39].

2.2.2 Smart Contracts: Automating Trust in Blockchain

Smart contracts, conceptualized by N. Szabo in 1994 [40], are automated digital protocols designed for executing contract terms autonomously, functioning as digital versions of traditional contracts on a blockchain. Encoded as software on a blockchain's distributed ledger, smart contracts autonomously manage and execute transactions based on predefined rules, negating the need for centralized intermediaries. The primary ad-vantages of smart contracts include heightened reliability due to their automatic operation, robust security stemming from the immutability and transparency of blockchain, and in-creased efficiency through automated execution, which streamlines the negotiation pro-cess.

Khan et al. [41] discussed blockchain's role in the financial industry, highlighting how it addresses trust issues in centralized systems. Smart contracts, as executable codes on blockchains, are pivotal in facilitating and automating agreements, especially between parties lacking mutual trust. Applications of smart contracts extend beyond finance into domains like smart city development. Ullah et al. [42] explored their use in managing real estate transactions in smart cities, proposing a framework for integrating blockchain and smart contracts in this sector. This approach facilitates more interactive and user-friendly contracting processes, potentially transforming traditional real estate into a model aligned with Industry 4.0 standards [3].

2.3 The theoretical Background of 5G Communication and Ceramic Antennas

2.3.1 5G Communication Technology: Unleashing the Next Era

5G, the fifth generation of mobile communication technology, marks a significant leap forward from its predecessors, 3G and 4G. It is characterized by substantially higher data speeds, reduced latency, and the ability to connect a vast array of devices simultaneously. These enhancements position 5G as a key driver of technological progress, particularly in areas requiring large-scale data transfer and Internet of Things (IoT) services.

A standout feature of 5G is its data transmission speed, reaching up to 20 Gbps. This speed is more than 20 times faster than 4G’s 1 Gbps and significantly surpasses 3G's 384

Kbps. Such rapid data handling capabilities enable smooth streaming of high-definition and 4K videos, and support real-time applications like gaming, virtual reality (VR), and augmented reality (AR) with minimal latency [43, 44]. Another critical aspect of 5G is its ultra-low latency, which can be reduced to less than 1 millisecond. This improvement is crucial for applications requiring high precision and stability, such as autonomous vehicles and industrial robotics. Furthermore, 5G's extensive connectivity allows for the simultaneous connection of thousands of devices, making it integral to the development of smart cities and the broader IoT ecosystem. As 5G continues to roll out globally, it is set to transform a wide range of industries, heralding a new era of techno-logical innovation and connectivity [45, 46].

2.3.2 Evolution of 5G Antenna Technology: Key Challenges and Innovations

The progression of 5G communications is intrinsically linked to the advancements in 5G antenna technology. These antennas are pivotal for efficiently transmitting and receiving wireless signals, necessitating support for broader bandwidths and faster data rates. Unlike traditional antennas with fixed directional capabilities, 5G antennas require enhanced flexibility to manage signals from various directions. Current research efforts are geared towards developing antennas with improved tuning abilities to ensure precise signal directionality and strength.

A significant challenge in this arena is the need for 5G antennas to be compact and lightweight. Traditionally, high-performance antennas tended to be larger, but the emerging trend is to employ multiple smaller antennas to handle data transmission and reception in different directions. This shift has spurred technical research focused on creating more compact antenna designs. The physical length (L) of a 5G antenna, for instance, is determined by the eq. (1), which incorporates factors like effective wavelength (λ_{eff}) as shown in eq. (2) [7].

$$L = \frac{\lambda_{\text{eff}}}{2} \tag{1}$$

Where, λ_{eff} can be expressed as eq. (2) [47, 48].

$$\lambda_{\text{eff}} = \frac{\lambda_0}{\sqrt{\mu_r \cdot \varepsilon_r}} \tag{2}$$

Where, λ_0 is the wavelength in a vacuum, μ_r is the relative permeability, and ε_r is the relative permittivity. By utilizing ceramic materials with a higher dielectric constant, such as 9.45 compared to the 3.47 in current PCB antennas, the size of the antenna can be effectively reduced [7].

Fig. 1 illustrates a 5G ceramic antenna array, comprising low-band (24.25 GHz to 29.5 GHz) and high-band (37 GHz to 40 GHz) unit chips suitable for specific frequency ranges. This configuration is integral to the development of a comprehensive 5G antenna system. Research is underway to reduce power consumption through innovative antenna designs, contributing to the sustainable advancement of 5G communication technologies.

Figure 1 A 5G ceramic chip antenna array, displaying a module that combines a high-band unit and a low-band unit ceramic chip antenna, arranged in a 1 x 5 configuration.

3 RESEARCH METHODOLOGY

3.1 Research Design Description

This study's research design encompasses delineated in Fig. 2 presents the AHP research framework employed in our investigation. The study focuses on three fundamental factors: Security, Standard & Regulation, and Efficiency. The intent was to evaluate how these factors influence the stabilization and cost-effectiveness of the manufacturing processes for 5G ceramic antennas, particularly with the integration of blockchain technology and smart contracts. The selection of these pivotal factors and their respective sub-components stems from an extensive review of efficient manufacturing practices for 5G ceramic antennas, as detailed in our previous work [8]. We conducted a thorough pairwise comparison to ascertain the relative importance of each primary and secondary factor, thus quantifying their impact within the AHP model. This methodological approach enabled a structured analysis of how each element contributes to the overall manufacturing goals.

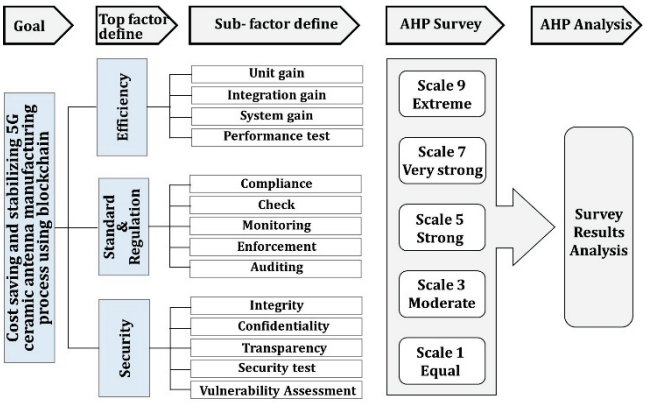


Figure 2 AHP research framework for evaluating key factors in 5G ceramic antenna manufacturing.

3.2 AHP analysis

AHP methodology was employed to facilitate decision-making in evaluating security, standard & law regulation, and efficiency in the blockchain-based 5G ceramic antenna manufacturing process. Data were gathered using a standardized questionnaire based on Saaty's 9-point scale [13], detailed in Tab. 1. This scale ranges from 1, representing equal importance between two factors, to 9, indicating a strong preference for one factor over another.

Table 1 The scoring system and meanings for Saaty's 9-point scale [13] used in the AHP analysis.

Scale	Definition	Explanation
9	Extreme	The evidence favouring one activity over another is of the highest possible order of affirmation
7	Very strong	An activity is favoured very strongly over another; its dominance demonstrated in practice
5	Strong	Experience and judgement strongly favour one activity over another
3	Moderate	Experience and judgement slightly favour one activity over another
1	Equal	Two activities contribute equally to the objective

The collected data underwent AHP analysis, leading to the development of a pair-wise comparison matrix (Eq. (3)). This approach is advantageous as it simplifies the evaluation process by allowing for direct comparisons between two factors at a time, which is generally more manageable for evaluators than multiple simultaneous comparisons.

Consistency in responses is crucial for the validity of AHP analysis. The significance of each response is determined by combining the results of individual evaluators, focusing on those with consistency ratio values (derived from the pairwise comparison matrix and geometric mean calculations) of less than 0.1. Responses with consistency ratios above this threshold are excluded to ensure the reliability of the analysis.

The pairwise comparison matrix $\boldsymbol{A}$, generated through the AHP process, reflects the relative importance of 'n' elements compared within each layer. For elements (A_1, A_2, ..., A_n), the result of pairwise comparisons between A_i and A_j is denoted as a_{ij}, forming the matrix $\boldsymbol{A} = (a_{ij})$. If the importance of each element is w_i ($i = 1, 2, \ldots, n$), the matrix $\boldsymbol{A} = (a_{ij})$ is used to calculate these weights, ensuring the accuracy and relevance of the derived values.

$$\boldsymbol{A} = (a_{ij}) = \begin{bmatrix} w_1/w_1 & w_1/w_2 & \cdots & w_1/w_n \\ w_2/w_1 & w_2/w_2 & \cdots & w_2/w_n \\ \vdots & \vdots & \cdots & \vdots \\ w_n/w_1 & w_n/w_2 & \cdots & w_n/w_n \end{bmatrix} \tag{3}$$

In the AHP framework, each element's relative importance is represented as a ratio. For instance, w_1/w_1 compares A_1 with itself, naturally yielding a value of 1. Similarly, w_1/w_2 represents the relative importance of A_1 compared to A_2, and w_1/w_n for A_1 compared to A_n. Multiplying eq. (3) by the column vector $w = [w_1, w_2, \ldots, w_n]$, which represents the approximate values of relative importance between evaluation items, results in eq. (4).

$$\begin{bmatrix} w_1/w_1 & w_1/w_2 & \cdots & w_1/w_n \\ w_2/w_1 & w_2/w_2 & \cdots & w_2/w_n \\ \vdots & \vdots & \cdots & \vdots \\ w_n/w_1 & w_n/w_2 & \cdots & w_n/w_n \end{bmatrix} \cdot \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix} = \begin{bmatrix} nw_1 \\ nw_2 \\ \vdots \\ nw_n \end{bmatrix} \tag{4}$$

Using the eigenvalue calculation method, it can be expressed as follows.

$$\boldsymbol{A} \cdot w = \lambda_{\max} \cdot w \tag{5}$$

The maximum eigenvalue of the pairwise comparison matrix $\boldsymbol{A}$, denoted as λ_{max}, is critical for determining the weights. It is found by solving the characteristic equation, which provides a non-zero solution in a system of n simultaneous equations, thereby determining the λ_{max} value that satisfies Eq. (6).

$$\left|\boldsymbol{A} - \lambda I\right| = 0 \tag{6}$$

The λ_{max} value is always greater than or equal to the number of elements, n. The closer the λ_{max} value is to n, the more consistent the pairwise comparison matrix A is considered. The consistency of these comparisons is quantified using the Consistency Index (CI) and the Consistency Ratio (CR), calculated as follows:

$$CI = \frac{\lambda_{\max} - n}{(n-1)} \tag{7}$$

$$CR = \frac{CI}{RI} \tag{8}$$

Where, n denotes the dimension of the matrix, and RI represents the Saaty Random Index [14], which varies with the matrix dimension. The RI is an average consistency index obtained from randomly generated reciprocal matrices, serving as a benchmark for acceptable consistency levels. The RI values are listed in Tab. 2. A CR of 0.1 or less indicates acceptable consistency in the survey responses, validating the reliability of the AHP analysis. For each survey item, the weights (w) were determined using the eigenvalue method, and then the overall weights were calculated using the geometric mean method. The values of the CI, CR, and the λ_{max} were determined by taking the arithmetic mean of the values obtained from each survey item to calculate the overall values.

Table 2 The Saaty Random Index (RI) values [14], which are used in the AHP to evaluate the consistency of pairwise comparisons.

n	3	4	5	6	7	8	9	10
RI	0.58	0.9	1.12	1.24	1.32	1.41	1.45	1.49

3.3 Research Process and Data Collection

For this study, an AHP questionnaire was developed based on the established model. The data collection phase spanned six weeks, from November 6 to December 15, 2023, targeting experts in the fields of blockchain and 5G ceramic antennas. The survey was disseminated through online platforms, email, and in-person visits, accompanied by comprehensive guidelines explaining the survey's objectives and key factors. Regarding the mix of in-person and email data collection, this approach was chosen to enhance the accessibility and response rate from experts in diverse locations, ensuring a comprehensive representation of opinions. Our sampling method was purposeful, targeting

individuals with specific qualifications (minimum master's degree and 10 years of experience) to gather informed perspectives into blockchain and 5G antenna technology. The survey questionnaire is detailed in the Appendix. From the 42 questionnaires collected, 30 were deemed suitable for analysis after excluding 8 due to non-response and 4 due to inconsistency, as determined by their consistency ratios. A threshold consistency ratio of 0.1 was applied to ensure the reliability of the responses.

Tab. 3 presents the demographics of the respondents. The majority (76.7 %) were male, with the largest age group being those in their 50s (53.3 %), followed by respondents in their 40s (33.3 %), and those in their 30s (13.3 %). Regarding professional experience, 64.5 % of the experts had at least 10 years in their field, 25.8 % had 15 years, and 9.7 % had over 20 years of experience. This distribution reflects the relatively recent development of blockchain and 5G technologies, resulting in a higher concentration of experts with 10 to 15 years of experience. The respondent pool comprised an equal split of 15 experts in blockchain and 15 in 5G antennas, maintaining a balanced representation of both fields.

Table 3 Demographic data of the respondents involved in the study. It includes information such as the distribution of gender, age groups, and professional experience.

Section	Characters	Frequency	Ratio
Gender	Male	23	76.7
	Female	7	23.3
	Total	30	100
Age	30s	4	13.3
	40s	10	33.3
	50s	16	53.4
	Total	30	100
Work experience in the related field	10-15years	20	64.5
	15-20years	8	25.8
	Over 20years	3	9.7
	Total	30	100
Professional area	Blockchain	15	50.0
	5G antenna	15	50.0
	Total	30	100

4 RESULTS AND DISCUSSION

Using AHP, we integrated blockchain and smart contracts to identify the relative importance of key factors affecting the 5G ceramic antenna manufacturing process, including 'security', 'standards and regulations', and 'efficiency' based on previous research [8]. It was decided as the top factor. The results of this comprehensive analysis are showcased in Tab. 4 and depicted in the visual format as Fig. 4. Among the trio of critical factors identified— 'Security', 'Standard & Regulation', and 'Efficiency'— 'Efficiency' emerged as the most significant, with a relative weight of 0.545. This dominant weight of 'Efficiency' underscores the potential for substantial cost savings and heightened production effective-ness in the manufacturing of 5G ceramic antennas.

Following 'Efficiency' in terms of importance, 'Security' received a considerable weight of 0.318, illustrating its vital role in the overall manufacturing framework. Despite being less influential than 'Efficiency' and 'Security', 'Standard & Regulation' held its ground with a weight of 0.137, indicating its necessary, albeit less prioritized, role in the production process. The pie chart in Fig. 4 provides a clear and concise visual representation of these weights, effectively communicating the hierarchy of factors in the decision-making process. The chart shows 'Efficiency' as the largest segment, colored in grey, denoting its predominant impact. The blue segment, representing 'Security', occupies a significant portion but is noticeably smaller than 'Efficiency'. The smallest slice, in orange, corresponds to 'Standard & Regulation', suggesting it has the least weight among the top factors. The precision of our AHP analysis is further supported by the calculated λ_{max} value of 3.035, with a *CI* of 0.017 and a *CR* of 0.033, against a *RI* of 0.58 for n equals 3. The *CR* value, being lower than 0.1, indicates that the expert opinions gathered for the study are consistent [49].

The strategic application of AHP, as evidenced by the detailed results in Tab. 4 and the illustrative Fig. 3, offers valuable perspectives into prioritizing efforts within the manufacturing sector. It highlights 'Efficiency' as a crucial lever in optimizing the production of 5G ceramic antennas, indicating that investments and improvements in this area are likely to yield the most significant returns in terms of cost-effectiveness and production capacity. This data-driven approach, enhanced by blockchain and smart contract technologies, paves the way for informed decision-making and strategic planning in the rapidly evolving telecommunications industry.

