TEHNIČKI GLASNIK

TEHNIČKI GLASNIK / TECHNICAL JOURNAL – GODIŠTE / VOLUME 19 – BROJ / ISSUE 1

OŽUJAK 2025 / MARCH 2025 – STRANICA / PAGES 1-172

/

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 1
Varaždin, March 2025 Pages 1-172

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):
February 4, 2025	February 6, 2025	March 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-20221010113555

Original scientific paper
Received: 2022-10-10, Accepted: 2024-02-07

Design and Development of Educational Modular Mobile Robot Platform

Denis Kotarski*, Petar Piljek, Tomislav Šančić

Abstract: This paper presents the design and experimental validation of a modular educational robot platform. The development of a novel modular unmanned ground vehicle (UGV) platform is based on open-source system architecture featuring carefully considered modules and is specifically tailored for application in STEM education. The proposed platform is comprehensive in terms of motion planning and modelling. Mathematical description of the non-holonomic and holonomic configuration of the robot is given, and a derived model is implemented in a software package in order to perform simulations and experiments. The design process resulted in a 3D printed novel brick-based robot chassis, offering advantages such as low manufacturing costs, easy repair and maintenance, and seamless expandability with additional modules. The proposed modular platform underwent validation across three robot configurations and three different control ecosystems.

Keywords: 3D printing; modular robot platform; STEM education; unmanned ground vehicle

1 INTRODUCTION

Nowadays, the field of robotics represents a very important, we could even say crucial, segment of industrial development and society as a whole. It can be said that robotics is a multidisciplinary field that includes various disciplines including mechanical engineering, electrical and electronic engineering, and computer science (software engineering). There is a need and great potential for integrating robotics into the education system through classes and extracurricular activities for primary and secondary school pupils with the goal of training technical experts. Robots are already used to some extent in introducing Science, Technology, Engineering, and Mathematics (STEM), as motivating and interesting tools for teaching engineering concepts and developing algorithmic thinking. Through practical exercises, laboratory classes and experiments, and work on projects, pupils and students develop creativity, critical thinking, teamwork, and other valuable skills. With the aim of the wider integration and increased accessibility, the goal of STEM education robotic platforms is to be open, cheap, and, among other things, multipurpose, like the modular robot presented in research [1].

By searching the literature with the keywords: modular educational robot, numerous works on this topic were found. In the vast majority of papers, mobile robots are discussed, which represent a unique challenge in STEM education, considering that they attract pupils and students in various disciplines. If we are talking about education, but also in general, from the aspect of control, numerous ecosystems are inherently modular, expandable, and most importantly open-source. The most widespread ecosystem is based on the Arduino open-source electronic prototyping platform, which is used in a wide range of education, from elementary school to engineering education. The paper [2] presents the design of an educational cheap mobile robot that is suitable as an educational tool in classrooms and laboratories and e-learning through Massive Open Online Courses (MOOC). Furthermore, [3] presents an open-source extensible robot platform whose control system is based on the widespread Raspberry Pi single-board computer, and whose software infrastructure is based on the Robot Operating System (ROS). In the aforementioned research, an overview of frequently used robotic platforms in education is presented, from which it is evident that mobile robots with differential drive configurations are mostly used.

In earlier research, educational mobile robots were considered for use in research and engineering education, such as the low-cost robotic vehicle described in the [4]. Considering the high price and limited functionality of the platforms used in the former phase of STEM education, [5] presents a cheap and very affordable robotic platform consisting of commercial components to facilitate robotic projects and further development of robotics. Commercial chassis that make up the robot systems were used in the mentioned works. In contrast to such an approach, [6] presents the hardware part of an educational mobile robot platform that enables the development of interesting and special applications using various expansion modules. Furthermore, [7] shows the creation of a robot chassis controlled by a mobile application that serves as a tool for teaching coding. In the above works, the mobile robot is based on a differential drive configuration. In addition to this most commonly used drive configuration, [8] presents a prototype of a mobile robot that represents an omnidirectional configuration. A robot rapid prototyping approach is proposed to implement different architectures and modules for robotic systems integrating hardware and software. Although the keywords in the listed references include modularity, this can be said mainly from the aspect of the software, i.e., the control part of the robot system.

In the full sense of the word modular, i.e., both from the software and hardware aspects, robotic structures that can be integrated for research and mostly engineering education at the graduate level have been researched and presented. Paper [9] describes a low-cost modular multipurpose robot intended for engineering students for research and education in the field of mechatronics and robotics. Such a platform is suitable because it encourages creativity and imagination and ensures the practice of fundamental control laws. Furthermore, [10] presents the development and integration

of the moreBots modular robotic system in education, which can be used as a tool in learning through play. The modular robotic system intended for creative learning, presented in [11], is considered for different user groups in the form of integration into everyday teaching. The system consists of independent modules that are equipped with actuators and sensors, and by assembling those users get a wide range of robotic designs. In paper [12], a new design of an inexpensive mobile reconfigurable modular robot is presented, which is intended for rapid development and testing of algorithms. Among other rapid prototyping technologies, mostly Additive Manufacturing (AM) was used in the development of the aforementioned modular robots. AM technologies have contributed to greater momentum in the field of educational robotics, as a new approach to the development of various robotic platforms and low-cost modules, such as the proposed 3D printed modules in [13]. The paper [14] presents the iterative process of developing a robotic educational platform for the Internet of Things (IoT) applications using 3D printing.

This paper discusses an educational modular robotic platform that can be used in a wide range of education, from elementary, and secondary school to engineering education. The open-source architecture of the system based on modules with selected take-off-the-shelf and low-cost components is presented. Experimental measurement of the drive components was carried out to identify the parameters of the modular robot drive configurations. Then a mathematical model of the robot, the so-called Unmanned Ground Vehicle (UGV), was developed, suitable for implementation in a software package with the aim of performing simulations and experiments. The main focus of the work is on the design and development of a novel modular brick-based robot chassis that can be manufactured cheaply, can be easily repaired and maintained, and is highly expandable with other modules. Chassis parts are constructed in the design phase using a CAD software package and then manufactured using AM technologies in the prototyping phase. After assembly, in order to validate the proposed modular robot platform, three versions of the robot with three different control ecosystems were tested.

2 EDUCATIONAL MODULAR ROBOT SYSTEM ARCHITECTURE

Given that mobile ground robots or UGVs are used in a wide range of applications, it is important to prepare users of STEM education for future jobs and challenges in the field of mechatronics and robotics. The most common configuration of such a robot is the mentioned differential drive, which is also widespread in education. One of the frequently used commercial platforms is the mBot robot [15], suitable for increasing the basic knowledge in programming and robotics at the elementary school level. But there are also real needs for the implementation of other configurations, such as so-called omnidirectional robots, given that their applications are widespread in the industry, such as KUKA omniMove robots used to assemble trains [16]. A fundamental requirement in the robot design phase is an open-source system architecture based on low-cost and take-off-the-shelf components that make up individual robot modules. The robot can be divided into a control module, a drive module, an energy module, a sensor module, and additional equipment that integrates with other robot modules. The chassis of the robot connects the components, where the main goal is to enable the easy and quick connection of modules into the functional assembly of the modular robot.

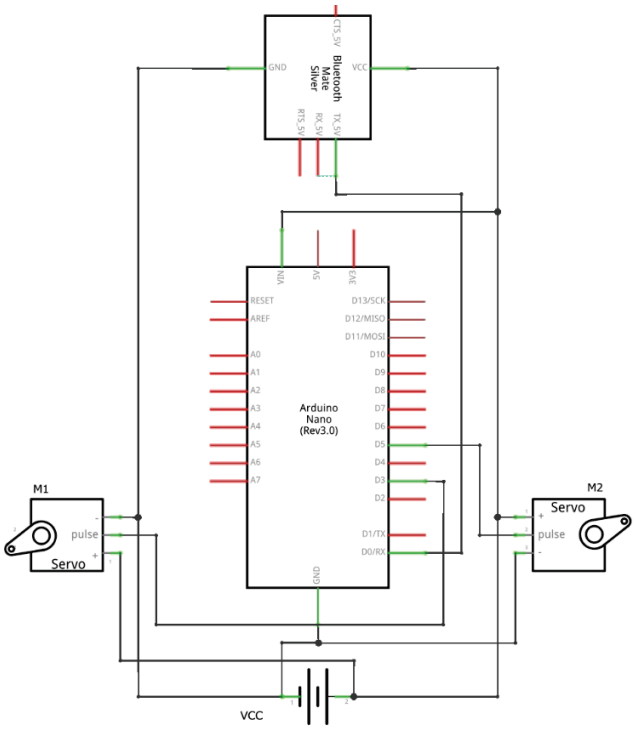


Figure 1 Diagram of differential configuration with remote control

The main component of the control module is the control unit, which essentially interprets the data from the sensors and, based on the control algorithm executed on it, sends control signals to the actuators of the drive module. This completes the basic control cycle sense-plan-act. Numerous open-source ecosystems are used in STEM education, and in this paper, in the testing phase of the modular platform, three types of control units that are suitable for integration into different levels of education are used, starting with BBC micro:bit, Arduino, and Pixhawk control unit, furthermore, ESP32 is considered for future application. Regardless, the control module encompasses remote control components, which communication methods rely on Bluetooth, radio control (across various frequencies), WiFi modules. More advanced control modules are further equipped with a telemetry link. There are also so-called shields that enable simple connection of different elements of the control module, sensor and drive components, and power supply with integrated different voltage levels. The sensor module should be compatible with the control module and power supply. Simpler versions of the sensor module, which are used mainly at the elementary school level of education, consist of distance sensors and IR sensors for line following. More advanced sensor modules suitable for use at a higher

level of education can consist of vision systems that enable Simultaneous Localization And Mapping (SLAM). The system diagram of the robot differential configuration, whose control module consists of an Arduino Nano board and a Bluetooth HC05 for remote control of the robot is shown in Fig. 1.

The locomotion mechanism affects the path planning, and the choice of mechanism often depends on the type of terrain for which the robot is intended. For the robot differential configuration, these are usually wheels or tracks for specific tasks. In this paper, the main focus is on the design of the modular chassis of the drive module, which, in addition to the standard differential configuration, enables the integration of an omnidirectional drive. The main component is the actuator, which has the task of converting the input electrical energy coming from the energy module (power supply) into mechanical outputs, i.e., angular velocity. In the proposed design, continuous servo motors were considered as they are available at a relatively low price and have a wide range of performances. Furthermore, they come in a very compact form, containing an electric motor, driver, and gearbox. Also, this type of actuator is used in other types of robots, such as educational robotic arms or Unmanned Aerial Vehicles (UAVs). The energy module essentially consists of batteries and additional elements for the distribution of electricity. For measurement purposes and in the validation phase of the proposed robot, two types of batteries are used, Nickel Metal Hydride (NiMH) and Lithium Polymer (LiPo) batteries. In order to identify the parameters of drive configurations, experimental measurements of the considered servo motors were carried out, the basic characteristics of which are shown in Tab. 1. Measurement results in the form of the characteristics of the Revolutions Per Minute (RPM) in relation to the Pulse Width Modulation (PWM) are shown in Fig. 2.