Table 4 The AHP results for the primary dimensions or top-level factors in the context of 5G ceramic antenna manufacturing. It ranks and assigns weights to each top-level factor, such as 'Security', 'Standard & Regulation', and 'Efficiency', demonstrating their relative importance.

Top factor	Rank	Weight
Security	2	0.318
Standard & Regulation	3	0.137
Efficiency	1	0.545
Sum		1.000
CI		0.017
CR		0.033
λ_{max}		3.035

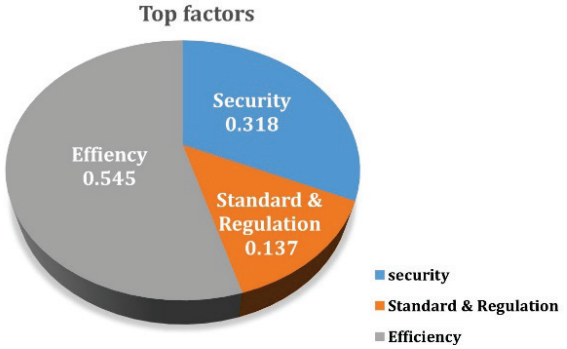


Figure 3 This pie chart representing the AHP results for the weights of top factors in the evaluation of 5G antenna manufacturing processes: 'Security', 'Standard & Regulation', and 'Efficiency'.

The AHP was employed to determine the weights of various sub-factors under the primary factor of 'Security', as

detailed in Tab. 5. This process involved a comprehensive evaluation of the relative importance of these sub-factors. Notably, within the hierarchy of 'Security'-related sub-factors, 'Integrity,' with a weight of 0.349, surfaced as the most crucial element. This high weighting underscores its significant role in the overall security framework for 5G antenna manufacturing. In contrast, 'Security Test' was assigned a comparatively lower weight of 0.126. This indicates that, while still important, it holds less significance relative to other sub-factors in the context of 'Security'. Such differentiation in the weights illustrates the varying levels of importance assigned to each sub-factor and their impact on the overall security of the manufacturing process.

Further validating the reliability and consistency of this AHP analysis are the calculated metrics: the λ_{max} value stood at 5.194, the *CI* was 0.049, and the *CR* was 0.044. These figures are significant because the *CR* is below the 0.1 threshold, implying a high degree of consensus among the expert opinions surveyed. Additionally, the *RI* value for a matrix of dimension 5 (n = 5) was 1.12 [14], further supporting the consistency and reliability of the findings. This thorough AHP approach not only quantifies the importance of various security aspects but also ensures the methodological robustness of the analysis.

Table 5 The AHP calculated weights for sub-factors under the top-level factor of 'Security' in the context of 5G ceramic antenna manufacturing. It ranks each sub-factor and assigns a weight to it, which cumulatively sum to 1.000. Additionally, it includes the λ_{max} value, *CI*, and *CR* to demonstrate the level of agreement among the surveyed experts' opinions, indicating a high level of consistency with a *CR* below 0.1.

Sub-factor	Rank	Weight
Integrity	1	0.349
Confidentiality	2	0.216
Transparency	4	0.143
Security test	5	0.126
Vulnerability assessment	3	0.167
Sum		1.000
CI		0.049
CR		0.044
λ_{max}		5.194

In conducting our study, we applied the AHP to meticulously evaluate the relative importance of various sub-factors falling under the overarching category of 'Standard & Regulation' within the context of 5G antenna manufacturing, as comprehensively delineated in Tab. 6. Our analysis brought to the forefront the sub-factor 'Compliance,' which conspicuously stood out as the most pivotal, evidenced by its highest weight value of 0.456. This prominence underscores 'Compliance's integral role in maintaining stringent standards and regulations in the manufacturing process. Further in the hierarchy of importance, 'Auditing' was assigned a notable weight of 0.174, reflecting its substantial role in ensuring adherence to established protocols and regulations. Closely following were the sub-factors 'Check' and 'Enforcement,' each with respective weights of 0.138 and 0.128. The nearly identical weights of these two sub-factors signify their parallel significance in the regulatory framework, highlighting the need for effective implementation and continuous oversight of standards and practices.

In contrast, the sub-factor 'Monitoring' was ascribed the lowest weight at 0.104. This lower weighting indicates its comparatively reduced priority in the spectrum of 'Standard & Regulation' sub-factors, albeit its relevance remains non-negligible in the overall regulatory landscape of antenna manufacturing. Crucially, the robustness of our AHP analysis is evidenced by the calculated λ_{max} value of 5.202, alongside a *CI* of 0.068 and a *CR* of 0.046. The *RI* value, calculated for an n of 5, stood at 1.12 [14]. The *CR*'s positioning well below the 0.1 benchmark is indicative of a high level of agreement among the expert responses, affirming the methodological soundness and reliability of our analysis. This meticulous assessment and quantification of the sub-factors under 'Standard & Regulation' provide a nuanced understanding of their hierarchical importance. It serves as a guiding framework for decision-makers in the 5G antenna manufacturing industry, enabling them to strategically prioritize and allocate resources to areas that are critical for maintaining high standards and regulatory compliance.

Table 6 The weights and ranks for the sub-factors within the top-level factor of 'Standard & Regulation' as determined by the AHP.

Sub-factor	Rank	Weight
Compliance	1	0.456
Check	3	0.138
Monitoring	5	0.104
Enforcement	4	0.128
Auditing	2	0.174
Sum		1.000
CI		0.068
CR		0.046
λ_{max}		5.202

In our detailed assessment encapsulated in Tab. 7, we meticulously computed the weights by evaluating the relative significance of various sub-factors associated with the top factor, 'Efficiency', in the realm of 5G ceramic antenna manufacturing. This evaluation led to revealing perspectives about the hierarchy of these sub-factors in terms of their influence on manufacturing efficiency. The sub-factor 'Unit Gain' distinctly stood out as the most influential, with a dominant weight of 0.568. This finding underscores the criticality of achieving high efficiency in the unit chip of 5G ceramic antennas, as 'Unit Gain' is a direct measure of an antenna's efficiency in its operational environment. The substantial weight of this sub-factor illuminates its pivotal role in the overall efficiency of the antenna manufacturing process.

Following 'Unit Gain' in terms of importance were 'Integration Gain and Performance Test,' with respective weights of 0.177 and 0.133. These sub-factors highlight the significance of integrating various system components effectively and the necessity of rigorous performance testing to ensure optimal antenna functionality. 'System Gain,' with a weight of 0.123, though ranked lower, still plays an integral role in the efficiency equation, reflecting the benefits accrued from system-level enhancements in the antenna design and manufacturing process.

The analytical robustness of this assessment is further corroborated by key AHP metrics: the λ_{max} value was determined to be 4.144, the *CI* stood at 0.048, and the *CR* was calculated as 0.054. The *RI* for a matrix size of $n = 4$ was 0.9 [14], as noted in Tab. 2. Importantly, with the *CR* being well below the 0.1 threshold, the survey's findings are validated as consistent, reinforcing the reliability of our AHP analysis. In essence, this comprehensive evaluation, as reflected in Tab. 7, not only delineates the relative weights of the Efficiency sub-factors but also provides a clear and consistent framework for prioritizing enhancements in the manufacturing process of 5G ceramic antennas, with a particular emphasis on unit chip efficiency.

Table 7 The AHP calculated weights for sub-factors within the top-level factor of 'Efficiency' in the con-text of 5G ceramic antenna manufacturing.

Sub-factor	Rank	Weight
Unit gain	1	0.568
Integration gain	2	0.177
System gain	4	0.123
Performance test	3	0.133
Sum		1.000
CI		0.048
CR		0.054
λ_{max}		4.144

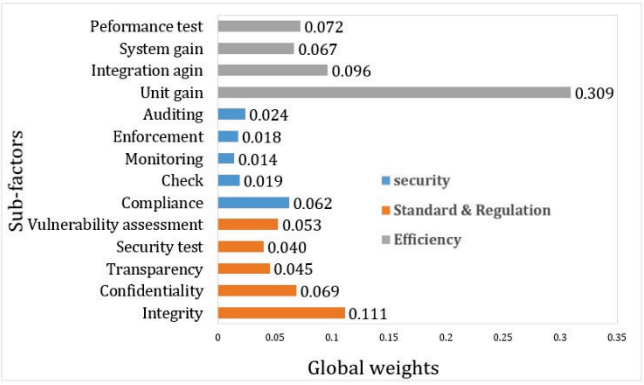


Figure 4 The global weights for various sub-factors as determined by the AHP in the context of 5G ceramic antenna manufacturing. These sub-factors are likely categorized under the main factors of 'Security', 'Standard & Regulation', and 'Efficiency'. The figure provides a visual representation of the relative importance or global weight of each sub-factor, highlighting how they contribute to the overall efficiency and effectiveness of the manufacturing process. This includes a focus on key aspects such as unit gain, integration gain, system gain, and performance test, thereby offering understanding into the areas that are most impactful for optimizing the manufacturing process.

Fig. 4 presents a comprehensive bar chart resulting from the AHP used to deter-mine the global weights of various sub-factors in the manufacturing process of 5G ceramic antennas, focusing on three main categories: 'Security', 'Standard & Regulation', and 'Efficiency'. According to the AHP results depicted in Fig. 4, 'Unit gain' emerges as the most influential sub-factor across all evaluated domains, holding the highest global weight value of 0.309, as detailed in Tab. 8. This sub-factor's prominence underscores its pivotal role in driving cost efficiencies within the antenna manufacturing industry. 'Integrity' follows in significance with a global weight of 0.111, occupying a central position in maintaining robust security protocols. Following in importance are 'Integration gain' and 'Performance test, with global weights of 0.0.096 and 0.0.072 respectively, ranking third and fourth. These factors underscore the pivotal role of effective integration and thorough performance testing in enhancing the efficiency of 5G ceramic antenna production.

Moreover, three sub-factors tied in the fifth rank, all falling under the 'Efficiency' category, reflect the highest weight observed in the top factor, 'Unit gain.' This conveys the significant impact that efficiency improvements in the unit chip of 5G ceramic antennas have on the cost-effectiveness and overall performance of the manufacturing process. These results are instrumental in highlighting the areas where strategic investments and improvements could optimize 5G antenna production, particularly emphasizing the substantial gains in cost reduction that can be achieved by enhancing efficiency in specific areas such as unit chip production, integration processes, and performance testing. The chart serves as a decision-making guide, illustrating the comparative importance of each sub-factor and their collective impact on advancing the manufacturing of 5G antennas.

Table 8 The AHP results, showing the weights of top factors and their associated sub-factors, along with the global weights and global ranks determined from the AHP analysis.

Top factor	Weight	Sub-factor	Local weight	Global weight	Global rank
Security	0.318	Integrity	0.349	0.111	2
		Confidentiality	0.216	0.069	5
		Transparency	0.143	0.045	9
		Security test	0.126	0.040	10
		Vulnerability assessment	0.167	0.053	8
Standard & Regulation	0.137	Compliance	0.456	0.062	7
		Check	0.138	0.019	12
		Monitoring	0.104	0.014	14
		Enforcement	0.128	0.018	13
		Auditing	0.174	0.024	11
Efficiency	0.545	Unit gain	0.568	0.309	1
		Integration gain	0.177	0.096	3
		System gain	0.123	0.067	6
		Performance test	0.133	0.072	4

5 CONCLUSIONS

This study focused on enhancing the efficiency of 5G ceramic antenna manufacturing through the integration of smart contracts and blockchain technology. Employing the AHP, we systematically prioritized key factors— 'Security', 'Standard & Regulation', and 'Efficiency'—in the manufacturing process. Our findings reveal that 'Efficiency is the most crucial factor, with a weight of 0.545, followed by 'Security' (0.318) and 'Standard & Regulation' (0.137). This underscores the importance of focusing on antenna efficiency to drive cost savings in the manufacturing process.

The AHP model identified 'Unit gain' as having the highest global weight, highlighting its critical role in the efficiency of 5G ceramic antenna manufacturing. Additionally, the use of blockchain and smart contracts was found to offer significant benefits in terms of trust enhancement, data integrity, and automation of contractual conditions. The detailed AHP analysis provided a quantitative understanding

of the impact of 'Security', 'Standard & Regulation', and 'Efficiency' on manufacturing costs. The in-depth evaluation of sub-factors within these top categories further refined our understanding of the priorities within each domain.

The consistency of the expert opinions, as evidenced by the λ_{max} values, *CI*, and *CR*, validates the reliability of the AHP methodology used in this study. These findings offer strategic implications for the application of blockchain technology in 5G antenna manufacturing, potentially fostering innovation and development within this sector.

Despite its contributions, this study is not without limitations. The specificity of the research context may limit the generalizability of the findings to different times or manufacturing environments. Variations in expert opinions in the AHP analysis could affect the consistency of the results across diverse expert groups. The AHP model's assumptions may not perfectly align with real-world scenarios, particularly concerning weight estimation and consistency evaluations. Additionally, the practical application of blockchain and smart contract technology may face technical challenges that impact the stability and reliability of these solutions. A drawback of AHP is the absence of an independent accuracy evaluation, meaning the results could be flawed due to the subjective assessments of the user.

External factors such as legal changes and technological advancements could also influence the outcomes, and controlling these variables is challenging. Recognizing these limitations is crucial for expanding the study's scope in future research. Future research will aim for a more nuanced understanding of the technical feasibility of implementing blockchain and smart contracts. We plan to explore solutions to realistic challenges and consider a wider range of technical aspects in 5G ceramic antenna manufacturing and blockchain integration. To validate the effectiveness of our model, empirical field research will be conducted to assess its practical applicability and efficacy.

Acknowledgement

This paper is written with support for research funding from aSSIST University.

6 REFERENCES

[1] Tokkozhina, U., Martins, A. L. & Ferreira, J. C. (2023). Uncovering dimensions of the impact of blockchain technology in supply chain management. *Operations Management Research*, 16, 99-125. https://doi.org/10.1007/s12063-022-00273-9

[2] Kumar, A., Kumar, V. R. P., Dehdasht, G., Mohandes, S. R., Manu, P. & Rahimian, F. P. (2023). Investigating barriers to blockchain adoption in construction supply chain management: A fuzzy-based MCDM approach. *Technological Forecasting & Social Change*, 196, 122849. https://doi.org/10.1016/j.techfore.2023.122849

[3] Hasan, M. K., Akhtaruzzaman, M., Kabir, S. R., Gadekallu, T. R., Islam, S., Magalingam, P. & Hassan, R. (2022). Evolution of Industry and Blockchain Era: Monitoring Price Hike and Corruption Using BIoT for Smart Government and Industry 4.0. *IEEE Trans-actions on Industrial Informatics*, 12, 9153-9161. https://doi.org/10.1109/TII.2022.3164066

[4] Albadri, H. (2022). Framework model to enhance the Effectiveness of Blockchain Technology through the Knowledge Management Processes. *Tehnički glasnik*, 16(3), 293-298. https://doi.org/10.31803/tg-20220305123321

[5] Kushwaha, S. S., Joshi, S., Singh, D., Kaur, M. & Lee, H. N. (2022). Systematic Review of Security Vulnerabilities in Ethereum Block-chain Smart Contract. *IEEE Access*, 10, 6605-6621. https://doi.org/10.1109/ACCESS.2021.3140091

[6] Zubaydi, H. D., Varga, P. & Molnar, S. (2023). Leveraging Blockchain Technology for Ensuring Security and Privacy Aspects in Inter-net of Things: A Systematic Literature Review. *Sensors*, 23, 788. https://doi.org/10.3390/s23020788

[7] An, S. Y. & Kim, B. (2023). Smallest form-factor multi-stacked ceramic patch antenna on package for 5G millimeter-wave mobile applications. *Photonics and Nanostructures – Fundamentals and Applications*, 57, 101189. https://doi.org/10.1016/j.photonics.2023.101189

[8] An, S. Y., Ngayo, G. & Hong, S. P. (2023). Efficient 5G Ceramic Antenna Manufacturing Architecture using Blockchain and Smart Con-tracts. *J. Adv. Navig. Technol*., 27(5), 594-609.