Table 1 Considered continuous servo motors

Continuous servo		Size (mm)	Weight (grams)	Operating voltage (V)
S1	HITEC HSR-1425CR	40.6 × 19.8 × 36.6	41.7	4.8 – 6.0
S2	FEETECH FS5103R	40.8 × 20.1 × 38.0	36.0	4.8 – 6.0
S3	PARALLAX High Speed	40.6 × 19.0 × 41.0	42.0	6.0 – 8.4
S4	PARALLAX Feedback 360	40.0 × 20.0 × 37.2	40.0	5.0 - 8.4

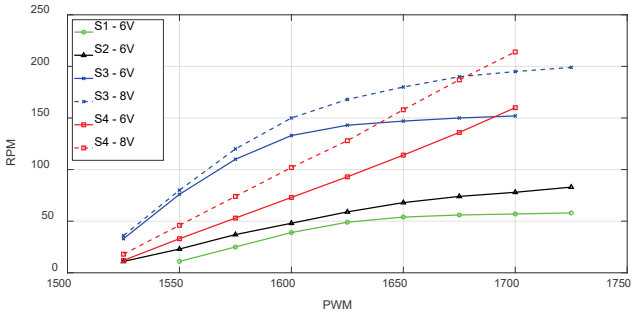


Figure 2 Characteristics of the four considered servo motors - revolutions per minute as a function of the input PWM signal

3 MATHEMATICAL DESCRIPTION OF THE UGV ROBOT

The modular UGV robot can be applied in a wide range of education, which is important from the aspect of preparing pupils, but also students for future technicians, engineers, and researchers. This type of robot exists in two-dimensional space and has three Degrees Of Freedom (DOF), position (x, y), and orientation (ψ) since the assumption is that it moves on a flat surface. For the purpose of mathematical representation, it is necessary to define two Cartesian reference coordinate systems. The base coordinate system (Earth frame) $\mathcal{F}^{\mathrm{E}}$, $\{\mathrm{O_E}, X_{\mathrm{E}}, Y_{\mathrm{E}}\}$, is attached to a stationary point on the Earth's surface where the position and orientation of the UGV robot are defined. Next, it is necessary to define the coordinate system of the robot, the so-called body frame $\mathcal{F}^{\mathrm{B}}$, $\{\mathrm{O_B}, X_{\mathrm{B}}, Y_{\mathrm{B}}\}$, which is attached to the robot chassis.

3.1 Kinematics of the Differential Drive Configuration

A mobile robot with a differential drive configuration due to the simple principle of operation is applied in a wide range of STEM education from the earliest age. Such a configuration most often consists of two actuators (electrical drives) whose arrangement enables rotation in place without changing position. The rotation of the robot, i.e., the heading angle, is determined by the differences in the angular velocities of the left and right actuators, on which, e.g., a wheel or a caterpillar drive element is mounted. A robot with a differential configuration belongs to non-holonomic robots [17] that do not have independent control of each of the variables (DOF), therefore, they perform complex manoeuvres to reach the desired state. Differential configuration robot parameters and variables, alongside reference coordinate systems, are shown in Fig. 3.

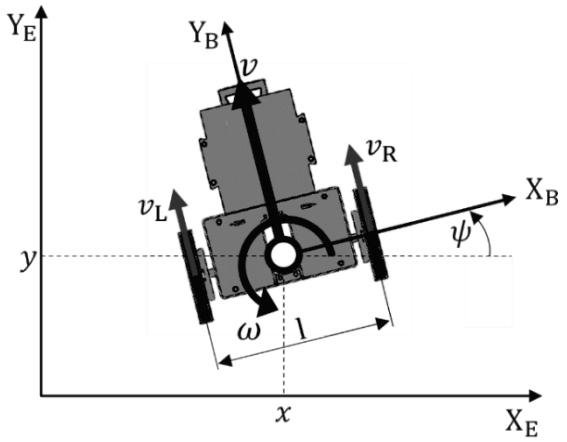


Figure 3 Differential drive configuration kinematic model

The kinematics of the differential configuration mobile robot is often described using the Unicycle-Type Wheeled robot model [18]. The robot velocities with respect to the earth frame are given with equations:

$$\dot{x} = v \cdot \cos\psi, \quad \dot{y} = v \cdot \sin\psi, \quad \dot{\psi} = \omega, \tag{1}$$

where v is the translational and ω is the rotational velocity of the robot with respect to the body frame, $\boldsymbol{v}^{\mathrm{B}} = [v \;\; \omega]^{\mathrm{T}}$. The problem of direct kinematics describes the mapping of the velocities of the left and right wheels $\boldsymbol{\Omega} = [\omega_L \;\; \omega_R]^{\mathrm{T}}$, which represent the input of the model, to the translational and rotational velocities of the robot, which represent the output of the model, and is given by the following expression:

$$\boldsymbol{v}^{\mathrm{B}} = \boldsymbol{\Gamma}_R \boldsymbol{\Omega} = \begin{bmatrix} \frac{d}{4} & \frac{d}{4} \\ -\frac{d}{2l} & \frac{d}{2l} \end{bmatrix} \begin{bmatrix} \omega_L \\ \omega_R \end{bmatrix}. \tag{2}$$

The mapping is described through the drive allocation matrix $\boldsymbol{\Gamma}_R$, which is defined by the configuration arrangement, in this case by the distance between the drive wheels l, and by the actuator parameters, in this case, wheel diameter d.

From the aspect of implementing robot control, it is necessary to map the translational and rotational velocities of the robot to the angular velocities of the left and right wheels, which represents the problem of inverse kinematics and is given by the following expression:

$$\boldsymbol{\Omega} = \boldsymbol{\Gamma}_R^{-1} \boldsymbol{v}^{\mathrm{B}} = \begin{bmatrix} \frac{2}{d} & \frac{l}{d} \\ -\frac{2}{d} & -\frac{l}{d} \end{bmatrix} \begin{bmatrix} v \\ \omega \end{bmatrix}. \tag{3}$$

Among different types of electric motor drives, in this proposed design, servo motors with continuous rotation were considered. When implementing system control, it is necessary to adjust the control variables to the signal sent to the actuator. The model presented in this way can be easily implemented in a software package for performing simulations and implementing experiments on robot control.

3.2 Kinematics of the Omnidirectional Drive Configuration with Mecanum Wheels

Holonomic robots have the ability to independently control each DOF variable. Unlike the conventional differential configuration, so-called omnidirectional robot configurations allow the performing of more complex movements in 2D space (on the surface). Such UGV configurations consist of a minimum of three actuators. Wheels that enable omnidirectional motion are the key assembly of the holonomic drive configuration. The geometric arrangement and wheel type depend on the number of actuators. The most common are configurations with three actuators consisting of so-called omnidirectional wheels [8], and configurations with four actuators consisting of mecanum wheels [19]. Due to the considered modular approach, the configuration with four actuators is further discussed in this paper. The velocities of the omnidirectional robot with respect to the earth frame, $\dot{\boldsymbol{\varepsilon}} = [\dot{x} \;\; \dot{y} \;\; \dot{\psi}]^{\mathrm{T}}$ are defined by the following expression:

$$\dot{\boldsymbol{\varepsilon}} = \boldsymbol{R}_{\psi} \boldsymbol{v}^{\mathrm{B}} = \begin{bmatrix} \cos\psi & -\sin\psi & 0 \\ \sin\psi & \cos\psi & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} v_x \\ v_y \\ \omega \end{bmatrix}, \tag{4}$$

where v_x is the translational robot velocity in the X_{B} axis, v_y is the translational velocity in the Y_{B} axis and ω is the rotational velocity of the robot with respect to the body frame as shown in Fig. 4.

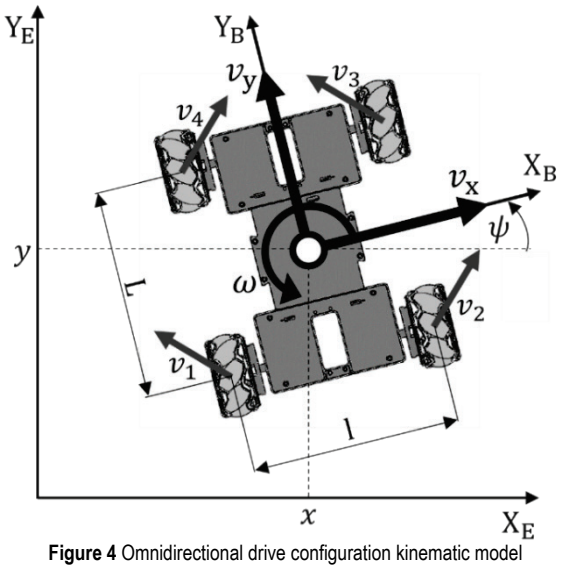


Figure 4 Omnidirectional drive configuration kinematic model

The translational and rotational velocities form the vector $\boldsymbol{v}^{\mathrm{B}} = [v_x \;\; v_y \;\; \omega]^{\mathrm{T}}$, and they depend on the geometric arrangement of the actuators, the parameters of the mecanum wheel, and the angular velocities of each wheel, $\boldsymbol{\Omega} = [\omega_1 \;\; \omega_2 \;\; \omega_3 \;\; \omega_4]^{\mathrm{T}}$, which represents the input of the model. More detailed descriptions of kinematics have been presented in numerous works, such as [20, 21]. The problem of direct kinematics of an omnidirectional configuration can be defined by the following expression:

$$\boldsymbol{v}^{\mathrm{B}} = \boldsymbol{\Gamma}_R \boldsymbol{\Omega} = \begin{bmatrix} -\frac{d}{8} & \frac{d}{8} & -\frac{d}{8} & \frac{d}{8} \\ \frac{d}{8} & \frac{d}{8} & \frac{d}{8} & \frac{d}{8} \\ -\frac{d}{4(L+l)} & \frac{d}{4(L+l)} & \frac{d}{4(L+l)} & -\frac{d}{4(L+l)} \end{bmatrix} \begin{bmatrix} \omega_1 \\ \omega_2 \\ \omega_3 \\ \omega_4 \end{bmatrix}, \tag{5}$$

where the distance between the drive wheels is l, the distance between axes is L, and by the actuator parameters, in this case, wheel diameter d, where the angle of the rollers is 45 degrees. For the inverse kinematics problem of the considered omnidirectional configuration, it is necessary to make a pseudoinverse since the allocation matrix is not square. Inverse kinematics can be defined by the following expression:

$$\boldsymbol{\Omega} = \begin{bmatrix} -\frac{2}{d} & \frac{2}{d} & -\frac{L+l}{d} \\ \frac{2}{d} & \frac{2}{d} & \frac{L+l}{d} \\ -\frac{2}{d} & \frac{2}{d} & \frac{L+l}{d} \\ \frac{2}{d} & \frac{2}{d} & -\frac{L+l}{d} \end{bmatrix} \begin{bmatrix} v_x \\ v_y \\ \omega \end{bmatrix}. \quad (6)$$

4 DESIGN PROCESS AND DEVELOPMENT OF PRINTABLE MODULAR ROBOTIC PLATFORM

In the design phase of the educational robotic platform, it is necessary to select the components and technologies for prototyping affordable low-cost robotic platforms. By choosing a control subsystem based on open-source ecosystems that can be programmed, the educational character of the platform is enabled. Groups of components form control, sensor, drive, and energy modules. The system itself is scalable, which enables the realization of various learning outcomes at different levels of STEM education and enables integrated education of engineering systems.

4.1 Design Considerations for a Modular Robotic Platform

The main focus of this work is on the design and development of a modular brick-based robot chassis, where each module combines 3D-printed parts with available standard electronic and mechanical components. The proposed approach enables the development of a highly expandable and versatile robotic platform. In order to realize the idea, it is necessary to carry out several phases, basically using software packages for 3D modelling and model preparation and a 3D printer for prototyping parts of the system. SolidWorks software package was used in the 3D modelling design phase. For the prototyping of parts, Fused Deposition Modelling (FDM) technology was considered since it is much cheaper compared to other AM technologies. Regardless of the AM technology, the process of designing and developing a mechatronic system is essentially the same as presented in [22]. The Prusa Slicer software was used for the preparation of parts and the Prusa i3 MK3S 3D printer for the parts prototyping. Key steps in the design phase of the robot chassis are shown in Fig. 5.