[9] Nahar, T. & Rawat, S. (2022). Efficiency enhancement techniques of microwave and millimeter-wave antennas for 5G communication: A survey. *Transactions on Emerging Telecommunications Technologie*, 33, 4530. https://doi.org/10.1002/ett.4530

[10] Watanabe, A. O., Ali, M., Sayeed, S. Y. B., Tummala, R. R. & Pulugurtha, M. R. (2021). A Review of 5G Front-End Systems Package Integration. *IEEE Transactions on Components, Packaging and Manufacturing Technology, 11*(1), 118-133. https://doi.org/10.1109/TCPMT.2020.3041412

[11] Saaty, T. (1980). *The Analytic Hierarchy Process*. McGraw-Hill: New York. https://doi.org/10.21236/ADA214804

[12] Saaty, T. (2000). *Fundamentals of Decision Making and Priority Theory with the Analytic Hierarchy Process*. RWS Publications; ISBN: 978-0962031762.

[13] Saaty, T. (2008). Decision making with the analytic hierarchy process. *International Journal of Services Sciences*, 1, 83-98. https://doi.org/10.1504/IJSSCI.2008.017590

[14] Saaty, T. (2004). Decision-Making with the AHP: Why Is the Principal Eigenvector Necessary. *Eur. J. Oper. Res*., 145, 85-91. https://doi.org/10.1016/S0377-2217(02)00227-8

[15] Saaty, T. (1989). *Group Decision Making and the AHP*, Springer; Berlin/Heidelberg. https://doi.org/10.1007/978-3-642-50244-6_4

[16] Saaty, T. (1990). An exposition of the AHP in reply to the paper: Remarks on the analytic hierarchy process. *Management Science*, 36(3), 259-268. https://doi.org/10.1287/mnsc.36.3.259

[17] Li, X. & Mo, X. (2020). Application of AHP Based on Mathematical Operational Research in Teaching Evaluation System. *Journal of Physics: Conference Series*, 1650, 032015. https://doi.org/10.1088/1742-6596/1650/3/032015

[18] Subramanian, N. & Ramanathan, R. A. (2012). Review of applications of Analytic Hierarchy Process in operations management. *Int. J. Production Economics*, 138, 215-241. https://doi.org/10.1016/j.ijpe.2012.03.036

[19] Lee, H. & Kim, B. (2022). Invigorating Care Farm Ecosystem Based on Public Service Innovation: Case of South Korea. *Adm. Sci*., 12, 58. https://doi.org/10.3390/admsci12020058

[20] Ahn, B. & Kim, B. (2023). A Decision-Making Model for Selecting Product Suppliers in Crop Protection Retail Sector. *Adm. Sci*., 13, 97. https://doi.org/10.3390/admsci13040097

[21] Choi, J. R., Sung, J. H., Kang, D. H., Choo, K. S. & Kim, B. S. (2023). A Study on the Priority Selection Method for Underground Dam Installation Considering Humanities and

Social Factors Using Fuzzy Analytic Hierarchy Process in Korea. *Water*, 15, 3296. https://doi.org/10.3390/w15183296
[22] Dozic, S., Babic, D., Kalic, M. & Zivojinovic, S. (2023). An AHP approach to airport choice by freight forwarder. *Sustainable Futures*, 5, 100106. https://doi.org/10.1016/j.sftr.2023.100106
[23] Samanaseh, V., Noor, Z. Z. & Mazlan, M. A. (2023). The application of analytic hierarchy process for innovative solution: A review. *IOP Conf. Series: Earth and Environmental Science*, 1143, 012022. https://doi.org/10.1088/1755-1315/1143/1/012022
[24] Moslem, S., Saraji, M. K., Mardani, A., Alkharabsheh, A., Duleba, S. & Esztergar-Kiss, D. (2023). A Systematic Review of Analytic Hierarchy Process Applications to Solve Transportation Problems: From 2003 to 2022. *IEEE Access*, 11, 11973. https://doi.org/10.1109/ACCESS.2023.3234298
[25] Liberatore, M. J. (1987). An Extension of the Analytic Hierarchy Process for Industrial R&D Project Selection and Resource Allocation. *IEEE Transactions on Engineering Management*, 34(1), 12-18. https://doi.org/10.1109/TEM.1987.6498854
[26] Tavana, M., Soltanifar, M. & Santos-Arteaga, F. J. (2023). Analytical hierarchy process: revolution and evolution. *Annals of Operations Research*, 326, 879-907. https://doi.org/10.1007/s10479-021-04432-2
[27] Liang, W. Y. (2003). The analytic hierarchy process in project evaluation An R&D case study in Taiwan. *Benchmarking: An International Journal*, 10(5), 445-456. https://doi.org/10.1108/14635770310495492
[28] Nakamoto, S. (2008). Bitcoin: A peer-to-peer electronic cash system. Available at https://bitcoin.org/en/bitcoin-paper
[29] Ballandies, M. C., Dapp, M. M. & Pournaras, E. (2022). Decrypting distributed ledger design—taxonomy, classification and blockchain community evaluation. *Cluster Computing*, 25, 1817-1838. https://doi.org/10.1007/s10586-021-03256-w
[30] Romano, D. & Schmid, G. (2021). Beyond Bitcoin: Recent Trends and Perspectives in Distributed Ledger Technology. *Cryptography*, 5, 36. https://doi.org/10.3390/cryptography5040036
[31] Chowdhury, E. K., Stasi, A. & Pellegrino, A. (2023). Blockchain Technology in Financial Accounting: Emerging Regulatory Issues. *Review of Economics and Finance*, 21, 862-868.
[32] Huang, Y. & Chen, J. (2023). Analysis of Development of Blockchain Technology Supply Chain Finance Based on PEST-SWOT Model. *Academic Journal of Business & Management*, 5(8), 27-33. https://doi.org/10.25236/AJBM.2023.050805
[33] Assiri, B. (2023). A Modified and Effective Blockchain Model for E-Healthcare Systems. *Appl. Sci*., 13, 12630. https://doi.org/10.3390/app132312630
[34] Siyal, A. A., Junejo, A. Z., Zawish, M., Ahmed, K., Khalil, A. & Soursou, G. (2019). Applications of Blockchain Technology in Medicine and Healthcare: Challenges and Future Perspectives. *Cryptography*, 3, 3. https://doi.org/10.3390/cryptography3010003
[35] Chen, C., Huang, H., Zhao, B., Shu, D. & Wang, Y. (2023). The Research of AHP-Based Credit Rating System on a Blockchain Application. *Electronics*, 12, 887. https://doi.org/10.3390/electronics12040887
[36] Grida, M. O., Elrahman, S. A. & Eldrandaly, K. A. (2023). Critical Success Factors Evaluation for Blockchain's Adoption and Implementing. *Systems*, 11, 2. https://doi.org/10.3390/systems11010002
[37] Lv, W., Qiu, X. & Meng, L. (2022). Blockchain localization spoofing detection based on fuzzy AHP in IoT systems. *J Wireless Com Network*, 2022, 13. https://doi.org/10.1186/s13638-022-02094-7
[38] Gutierrez O. C. U. & Xu, G. (2023). Blockchain and Smart Contracts to Secure Property Transactions in Smart Cities, *Appl. Sci*., 13, 66. https://doi.org/10.3390/app13010066
[39] Trivedi, P., Shah, J., Moslem, S. & Pilla, F. (2023). An application of the hybrid AHP-PROMETHEE approach to evaluate the severity of the factors influencing road accidents. *Heliyon*, 9, e21187. https://doi.org/10.1016/j.heliyon.2023.e21187
[40] Szabo, N. (1994). Smart Contracts. Available online: https://nakamotoinstitute.org/smart-contracts/.
[41] Khan, S. N., Loukil, F., Ghedira-Guegan, G., Benkhelifa, E. & Bani-Hani, A. (2021). Blockchain smart contracts: Applications, challenges, and future trends. *Peer-to-Peer Networking and Applications, 14*, 2901-2925. https://doi.org/10.1007/s12083-021-01127-0
[42] Ullah, F. & Al-Turjman, F. (2023). A conceptual framework for blockchain smart contract adoption to manage real estate deals in smart cities. *Neural Computing and Applications*, 35, 5033-5054. https://doi.org/10.1007/s00521-021-05800-6
[43] Kaur, K., Kumar, S. & Baliyan, A. (2020). 5G: a new era of wireless communication. *International Journal of Information Technology*, 12, 619-624. https://doi.org/10.1007/s41870-018-0197-x
[44] Sharma, S., Deivakani, M., Reddy, K. S., Gnanasekar, A. K. & Aparna, G. (2021). Key Enabling Technologies of 5G Wireless Mobile Communication. *Journal of Physics: Conference Series*, 1817, 012003. https://doi.org/10.1088/1742-6596/1817/1/012003
[45] Tarver, C., Tonnemacher, M., Chen, H., Zhang, J. & Cavallaro, J. R. (2021). GPU-Based, LDPC Decoding for 5G and Beyond. *Circuits and Systems*, 2, 278-290. https://doi.org/10.1109/OJCAS.2020.3042448
[46] Wali, S. Q., Sali, A., Allami, J. K. & Osman, A. F. (2022). RF-EMF Exposure Measurement for 5G over Mm-Wave Base Station with MIMO Antenna. *IEEE Access*, 10, 9048. https://doi.org/10.1109/ACCESS.2022.3143805
[47] Lee, J., Hong, Y. K., Bae, S., Abo, G. S., Seong, W. M. & Kim, G. H. (2011). Miniature Long-Term Evolution (LTE) MIMO Ferrite Antenna. *IEEE Antennas and Wireless Propagation Letters*, 10, 603-606. https://doi.org/10.1109/LAWP.2011.2159468
[48] Lee, W., Hong, Y. K., Park, J., LaRochelle, G. & Lee, J. (2016). Low-loss Z-type hexaferrite (Ba3Co2Fe24O41) for GHz antenna applications. *J. Magn. Magn. Mater*., 414, 194-197. https://doi.org/10.1016/j.jmmm.2016.04.075
[49] Thakkar, J. J. (2021). *Analytic Hierarchy Process (AHP). In: Multi-Criteria Decision Making. Studies in Systems, Decision and Control*, 336, 33-62. Springer, Singapore. https://doi.org/10.1007/978-981-33-4745-8_3

Authors' contacts:

Sung Yong An, PhD
Seoul Business School, aSSIST University,
46 Ewhayeodae 2-gil, Fintower, Sinchon-ro, Seodaemun-gu, Seoul, 03767, Korea
sung.an@stud.assist.ac.kr

Guy Ngayo, Prof.
School of Business, Franklin University Switzerland,
Via Ponte Tresa 29 • 6924 Sorengo (Lugano), Switzerland
guyngayo@hotmail.com

Seng-Phil Hong, Prof.
(Corresponding author)
School of Computing and AI, HanShin University,
137 Hanshindae-gil, Osan-si, Gyenggi-Do, 18101, Korea
sphong@hs.ac.kr

ISSN 1846-6168 (Print), ISSN 1848-5588 (Online)
https://doi.org/10.31803/tg-20240229101629

Preliminary communication
Received: 2024-02-29, Accepted: 2024-04-15

Utilizing Digital Therapeutics Technology to Improve Willingness to Actively Participate in Gait Rehabilitation Training and Increase Interest in Treatment

Hwamok Koo, Junsik Moon, Sung Teac Hwang, Dong Keun Kim*

Abstract: In this study, we demonstrated the feasibility of 'Morning Walk S200' as a rehabilitation training tool using an eXtended Reality (XR) system for those who are in need of gait rehabilitation. The lower limb rehabilitation training robot performs repetitive training according to the specified input gait trajectory. While the robot constructs passive training, it is crucial to encourage patients to engage in active training in order to enhance treatment effectiveness. 'Morning Walk S200' utilizes a diverse range of VR (Virtual Reality) experiences—such as walking through Saryeoni Forest in Jeju Island, landscaping, a baseball batting game, and the Blue and White Flag game—all designed to improve patients' active participation and increase interest in their treatment. Additionally, this study includes several papers that validate the technical efficiency of "Morning Walk S200" and examine the VR content. We suggest that developing innovative and engaging treatment options is essential, highlighting the need to utilize VR content effectively to enhance the outcomes of gait rehabilitation training.

Keywords: Gait; Rehabilitation Robot; Robot-Assisted Training; Virtual Reality (VR)

1 INTRODUCTION

Digital therapeutics refers to a specific type of software that is designed to treat diseases and improve health. This generally includes bio-healthcare or medical devices that utilize applications, games, and virtual reality or augmented reality (VR/AR) technologies [1]. Numerous studies have been conducted on the treatment and rehabilitation of vestibular dysfunction using virtual reality and Head-Mounted Displays (HMDs). A study has been published showing the effectiveness and safety of vestibular rehabilitation training using a gaming display created with Unity [2]. Dascal's research also validated the utility and efficacy of VR for three major medical conditions: pain management, eating disorders/obesity, and cognitive/rehabilitative exercises among hospitalized patients from 2005 to 2015 [3]. A study on "Effects of a virtual reality exercise program using video games on the motor performance of the elderly", proved that virtual reality exercise programs are significantly effective in improving strength, balance, and gait in the elderly [4]. Therefore, this paper aims to introduce the 'Morning Walk S200' developed by Curexo™, along with various VR contents used for VR rehabilitation therapy, one of the main technologies in modern rehabilitation therapy.

2 THEORETICAL BACKGROUND

2.1 Digital Therapeutics (DTx)

An analysis of domestic research trends related to the development of Digital Therapeutics (DTx) in the field of communication disorders defines digital therapeutics as software medical devices that provide evidence-based therapeutic interventions to patients to prevent, manage, and treat medical disorders or diseases [24]. According to Samjung KPMG's Investment Trends and Future Strategies for Third-Generation New Drug Digital Therapeutics, the following five technologies are considered the key digital therapeutics [5]:

1) Mobile and PC-based Apps: Therapeutic applications are provided as mobile and PC-based applications with a variety of content. These apps represent the most basic form of digital therapy.
2) Virtual and Augmented Reality (VR and AR): Virtual Reality Therapy (VRT) employs the use of VR technology for therapeutic purposes, including psychotherapy, neurological rehabilitation, and musculoskeletal rehabilitation.
3) AI/Big Data: Healthcare big data can be analyzed to recognize specific patterns that help predict and diagnose diseases or provide personalized treatment plans for patients.
4) Brain-Computer Interface (BCI): A collection of interface technologies that utilize brain waves to control computers or machines. The goal is to improve neurological communication between the brain and muscles in conditions such as Lou Gehrig's disease, stroke, spinal cord injury, and cerebral palsy.
5) Games: Utilizes adaptive algorithms to provide a personalized treatment and rehabilitation experience for patients, incorporating various effects and rewards to engage patients and enhance their motivation.