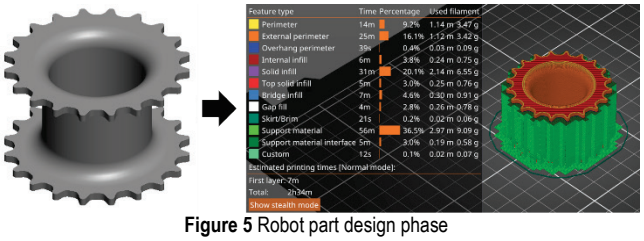


Figure 5 Robot part design phase

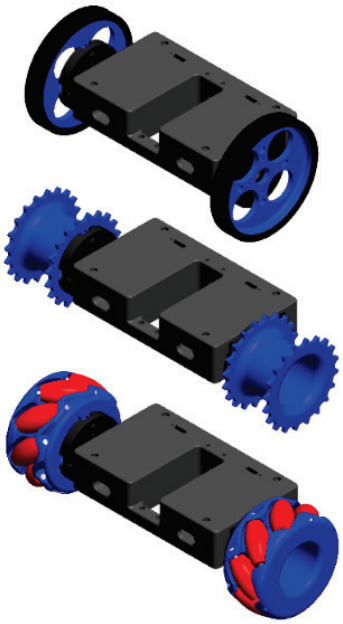


Figure 6 Drive module brick configurations

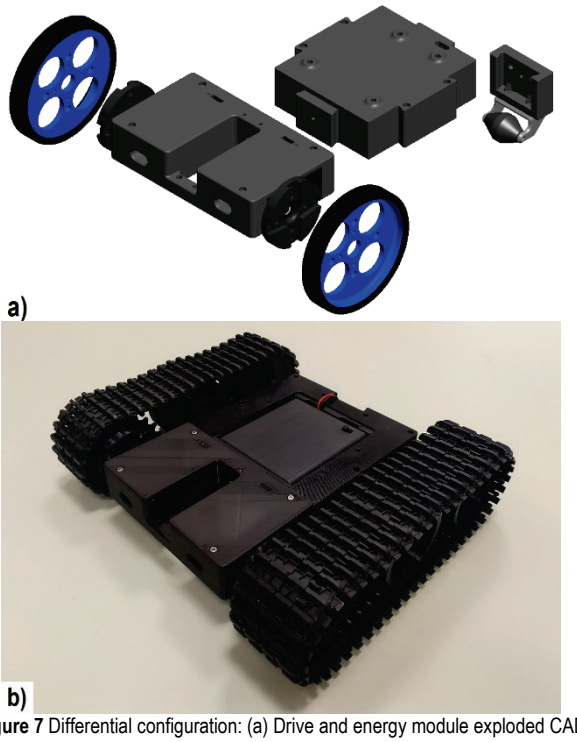


Figure 7 Differential configuration: (a) Drive and energy module exploded CAD model, (b) Assembled caterpillar platform

An approach to modular design in an educational sense can be presented through two production paradigms: mass customization and mass individualization. Paradigms refer to both, software and hardware, i.e., the structural parts of the assembly. In the field of STEM education, the second paradigm of mass individualization/personalization is of interest, which is enhanced by the use of 3D printing. Therefore, pupils/students will be able to participate in improving the design and creating their own constructions according to their needs, perfecting their knowledge. In order

to obtain complete open architecture products, not only product modularization is needed, but also interface compatibility. The platform and the module must have certain constraints and relationships, for the system to be valid. By replacing or adding a module the functionality of the robot changes, e.g., the different types of drive configurations can be used to adapt to different types of terrain (Fig. 6). Furthermore, different sensors can be used that can communicate with different control ecosystems, which changes the functionality of the overall system.

The development of the platform is actually an iterative process because the stages of construction, prototyping, and testing are repeated in order to eliminate design flaws and make further improvements. The basic assembly of the platform consists of a drive module with associated actuators, an energy module that basically consists of a battery, and a control and sensor module, depending on the level of education. The proposed design of the robot platform enables the rapid assembly of the functional robot. Exploded CAD model of the conventional differential configuration drive and energy module is shown in Fig. 7a. The differential configuration with the wheels consists of a freewheel that can be purchased, or it can also be prototyped using FDM technology. Such a basic platform can be further expanded with caterpillar elements. To assemble such a platform, it is necessary to replace the wheels with the caterpillar drive element, which is shown in Fig. 5, and it is necessary to expand the assembly with additional wheels that rotate freely and carry the tracks. Caterpillar tracks are affordable, a pair costs less than 10 euros, but they can also be prototyped. Assembled caterpillar platform ready for testing is shown in Fig. 7b.

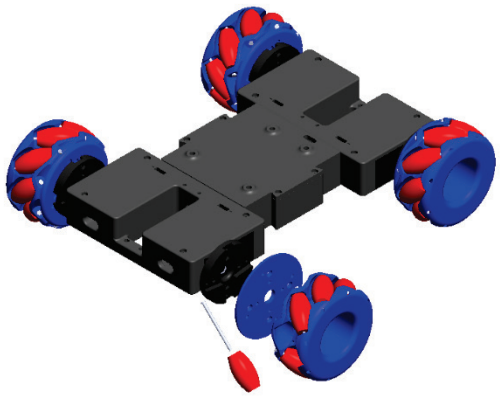


Figure 8 Omnidirectional configuration drive and energy module CAD model

The omnidirectional configuration of the robotic platform consists of two drive modules equipped with mecanum wheels. For this configuration, a mecanum wheel assembly is presented, consisting of printed parts and metal pins around which the wheel rollers rotate freely. The rollers are placed at an angle of 45 degrees, and the prototyped wheel consists of a total of eight rollers. The presented design significantly lowers the price of the robot, considering that such wheels are relatively expensive, and additional adaptation is required to fit the existing elements of the drive module. It is important to note that the layout of the connections of the drive module is considered, which in further work will enable the connection with the control and energy module through an element that represents a shield. The omnidirectional robot CAD model of the drive and energy module is shown in Fig. 8.

The price of the platform depends on the selected components. For instance, the servo brick can include two continuous servo motors (S2 – Tab. 1), priced at around 10 euros each, or two S4 motors, around 30 euros each. The materials for printing parts of the drive module brick, as seen in the Fig. 6, vary from three to six euros per assembly. The overall platform cost, for instance based on the ESP32 development ecosystem with the accompanying basic sensor package, the drive module brick module featuring two S2 motors and drive wheels, and the associated battery module, is below 50 euros per unit. Control modules utilizing Arduino and BBC micro:bit ecosystems are two to three times more expensive. Nevertheless, the material cost for proposed robotic module parts represents only a fraction of the total price.

4.2 Experimental Validation of Modular Robotic Platform Prototype

After assembly, in order to validate the proposed modular robot platform, tests of three versions of the robot with three different control ecosystems were carried out, as shown in Fig. 9. First, the basic configuration of the differential wheel drive in combination with the BBC micro:bit ecosystem was tested. A sensor module was used to track the line (Fig. 9a), which is one of the characteristic tasks at the elementary school level of education. This ecosystem can be programmed by block-based visual programming language (SCRATCH) or using Python, so it is suitable for a wider spectrum of integration into education. Furthermore, the caterpillar platform was tested in combination with the Arduino ecosystem, as shown in Fig. 9b, where remote control of the robot using a Bluetooth module and an Android application is utilized.

Finally, an omnidirectional robot was tested in combination with the PX4 module, which was used in previous work as a control module for UAV [23]. In the MATLAB software package, an experimental firmware was created in which the allocation for the omnidirectional configuration was implemented and the remote control (radio control) was tested, as shown in Fig. 9c.

It is planned to hold a workshop for pupils of the 7th and 8th grades of elementary school, where the differential configuration based on the BBC micro:bit ecosystem will be used. The preliminary demonstration workshop was successfully held. To conduct workshops for pupils of technical secondary schools, a platform based on the Arduino ecosystem with associated sensors and modules will be used. Regarding engineering education, the plan is to implement a modular robot platform through the mobile robots course, where the control module of the robot will be based on the ESP32 ecosystem with a companion Raspberry Pi on-board computer.

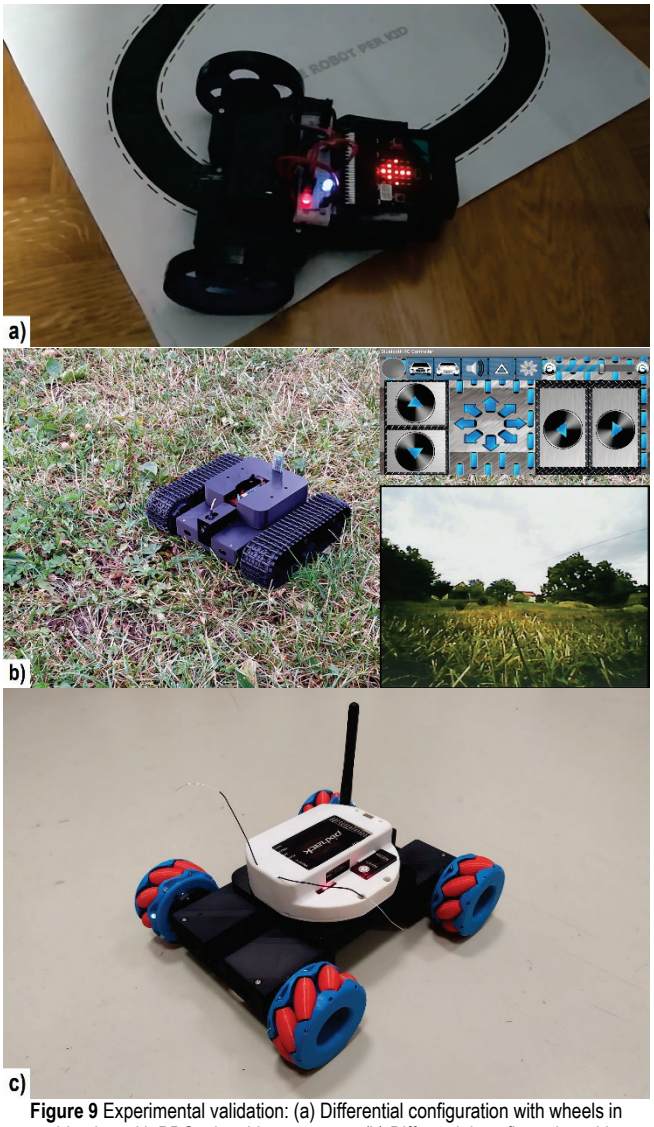


Figure 9 Experimental validation: (a) Differential configuration with wheels in combination with BBC micro:bit ecosystem, (b) Differential configuration with a caterpillar in combination with Arduino ecosystem, (c) Omnidirectional configuration in combination with PX4 ecosystem

5 CONCLUSION

In this paper, the design process of a modular UGV robotic platform intended for STEM education is proposed. Besides modularity, the convenience of the proposed educational robotic platform is the usage of commercial off-the-shelf and open-source components. A prerequisite for the implementation of simulations and real tasks is knowledge of system parameters, which is particularly important for the proper platform control and selection of drive actuators. The characteristics of considered servo motors were presented based on conducted experimental measurements. A novel modular UGV robot platform was mathematically described, and the model implemented into the software package enabled the execution of simulations and experiments. The parts of the chassis that make up the robot assembly were made using FDM technology, and experimental testing was carried out. The presented novel modular platform is validated for three drive module assemblies in combination with three control ecosystems. In future work, it is planned to implement workshops at three levels of education: elementary school, secondary school, and undergraduate degree.

6 REFERENCES

[1] Chatzopoulos, A., Kalogiannakis, M., Papadakis, S. & Papoutsidakis, M. (2022). A Novel, Modular Robot for Educational Robotics Developed Using Action Research Evaluated on Technology Acceptance Model. *Education Sciences, 12*(4). https://doi.org/10.3390/educsci12040274

[2] López-Rodríguez, F. M. & Cuesta, F. (2016). Andruino-A1: Low-Cost Educational Mobile Robot Based on Android and Arduino. *Journal of Intelligent & Robotic Systems, 81,* 63-76. https://doi.org/10.1007/s10846-015-0227-x

[3] Karalekas, G., Vologiannidis, S. & Kalomiros, J. (2020). EUROPA: A case study for teaching sensors, data acquisition and robotics via a ROS-based educational robot. *Sensors, 20*(9). https://doi.org/10.3390/s20092469

[4] Piperidis, S., Doitsidis, L., Anastasopoulos C. & Tsourveloudis N. C. (2007). A Low Cost Modular Robot Vehicle Design for Research and Education. *Proceedings of the 2007 Mediterranean Conference on Control & Automation*, 1-6. https://doi.org/10.1109/MED.2007.4433675

[5] Lumsden J. & Ortega-Sanchez, C. (2010). Modular Autonomous Robotics Platform for Educational Use. *Proceedings of the TENCON 2010 - 2010 IEEE Region 10 Conference*, 1577-1582. https://doi.org/10.1109/TENCON.2010.5686047

[6] Kochlán, M. & Hodon, M. (2014). Open Hardware Modular Educational Robotic Platform – Yrobot. *Proceedings of the 2014 23rd International Conference on Robotics in Alpe-Adria-Danube Region (RAAD)*, 1-6. https://doi.org/10.1109/RAAD.2014.7002246

[7] Yuen, M. C., Chan, K. K. & Li, K. S. (2021). Mobile App Controlled Modular Combat Robot for STEM Education. *Proceedings of the 2021 International Conference on COMmunication Systems & NETworkS (COMSNETS)*, 64-68. https://doi.org/10.1109/COMSNETS51098.2021.9352927

[8] Bonarini, A., Matteucci, M., Migliavacca, M. & Rizzi, D. (2014). R2P: An open source hardware and software modular approach to robot prototyping. *Robotics and Autonomous Systems, 62*, (7), 1073-1084. https://doi.org/10.1016/j.robot.2013.08.009

[9] Fang, Z., Fu, Y. & Chai, T. (2009). A low-cost modular robot for research and education of control systems, mechatronics and robotics. *Proceedings of the 4th IEEE Conference on Industrial Electronics and Applications*, 2828-2833. https://doi.org/10.1109/ICIEA.2009.5138725

[10] Le, V. T. & Ngo, T. D. (2017). moreBots: System development and integration of an educational and entertainment modular robot. *Proceedings of the 2017 IEEE International Symposium on Robotics and Intelligent Sensors (IRIS)*, 74-80, https://doi.org/10.1109/IRIS.2017.8250101

[11] Pacheco, M., Fogh, R., Lund, H. H. & Christensen, D. J. (2015). Fable II: Design of a modular robot for creative learning. *Proceedings of the 2015 IEEE International Conference on Robotics and Automation (ICRA)*, 6134-6139. https://doi.org/10.1109/ICRA.2015.7140060

[12] Wolfe, K. C., Moses, M. S., Kutzer, M. D. M. & Chirikjian, G. S. (2012). M3Express: A low-cost independently-mobile reconfigurable modular robot. *Proceedings of the 2012 IEEE*

International Conference on Robotics and Automation, 2704-2710. https://doi.org/10.1109/ICRA.2012.6224971

[13] Mitsuhashi, K. & Ohyama, Y. (2012). Suggestion and Verification of the Modular Robot Education. *Proceedings of the 2017 11th Asian Control Conference (ASCC)*, 182-186. https://doi.org/10.1109/ASCC.2017.8287373

[14] Dimitrios, S., Fotios, G., Emmanouil, S., Areti, P., Dimitris, R. & Christos, S. C. (2021). A novel, fully modular educational robotics platform for Internet of Things Applications. *Proceedings of the 2021 1st Conference on Online Teaching for Mobile Education (OT4ME)*, 138-145. https://doi.org/10.1109/OT4ME53559.2021.9638892

[15] Lee, B. Y., Liew, L. H., Khan, M. Y. B. M. A. & Narawi, A. (2020). The Effectiveness of Using mBot to Increase the Interest and Basic Knowledge in Programming and Robotic among Children of Age 13. *Proceedings of the the 6th International Conference on E-Business and Applications*, 105-110. https://doi.org/10.1145/3387263.3387275

[16] KUKA AG, https://www.kuka.com/en-de/products/mobility/mobile-platforms/kuka-omnimove

[17] Hassan Zarabadipour, H. & Yaghoubi, Z. (2019). Control of a non-holonomic mobile robot system with parametric uncertainty. *Technical Journal, 13*(1), 43-50. https://doi.org/10.31803/tg-20190116100550

[18] Kim, B. M. & Tsiotras, P. (2002). Controllers for Unicycle-Type Wheeled Robots: Theoretical Results and Experimental Validation. *IEEE Transactions on Robotics and Automation. 18*, (3), 294-307. https://doi.org/10.1109/TRA.2002.1019460

[19] Alshorman, A. M., Alshorman, O., Irfan, M., Glowacz, A. Muhammad, F. & Caesarendra, W. (2020). Fuzzy-Based Fault-Tolerant Control for Omnidirectional Mobile Robot. *Machines, 8*(3). https://doi.org/10.3390/machines8030055

[20] Muir P. & Neuman, C. (1987). Kinematic modeling for feedback control of an omnidirectional wheeled mobile robot. *Proceedings of the 1987 IEEE International Conference on Robotics and Automation*, 1772-1778. https://doi.org/10.1109/ROBOT.1987.1087767

[21] Röhrig, C., Heß, D. & Künemund, F. (2017). Motion controller design for a mecanum wheeled mobile manipulator. *Proceedings of the 2017 IEEE Conference on Control Technology and Applications (CCTA)*, 444-449. https://doi.org/10.1109/CCTA.2017.8062502

[22] Piljek, P., Krznar, N., Krznar, M. & Kotarski, D. (2022). Framework for Design and Additive Manufacturing of Specialised Multirotor UAV Parts. In Răzvan Păcurar (Ed.), Trends and Opportunities of Rapid Prototyping Technologies [Working Title]. IntechOpen. https://doi.org/10.5772/intechopen.102781

[23] Kotarski, D., Piljek, P., Pranjic, M., Grlj, C. G. & Kasac, J. (2021). A Modular Multirotor Unmanned Aerial Vehicle Design Approach for Development of an Engineering Education Platform. *Sensors, 21*(8). https://doi.org/10.3390/s21082737

Authors' contacts:

Denis Kotarski, PhD
(Corresponding author)
Trivium STEM Edu association,
Ciglenica Zagorska 10, 49223 Sveti Križ Začretje, Croatia
denis.kotarski@gmail.com

Petar Piljek, PhD, Assistant Professor
University of Zagreb, Faculty of Mechanical Engineering and Naval Architecture,
Trivium STEM Edu association,
Ivana Lučića 5, 10002 Zagreb, Croatia
petar.piljek@fsb.hr

Tomislav Šančić
Karlovac University of Applied Sciences,
Josipa Jurja Strossmayera 9, 47000, Karlovac, Croatia
tsancic@vuka.hr

ISSN 1846-6168 (Print), ISSN 1848-5588 (Online)
https://doi.org/10.31803/tg-20230715000112

Original scientific paper
Received: 2023-07-15, Accepted: 2024-01-02

Identification of LSA Data Retrieval Method and Temporal Graph for Document Retrieval

Shahla Rezvani, Nader Naghshineh*, Ahmad Khalilijafarabad

Abstract: The field of expert finding has seen a large number of approaches proposed both in universities and in industries, using a variety of new techniques in relevant data fields. This study tends to identify information retrieval method of latent semantic analysis and temporal graph for document retrieval. In this study, citation occurrence and author occurrence are independent variables and scales of expert author finding are dependent variables. The method used to evaluate judgment of document and author relevance in the test set formation phase is more similar to survey methods. Library method is used to study theoretical foundations and judge literature. This study has three populations: a) test set documents; b) people who make queries and judge relevance of retrieved documents; c) people who judge relevance of the retrieved experts. To measure judgments of document relevance, a method similar to peer tests is used. Among the retrieved results, repeated results are placed to determine accuracy and reliability of the judge. The degree of correlation obtained in this method is very high (0.98), indicating the reliability of the results. Regarding the results of the current study on application of latent semantic indexing (LSA) information retrieval model, which was ultimately used to retrieve expert authors, the performance of LSA-based retrieval model outperformed the baseline model. This was evident from the obtained metrics, including precision at the top 5 results (p@5) with a value of 0.895, mean average precision (MAP) of 0.839, and mean reciprocal rank (MRR) of 0.909. The improved retrieval performance can be attributed to the superior performance of the dimension reduction method compared to keyword matching.

Keywords: expert finding; expert retrieval; information retrieval; information systems; temporal graph; test set

1 INTRODUCTION

Latent semantic indexing (LSA) was one of the solutions that emerged to solve problems of vector space model. Vector space model is one of several methods for detecting similarity between two documents, which was developed by Salton in 1975 [1]. Vector space model treats dissimilar terms as irrelevant items. This, i.e. mismatch of terms, is the main problem of vector space model. The second drawback of these models is formation of a large and scattered phrase-document matrix, occurring in large document sets which requires a large storage space, and its processing and calculation time is also long. To solve these problems, one of the branches of this model called LSA is used, which has become popular in recent years [2, 3]. Latent indexing was first used by a group of scholars, Deerwoster et al. at Bell Corporation, for information retrieval [4] and was called latent indexing. The reason why scholars use the word 'latent' is that new statements that represent semantic information are not found directly from documents, but are the result of examining the set of documents and using the mathematical method called singular value decomposition (SVD). LSA assumes that usually the entire semantic content of a text such as paragraph, abstract or entire document is approximately equal to sum of the meanings of its words. Stable meaning of word representations can also be obtained from a large document set by considering each text as a linear equation of the entire set of documents in the form of a system of concurrent equations [5-7]. Yih et al. [8] showed that LSA improves vector space models by preferring the semantic concept of evidence over its words. In LSA, unlike document vector models that assume independent terms and words, different levels of correlation, dependence or connection are considered for them, and these connections between terms are characterized by formation of a new set of statements using statistical SVD method. In latent semantic space, a question and a document can also have a great cosine similarity even if they do not have a common statement. Because their statements are semantically similar, while this is not possible in vector models. The most important strength of LSA is retrieval efficiency based on user question, which is achieved through matrix calculation. As by using this method, relevant documents are retrieved even if the words of document content do not match with each other [9, 10].

Temporal information retrieval is presently a subject of study in the realm of information retrieval. Given the extensive amount of data available on the internet and the significant impact of time on document contents, procuring pertinent information becomes a formidable task. Traditional information retrieval approaches that are based on thematic similarity are not sufficient for searching the set of temporal documents. Time dimension of the documents available in the documents should be associated with the ranking of the documents for efficient retrieval. The objective of an information system is to detect documents that are relevant to a given query. However, the challenge lies in the fact that documents are influenced by time and continue to accumulate, resulting in a substantial number of irrelevant documents in the retrieved set. Consequently, users are compelled to invest more time in searching for documents that fulfill their information requirements. With the continuous influx of information in the digital realm, the incorporation of time as a crucial factor becomes significant for a wide range of searches [11, 12]. Kanhabua and Nørvåg [13] analyzed query logs and showed that a significant fraction of queries is temporal, that is, time-dependent relevance, and temporal queries play an important role in many fields such as digital libraries and document archives. There are two categories of temporal queries: 1) Those that provide a certain time criterion; and 2) those that are presented without time criterion. In this particular case, candidates who have retired or are no longer alive will not be retrieved. The introduction of the concept of graph and the subsequent development of graph theory are considered

groundbreaking accomplishments in the fields of mathematics and various disciplines over the past few centuries. There are innumerable and numerous applications for graphs across a wide range of scientific disciplines, ranging from maths to computer science to chemistry to biology, to name a few. These domains inherently present significant problems that can be effectively represented and explored through the use of graphs. A temporal graph, as defined by [14], is a data structure comprising nodes and edges that are associated with time labels. There are two types of temporal graphs: those that undergo changes over time and those that do not. The temporal graph can be thought of as one in which time is discretized, in which only the relationships between participants may change instead of the different types of entities [15]. In the current study, the candidates who have retired may no longer have any publications in the considered field, and as a result, updated information is not available to the users. Considering the rapid changes that occur in the world of information and the passing of a few years reduces useful life of information, it is necessary to retrieve candidates who have up-to-date publications on the subject needed by users. This also applies to non-living candidates.

Latent semantic analysis [4] converts queries and documents into a latent semantic space. Cosine similarities can be found between queries and documents (cosine of the angle between two vectors, the closer it is to 1, the smaller the angle between two vectors), even though they have no common term; but as long as their terms have semantic similarity, this applies. Latent semantic space has less dimensions than latent space [16, 17]. Therefore, LSA is a method to reduce dimension. A temporal graph for document retrieval is presented in this study in order to identify the LSA information retrieval method.

2 LITERATURE REVIEW

Graphs easily present a series of objects and a set of dual relationships between them. These graphs commonly display bidirectional relationships, accompanied by supplementary details. As an illustration, consider a graph that represents a collection of cities interconnected by roads, where each edge (C1, C2) includes information on the average travel time from city C1 to C2. Similarly, in a graph designed to establish connections between atoms within a molecule, edges may contain additional data on bond order or bond strength. Such scenarios can be modeled using weighted, or more generally, labeled graphs, where each edge (and sometimes node) is assigned values from specific domains, such as the set of natural numbers. An example of a well-studied and extensively explored type of labeled graph is graph-coloring regions [18-20, 39]. A temporal graph is also known as a dynamic and evolutionary graph and can be informally described as a graph that changes with time. In terms of modeling, they can be considered as a special case of labeled graphs in which the labels include some time measurements.