As such, the continuous evolution of devices and networks has enabled VR rehabilitation therapy to be delivered. Digital Therapeutics (DTx) has advantages over classical therapies [22]. First, DTx can be delivered digitally, free from space and time constraints. Second, as a software-based therapy, DTx can quantify and standardize both the data from the patient and the data from the expert physician. Third, DTx is a personalized service and treatment.

2.2 Extended Reality (XR)

Immersive content is defined as digital media that employs immersive technologies to engage the human senses, providing information and creating a nearly realistic sense of presence for users [23]. Extended Reality is one form of immersive content, which combines Virtual Reality and Augmented Reality to fully immerse users within a virtual

environment. In addition to VR and AR, games and PC-based apps are the key technological elements of digital therapeutics. When these elements are fused and expanded, they leverage XR technologies. XR digital therapies show great promise for rehabilitation training, particularly for people who acquire gait and mobility abnormalities.

2.3 HMD

To experience VR, AR, and MR, you need a specialized device. Among the various devices, the most popular and fastest-growing device is the Head Mounted Display. These devices are worn over the head, immersing you through a new virtual world right in front of your eyes. Google was the first company to enter the HMD market, followed by Microsoft, Meta (formerly Facebook), Samsung, and others. Additionally, various innovative HMD devices are being developed exponentially. The platform market is evolving through a sequence of advancements in computers, the Internet, and mobile devices, with HMDs now emerging as the fourth major technological platform. VR and AR technologies are developing at a rapid pace and converging across various fields. In the healthcare sector, these VR and AR technologies are being utilized, including the following [25]:

- Life Care Sector: Healing VR content, Sports VR, Health Avatars, etc.
- Healthcare: 360-degree Surgical Video Services, Dental Simulators, Rehabilitation, etc.

Through these HMD devices, 'Morning Walk S200' offers users an enhanced sense of immersion in virtual reality, potentially increasing their interest and commitment to rehabilitation training.

2.4 Robot Rehabilitation

Rehabilitation is divided into physical therapy, occupational therapy, cognitive therapy, speech therapy, psychotherapy, pain therapy, and robotic therapy. Furthermore, robotic rehabilitation is divided into therapeutic goals, as shown in Fig. 1.

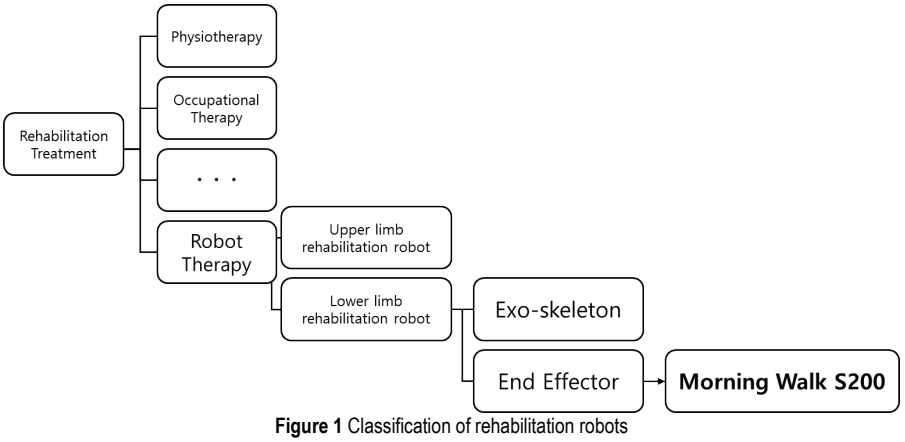


Figure 1 Classification of rehabilitation robots

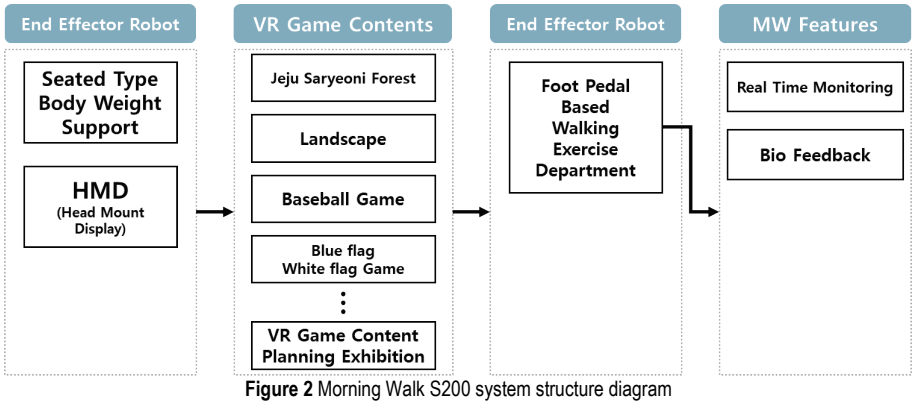


Figure 2 Morning Walk S200 system structure diagram

Rehabilitation robots are categorized into two types, upper-limb rehabilitation robots and lower-limb rehabilitation robots. One example of a lower-limb rehabilitation robot is ‘Morning Walk S200’, which focuses on gait rehabilitation targeting the lower limb. This training is conducted through an input protocol programmed into the robot. In order to perform repetitive and intense gait training without additional space, it is essential for the patient to actively participate in the training as the robot follows a specific gait trajectory. To encourage active participation, ‘Morning Walk S200’ utilizes VR content, which enhances patient participation and increases their interest in therapy [11-13]. This study introduces ‘Morning Walk S200’ and describes how it utilizes VR as a key technological element of digital therapeutics in rehabilitation.

3 SYSTEM DESIGN

3.1 System Structure

Fig. 2 is a simplified system block diagram of the 'Morning Walk S200'. 'Morning Walk S200' helps patients of all ages, from those who cannot walk to those who need gait rehabilitation training. After recognizing the user's motion using the Kinetic system camera loaded on the 'Morning Walk S200', the user must board on top of the robot. The user (patient) can receive VR therapy through a separate HMD and a screen placed in front of the user.

3.2 Use Case Diagram

Through 'Morning Walk S200', users can experience various VR game contents to enhance the effectiveness of rehabilitation therapy. Fig. 3 is a diagram of the use cases and requirements of these users. It is a simple summary of the requirements that may arise between professionals and patients who assist in rehabilitation training.

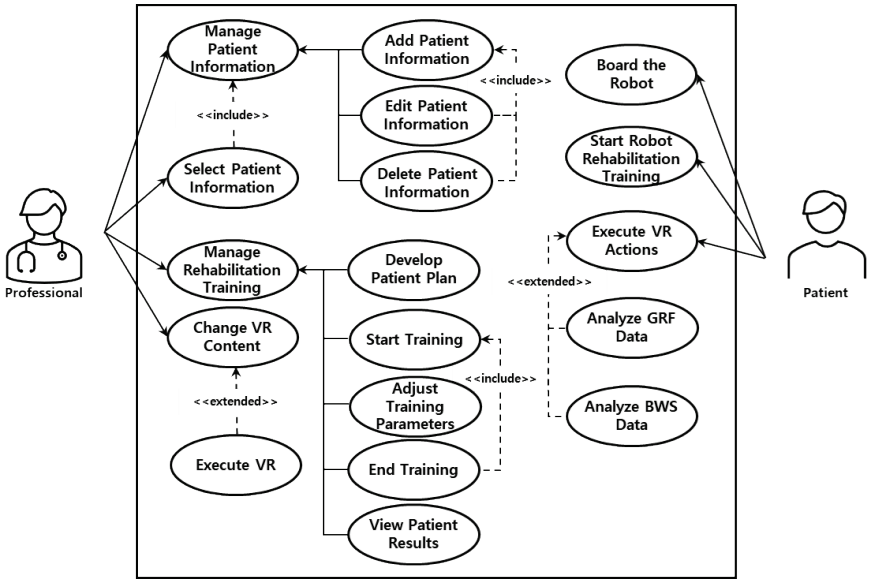


Figure 3 Use case diagram: user requirements

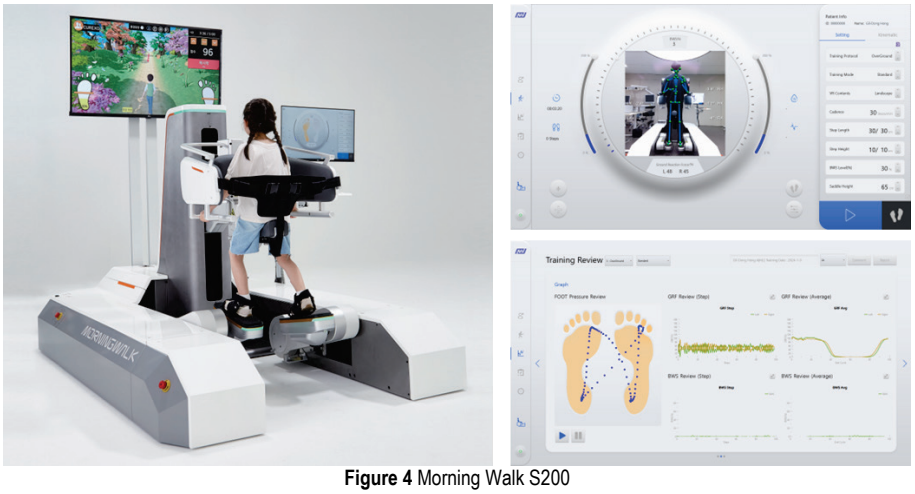


Figure 4 Morning Walk S200

The therapist conducts four requirements (Fig. 3): Manage patient information, select patient information, manage training, and change VR content. The therapist can *Add/Edit/Delete Patient Information* by himself. They can also *Plan/Start/Adjust the Parameters*, End the Training, and *View Results* for further *Training Management*. The patient executes three requirements: *Boarding the robot, training the robot, and Participating the VR events*. By boarding the robot, the patient provides GRF (Ground Reaction Force) and BWS (Body Weight Support) data to the robot. Then, the robot passes the information to the VR content, and the VR content analyses the data. Lastly, based on the analyzed results, the VR content events implement.

3.3 Morning Walk S200

Morning Walk provides a real-time monitoring and biofeedback system. You can check joint angle values for hip, knee, and ankle in real time [7, 14]. You can also monitor GRF, BWS, plantar pressure distribution, and balance through the Butterfly Diagram [7].

Fig. 4 is 'Morning Walk S200', a robotic orthotic

exercise device and walking rehabilitation robot developed by Curexo. It is a compound word of Morning + Walk which signifies the meaning of presenting (rewarding) a comfortable morning walk [6]. It is a medical robot developed to help patients with walking difficulties recover and provides optimized training for children and adults [19, 21]

There are two types of lower limb rehabilitation robots: the exoskeleton type, which is worn directly on the patient's body, and the end effector type, which simply connects the end of the device to the patient. 'Morning Walk S200' is an end effector type, which means it has a high degree of freedom in the movement of the center of gravity, unlike the exoskeleton type, which limits the range of motion of the center of gravity during gait training. End-effector-type walking robots are more suited and safer for various neurological diseases [8, 18]. Gait training with Morning Walk S200 has been proven to improve mobility and balance in subacute and hemiplegic patients [9, 16, 17], and the scaffold-based gait component has been proven to be a valid rehabilitation method for lower extremity movement [10, 15, 20, 26].

4 VR CONTENT ON THE MORNING WALK S200

Fig. 5 shows the VR contents that are officially supported by Morning Walk S200. We currently support five VR content and are continuously conducting content research and development. The VR content utilizes the trainer's foot pressure and BWS values to generate events and interact with the user. The VR content we currently support was created based on ideas developed internally at Curexo in collaboration with VR development companies.

Figure 5 Morning Walk S200 - officially supported VR content

Figure 6 Morning Walk S200 - training with an HMD

As shown in Fig. 6, the patient checks the VR content on the front screen after boarding the Morning Walk. The VR content is linked to 'Morning Walk S200' and receives data input in real-time. Each VR content has a 'Foot Calibration' function, which analyzes the patient's average foot pressure and sets that value as the default value. The foot pressure value is then read in real-time and used to trigger events in each content.

Character movement in certain VR content can be set to behave normally when the patient's foot pressure is higher than the average foot pressure. The patient actively participates in the training to activate the normal behavior of the VR content character. In addition, there are events such as eating items in the content map and completing timely events to enhance the patient's interest. We are researching various events to actively engage patients. Morning Walks are conducted after wearing the HMD. However, patients who are not familiar with 3D content may experience discomforts such as motion sickness, so it is used selectively. Rehabilitation training is conducted through 360° VR content when wearing an HMD. The screen that the patient wearing the HMD sees is simultaneously transmitted in real-time to the TV in front of the Morning Walk equipment.

Patients who have difficulty performing normal movements can adjust the level of training by lowering the level of abnormal gait or adjusting the robot's movement speed in the VR content settings. Fig. 7 is a VR content that depicts the Saryeoni Forest in Jeju Island. It is possible to enjoy the scenery without any specific events, so even patients with difficulty walking can enjoy VR content. The map progresses in the following order: Rain Forest, Pine tree,

Redwood, Valley, Flower field, Winding path, Waterfall, or Lake. When you first start, you can specify where you want to start and progress through the training.

Figure 7 Morning Walk VR - Saryeoni Forest, Jeju Island

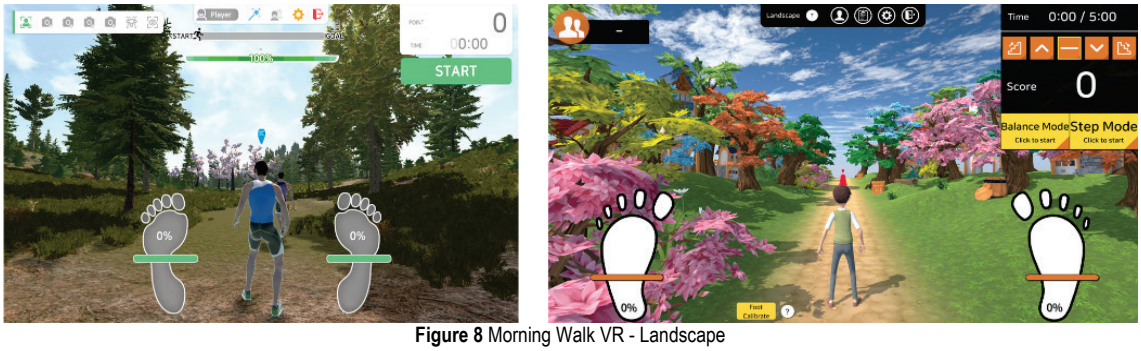


Figure 8 Morning Walk VR - Landscape

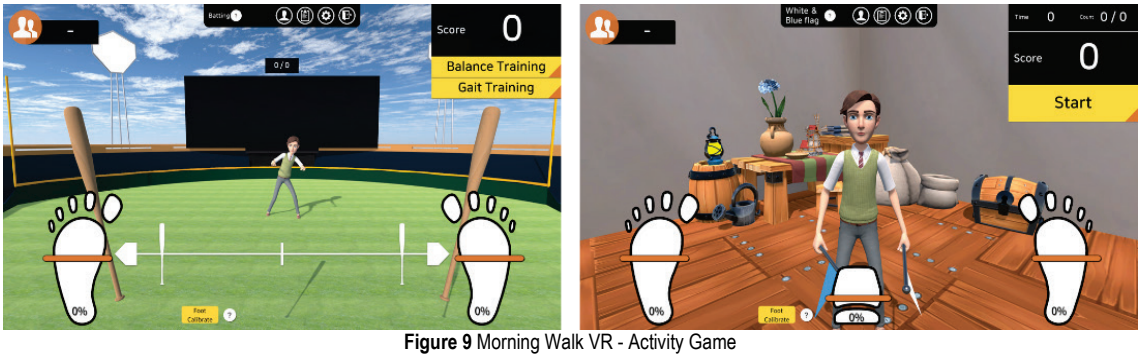


Figure 9 Morning Walk VR - Activity Game

Fig. 8 is the Morning Walk VR content 'Landscape'. Morning Walk has a total of five protocols (Over Ground, Stair Up, Stair Down, Slope Up, Slope Down). The VR content automatically changes to the appropriate terrain for the protocol when the protocol is changed. The patient gets the feeling that they are actually performing the protocol in the robot's behavior as the appropriate terrain appears. Stair Up/Down provides the ability to adjust the stair height value and Slope Up/Down provides the ability to adjust the slope angle value, providing a wide range of training protocols for patients.