Numerous applications and areas of research have identified the potential benefits associated with the development of a comprehensive collection of results, tools, and techniques for temporal graphs. Various types of networks, including but not limited to information and communication networks, social networks, transportation networks, as well as several physical systems, lend themselves naturally to being represented as temporal graphs.

A graph includes two sets; a non-null set of nodes or vertices and a set of edges that connect the vertices. Assume cities of a country as vertices and the roads between them as edges of a graph. A name is assigned to each vertex or each edge of the graph. A null graph is a graph that contains only vertices and its set of edges is empty, that is, it has no edges. A graph can be directed or undirected. A directed graph is a graph in which direction of each edge is determined. In a directed graph, order of the vertices in each edge is important, and the edges are drawn with arrows from the first vertex to the end vertex. In an undirected graph, one can move between vertices in both directions, and order of the vertices does not matter (Fig. 1).

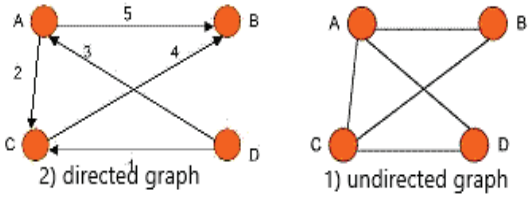


Figure 1 Types of graphs

Maximum number of edges in a simple directed graph with n vertices is $n \times (n-1)$.

Maximum number of edges in a simple undirected graph with n vertices is equal to $n \times (n-1)/2$.

2.1 Expert Finding

If technology is to strengthen the effective use of a wide range of knowledge, organizations should be able to not only use access to open and documented knowledge, but more importantly, they can also take advantage of tacit knowledge that other people have [35-38, 40-43]. By increasing the visibility and traceability of such knowledge, the knowledge can be analyzed and shared for strengthening the formation and sustainability of virtual organizations and companies, scientific communities, expert networks, etc. Yimam-Seid and Kobsa [21] identified two motives for expert finding, which are: 1) expert as a source of information; 2) expert as someone who is able to perform organizational or social role. Different situations where an expert is searched as a source of information are as follows:

1) Access to undocumented information. All the information within the organization is not fully documented; most of important information is gained through internships, experiences, and informal conversations. In many new situations, documented information is rarely helpful. Sometimes, the required information is not available to the public due to various economic, social and political reasons.
2) Specification of information needs. The information that users need is often not clear and specific. Therefore, it is necessary for them to consult with an expert to determine exactly what their information needs are.

3) Relying on expertise of others. Users often tend to spend less effort and time to find the information they need. This makes them use experts to select useful information from the huge amount of information.
4) Needs interpretation. In many cases, users are not interested in the information itself, but rather in its applications or interpretations. In some cases, they are not able to understand the retrieved information. In this situation, they turn to experts to interpret or understand the information.
5) The need to socialize. Users may raise their information needs with humans instead of interacting with documents and computers.

To satisfy the second type of motivation, those experts are required who are able to perform a specific task or role in the organization. In these cases, there is a need for those who have a certain type of expertise necessary to play a role in special situations. This occurs when inclusion of an expert in a specific activity or continuous cooperation between him and the searcher is considered. Examples include:

1) Searching for a consultant, employee or contractor,
2) Searching for a colleague, team member, committee member, or judge of journal and conference articles,
3) Searching for an expert speaker, lecturer, researcher, and interviewee for the media.

Searching for experts as a source of information is considered "information need", and searching for experts for the purpose of entrusting them with a role or task is considered "expertise need".

Table 1 Relevant research studies

No.	Authors	Result
1	Karimzadehgan et al. [22]	An algorithm was proposed to enhance the performance of expert finding systems by considering not only the expertise of an individual but also the expertise of their peers. The experimental results demonstrated that incorporating this supplementary information resulted in improved retrieval performance for the expert finding system.
2	Macdonald [23]	Effective document retrieval approaches have a positive effect on performance of voting techniques. Finally, this thesis states that the proposed model can be used to search for other people such as bloggers in the environment of blogs, and to suggest judges for papers submitted to conferences.
3	Daud et al. [24]	Occurrence of topics (probabilistic semantic classification of words) and correlation changes throughout time, while meaning of a specific topic remains almost unchanged. The results significantly show superiority of subject time general modeling approach over modeling approach that was not based on effect of conference and information time.
4	Smirnova & Balog [25]	To test their proposed algorithm, researchers used a real test set created from interactions of employees at the Tilburg University level, and the results showed that substantial progress has been made in all retrieval measures in the basic approach.
5	Omidvar et al. [26]	Novelty of this study is to find semantic relevance of the posts using text mining technique and semantic similarity provided by using WordNet. Evaluation tests were used to calculate precision of the proposed method and compare with other methods, and the results showed that it performed better.
6	Attapur [27]	In document retrieval stage, performance of other models was not significantly different from each other except for the Hiemstra linguistic model, which had a weaker performance compared to other models. Therefore, documents retrieved by any of the vector space models, Dirichlet language model, Okapi BM25 probabilistic model, PL2 model, and DLH13 model can be used to extract and rank authors.
7	Lipani [28]	The research article focused on the aim of studying biases in information retrieval systems. The thesis emphasized the reduction of retrieval systems to filters or sampling processes as a means to systematically investigate these biases. By approaching retrieval systems from that perspective, the researchers aimed to uncover and understand the various biases present in these systems. That knowledge could lead to the development of more effective and unbiased retrieval systems in the future.
8	Wu et al. [29]	This research paper aimed to address the retrieval of interns from CQA (Community Question Answering) websites. To identify suitable candidates for internship programs in companies, it introduced the notions of generalist and shape of expertise. The researchers conducted experiments using three test collections extracted from StackOverflow, comparing the effectiveness of their models against several baseline approaches. They presented retrieval models and assessed their performance using specific measures, highlighting their effectiveness compared to baseline methods.

It is shown in Tab. 1 that all works examined estimate the relationship between search queries and supporting documents concerning expertise retrieval using the occurrence of query terms in those supporting documents for the search queries. These models are not capable of semantic relevance. By using these models, if there are candidates really related to query and query words are not used in their respective supporting documents, but they contain words synonymous with the query words, they will not be considered as expert in that query. Subject-based models tend to solve this problem. The goal of these models is to somehow include the concept of meaning in the retrieval process. Two people with different words may describe a subject. Because of this, different people are introduced as experts for a specific query raised by different people. Usually, to calculate the degree of relevance of a document to a query, frequency of occurrence of the words of a query in the document is considered. If words other than query words with the same concept are used in the text, they are not considered as relevant words. This has a negative effect on calculation of relevance of a document to the query. As a result, it is necessary to consider multiplicity of meanings (one word with multiple meanings) or synonyms (different words with the same meaning) for words when calculating the degree of relevance of a document to a query. These descriptions show the challenges that exist in the expert finding process based on subject.

In this study, the supporting documents are the articles published by the authors, and the time factor or the date of publication of these articles is also considered as an important feature, which can be similar to giving different weights to different sources of expertise; of course, a time graph was used to include it.

3 RESEARCH METHODOLOGY

In this study, citation occurrence and author occurrence are independent variables and expert finding scales are dependent variables. In a typical classification, experimental research is divided into three types: real, preliminary and semi-experimental. The current study is also consistent with characteristics of a single-stage case study of preliminary experimental research subsets; it can be considered preliminary experimental research [27]. The method used to evaluate judgement of document and author relevance in the test set formation phase is more similar to survey methods. Library method is used to study theoretical foundations and literature.

This study has three populations: a) test set documents; b) people who make queries and judge relevance of the retrieved documents; c) people who judge relevance of the retrieved experts. The test set consists of a huge set of documents and has three components: set of documents, a set of subject elements (queries), and a set of relevance judgements [30]. Document set that makes up the main body of the test set is the first population. These documents consist of English articles in the field of information science and librarianship that have been indexed under the subject of information science and librarianship in the Web of Science database from 1989 to 2018 (N = 126924). Queries created by users were presented to all these articles and no sampling was done among the above articles. In order to make queries and make relevant judgements about the retrieved documents and authors, people should be selected who have knowledge of the query subject area and expert authors and have the ability to make relevant judgements about the retrieved documents. For this reason, graduates and postgraduate students of information science and epistemology of Tehran University will be selected as one of the populations.

The third population consists of people who judge relevance about expert authors retrieved for each query. These people should have a comprehensive understanding of the query as well as experts in that subject. To judge relevance of the retrieved expert authors, ten queries are randomly selected from each of the queries and presented to eight people who are introduced by people of the second population.

Various methods are used to calculate reliability coefficient of the measurement instrument, including re-run (retest method), parallel method (peer), split-half and Kudrichardson's method and Cronbach's alpha. To measure judgments of relevance of documents, a method similar to peer tests is used. For this purpose, repeated results are placed within the retrieved results to determine precision and reliability of the judge. The degree of correlation obtained in this method is very high (0.98) and indicates reliability of the results. To measure the relevance judgment of expert authors, parallel tests (simultaneous judgment of several people) is used.

4 RESULTS

To select the required number of components that represent maximum information available in the matrix, scree plot diagram is used, which according to the article presented for 250 components is as shown in Fig. 2. According to elbow position in Fig. 2, n = 100 for the number of components will be statistically precise enough to separate the subjects.

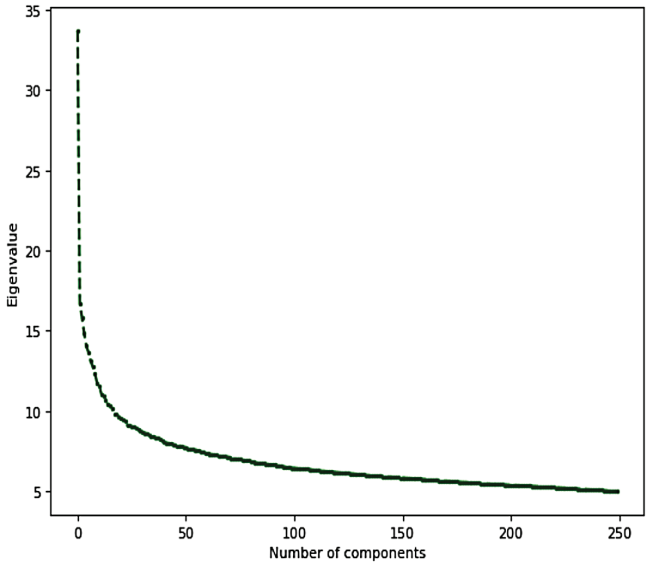


Figure 2 Scree plot for determining the number of components

The next step in this project is applying queries in the form of vector multiplication and finding relevant documents. In this method, the query is transformed into a vector with dimensions as long as the number of components of the above matrix based on the words in it and multiplied in each row of the result_transform matrix. The result gives us a numerical value, the larger it is, the closer it is to the relevant subject and, as a result, closer to content of that document (Tab. 2).

Table 2 Comparison of three measures in LSA model and the basic model

p@5 in LSA	MAP in LSA	MRR in LSA	p@5 in DLH13	MAP in DLH13	MRR in DLH13
0.895	0.839	0.909	0.887	0.567	0.903

After inserting the stage packages, including Numpy and Pandas, as well as JGraph package for drawing the graphs, root folder is determined, that is, address of the storage location, and then results file is read again in the form of a data frame by Pandas.

The "create-array" function converts the same variables into a list. Inside the parentheses is the query number. For example, if the number 100 is placed inside the parentheses, an array of author names, date of publication and citations of articles relevant to this query will be calculated.

In "get-names-year" function, main author name and year of publication are extracted from the references.

In "get-auth-names" function, the list of all authors of the article is separated because it is in the form of a string.

The "get-years" function extracts publication year from string mode and transforms it to list mode.

The "get-unique-name" function removes duplicate names from the list and provides a set of non-duplicate names.

The "draw_raph" function is for drawing graphs. Its sub-function i.e. "ith_graph" calculates the i^{th} function.

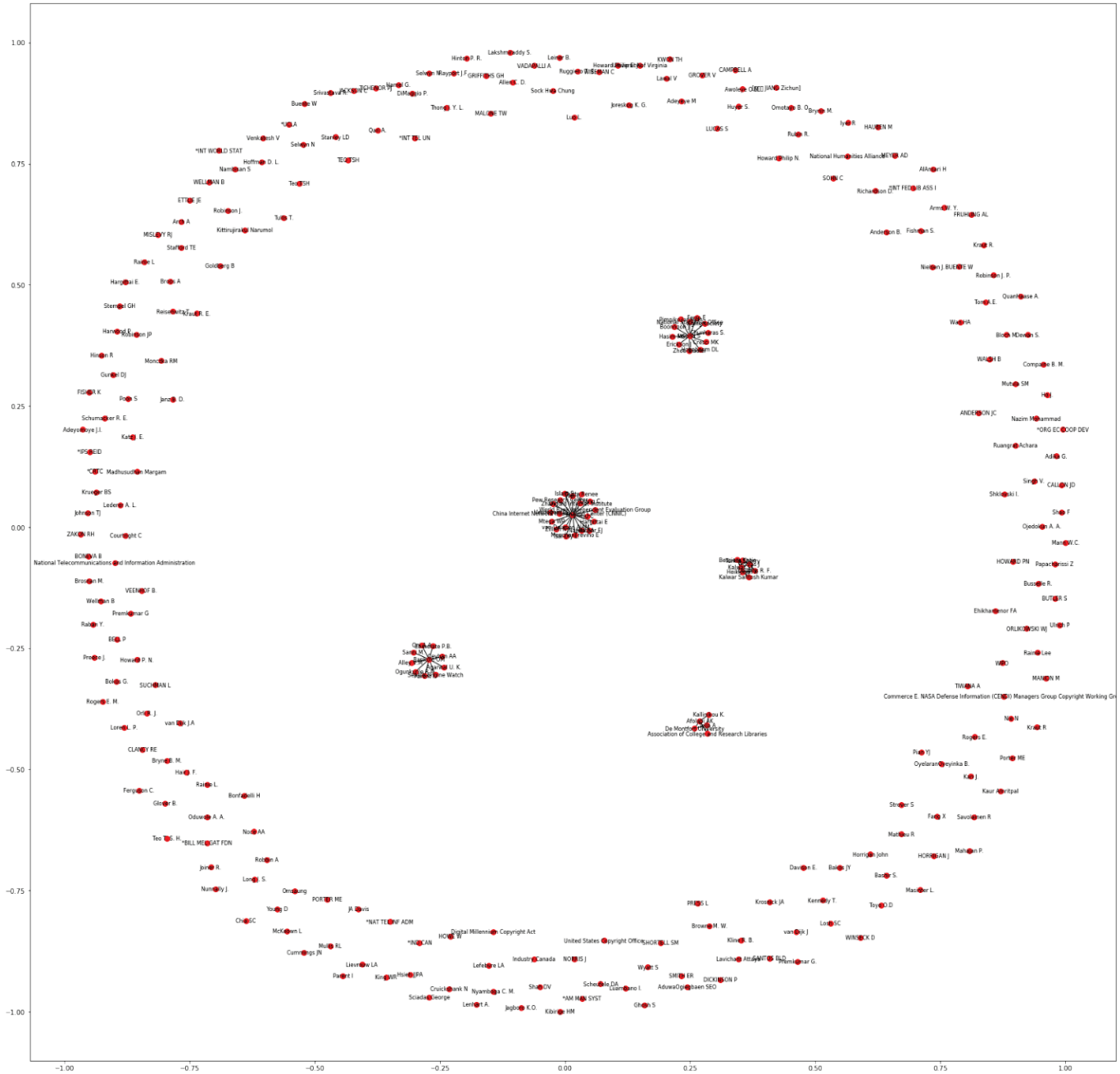


Figure 3 Temporal graph

As shown in Fig. 3, authors whose publication year was not in the period from 2008 to 2018 were removed from the graph.

In the following, authors should be retrieved whose works are published in the last ten years, because they lose their usefulness after ten or fifteen years, depending on the subject. Sciences that are mostly theoretical, such as mathematics, have a long half-life, and sciences that depend on new, up-to-date topics and technology, such as medicine, have a short half-life [44]. The minimum year is calculated from 2008. Then centrality measures, i.e. degree, closeness, betweenness and eigenvector centrality, followed by busy author, are calculated in the first hundred results.

The number of 10 queries from each study, a total of 20 queries, were randomly selected and the specified experts of each study were given a score of zero or one by the third statistical population, which was consisted of 8 people. The basic model was given value 4 by the 1st person, value 5 by the 2nd person, value 2 by the 3rd person, value 4 by the 4th person, value 4 by the 5th person, value 5 by the 6th person, value 8 by the 7th person and value 5 by the 8th person. Time inclusion and temporal graph were given value 8 by the 1st person, value 7 by the 2nd person, 10 by the 3rd person, 9 by the 4th person, 9 by the 5th person, 9 by the 6th person, 7 by the 7th person and 8 by the 8th person. The results of expert relevance judgment are shown in Tab. 3 and Fig. 4.

Table 3 Comparison of relevance judgement of the retrieved experts in the current model and basic model

Current model	*N*	Basic model	*N*
1st person	7	1st person	4
2nd person	8	2nd person	5
3rd person	10	3rd person	2
4th person	9	4th person	4
5th person	9	5th person	4
6th person	9	6th person	5
7th person	7	7th person	8
8th person	8	8th person	5
Sum	67	Sum	46

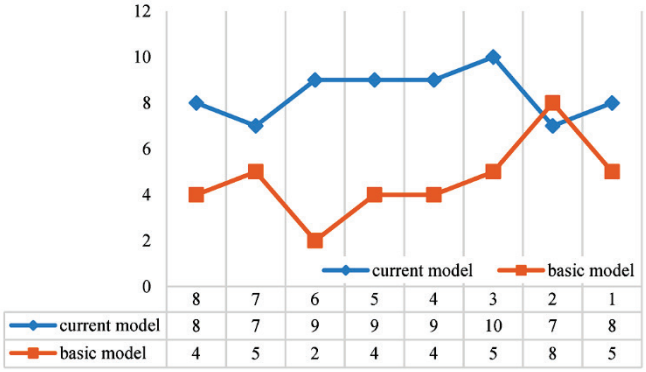


Figure 4 Comparison of the retrieved expert judgment in two models

According to the diagram, the retrieved experts have a higher relevance judgment in the current model, except for one point.

5 CONCLUSION

This study tended to identify LSA information retrieval method and temporal graph for document retrieval. According to the findings, LSA retrieval model outperformed the basic model in terms of p@5 (0.895), MAP (0.839) and MRR (0.909). The reason for the improved retrieval performance using dimension reduction methods compared to keyword matching lies in the use of latent semantic indexing (LSI), which is a type of conceptual indexing that utilizes statistical methods such as least squares estimation. The aforementioned indexing is extracted by employing this statistical approach. As we know, there are various ways to express a word (synonyms), so it is possible for query words to not be matched with document words. Additionally, many words have multiple meanings (polysemy), making information retrieval based on the concept and meaning of a document a better approach. LSI assumes that there are hidden structures in the usage of words that are partially captured by selecting diverse words. Singular value decomposition (SVD) is used to estimate these structures. The vectors obtained statistically enhance the representation of meanings more than individual words. Other research results also indicate that document retrieval using keyword matching is weaker compared to other methods. Furthermore, the performance of proposed model in retrieving documents is more pronounced in larger document sets compared to smaller ones [31]; however, this was not investigated in the current study. The results showed LSA has a better performance [32, 33].

One of the limitations of this study was that full text of the test set documents was not reviewed. The presence of full text of articles may have a positive effect on performance of information retrieval models. However, Bogers, Kox & Van Den Bosch [34] showed that performance of information retrieval models is better when indexed documents contain only abstracts than when indexed documents contain full text. This greatly reduces the problem of negative effect of not using the full text. The lack of a thesaurus in the document retrieval system is not a limitation due to the nature of LSA retrieval model explained above. In the document retrieval stage, relevance of the documents was judged based on title, abstract and keywords of the articles, which are common in most information retrieval systems. Since the present study tended to use the retrieved documents in the next stage to find expert authors, presenting the author names might lead to bias of the judges. For this reason, author names were not included in the relevance judgment files. However, information retrieval systems can consider the effect of the presence of these items on relevance judgment of users.

Based on the results of this research, the incorporation of temporal factors in expert finding and the utilization of social network indicators significantly enhance the performance of the method. The use of temporal factors to prevent the retrieval of individuals who are no longer active or have not published relevant works for an extended period has shown a significant improvement in the performance of the expert finding method compared to the baseline model. These findings can contribute to the enhancement of expert finding methods in the field of library and information science. Additionally, the use of social network centrality measures such as degree centrality, betweenness centrality, closeness, and eigenvector centrality as determining factors has also demonstrated a considerable improvement in the performance of the expert finding method compared to the baseline model. In the present study, ten queries were designed and sent to eight selected participants from the research population, and the results indicated that the use of the temporal graph and expert finding methods incorporating the factor of the highest number of published relevant works and the factor of social network centrality indicators yielded better performance.

Developing an organizational expert finding system that relies on expert profiles involves leveraging the organizational hierarchy [45]. This approach allows the

prediction of relationships between managers, subordinates, and peers, as well as the identification of members with limited or no available information. The algorithm employed in this system takes into account not only the expertise of the individual members but also the expertise of their peers.

6 REFERENCES

[1] Salton, G., Wong, A. & Yang, C. S. (1975). A vector space model for automatic indexing. *Communications of the ACM*, *18*(11), 613-620. https://doi.org/10.1145/361219.361220

[2] Rezvani, S., Naghshineh, N. & Khalilijafarabad, A. (2023). Implementation of Experts' Retrieval Model Using Latent Semantic Indexing (LSA) Method and Temporal Graph. *Library and Information Science Research*.

[3] Ramezani, M., Shahryari, M. S., Feizi-Derakhshi, A. R. & Feizi-Derakhshi, M. R. (2023). Unsupervised broadcast news summarization; a comparative study on maximal marginal relevance (MMR) and latent semantic analysis (LSA). In *The 28th IEEE International Computer Conference, Computer Society of Iran (CSICC)*, 1-7. https://doi.org/10.1109/CSICC58665.2023.10105403

[4] Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K. & Harshman, R. (1990). Indexing by latent semantic analysis. *Journal of the American Society for Information Science*, *41*(6), 391-407. https://doi.org/10.1002/(SICI)1097-4571(199009)41:6%3C391::AID-ASI1%3E3.0.CO;2-9

[5] Cui, J., Wang, Z., Ho, S. B. & Cambria, E. (2023). Survey on sentiment analysis: evolution of research methods and topics. *Artificial Intelligence Review*, 1-42. https://doi.org/10.1007/s10462-022-10386-z

[6] Landauer, T. K., Laham, D. & Derr, M. (2004). From paragraph to graph: Latent semantic analysis for information visualization. *Proceedings of the National Academy of Sciences*, *101*(suppl_1), 5214-5219. https://doi.org/10.1073/pnas.0400341101

[7] Ahmad, S. N. & Laroche, M. (2023). Extracting marketing information from product reviews: a comparative study of latent semantic analysis and probabilistic latent semantic analysis. *Journal of Marketing Analytics*, 1-15. https://doi.org/10.1057/s41270-023-00218-6

[8] Yih, W. T., Zweig, G., & Platt, J. C. (2012). Polarity inducing latent semantic analysis. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, 1212-1222.