Fig. 9 is an Activity Game from Morning Walk. On the left is a baseball game, and on the right is a blue-and-white flag game. The baseball game tells you when the ball will fly during the exercise, and if you step on the footstep at the right time, you will hit the ball. In Blue and White Flag, you will see phrases like "Blue Flag up - Blue Flag down" and "White Flag up - White Flag down" on the screen. If the player performs the command within a certain amount of time after the phrase appears, they will succeed. Active games help patients focus on their rehabilitation training. Unlike other games, the Blue-White Flag game has a saddle event: sit/stand up.

Depending on the patient's severity level, the intensity, duration, and number of exercises vary. Each VR content repeats a unique event without any specific reward, such as

giving items to customize the character or providing in-game currency to purchase specific items. In this case, patients who are training for an extended period may feel bored with the VR content. Therefore, in order to provide a variety of VR content, Curexo is developing its own VR content as well as holding VR content planning contests.

Figure 10 Winners of the Morning Walk VR game content planning contest

Fig. 10 is the winner of the first Morning Walk VR game content planning contest. Through the content planning contest, we got various ideas such as a game that catches grapes falling from the sky in a bowl, a skiing game, and a painting game. A total of forty-five teams participated in the contest, and eight teams were selected to receive awards. These ideas will be utilized for future Morning Walk VR content development.

5 CONCLUSION AND FUTURE RESEARCH

The "Morning Walk S200" is a walking rehabilitation robot developed by Curexo, designed as an end-effector type. It features a weight-bearing seat, a footrest for walking exercises, real-time monitoring, biofeedback, and the provision of virtual reality (VR) content. The VR content aims to engage patients actively in their training and enhance their interest in therapy. In this paper, we introduce the "Morning Walk S200" and discuss the utilization of VR as a key technological element in digital therapies for rehabilitation.

As the aging society accelerates around the world and the elderly population increases, the number of falling accidents, cognitive dissonance, and various chronic diseases is increasing. Thus, hospitals and several research institutes are developing various rehabilitation treatment devices that combine IT technology and Rehabilitation robots, and many researchers have established to prove the effectiveness of these treatments [27 - 29]. Furthermore, as the market develops for contents installed in treatment devices, it creates greater benefits that could increase treatment effects [30 - 34]. Hence, we hope this paper will be used as future reference research material to follow when developing digital healthcare industry content.

Based on this study, we set three future research directions. First, research on design techniques for VR content in rehabilitation therapy. As the purpose of utilizing VR content is to improve patients' willingness to actively participate in rehabilitation training and enhance their interest in treatment, the content should not be complex or colorful to lose its main purpose. Therefore, we would like to conduct a study on how to design VR content that can enhance patients' active participation in training and interest in treatment without losing the main purpose of rehabilitation training.

Second, we will study how to design XR content optimized for rehabilitation therapy. Patients undergoing rehabilitation therapy board a morning walker and train while looking only at the VR content on the monitor. It is believed that providing an XR world rather than just a VR world will allow patients to participate more actively in rehabilitation and increase their interest in treatment than VR alone. Therefore, we would like to conduct research on XR content design techniques optimized for general-purpose gait rehabilitation training.

Third, we need to create a formal evaluation method for VR or XR content. Currently, the experience of using Morning Walk's VR content is only conveyed verbally to patients and therapists. Without a structured evaluation method in place, it is challenging to enhance the quality of the content. Therefore, we propose conducting a study to develop a structured evaluation method for VR or XR content, which will help us design content optimized for rehabilitation therapy.

6 REFERENCES

[1] Koo, Y.-D. (2022). Digital Therapeutics. Korea Institute of Science and Technology Information. https://repository.kisti.re.kr/bitstream/10580/18042/1/ASTI%20MARKET%20INSIGHT%20095%280712%29.pdf

[2] Nehrujee, A. et al. (2019). A Smartphone-based Gaming System for Vestibular Rehabilitation: A Usability Study. *J*

Vestib Res., 29(2-3), 147-160. https://doi.org/10.3233/VES-190660
[3] Dascal, J., Reid, M., IsHak, W. W., Spiegel, B., Recacho, J., Rosen, B. & Danovitch, I. (2017). Virtual Reality and Medical Inpatients: A Systematic Review of Randomized, Controlled Trials. *Innovations in Clinical Neuroscience, 14*(1-2), 14-21.
[4] Song, C.-H., Shin, W.-S., Lee, K.-J. & Lee, S.-W. (2009). Effects of a virtual reality exercise program using video games on muscle strength, balance, and gait in older adults. *Korean Gerontology, 29*(4), 1261-1275.
[5] Samjung KPMG. (2023). Investment Trends and Future Strategy for 3rd Generation Reagent Digital Therapeutics. https://assets.kpmg/content/dam/kpmg/kr/pdf/2023/business-focus/kr-bf-2023-digital-therapeutics_20230105.pdf
[6] CUREXO INC. Gait Rehabilitation Robot Overview. https://www.curexo.com/medical/sub02.php?kind=1
[7] CUREXO INC. Gait Rehabilitation Robot Product Composition and Features. https://www.curexo.com/medical/sub02.php?kind=2
[8] Choi, S.-J. et al. (2017). Feasibility of Robot assisted Gait Training with End effector Type in Various Neurologic Disorders. *Korean Academy of Rehabilitation Medicine Fall Conference*, OP4-2-7,
[9] Kim, J.-Y. et al. (2017). The Effect of Morning Walk Assisted Gait Training for Patients with Hemiparesis due to Stroke. *Korean Academy of Rehabilitation Medicine Fall Conference*, OP2-4-9.
[10] Mehrholz, J. & Pohl, M. (2012). Electromechanically-assisted gait training after stroke: a systematic review comparing end-effector and exoskeleton devices. *J Rehabil Med., 44*, 193-199. https://doi.org/10.2340/16501977-0943
[11] Park, M. G., Ha, Y. M., Cho, H. J. & Jeon, M. Y. (2021). The Effects of Robot-Assisted Rehabilitation on the Gait Ability of Stroke Patients with Hemiplegia: A Mixed Methods Research Study. *Korean Society of Biological Nursing Science, 23*(1), 72-82. (in Korean) https://doi.org/10.7586/jkbns.2021.23.1.72
[12] Rhee, S. Y., Jeon, H., Kim, S. W. & Lee, J. S. (2020). The effect of an end-effector type of robot-assisted gait training on patients with Guillain-Barre syndrome: a cross-sectional study. F1000Res. https://doi.org/10.12688/f1000research.26246.1
[13] Park, S.-Y. (2015). Effectiveness of robotic gait rehabilitation in stroke patients: A meta-analysis. *Korean Journal of Specialized Physical Therapy, 22*(2), 30-40. https://doi.org/10.12674/ptk.2015.22.2.030
[14] Bessler, J., Prange-Lasonder, G. B., Schulte, R. V., Schaake, L., Prinsen, E. C. & Buurke, J. H. (2020). Occurrence and Type of Adverse Events during the Use of Stationary Gait Robots-A Systematic Literature Review. *Frontiers in robotics and AI, 7*, 557606. https://doi.org/10.3389/frobt.2020.557606
[15] Hwang, C. H. (2019). Immediate muscle strengthening by an end-effector type gait robot with reduced real-time use of leg muscles: A case series and review of literature. *World journal of clinical cases, 7*(19), 2976-2985. https://doi.org/10.12998/wjcc.v7.i19.2976
[16] Jung, M., Kim, D. & Chun, M. (2018). Effect of Robot-Assisted Gait Training in Patients with Gait Disturbance Caused by Brain Tumor: a Case Series. *Brain & NeuroRehabilitation, 11*(2), e21. https://doi.org/10.12786/bn.2018.11.e21
[17] Do, K.-H. & Chun, M.-H. (2017). Clinical Use of Robots as a Part of Rehabilitation Medicine. *Brain & NeuroRehabilitation, 10*(1), e7. https://doi.org/10.12786/bn.2017.10.e7
[18] Mehrholz, Jan & Pohl, Marcus. (2012). Electromechanical-assisted gait training after stroke: A systematic review comparing endeffector and exoskeleton devices. *Journal of rehabilitation medicine: official journal of the UEMS European Board of Physical and Rehabilitation Medicine, 44*, 193-199. https://doi.org/10.2340/16501977-0943
[19] Koo, K. I. & Hwang, C. H. (2020). Five-day rehabilitation of patients undergoing total knee arthroplasty using an end-effector gait robot as a neuromodulation blending tool for deafferentation, weight offloading and stereotyped movement: Interim analysis. *PloS one, 15*(12), e0241117. https://doi.org/10.1371/journal.pone.0241117
[20] Song, K. J., Chun, M. H., Lee, J., & Lee, C. (2021). The effect of robot-assisted gait training on cortical activation in stroke patients: A functional near-infrared spectroscopy study. *NeuroRehabilitation, 49*(1), 65-73. https://doi.org/10.3233/NRE-210034
[21] Rhee, S. Y., Jeon, H., Kim, S. W. & Lee, J. S. (2020). The effect of an end-effector type of robot-assisted gait training on patients with Guillain-Barre syndrome: a cross-sectional study. F1000Res. https://doi.org/10.12688/f1000research.26246.1
[22] Son, H., Lee, J., Choi, J. & Jung, T.-M. (2022). Digital therapy for ADHD diagnosis using VR. *Broadcasting and Media, 27*(1), 38-48.
[23] Kim, K.-S. (2020). A study on the use of 'extended reality psychotherapy' (XR Therapy) for adolescents with post-traumatic stress disorder (PTSD): Focusing on 'Frozen' character Elsa. *Animation Studies, 16*(4), 39-58. https://doi.org/10.51467/ASKO.2020.12.16.4.39
[24] Yoon, E. & Lim, I. (2022). Analysis of domestic research trends related to the development of digital therapeutics (DTx) in the field of communication disorders. *Speech and Language Sciences, 14*(4), 57-66. https://doi.org/10.13064/KSSS.2022.14.4.057
[25] Lee, J.-N. (2019). Development Direction and Implications of Virtual Augmented Reality (AR-VR) Industry. Industrial Research Institute.
[26] Sung, Y.-B., Lee, D.-H. & Lee, J.-H. (2020). Descriptive Study on Rehabilitation Treatment Methods for Improving Upper Limb Function in Stroke Patients. *International Journal of Advanced Nursing Education and Research, 5*(2), 19-28, https://doi.org/10.21742/IJANER.2020.5.2.03
[27] Kwon, S. H., Park, J. K., & Koh, Y. H. (2023). A systematic review and meta-analysis on the effect of virtual reality-based rehabilitation for people with Parkinson's disease. *Journal of neuroengineering and rehabilitation, 20*(1), 94. https://doi.org/10.1186/s12984-023-01219-3
[28] Wenk, N., Penalver-Andres, J., Buetler, K. A., Nef, T., Müri, R. M., & Marchal-Crespo, L. (2023). Effect of immersive visualization technologies on cognitive load, motivation, usability, and embodiment. *Virtual Reality, 27*(1), 307-331. https://doi.org/10.1007/s10055-021-00565-8
[29] Hong, R., Li, B., Bao, Y., Liu, L. & Jin, L. (2024). Therapeutic robots for post-stroke rehabilitation. *Medical Review, 4*(1), 55-67. https://doi.org/10.1515/mr-2023-0054
[30] Hao, J., Pu, Y., Chen, Z. & Siu, K. C. (2023). Effects of virtual reality-based telerehabilitation for stroke patients: A systematic review and meta-analysis of randomized controlled trials. *Journal of Stroke and Cerebrovascular Diseases, 32*(3), 106960. https://doi.org/10.1016/j.jstrokecerebrovasdis.2022.106960
[31] Denche-Zamorano, A., Rodriguez-Redondo, Y., Barrios-Fernandez, S., Mendoza-Muñoz, M., Castillo-Paredes, A., Rojo-Ramos, J., ... & Adsuar, J. C. (2023). Rehabilitation is the main topic in virtual and augmented reality and physical activity research: A bibliometric analysis. *Sensors, 23*(6), 2987. https://doi.org/10.1038/s41598-023-31477-z
[32] Liu, H., Cheng, Z., Wang, S. & Jia, Y. (2023). Effects of virtual reality-based intervention on depression in stroke patients: a

meta-analysis. *Scientific Reports, 13*(1), 4381. https://doi.org/10.1038/s41598-023-31477-z

[33] Gulcan, K., Guclu-Gunduz, A., Yasar, E., Ar, U., Sucullu Karadag, Y. & Saygili, F. (2023). The effects of augmented and virtual reality gait training on balance and gait in patients with Parkinson's disease. *Acta Neurologica Belgica, 123*(5), 1917-1925. https://doi.org/10.1007/s13760-022-02147-0

[34] Chang, Y. H., Jeong, B. R., Kang, J. S. & Kim, G. S. (2020). The Influence on Proprioception, Balance and Agility of Transtibial Amputees by Immersive Virtual Reality Game Training. *Journal of Rehabilitation Welfare Engineering & Assistive Technology, 14*(4), 238-248, https://doi.org/10.21288/resko.2020.14.4.238

Authors' contacts:

Hwamok Koo
Department of Intelligent Information Engineering,
Graduate School, Sangmyung University,
20, Hongjimun 2-gil, Jongno-gu, Seoul, Republic of Korea
hwamokkoo@gmail.com

Junsik Moon
CUREXO INC., 3rd-4th Floor,
480, Wiryesunhwan-ro, Songpa-gu, Seoul, Republic of Korea
jsmoon@curexo.com

Sung Teac Hwang
CUREXO INC., 3rd-4th Floor,
480, Wiryesunhwan-ro, Songpa-gu, Seoul, Republic of Korea
hwang.st@curexo.com

Dong Keun Kim
(Corresponding author)
Department of Human-Centered Artificial Intelligent,
Sangmyung University,
20, Hongjimun 2-gil, Jongno-gu, Seoul, Republic of Korea
dkim@smu.ac.kr

ISSN 1846-6168 (Print), ISSN 1848-5588 (Online)
https://doi.org/10.31803/tg-20250105140152

Subject review
Received: 2025-01-05, Accepted: 2025-01-08

Use of Artificial Intelligence (AI) in the Workplace Ergonomics of Industry 5.0

Maja Trstenjak*, Tihomir Opetuk, Goran Đukić, Hrvoje Cajner

Abstract: Industry 5.0 emphasizes human-centricity, sustainability, and resilience as its core characteristics, with a focus on developing socio-technical systems that enhance human health, safety, and well-being while fostering sustainable societal practices. The human-centric perspective places significant importance on human factors and ergonomics, aiming to align technological advancements with the needs and capabilities of individuals. In this context, artificial intelligence (AI) emerges as a transformative tool for advancing human factors and ergonomics by optimizing workplace conditions and supporting human-centered design principles. This paper conducts a literature review to explore the applications and potential of AI in addressing human factors and ergonomics challenges, providing insights into its role in shaping the future of human-centric systems within Industry 5.0.