[9] Börner, K., Chen, C. & Boyack, K. W. (2003). Visualizing knowledge domains. *Annual review of information science and technology*, *37*(1), 179-255. https://doi.org/10.1002/aris.1440370106

[10] Duong, Q. H., Zhou, L., Meng, M., van Nguyen, T., Ieromonachou, P. & Nguyen, D. T. (2022). Understanding product returns: A systematic literature review using machine learning and bibliometric analysis. *International Journal of Production Economics*, *243*, 108340. https://doi.org/10.1016/j.ijpe.2021.108340

[11] Ma, K., Tan, Y., Tian, M., Xie, X., Qiu, Q., Li, S. & Wang, X. (2022). Extraction of temporal information from social media messages using the BERT model. *Earth Science Informatics*, *15*(1), 573-584. https://doi.org/10.1007/s12145-021-00756-6

[12] Chekuri, S., Chandrasekar, P., Banerjee, B., Park, S. H., Masrourisaadat, N., Ahuja, A., ... & Fox, E. A. (2023, June). Integrated digital library system for long documents and their elements. In *ACM/IEEE Joint Conference on Digital Libraries (JCDL2023)*, 13-24. https://doi.org/10.1109/JCDL57899.2023.00012

[13] Kanhabua, N. & Nørvåg, K. (2010, September). Determining time of queries for re-ranking search results. In *International conference on theory and practice of digital libraries* (pp. 261-272). Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-15464-5_27

[14] Lightenberg, W., Pei, Y., Fletcher, G. & Pechenizkiy, M. (2018). Tink: A temporal graph analytics library for apache flink. In *Companion Proceedings of the Web Conference 2018*, 71-72. https://doi.org/10.1145/3184558.3186934

[15] Michail, O. (2016). An introduction to temporal graphs: An algorithmic perspective. *Internet Mathematics*, *12*(4), 239-280. https://doi.org/10.1080/15427951.2016.1177801

[16] Hofmann, T. (1999). Probabilistic latent semantic indexing. In *Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval*, 50-57. https://doi.org/10.1145/312624.312649

[17] Hu, X., Ma, W., Chen, C., Wen, S., Zhang, J., Xiang, Y. & Fei, G. (2022). Event detection in online social network: Methodologies, state-of-art, and evolution. *Computer Science Review*, *46*, 100500. https://doi.org/10.1016/j.cosrev.2022.100500

[18] Elumalai, A. (2020). Graph coloring and labeling applications in computer science. *Malaya Journal of Mathematic, S*(2), 2039-4041.

[19] Molloy, M. & Reed, B. (2002). *Graph colouring and the probabilistic method* (Vol. 23). Springer Science & Business Media. https://doi.org/10.1007/978-3-642-04016-0

[20] Vinutha, M. S. & Arathi, P. (2017). Applications of graph coloring and labeling in computer science. *International Journal on Future Revolution in Computer Science and Communication Engineering*, *3*(8), 14-16.

[21] Yimam-Seid, D. & Kobsa, A. (2003). Expert-finding systems for organizations: Problem and domain analysis and the DEMOIR approach. *Journal of Organizational Computing and Electronic Commerce*, *13*(1), 1-24. https://doi.org/10.1207/S15327744JOCE1301_1

[22] Karimzadehgan, M., White, R. W. & Richardson, M. (2009). Enhancing expert finding using organizational hierarchies. In *Advances in Information Retrieval: 31th European Conference on IR Research, ECIR 2009, Toulouse, France, April 6-9, Proceedings 31*, Springer Berlin Heidelberg, 177-188. https://doi.org/10.1007/978-3-642-00958-7_18

[23] Macdonald, C. (2009). The voting model for people search. *Doctoral dissertation*, University of Glasgow. https://doi.org/10.1145/1670598.1670616

[24] Daud, A., Li, J., Zhou, L. & Muhammad, F. (2010). Temporal expert finding through generalized time topic modeling. *Knowledge-Based Systems*, *23*(6), 615-625. https://doi.org/10.1016/j.knosys.2010.04.008

[25] Smirnova, E. & Balog, K. (2011, April). A user-oriented model for expert finding. In *European Conference on Information Retrieval* (pp. 580-592). Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-20161-5_58

[26] Omidvar, A., Garakani, M. & Safarpour, H. R. (2014). Context based user ranking in forums for expert finding using WordNet dictionary and social network analysis. *Information Technology and Management*, *15*, 51-63. https://doi.org/10.1007/s10799-013-0173-x

[27] Attapur, H. (2015). The study of improving the efficiency of expert author finding model, focusing on the relationship between documents and people. *PhD Thesis*, University of Tehran, Tehran, Iran.

[28] Lipani, A. (2019). On Biases in Information retrieval models and evaluation. In *ACM SIGIR Forum*, 52(2), 172-173). https://doi.org/10.1145/3308774.3308804
[29] Wu, D., Fan, S. & Yuan, F. (2021). Research on pathways of expert finding on academic social networking sites. *Information Processing & Management*, *58*(2), 102475. https://doi.org/10.1016/j.ipm.2020.102475
[30] Sanderson, M. (2010). Test collection based evaluation of information retrieval systems. *Foundations and Trends® in Information Retrieval*, *4*(4), 247-375. https://doi.org/10.1561/1500000009
[31] Song, C., Guo, J., Gholizadeh, F. & Zhuang, J. (2022). Quantitative Analysis of Food Safety Policy—Based on Text Mining Methods. *Foods, 11*(21), 3421. https://doi.org/10.3390/foods11213421
[32] Berry, M. W., Dumais, S. T. & O'Brien, G. W. (1995). Using linear algebra for intelligent information retrieval. *SIAM review*, *37*(4), 573-595. https://doi.org/10.1137/1037127
[33] Evangelopoulos, N. E. (2013). Latent semantic analysis. *Wiley Interdisciplinary Reviews: Cognitive Science*, *4*(6), 683-692. https://doi.org/10.1002/wcs.1254
[34] Bogers, T., Kox, K., & van den Bosch, A. (2008, April). Using citation analysis for finding experts in workgroups. In *Proc. DIR*, 21-28.
[35] 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
[36] Noruzy, A., Dalfard, V. M., Azhdari, B., Nazari-Shirkouhi, S. & Rezazadeh, A. (2013). Relations between transformational leadership, organizational learning, knowledge management, organizational innovation, and organizational performance: an empirical investigation of manufacturing firms. *The International Journal of Advanced Manufacturing Technology, 64*, 1073-1085. https://doi.org/10.1007/s00170-012-4038-y
[37] Paksaz, A. M., Salamian, F. & Jolai, F. (2020). Waste collection problem with multi-compartment vehicles and fuzzy demands. *The 2nd National Conference on Industrial Engineering, Management, Economy and Accounting*. September 2020, Oslo, Norway. 1-12.
[38] Nazari-Shirkouhi, S., Keramati, A. & Rezaie, K. (2015). Investigating the effects of customer relationship management and supplier relationship management on new product development. *Tehnički vjesnik, 22*(1), 191-200. https://doi.org/10.17559/TV-20140623130536
[39] Neshatfar, S., Magner, A. & Sekeh, S. Y. (2023). Promise and Limitations of Supervised Optimal Transport-Based Graph Summarization via Information Theoretic Measures. *IEEE Access, 11*, 87533-87542. https://doi.org/10.1109/ACCESS.2023.3302830
[40] Khorsandi, H. & Bayat, M. (2022). Prioritizing operational strategies of saman bank. *International Journal of Health Sciences, 6*(S7), 1442-1453. https://doi.org/10.53730/ijhs.v6nS7.11548
[41] Gholami, M. H., Asli, M. N., Nazari-Shirkouhi, S. & Noruzy, A. (2013). Investigating the influence of knowledge management practices on organizational performance: an empirical study. *Acta Polytechnica Hungarica, 10*(2), 205-216. https://doi.org/10.12700/APH.10.02.2013.2.14
[42] Azimi Asmaroud, S. (2022). Preservice Elementary Teachers' Categorical Reasoning and Knowledge Transfer on Definition Tasks with Two Dimensional Figures. *Theses and Dissertations*. 1588. https://ir.library.illinoisstate.edu/etd/1588
[43] Zarei, P. & Dobakhti, L. (2023). An Exploration of Iranian Teachers' Professional and Institutional Identities and their Enactment of Critical Pedagogy. *Teaching English Language*, 1-25.
[44] Rasekh Eslami, Z. & Zohoor, S. (2023). Second Language (L2) Pragmatics and Computer Assisted Language Learning (CALL). *Technology Assisted Language Education, 1*(2), 1-17.
[45] Zandi, J. & Naderi-Afooshteh, A. (2015). LRBAC: Flexible function-level hierarchical role based access control for Linux. In *The 12th IEEE International Iranian Society of Cryptology Conference on Information Security and Cryptology (ISCISC2015)*, 29-35. https://doi.org/10.1109/ISCISC.2015.7387894

Authors' contacts:

Shahla Rezvani
Department of Information Science and Knowledge Management,
Faculty of Management, University of Tehran, Tehran, Iran
E-mail: shahla.rezvani.gi@ut.ac.ir

Nader Naghshineh, Associate Professor
(Corresponding author)
Department of Information Science and Knowledge Management, Faculty of Management, University of Tehran, Tehran, Iran
E-mail: Nnaghsh@ut.ac.ir

Ahmad Khalilijafarabad, PhD
Department of IT Management, Faculty of Management,
University of Tehran, Tehran, Iran
E-mail: Ahmad.khalili@ut.ac.ir

ISSN 1846-6168 (Print), ISSN 1848-5588 (Online)
https://doi.org/10.31803/tg-20231018184747

Original scientific paper
Received: 2023-10-18, Accepted: 2024-01-08

Improving Freedom of Visually Impaired Individuals with Innovative EfficientNet and Unified Spatial-Channel Attention: A Deep Learning-Based Road Surface Detection System

Amit Chaudhary*, Prabhat Verma

Abstract: Individuals with visual impairments often encounter substantial challenges navigating outdoor spaces due to their inability to perceive road-surface conditions. This study introduces an innovative method that harnesses deep learning to identify and categorize road surfaces, aiming to enhance the independence and mobility of the visually impaired. Leveraging the EfficientNetB0 model as a foundational framework and employing unified spatial-channel attention, we classified road surface images captured from a wearable camera. Through rigorous training and evaluation on a substantial dataset of road images, our modified system exhibited remarkable performance, accurately identifying road surfaces with an impressive 99.39% accuracy rate. This deep learning-driven approach holds promise as a pivotal tool for improving the autonomy and safety of individuals with visual challenges by providing instantaneous feedback on road conditions.

Keywords: attention mechanism; deep learning network; EfficientNet-B0; pedestrian with vision limitations

1 INTRODUCTION

The World Health Organization (WHO) states that approximately 2.2 billion individuals experience visual impairments [1]. This poses significant challenges for visually impaired individuals, who struggle to navigate and interact with their surroundings. They lack environmental information and have difficulty recognizing objects and people, limiting their independence and access to important services. To tackle this issue, various navigation solutions have been proposed that cater to different GPS-based, audio-based, and smartphone-based systems. However, GPS-based systems can be unreliable in urban areas and indoors owing to their poor signal quality [2]. Audio-based systems struggle in noisy environments [3] and do not provide object-location information [4]. Smartphone-based systems have limitations in terms of their battery life and indoor functionality. Given these obstacles, there is a growing emphasis on research to enhance the freedom and movement of those with visual impairments through deep learning-based road surface detection. By utilizing deep learning techniques to detect and classify road surfaces, this study aims to significantly enhance the lives of visually impaired individuals. This technology enables them to navigate safely and efficiently by providing real-time information about road surfaces, including detecting obstacles and changes in elevation. It also improves accessibility to public spaces and buildings, making it easier to access services and to engage in community activities. Ultimately, the following [5] Studies can significantly enhance the quality of life for those with visual challenges by boosting their freedom and movement.

We have divided assistive devices that assist visually impaired people for a better understanding of Fig. 1. Assistive devices have evolved on par with technologies and have become increasingly advanced. Physical devices such as white cans and guide dogs have been helping visually impaired people for a long time. As research progresses, various sensor-based systems use built-in sensors and GPS of smartphones to provide location information and directions. The. It has been used to develop various location information and directions, but these can be limited by the battery life of the smartphone, and they may not work indoors.