Keywords: ergonomics; human-centered; human factors; Industry 5.0; management; manufacturing; optimization; organization; workplace

1 INTRODUCTION

Industry 5.0 represents a human-centered, resilient, and sustainable paradigm designed to address the shortcomings of Industry 4.0. While Industry 4.0 focused on automation and digitalization, it often neglected the human role and perspective, leading to challenges in its adoption within the manufacturing sector. In contrast, Industry 5.0 emphasizes the integration of socio-technical systems where humans are placed at the center, fostering effective communication and collaboration with machines. This shift is critical in overcoming challenges posed by global uncertainties and volatile market conditions, which demand resilience and sustainability with a broader societal impact [1, 2].

In Industry 5.0, humans remain the key drivers of production systems, necessitating technological solutions tailored to their needs. The evolution of technology has shifted human roles toward more complex tasks, emphasizing interdisciplinary collaboration and the integration of ergonomics to enhance productivity while mitigating workplace illnesses and injuries [3]. Providing workers with both physical and cognitive support tools is essential for enabling efficient and safe operations. Collaborative machines, for example, can reduce ergonomic and cognitive strain, enhance safety, improve process monitoring, and boost productivity. These advancements result in higher product quality and increased competitiveness for businesses [4].

However, many small and medium-sized enterprises (SMEs) face barriers to adopting collaborative technologies, such as limited in-house expertise, high upfront costs, and employee resistance to change. To overcome these challenges, SMEs can collaborate with technology providers for support and training or invest in employee skills development. This process is further strengthened by applying a human-centered design approach that incorporates human factors into production systems, addressing the physical, psychological, social, and cultural needs of workers. Such strategies ensure not only successful implementation but also sustainable growth and resilience in the Industry 5.0 era [5].

Artificial intelligence (AI) refers to the simulation of human intelligence in machines designed to perform tasks that typically require cognitive abilities such as learning, reasoning, problem-solving, decision-making, and understanding language. AI systems leverage algorithms, data analysis, and computational models to process information, adapt to new inputs, and make autonomous or semi-autonomous decisions, often with capabilities exceeding human efficiency in specific domains [6].

In the context of Industry 5.0, AI plays a pivotal role in enhancing human-machine collaboration by enabling smarter, more adaptive, and human-centric production systems. AI drives innovation by [7]: Supporting Human Decision-Making; Enhancing Ergonomics and Safety; Enabling Personalized and Adaptive Systems; Improving Productivity and Quality; Facilitating Resilience and Sustainability; and Bridging Skill Gap. By fostering collaboration between humans and machines, AI empowers a human-centric approach in Industry 5.0, prioritizing worker health, safety, and well-being while driving technological and societal progress. Artificial intelligence (AI) plays a critical role in autonomous systems, yet its integration introduces complex safety challenges, especially in human-AI interactions. These challenges arise as human roles evolve alongside increasingly autonomous machines era [8].

2 METHODOLOGY

To gain insight into the most relevant work published in the literature, the Web of Science database was browsed. The objective was to understand the characteristics of the current use of AI tools in ergonomic design of workplace and organization but also it is potential for the future use. This is why the database Web of Science was searched by keyword "AI in Ergonomics". Among the 674 records, found 35 were chosen for the detailed analysis.

3 RESULTS

The results were grouped by the common topic in three different groups among which use of AI in ergonomics of the

workplace is found: (1) Posture prediction, (2) Risk analysis and assessment and (3) Organizational challenges.

3.1 Posture Prediction

Traditional motion capture uses specialized sensors and cameras, but the more accessible method is conventional video recording. Evaluation of AI-based computer vision techniques was used to automate human movement annotation from video. Four machine learning algorithms (random forest, K neighbors, neural network, and decision tree) were trained on performance datasets, producing automated annotations across Laban's four dimensions—effort, space, shape, and body. Results showed that these AI-generated annotations were accurate compared to manual Laban annotations, offering a promising tool for systematic movement analysis from video [11].

Traditional ergonomics assessment methods in digital human modeling (DHM) tools primarily focus on observing work characteristics, with direct measurement methods designed for easy integration into DHM. However, these methods have historically lacked action levels, which are thresholds for necessary ergonomic interventions. This is why there is a solution that integrates recent physical load exposure calculations and action level recommendations into a DHM tool, using the IPS IMMA tool and Xsens MVN motion capture. In two use cases—one with simulated human motions and one with real motion capture data—the demonstrator calculated and color-coded exposure data for specific postures and velocities, highlighting extreme action levels. The results suggest that DHM tools can effectively incorporate automated ergonomics assessments, aligning with Industry 4.0 and 5.0 goals for automation and digitalization in ergonomics [9].

Collaborative robots (cobots), which are part of Industry 5.0 concept, offer a way to reduce operator workload and lower the risk of occupational injuries, such as musculoskeletal disorders (MSDs). Innovative ergonomics optimization framework provide human tracking to monitor and assess operator posture, identifying ergonomic risks and suggesting improved poses. A feedback interface notifies users of non-ergonomic postures and recommends adjustments, while a workpiece position controller adjusts the cobot's end-effector to help improve operator posture. A user study conducted on a human-robot polishing task showed promising results, with positive feedback from users and improved REBA (Rapid Entire Body Assessment) scores, indicating reduced ergonomic risk [10]. Another study demonstrates the value of AI in ergonomic posture assessments, especially for methods like the Ovako Working Posture Assessment System (OWAS), by training algorithms on Xsens MVN MOCAP data. The results indicate that AI can accurately predict postures, offering a foundation for AI-assisted ergonomic assessments and establishing a specialized database to enhance future AI training [13]. MOCAP-/AI-based framework demonstrated improved accuracy, consistency, and efficiency compared to manual methods. This advancement supports the design of healthier, safer, and more productive job tasks and work environments [14].

ERG-AI is an innovative AI/ML pipeline developed to address the gap in sustainable, sensor-based, uncertainty-aware posture prediction combined with large language models (LLMs) for communicating occupational health risks and recommendations. Designed to predict extended worker postures using data from multiple wearable sensors, ERG-AI enables personalized health risk assessment based on individual worker performance. Utilizing LLMs such as GPT-4 and LLAMA-2, the model translates posture predictions and uncertainty estimates into clear occupational health insights. ERG-AI demonstrated robust data handling, enhancing posture prediction accuracy in real-world applications. Findings indicated that while basic postures were effectively identified, complex movements like kneeling and stair climbing posed prediction challenges, underscoring the importance of specific sensors, such as those positioned on the thigh, for balancing system cost and accuracy. Feedback from occupational health experts pointed to a need for even more personalized recommendations tailored to individual attributes and job roles, suggesting a pathway for further refinement of LLMs to provide customized ergonomic guidance [12].

3.2 Risk Assessment

Integrating AI and other digital technologies with a focus on human factors will be essential for advancing biomechanical risk assessment and facilitating technology adoption. Additionally, successfully incorporating these technologies into industrial settings may require workforce reskilling, upskilling, and efficient system design to manage information flow and improve user-technology interactions [16].

Traditional human-based risk assessments for muscle injuries can lead to inaccuracies and even cause injuries due to expertise requirements. Risk Assessment System for Muscle Injuries (RASMI) using AI technology is used to evaluate electric welders' postures based on Rapid Entire Body Assessment (REBA) standards. RASMI identifies potential causes of muscle injuries and issues warnings when a welder's posture poses a risk, providing precise, cost-effective assessments. Results show that RASMI effectively evaluates injury risks and is positively received by workers, who appreciate its role in promoting long-term health and well-being through posture adjustment and behavior modification alerts [15].

Holistic job improvement framework, which automatically performs root cause analysis and recommends control strategies to mitigate musculoskeletal disorder (MSD) risks was also presented. Using deep learning-based Natural Language Processing (NLP) techniques like Part of Speech (PoS) tagging and dependency parsing, the system analyzes textual descriptions of job actions (e.g., "pushing") and objects (e.g., "cart") to infer root causes of MSD risks (e.g., excessive shoulder forces due to small caster size). These insights guide an expert-based Machine Learning (ML) system to identify specific work-related causes and recommend targeted solutions, such as larger diameter casters, to lower risks. Unlike existing AI-based MSD risk assessment tools, which focus solely on scoring, this framework extends beyond scoring to diagnose root causes

and provide actionable controls, combining AI, computer science, and ergonomics. The system's robustness stems from integrating action-object inferences from text with risk scores and exposure types from video analysis. Even without text, the system can utilize motion capture from videos to identify root causes, although this leads to less efficient recommendations. This approach streamlines and enhances the job improvement process, making it more efficient and effective in reducing MSD risks [17].

Understanding human cognitive and behavioral responses, such as vigilance, processing intensity, gaze patterns, and visual scanning efficiency, is essential for designing effective AI-assisted inspection systems. Results of one study show that these cognitive factors impact inspection performance, emphasizing the need for protocols, drones, and AI systems tailored to reduce cognitive overload and prevent errors. Additionally, insights into gaze and scanning patterns associated with missed information provide practical guidance for inspectors to improve their performance [18].

The Level of Preventive Action (Lpac) methodology, adapted for construction sites, establishes preventive action levels by monitoring quantitative data related to physical and behavioral conditions using sensors in both the construction environment and on workers. Integrating Lpac with BIM (Building Information Modeling) technology allows for real-time data collection on environmental conditions, safety systems, worker behavior, and emotional states. This setup enables comprehensive preventive action controls and immediate communication of safety measures to workers through a mix of direct and AI-driven methods. By using body sensors, location sensors, and AI, Lpac supports dynamic safety assessments, adjusting preventive actions based on real-time data rather than estimates. This approach enhances safety coordination, improves the safety climate, and optimizes construction site conditions to prevent accidents. Lpac is applicable in both known and unknown environments, adapting flexibly to varying construction conditions and communicating risk levels efficiently, helping foster a collaborative safety culture with worker participation [19].

Knowledge acquisition through domain experts and historical data remains crucial for effective human reliability and human factor analysis across fields. However, fully utilizing this knowledge requires implementing fuzzy expert systems and AI models to enhance HFA outcomes [20].

A cost-effective, vision-based method for automatically monitoring ergonomic and fatigue risks was developed using a 3D camera system and AI-driven posture analysis to track body movements and repetitive motions. Laboratory trials demonstrated that this method achieves joint motion tracking with 3.5° accuracy, performs similarly to human operators in ergonomic risk assessments, and effectively monitors repetitive tasks. The system supports data visualization, real-time analysis, and report generation, making it a valuable tool for enhancing manufacturing environments [21].

3.3 Organizational Challenges

The integration of Artificial Intelligence (AI) into ergonomics is transforming workplace optimization by enhancing worker well-being and operational efficiency. Bibliometric analyses revealed AI's significant impact on ergonomics and safety, establishing this synergy as a key element in the evolution of Industrial Engineering. The report advocates for strategic funding, interdisciplinary collaboration, and workforce development to support this field, emphasizing AI's potential to create safer, efficient, and ethically focused workplaces [22]. It is crucial for researchers to develop AI teammates that incorporate an understanding of human team members' needs into their adaptive behaviors, ensuring effective and supportive collaboration [23].

Traditionally, productivity and worker well-being optimizations are handled separately, often resulting in suboptimal solutions and prioritizing one objective over the other. One study uses data mining methods on real-world multi-objective optimization data to extract actionable insights for optimizing both productivity and worker well-being in workstation design. By analyzing a welding gun workstation, the study identified critical design constraints that, when removed, could improve work efficiency and accommodate diverse anthropometric needs. The findings suggest that using data-driven rules and insights from previous cases can streamline future workstation designs across various tasks, such as assembly lines, improving overall productivity and worker health. Implementing the optimization and knowledge discovery process in a user-friendly digital tool could further support engineers by guiding them through the process based on their expertise [31]. The integration of artificial intelligence (AI) into workplaces offers significant potential for productivity and progress, but also raises critical occupational safety and health (OSH) concerns. AI can increase risks such as stress, discrimination, precarious employment, musculoskeletal disorders, and job intensification, especially when used for heightened monitoring and micro-management. Key risks include unfair treatment in AI-augmented HR decisions, overwork due to insufficient training, privacy issues from intensified surveillance, and the deskilling of jobs, particularly in manufacturing and gig work. To harness AI's benefits without compromising OSH, the report suggests focusing on assistive and collaborative AI, rather than universal AI, to enhance supportive roles. Benefits of AI include improving workplace relationships, aiding decision-making, and allowing workers more time for personal and career development by handling repetitive tasks. For safe AI integration, recommendations include comprehensive training, consistent oversight by OSH authorities, and a worker-centered approach that prioritizes human command over AI. It concludes that OSH risks stem not from AI itself, but from the ways in which it is implemented, underscoring the importance of regulatory oversight and worker involvement [38].

The field of Explainable AI (XAI) seeks to make AI decisions understandable, yet often overlooks the human element. Unlike typical AI-centric XAI approaches, this framework lets users define and rank features according to their preferences, compare these with the AI's weighted features, and explore the rationale behind AI decisions through a chatbot. The study found that agreement between user and AI factors depends on shared knowledge and experience, emphasizing the need for alignment in human-

centered XAI to enhance understanding and trust. These insights offer a foundation for future human-centered XAI advancements [24].

Integrating the Worker Fatigue Model into Airbus's industrial system architecture enhanced their ability to predict system performance based on workforce composition, which may include human-robot teams or a mix of experienced and less experienced workers. The model, adapted from existing fatigue models, accounts for worker characteristics, tasks, and robot assistance, simulating scenarios like fully manual and semi-automated work. Variables such as worker skill, age, and motivation were used to measure fatigue and error probabilities. Results indicated that workforce composition—particularly higher ratios of high-skilled workers—significantly reduced fatigue and improved system performance, demonstrating the model's value for workforce planning and technology integration [25].

How people build trust with AI partners compared to human partners is through a two-stage process: initial trust and feedback-based iteration. Using repeated trust games, the study measured investment behaviors, emotional responses, and neuroactivity. Results indicated that initial trust in AI partners had a stronger influence on final trust than with human partners, while feedback from human partners had a greater impact on trust and emotional arousal than feedback from AI. Positive emotions were more pronounced in interactions with humans, and neuroactivity in the prefrontal cortex was higher when participants made investment decisions with human partners, underscoring the stronger influence of human feedback on decision-making and emotional responses [27].

Human Factors and Ergonomics (HFE) should contribute to the sustainable and ethical development of AI: deciding when to automate versus augment human work, ensuring control and accountability in AI outcomes, and addressing power imbalances among AI stakeholders. Suggested actions for the HFE community include converting ethical considerations into design principles, embracing broader, multi-stakeholder perspectives, and fostering interdisciplinary collaboration within a design science framework. These measures aim to enhance HFE's role in creating AI systems that are socially and economically beneficial [29].

The integration of AI into Nuclear Power Plant (NPP) operator support systems has shown significant potential for reducing human error, enhancing safety, and improving operational efficiency. AI-powered systems—ranging from decision support and fault detection to predictive maintenance and automated text analysis—can aid plant operators in rapid situation assessment and response, support continuous monitoring of worker safety, and enable early detection of emerging issues. Despite these advancements, challenges remain in scaling these technologies, ensuring explainability, and bolstering cybersecurity in mission-critical NPP environments. Further research and novel approaches are needed to optimize human-automation interaction, balance task allocation between operators and AI, and design interfaces that support seamless collaboration. While full autonomy in NPPs is a distant goal, AI-enabled intelligent support systems promise to pave the way toward safer, more reliable, and efficient NPP operations in the future [6].

As machine learning (ML) becomes prevalent in industrial systems, Human Factors must innovate to help users understand and interact with autonomous ML capabilities. This includes advancing cooperation between humans and ML algorithms, integrating human sensing into ML, and leveraging ML for assessing human states and capabilities. This study addresses ergonomic injury in semiconductor manufacturing, particularly in the etching process, where repetitive tasks heighten injury risk. Using visual recognition technology and Convolutional Neural Networks (CNNs), the study achieved 95% accuracy in identifying unsafe actions among maintenance personnel, significantly reducing the non-conformity rate from 40% to 15%. These results, communicated via a cloud-based alert system, demonstrate ML's potential to improve ergonomic safety and reduce workplace injuries in semiconductor manufacturing [35].

The new protocol optimizes human-robot positioning and orientation to improve human ergonomics by using an RGB-D camera that monitors joint angles in real-time and assesses ergonomic states. The system identifies six main causes of poor ergonomics and enables six corresponding robot responses to help users achieve an optimal ergonomic posture. This adaptive algorithm allows the robot to adjust the work environment, such as modifying the position of a workpiece, to enhance user comfort continuously. An experimental study validated this approach, demonstrating significant ergonomic improvements through real-time human-robot interaction adjustments [39]. The researchers also developed Contextual Ergonomics Models, using Gaussian Process Latent Variable Models trained on high-dimensional musculoskeletal simulations for specific task contexts. These models enable ergonomic optimization in a low-dimensional latent space, allowing for efficient reconstruction of high-dimensional musculoskeletal models without extensive computational cost. Experiments with eight subjects performing a drilling task demonstrated that optimizing with Contextual Ergonomics Models significantly reduced muscle activation, showcasing the potential of these models in enhancing ergonomic outcomes in robotic systems [41].

4 CONCLUSION

The use of artificial intelligence (AI) in workplace ergonomics holds significant promise, particularly within the context of Industry 5.0. Through this literature review, the topic has been systematically explored and categorized into three primary groups: posture prediction, risk assessment, and organizational challenges. Each of these domains highlights the potential of AI to advance ergonomic practices by leveraging data-driven approaches to address traditional limitations. AI-driven tools in posture prediction enable the identification and correction of potentially harmful movements, contributing to proactive injury prevention. Similarly, AI-based risk assessment methodologies provide insights into workplace hazards, enhancing the ability to mitigate risks effectively. Addressing organizational challenges requires integrating AI to foster adaptive and

personalized solutions that not only ensure worker safety and health but also improve overall well-being and job satisfaction. Despite these advancements, this field remains at an early stage, with significant untapped potential. To fully realize the benefits of AI in workplace ergonomics, future research should focus on developing novel frameworks tailored to the unique demands of manufacturing environments. These frameworks should integrate AI's predictive, analytical, and adaptive capabilities to create comprehensive, human-centric solutions that align with the principles of Industry 5.0. In conclusion, the integration of AI in workplace ergonomics is crucial for shaping the future of safe, efficient, and worker-oriented industries. Continued interdisciplinary efforts are essential to advance this field, paving the way for smarter, healthier, and more sustainable workplaces.

5 REFERENCES

[1] Borchardt, M., Pereira, G. M., Milan, G. S., Scavarda, A. R., Nogueira, E. O. & Poltosi, L. C. (2022). Industry 5.0 beyond Technology: An Analysis through the Lens of Business and Operations Management Literature. *Organizacija, 55*(4), 305-321. https://doi.org/10.2478/orga-2022-0020

[2] Javaid, M. & Haleem, A. (2020). Critical Components of Industry 5.0 towards a Successful Adoption in the Field of Manufacturing. *J. Ind. Intg. Mgmt., 05*(03), 327-348. https://doi.org/10.1142/S2424862220500141

[3] Trstenjak, M., Opetuk, T. & Đukić, G. (2024). Human-centric Manufacturing System of Industry 5.0 by Principles of Human Factors and Ergonomics-A Systematic Literature Review. In *4th LA Conference on Sustainable Development of Energy, Water and Environment Systems*, 103-103. Available: https://www.croris.hr/crosbi/publikacija/prilog-skup/833640

[4] Trstenjak, M., Đukić, G. & Opetuk, T. (2023). Human Factors and Ergonomics in Logistics Activities by Principles of Industry 5.0. *This document and the underlying event have been planned and executed as a part of the JOINOLOG research project*, p. 64.

[5] Trstenjak, M., Gregurić, P., Janić, Ž. & Salaj, D. (2024). Integrated Multilevel Production Planning Solution According to Industry 5.0 Principles. *Applied Sciences*, 14(1). https://doi.org/10.3390/app14010160

[6] Sethu, M. et al. (2023). Application of Artificial Intelligence in Detection and Mitigation of Human Factor Errors in Nuclear Power Plants: A Review. *Nuclear Technology, 209*(3), 276-294. https://doi.org/10.1080/00295450.2022.2067461

[7] Nosova, S., Norkina, A., Makar, S., Gerasimenko, T. & Medvedeva, O. (2022). Artificial intelligence as a driver of business process transformation. *Procedia Computer Science, 213*, 276-284. https://doi.org/10.1016/j.procs.2022.11.067

[8] Karvonen, H., Heikkilä, E. & Wahlström, M. (2020). Safety Challenges of AI in Autonomous Systems Design – Solutions from Human Factors Perspective Emphasizing AI Awareness. In Harris, D. & Li, W.-C., Eds. *Engineering Psychology and Cognitive Ergonomics. Cognition and Design*, Cham: Springer International Publishing, 147-160. https://doi.org/10.1007/978-3-030-49183-3_12

[9] Hanson, L., Högberg, D., Iriondo, A., Brolin, A., Brolin, E. & Lebram, M. (2022). Integrating Physical Load Exposure Calculations and Recommendations in Digitalized Ergonomics Assessment Processes. https://doi.org/10.3233/ATDE220142

[10] El Makrini, I., Mathijssen, G., Verhaegen, S., Verstraten, T. & Vanderborght, B. (2022,). A Virtual Element-Based Postural Optimization Method for Improved Ergonomics during Human-Robot Collaboration. *IEEE Transactions on Automation Science and Engineering, 19*(3), 1772-1783. https://doi.org/10.1109/TASE.2022.3147702

[11] Guo, W. et al. (2022). AI-Driven Human Motion Classification and Analysis Using Laban Movement System. In Duffy, V. G., Ed., *Digital Human Modeling and Applications in Health, Safety, Ergonomics and Risk Management. Anthropometry, Human Behaviour, and Communication*, Cham: Springer International Publishing, 201-210. https://doi.org/10.1007/978-3-031-05890-5_16

[12] Sen, S., Gonzalez, V., Husom, E. J., Tverdal, S., Tokas, S. & Tjøsvoll, S. O. (2024). ERG-AI: enhancing occupational ergonomics with uncertainty-aware ML and LLM feedback. *Appl Intell*, 54(23), 12128-12155. https://doi.org/10.1007/s10489-024-05796-1

[13] Igelmo, V., Syberfeldt, A., Högberg, D., García Rivera, F. & Peréz Luque, E. (2020). Aiding Observational Ergonomic Evaluation Methods Using MOCAP Systems Supported by AI-Based Posture Recognition. *The 6th International Digital Human Modeling Symposium*, August 31 – September 2, Skövde, Sweden, IOS Press, 419-429. Available: https://urn.kb.se/resolve?urn=urn:nbn:se:his:diva-19002

[14] Aliasgari, R., Fan, C., Li, X., Golabchi, A. & Hamzeh, F. (2024). MOCAP and AI-Based Automated Physical Demand Analysis for Workplace Safety. *Journal of Construction Engineering and Management, 150*(7), p. 04024060. https://doi.org/10.1061/JCEMD4.COENG-13811

[15] Ruengdech, C., Howimanporn, S., Intarakumthornchai, T. & Chookaew, S. (2024). Implementing a Risk Assessment System of Electric Welders' Muscle Injuries for Working Posture Detection with AI Technology. *International Journal of Online and Biomedical Engineering (iJOE), 20*(04), Art. no. 04. https://doi.org/10.3991/ijoe.v20i04.46465

[16] Anacleto Filho, P. C., Colim, A., Jesus, C., Lopes, S. I. & Carneiro, P. (2024). Digital and Virtual Technologies for Work-Related Biomechanical Risk Assessment: A Scoping Review," *Safety*, 10(3), Art. no. 3. https://doi.org/10.3390/safety10030079

[17] Parikh, P., Penfield, J., Barker, R., McGowan, B. and Mallon, J. R. (2024). NLP-based ergonomics MSD risk root cause analysis and risk controls recommendation. *Ergonomics*, 0(0), 1-13. https://doi.org/10.1080/00140139.2024.2394510

[18] Dalilian, F. & Nembhard, D. (2024). Cognitive and behavioral markers for human detection error in AI-assisted bridge inspection. *Applied Ergonomics, 121*, p. 104346. https://doi.org/10.1016/j.apergo.2024.104346

[19] Carpio de los Pinos, A. J., González García, M. de las N., Baptista, J. S. & Fernández Álvarez, M. (2024). Linguistic analysis for occupational risk assessment communication using the Lpac in an AI environment. *Theoretical Issues in Ergonomics Science*, 0(0), 1-14, https://doi.org/10.1080/1463922X.2024.2318660

[20] Zarei, E., Khan, F. & Abbassi, R. (2023). How to account artificial intelligence in human factor analysis of complex systems? *Process Safety and Environmental Protection*, 171, 736-750. https://doi.org/10.1016/j.psep.2023.01.067

[21] Kunz, M., Shu, C., Picard, M., Vera, D., Hopkinson, P. & Xi, P. (2022). Vision-based Ergonomic and Fatigue Analyses for Advanced Manufacturing. In *IEEE 5th International Conference on Industrial Cyber-Physical Systems (ICPS2022)*, 01-07. https://doi.org/10.1109/ICPS51978.2022.9816946

[22] Somaraju, P., Kulkarni, S. S., Duffy, V. G. & Kanade, S. (2024). Artificial Intelligence and Mobile Computing: Role of AI in Ergonomics. In Duffy, V. G. Ed., *Digital Human Modeling and Applications in Health, Safety, Ergonomics and Risk Management*, Cham: Springer Nature Switzerland, 265-281. https://doi.org/10.1007/978-3-031-61066-0_16

[23] Hauptman, A. I., Mallick, R., Flathmann, C. & McNeese, N. J. (2024). Human factors considerations for the context-aware design of adaptive autonomous teammates. *Ergonomics*, 1-17. https://doi.org/10.1080/00140139.2024.2380341

[24] Baber, C., Kandola, P., Apperly, I. & McCormick, E. (2024). Human-centred explanations for artificial intelligence systems. *Ergonomics*, 0(0), 1-15. https://doi.org/10.1080/00140139.2024.2334427

[25] Holman, M., Walker, G., Lansdown, T. & Hulme, A. (2020). Radical systems thinking and the future role of computational modelling in ergonomics: an exploration of agent-based modelling. *Ergonomics, 63*(8), 1057-1074. https://doi.org/10.1080/00140139.2019.1694173

[26] Naikar, N., Brady, A., Moy, G. & Kwok, H.-W. (2023). Designing human-AI systems for complex settings: ideas from distributed, joint, and self-organising perspectives of sociotechnical systems and cognitive work analysis. *Ergonomics, 66*(11), 1669-1694. https://doi.org/10.1080/00140139.2023.2281898

[27] Yang, M. & Rau, P.-L. P. (2024). Trust building with artificial intelligence: comparing with human in investment behaviour, emotional arousal and neuro activities. *Theoretical Issues in Ergonomics Science, 25*(5), 593-614. https://doi.org/10.1080/1463922X.2023.2290019

[28] McLean, S. et al. (2023). Forecasting emergent risks in advanced AI systems: an analysis of a future road transport management system. *Ergonomics, 66*(11), 1750-1767. https://doi.org/10.1080/00140139.2023.2286907

[29] Grote, G. (2023). Shaping the development and use of Artificial Intelligence: how human factors and ergonomics expertise can become more pertinent. *Ergonomics, 66*(11), 1702-1710. https://doi.org/10.1080/00140139.2023.2278408

[30] Mazarakis, A., Bernhard-Skala, C., Braun, M. & Peters, I. (2023). What is critical for human-centered AI at work? - Toward an interdisciplinary theory. *Front Artif Intell, 6*, p. 1257057. https://doi.org/10.3389/frai.2023.1257057

[31] Iriondo Pascual, A., Smedberg, H., Högberg, D., Syberfeldt, A. & Lämkull, D. (2022). Enabling Knowledge Discovery in Multi-Objective Optimizations of Worker Well-Being and Productivity. *Sustainability, 14*(9), Art. no. 9. https://doi.org/10.3390/su14094894

[32] Saßmannshausen, T., Burggräf, P., Wagner, J., Hassenzahl, M., Heupel, T. & Steinberg, F. (2021). Trust in artificial intelligence within production management – an exploration of antecedents. *Ergonomics, 64*(10), 1333-1350. https://doi.org/10.1080/00140139.2021.1909755

[33] Bitkina, O. Vl., Jeong, H., Lee, B. C., Park, J., Park, J. & Kim, H. K. (2020). Perceived trust in artificial intelligence technologies: A preliminary study. *Human Factors and Ergonomics in Manufacturing & Service Industries, 30*(4), 282-290. https://doi.org/10.1002/hfm.20839

[34] Illankoon, P. & Tretten, P. (2021). Collaborating AI and human experts in the maintenance domain. *AI & Soc, 36*(3), 817-828. https://doi.org/10.1007/s00146-020-01076-x

[35] Lau, N., Hildebrandt, M. & Jeon, M. (2020). Ergonomics in AI: Designing and Interacting with Machine Learning and AI. *Ergonomics in Design, 28*(3), 3-3. https://doi.org/10.1177/1064804620915238

[36] Enebechi, C. N. & Duffy, V. G. (2020). Virtual Reality and Artificial Intelligence in Mobile Computing and Applied Ergonomics: A Bibliometric and Content Analysis. In Duffy, V. G., Ed. *Digital Human Modeling and Applications in Health, Safety, Ergonomics and Risk Management. Human Communication, Organization and Work*, Cham: Springer International Publishing, 334-345. https://doi.org/10.1007/978-3-030-49907-5_24

[37] Aubin, F. & Prevost, M.-C. (2019). Using Prospective Ergonomics to Identify Opportunities from Recent Technological Advances in AI: The Case of a West African Bank. In Bagnara, S., Tartaglia, R., Albolino, S., Alexander, T. & Fujita, Y., Eds. *Proceedings of the 20th Congress of the International Ergonomics Association (IEA 2018)*, Cham: Springer International Publishing, 1365-1371. https://doi.org/10.1007/978-3-319-96071-5_138

[38] Moore, P. V. (2019). OSH and the Future of Work: Benefits and Risks of Artificial Intelligence Tools in Workplaces. In Duffy, V. G., Ed. *Digital Human Modeling and Applications in Health, Safety, Ergonomics and Risk Management. Human Body and Motion*, Cham: Springer International Publishing, 292-315. https://doi.org/10.1007/978-3-030-22216-1_22

[39] Shafti, A., Ataka, A., Lazpita, B. U., Shiva, A., Wurdemann, H. A. & Althoefer, K. (2019). Real-time Robot-assisted Ergonomics. In *International Conference on Robotics and Automation (ICRA2019)*, 1975-1981. https://doi.org/10.1109/ICRA.2019.8793739

[40] Kistan, T., Gardi, A. & Sabatini, R. (2018). Machine Learning and Cognitive Ergonomics in Air Traffic Management: Recent Developments and Considerations for Certification. *Aerospace*, 5(4), Art. no. 4. https://doi.org/10.3390/aerospace5040103

[41] Marin, A. G., Shourijeh, M. S., Galibarov, P. E., Damsgaard, M., Fritzsch, L. & Stulp, F. (2018). Optimizing Contextual Ergonomics Models in Human-Robot Interaction. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS2018)*, 1-9. https://doi.org/10.1109/IROS.2018.8594132

Authors' contacts:

Maja Trstenjak, dr. sc.
(Corresponding author)
Faculty of Mechanical Engineering and Naval Architecture, University of Zagreb,
Ivana Lucica 5, 10 000, Zagreb, Croatia
maja.trstenjak@fsb.unizg.hr

Tihomir Opetuk, Associate Professor, dr. sc.
Faculty of Mechanical Engineering and Naval Architecture, University of Zagreb,
Ivana Lucica 5, 10 000, Zagreb, Croatia
tihomir.opetuk@fsb.unizg.hr

Goran Đukić, Full Professor, dr. sc.
Faculty of Mechanical Engineering and Naval Architecture, University of Zagreb,
Ivana Lucica 5, 10 000, Zagreb, Croatia
goran.dukic@fsb.unizg.hr

Hrvoje Cajner, Associate Professor, dr. sc.
Faculty of Mechanical Engineering and Naval Architecture, University of Zagreb,
Ivana Lucica 5, 10 000, Zagreb, Croatia
hrvoje.cajner@fsb.unizg.hr

ISSN 1846-6168 (Print), ISSN 1848-5588 (Online)
https://doi.org/10.31803/tg-20.....

Categorization of paper
Received: y-m-d, Accepted: y1-m1-d1

14pt
14pt

Article Title Only in English (Style: Arial Narrow, Bold, 14pt)

14pt

Ivan Horvat, Thomas Johnson, Marko Marić (Style: Arial Narrow, Normal, 10pt)

14pt

Abstract: Article abstract contains maximum of 150 words and is written in the language of the article. The abstract should reflect the content of the article as precisely as possible. TECHNICAL JOURNAL is a trade journal that publishes scientific and professional papers from the domain(s) of mechanical engineering, electrical engineering, civil engineering, multimedia, logistics, etc., and their boundary areas. This document must be used as the template for writing articles so that all the articles have the same layout. (Style: Arial Narraw, 8pt)

8pt

Keywords: keywords in alphabetical order (5-6 key words). Keywords are generally taken from the article title and/or from the abstract. (Style: Arial Narraw, 8pt)

10pt
10pt

1 INTRODUCTION (Article Design)

(Style: Arial Narrow, Bold, 10pt)

10pt

(Tab 6 mm) The article is written in Latin script and Greek symbols can be used for labelling. The length of the article is limited to eight pages of international paper size of Letter (in accordance with the template with all the tables and figures included). When formatting the text the syllabification option is not to be used.

10pt

1.1 Subtitle 1 (Writing Instructions)

(Style: Arial Narrow, 10pt, Bold, Align Left)

10pt

The document format is Letter with margins in accordance with the template. A two column layout is used with the column spacing of 10 mm. The running text is written in Times New Roman with single line spacing, font size 10 pt, alignment justified.

Article title must clearly reflect the issues covered by the article (it should not contain more than 15 words).

Body of the text is divided into chapters and the chapters are divided into subchapters, if needed. Chapters are numbered with Arabic numerals (followed by a period). Subchapters, as a part of a chapter, are marked with two Arabic numerals i.e. 1.1, 1.2, 1.3, etc. Subchapters can be divided into even smaller units that are marked with three Arabic numerals i.e. 1.1.1, 1.1.2, etc. Further divisions are not to be made.

Titles of chapters are written in capital letters (uppercase) and are aligned in the centre. The titles of subchapters (and smaller units) are written in small letters (lowercase) and are aligned left. If the text in the title of the subchapter is longer than one line, no hanging indents.

10pt

Typographical symbols (bullets), which are being used for marking an item in a list or for enumeration, are placed at a beginning of a line. There is a spacing of 10pt following the last item:

- Item 1
- Item 2
- Item 3

10pt

The same rule is valid when items are numbered in a list:

1) Item 1
2) Item 2
3) Item 3

10pt

1.2 Formatting of Pictures, Tables and Equations

(Style: Arial Narrow, 10pt, Bold, Align Left)

10pt

Figures (drawings, diagrams, photographs) that are part of the content are embedded into the article and aligned in the centre. In order for the figure to always be in the same position in relation to the text, the following settings should be defined when importing it: text wrapping / in line with text.

Pictures must be formatted for graphic reproduction with minimal resolution of 300 dpi. Pictures downloaded from the internet in ratio 1:1 are not suitable for print reproduction because of unsatisfying quality.

10pt

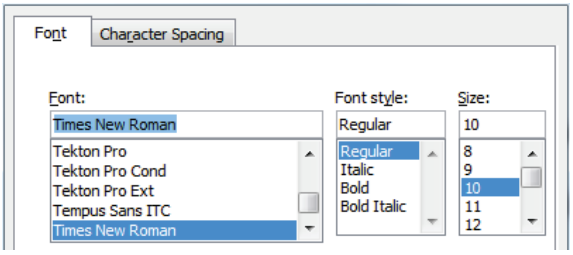


Figure 1 Text under the figure [1]
(Style: Arial Narraw, 8pt, Align Centre)

10pt

The journal is printed in black ink and the figures have to be prepared accordingly so that bright tones are printed in a satisfactory manner and are readable. Figures are to be in colour for the purpose of digital format publishing. Figures in the article are numbered with Arabic numerals (followed by a period).

Text and other data in tables are formatted - Times New Roman, 8pt, Normal, Align Center.

When describing figures and tables, physical units and their factors are written in italics with Latin or Greek letters, while the measuring values and numbers are written upright.

10pt

Table 1 Table title aligned centre
(Style: Arial Narraw, 8pt, Align Centre)

	1	2	3	4	5	6
ABC	ab	ab	ab	ab	ab	ab
DEF	cd	cd	cd	cd	cd	cd
GHI	ef	ef	ef	ef	ef	ef

10 pt

Equations in the text are numbered with Arabic numerals inside the round brackets on the right side of the text. Inside the text they are referred to with equation number inside the round brackets i.e. ".... from Eq. (5) follows ...." (Create equations with MathType Equation Editor - some examples are given below).

10pt

$$F_{\text{avg}}(t, t_0) = \frac{1}{t}\int_{t_0}^{t_0+t} F\left[q(\tau), p(\tau)\right] \mathrm{d}\tau, \tag{1}$$

$$\cos\alpha + \cos\beta = 2\cos\frac{\alpha+\beta}{2}\cdot\cos\frac{\alpha-\beta}{2}, \tag{2}$$

$$(\boldsymbol{AB})^{\mathrm{T}} = \boldsymbol{B}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}, \tag{3}$$

$$AAMC = \frac{1}{n}\sum_{i=1}^{n} PVMC_i. \tag{4}$$

10pt

Variables that are used in equations and also in the text or tables of the article are formatted as *italics* in the same font size as the text.

10pt

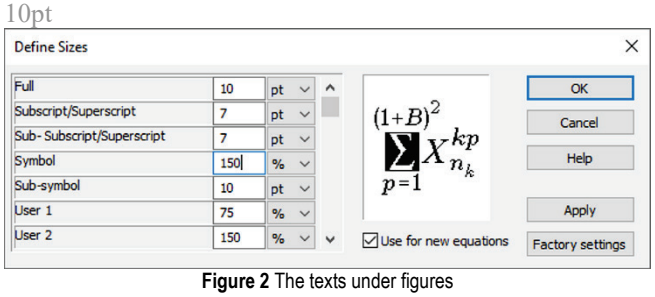


Figure 2 The texts under figures
(Style: Arial Narraw, 8pt, Align Centre)

10pt

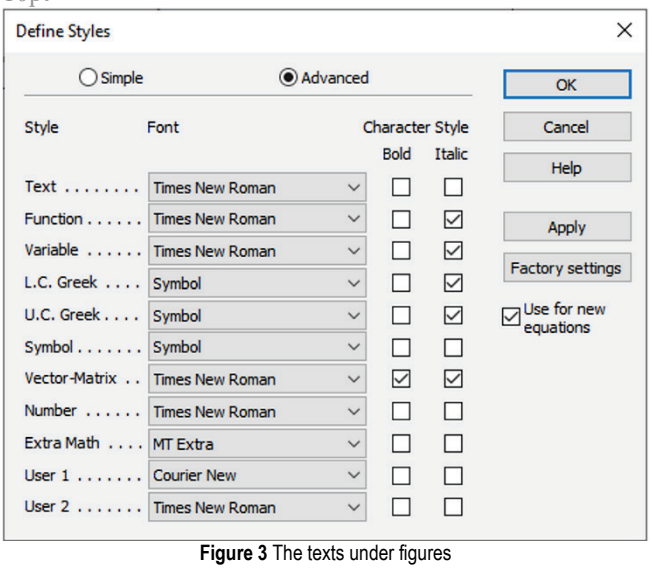


Figure 3 The texts under figures
(Style: Arial Narraw, 8pt, Align Centre)

Figures and tables that are a part of the article have to be mentioned inside the text and thus connected to the content i.e. " … as shown in Fig. 1…" or "data from Tab. 1…" and similar.

10pt

2 PRELIMINARY ANNOTATION

10pt

Article that is offered for publication cannot be published beforehand, be it in the same or similar form, and it cannot be offered at the same time to a different journal. Author or authors are solely responsible for the content of the article and the authenticity of information and statements written in the article.

Articles that are accepted for publishing are classified into four categories: original scientific papers, preliminary communications, subject reviews and professional papers.

Original scientific papers are articles that according to the reviewer and the editorial board contain original theoretical or practical results of research. These articles need to be written in such a way that based on the information given, the experiment can be repeated and the results described can be achieved together with the author's observations, theoretical statements or measurements.

Preliminary communication contains one or more pieces of new scientific information, but without details that allow recollection as in original scientific papers. Preliminary communication can give results of an experimental research, results of a shorter research or research in progress that is deemed useful for publishing.

Subject review contains a complete depiction of conditions and tendencies of a specific domain of theory, technology or application. Articles in this category have an overview character with a critical review and evaluation. Cited literature must be complete enough to allow a good insight and comprehension of the depicted domain.

Professional paper can contain a description of an original solution to a device, assembly or instrument, depiction of important practical solutions, and similar. The article need not be related to the original research, but it should contains a contribution to an application of known scientific results and their adaptation to practical needs, so it presents a contribution to spreading knowledge, etc.

Outside the mentioned categorization, the Editorial board of the journal will publish articles of interesting content in a special column. These articles provide descriptions of practical implementation and solutions from the area of production, experiences from device application, and similar.

10pt

3 WRITING AN ARTICLE

10pt

Article is written in the English language and the terminology and the measurement system should be adjusted to legal regulations, standards and the International System of Units (SI) (Quantities and Units: ISO 80 000 - from Part 1 to Part 14). The article should be written in third person.

Introduction contains the depiction of the problem and an account of important results that come from the articles that are listed in the cited literature.

Main section of the article can be divided into several parts or chapters. Mathematical statements that obstruct the reading of the article should be avoided. Mathematical statements that cannot be avoided can be written as one or more addendums, when needed. It is recommended to use an example when an experiment procedure, the use of the work in a concrete situation or an algorithm of the suggested method must be illustrated. In general, an analysis should be experimentally confirmed.

Conclusion is a part of the article where the results are being given and efficiency of the procedure used is emphasized. Possible procedure and domain constraints where the obtained results can be applied should be emphasized.

AI and AI-assisted tools do not qualify for authorship under TG/TJ's authorship policy. Authors who use AI or AI-assisted tools during the manuscript writing process are asked to disclose their use in a separate section of the manuscript. The publishing agreement process works as usual, with the authors keeping the copyright to their own work.

10pt

4 RECAPITULATION ANNOTATION

10pt

In order for the articles to be formatted in the same manner as in this template, this document is recommended for use when writing the article. Finished articles written in MS Word for Windows and formatted according to this template must be submitted using our The Paper Submission Tool (PST) (https://tehnickiglasnik.unin.hr/authors.php) or eventually sent to the Editorial board of the Technical Journal to the following e-mail address: tehnickiglasnik@unin.hr

The editorial board reserves the right to minor redaction corrections of the article within the framework of prepress procedures. Articles that in any way do not follow these authors' instructions will be returned to the author by the editorial board. Should any questions arise, the editorial board contacts only the first author and accepts only the reflections given by the first author.

10pt

5 REFERENCES (According to APA)

10pt

The literature is cited in the order it is used in the article. No more than 35 references are recommended. Individual references from the listed literature inside the text are addressed with the corresponding number inside square brackets i.e. "… in [7] is shown …". If the literature references are web links, the hyperlink is to be removed as shown with the reference number 8. Also, the hyperlinks from the e-mail addresses of the authors are to be removed. In the literature list, each unit is marked with a number and listed according to the following examples (omit the subtitles over the references – they are here only to show possible types of references):

9pt

[1] See http://www.bibme.org/citation-guide/apa/

[2] See http://sites.umuc.edu/library/libhow/apa_examples.cfm

[3] (Style: Times New Roman, 9pt, according to APA)

[4] Amidzic, O., Riehle, H. J. & Elbert, T. (2006). Toward a psychophysiology of expertise: Focal magnetic gamma bursts as a signature of memory chunks and the aptitude of chess players. *Journal of Psychophysiology, 20*(4), 253-258. https://doi.org/10.1027/0269-8803.20.4.253

[5] Reitzes, D. C. & Mutran, E. J. (2004). The transition to retirement: Stages and factors that influence retirement adjustment. *International Journal of Aging and Human Development, 59*(1), 63-84. Retrieved from http://www.baywood.com/journals/PreviewJournals.asp?Id=0091-4150

[6] Jans, N. (1993). *The last light breaking: Life among Alaska's Inupiat Eskimos*. Anchorage, AK: Alaska Northwest Books.

[7] Miller, J. & Smith, T. (Eds.). (1996). *Cape Cod stories: Tales from Cape Cod, Nantucket, and Martha's Vineyard*. San Francisco, CA: Chronicle Books.

[8] Chaffe-Stengel, P. & Stengel, D. (2012). *Working with sample data: Exploration and inference*. https://doi.org/10.4128/9781606492147

[9] Freitas, N. (2015, January 6). People around the world are voluntarily submitting to China's Great Firewall. Why? Retrieved from http://www.slate.com/blogs/future_tense/2015/01/06/tencent_s_wechat_worldwide_internet_users_are_voluntarily_submitting_to.html (Style: Times New Roman, 9pt, according to APA)

10pt

10pt

Authors' contacts:

8pt

Full Name, title
Institution, company
Address
Tel./Fax, e-mail

8pt

Full Name, title
Institution, company
Address
Tel./Fax, e-mail

Note: Gray text should be removed in the final version of the article because it is for guidance only.

TEHNIČKI GLASNIK / TECHNICAL JOURNAL – GODIŠTE / VOLUME 19 – BROJ / ISSUE 2

LIPANJ 2025 / JUNE 2025 – STRANICA / PAGES 173-340

Sveučilište Sjever

SVEUČILIŠTE SJEVER / UNIVERSITY NORTH – CROATIA – EUROPE

ISSN 1846-6168 (PRINT) / ISSN 1848-5588 (ONLINE)

TEHNICKIGLASNIK@UNIN.HR – HTTP://TEHNICKIGLASNIK.UNIN.HR