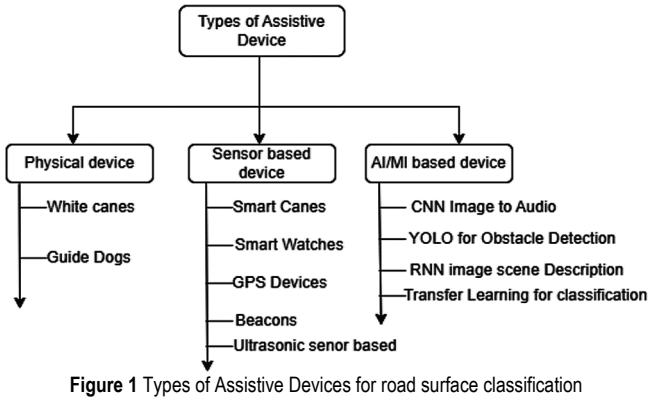


Figure 1 Types of Assistive Devices for road surface classification

Enhancing mobility and independence of visually impaired individuals through artificial intelligence-based road surface detection is an important and significant research area because it aims to improve the lives of visually impaired individuals by utilizing deep learning techniques to detect and classify road surfaces. This technology can help visually impaired individuals navigate more safely and efficiently by providing them with real-time information about road surfaces, such as the presence of obstacles or changes in elevation. Additionally, it can improve the accessibility of public spaces and buildings for visually impaired individuals, making it easier for them to access services and participate in community activities. Overall, this research has the ability to have a positive impact on the daily lives of visually impaired individuals by improving their mobility and independence.

This paper makes the following contributions to assisting visually impaired people.

- Develop a deep-learning-based road surface detection system to assist in the navigation and independence of visually impaired individuals.

- Enhances the precision of detecting road surfaces, diminishing accident risks, and bolstering safety for those with visual impairments.
- Provides an immediate and adaptable method for identifying road conditions, assisting those with visual challenges to traverse unknown terrains more comfortably and assuredly.

The objective of this research is to leverage the capabilities of a pre-trained deep learning model equipped with cutting-edge techniques to detect and classify various road surfaces, in order to aid visually impaired individuals in navigating their surroundings. The author has further enhanced the model's accuracy by incorporating a novel attention mechanism. This system can be seamlessly integrated into mobile devices, such as smartphones, canes, or other assistive devices.

This research will likely involve both theoretical and practical components, such as developing and training deep-learning models, collecting and labelling data, and evaluating the performance of the system in real-world scenarios. Additionally, this research might further examine how the suggested system influences the freedom and movement of those with visual impairments.

The following part reviews previous research on this topic, including deep learning and image recognition. The proposed model is introduced in the Methodology section. The study then addresses the model setting, provides the data information, performs the trials, and compares the outcomes. Finally, the paper concludes with a conclusion.

2 LITERATURE REVIEW

Several authors have emphasized the significance of assistive devices for individuals with visual impairment. In a subsequent paper, the author expands on this topic by conducting four focus groups with assistive technology computer users who are blind or visually impaired to gain broader insights [6]. The objective is to better comprehend how these individuals obtain information about assistive devices and to identify the specific types of information they may possess. The author presents two experiments on Social Interaction Assistants, one of which aims to reduce stereotypical body mannerisms that impede social interactions, while the other is designed to provide individuals with assistive technology to interpret the facial expressions of those they interact with [7].

Advances in CNNs have assisted visually impaired people by enhancing the accuracy and applicability of computer vision systems that are designed to assist them. In recent years, CNNs have been widely used for object recognition [8], visual place recognition [9], verification of CCTV image data through unsupervised learning [10], text classification based on neural networks [11, 12], ANN for estimating cutting forces during helical end milling of metal materials deposited by laser [13], detecting coins and banknotes, and many other applications [14].

A few of these integrate navigation and recognition capabilities into their systems. Based on the above requirements, an assistive device is presented that achieves both capabilities to aid the Visually Impaired person to navigate safely from his/her current location (pose) to a desired destination in an unknown environment and to recognize their surrounding objects. The author described a wearable device designed to help visually impaired individuals navigate unfamiliar environments [15]. The device takes the shape of a pair of eyeglasses, and can help users move safely and efficiently. Additionally, it can help interpret complex surroundings and automatically provide directions on how to move. This study aimed to create a new system that employs OCR and machine learning to assist individuals with visual impairments [16]. Specifically, it develops an indoor item identification system that utilizes a framework based on deep convolutional neural networks. Our objective was to create a robust and reliable solution that can provide visually impaired individuals with an enhanced perception of their surroundings [17].

A new streamlined Convolutional Neural Network (CNN) design was created for the swift recognition of Indian currency notes on web and mobile platforms [18]. The author proposed a walking stick design to help the visually impaired commute to their livelihood [19]. Numerous methods are available to aid blind individuals in navigating their surroundings, including technologies utilizing radio frequency identification (RFID), GPS, and computer vision modules. In a following paper, the author introduced a method for estimating depth from a solitary image, leveraging a local depth assumption without the need for user input. This solution, aimed at aiding individuals with visual impairments, is tailored exclusively for indoor environments such as homes, offices, and businesses [20]. A new system for NAVI was presented based on visual and range information [21]. The author suggested a system that utilizes smartphones to provide navigation assistance, specifically turn-by-turn guidance, through precise and current localization across vast areas [22].

This passage explores several novel strategies aimed at aiding individuals with visual impairments in navigating indoor environments without assistance [23]. One of these strategies is an ambient navigation system that enables free movement without relying on assistance. Another approach utilizes a classification system that employs a Deep Convolutional Neural Network (DCNN) model to identify indoor objects, and this system can be integrated into mobile devices. Moreover, a wearable assistive device shaped like a pair of glasses was presented, which can enhance the user's perception of their surroundings and provide guidance on the direction of movement. Finally, a new indoor object detector was developed using a deep convolutional neural-network-based framework.

Upon analyzing the collected papers, it has been determined that navigation for the visually impaired is a vital area of research, and that deep learning possesses the potential to significantly aid visually impaired individuals in navigating outdoor environments. The results of the literature review indicate that the most recent and pertinent papers

provide invaluable insights that will prove beneficial to our research endeavors. Our thorough examination of these papers has led us to conclude that navigation for visually impaired individuals is an area of utmost importance that warrants further investigation, and that deep learning can play a pivotal role in assisting visually impaired individuals in navigating outdoor environments.

3 PROPOSED METHODOLOGY

This study presents a road-surface detection technique that employs Efficient-Net and a Unified Spatial-Channel attention mechanism [24]. The approach is centered on developing a classification model and incorporating techniques, such as transfer learning and data augmentation, to attain precise automatic categorization of road surfaces. A flow diagram of the proposed method is shown in Fig. 2.

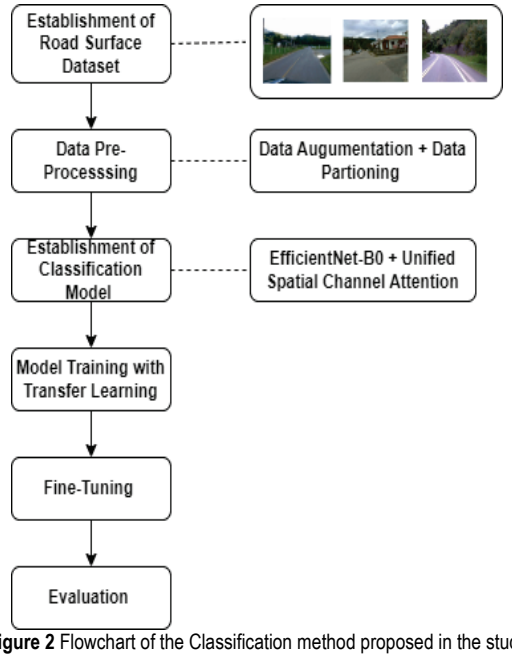


Figure 2 Flowchart of the Classification method proposed in the study

The flowchart in Fig. 2 illustrates a procedure that unites EfficientNet B0, a pre-trained model on ImageNet, with Unified Spatial Channel Attention (USCA) to categorize road surfaces. EfficientNet B0 functions as the foundation for extracting features from raw road-surface images. Subsequently, USCA is integrated, comprising two primary components: Spatial Attention, which concentrates on specific spatial regions, and Channel Attention, which accentuates vital channels. This mechanism improves the feature representation by examining spatial and channel dependencies, resulting in a more refined and focused feature map. The refined features undergo either fine-tuning or transfer learning depending on the chosen strategy. In transfer learning, the knowledge of the pre-trained model is adapted to the road surface dataset, whereas in fine-tuning, the model parameters are adjusted to further refine the pre-trained features for better performance on the road surface images.

Finally, the refined features were classified using a set of fully connected layers, which provided predictions for diverse types of road surfaces. The combination of EfficientNet-B0, USCA, transfer learning, and fine-tuning creates a robust and accurate pipeline for road-surface classification.

3.1 Dataset

We acquired images from the RTK dataset, which were captured using low-cost cameras such as the HP Webcam HD-4110, under real-world conditions [25]. The dataset comprised 77,547 frames from various conditions, including asphalt roads, unpaved roads, and paved roads. From the RTK dataset, we created a dataset consisting of 5,558 images and classified all images into seven different classes. The three classes are asphalt, paved, and unpaved. The dataset was divided into Training, Validation, and Testing sets as shown in Tab. 1. Approximately 70% of the data were in the training folder (4015), 20% were in the testing folder (986), and 10% were in the validation folder (557). The RTK dataset contains real-world images of complex environmental scenarios, such as roads with different vehicles, potholes, and road damage, as shown in Fig. 3.

Figure 3 Sample images from dataset for road surface detection

All images were collected during the daytime with a variety of brightness, texture, and other features. In each roadcategory, there is a slight difference in the surface patterns, such as paved roads that are lighter in color and asphalt roads that are darker in color. We have considered that asphalt roads are roads that do not have any sort of bumps, potholes, or other damage, such as highways and expressways. Unpaved roads are considered bad to walk on because they are not madeup of hard smooth surfaces and have different types of road anomalies. These roads are full of dirt, which is composed of native material on the land

surface. Paved roads are composed of concrete blocks or interlocking. They had different types of patterns on their surfaces. Most pedestrian ways are paved or concrete. We did not perform any type of cropping because we did not want to put extra overhead on computation, which can lead to difficulty in deploying the model in real-time usage.

Table 1 Summary of dataset for road surface detection

Name of class	Train	Test	Valid
Asphalt	1417	343	197
Paved	1386	359	204
Unpaved	1212	284	156

We pre-processed the images before feeding them into the network. We augmented the images to prevent overfitting. Various augmentation techniques, such as geometric transformation, color and contrast adjustments, noise addition, crop, and resize are used, as mentioned in Tab. 2.

Table 2 Different Data Augumentation Techniques Applied

Type	Details
Random Flip	Horizontal and Vertical
Random Rotation	–0.2, 0.2
Random Zoom	0.2
Random Contrast	0.2
Random Translation	0.2, 0.2
Random Height	–0.2,0.2
Random Width	–0.2, 0.2

The implementation of augmentation techniques will enhance the intricacy of our dataset, thereby enabling our model to exhibit more effective generalization capabilities with respect to unseen data. These techniques produce additional variations within an existing dataset without altering the total number of images.

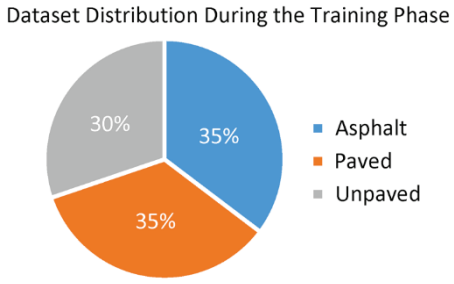


Figure 4 Distribution of dataset for road surface detection for the visually impaired individual

The above pie chart, as shown in Fig. 4, ensures that our dataset is balanced. Balanced dataset analysis has many benefits, such as better generalization to the unseen dataset than to unbalanced data, reduced overfitting, improved model performance, and faster convergence.

3.2 Efficient Net Neural Network

Many pre-trained models are available, but each one is used according to the specific problem domain. The author deals with navigation for visually impaired people, which requires the model to have a smaller size and fewer parameters, making it suitable for a low-edge embedded device such as a smartphone for real-time navigation. The proposed model is based on the EfficientNetB0 architecture [26]. The EfficientNetB0 architecture is a well-known and extensively used network architecture designed for computer vision applications. It is lightweight and can be deployed effortlessly on embedded devices, making it a popular choice in many applications.

Table 3 Size and parameter of different models

Model	No. of Parameters (million)	Size (MB)
EfficientNet-B0 [26]	5.3	350
ResNet-50 [27]	25.6	100
Vgg-16 [28]	138	553
DenseNet-121 [29]	8.8	100

From Tab. 3 EfficientNet-B0 can act as a suitable model for assisting visually impaired people because of its size and the number of parameters used, which makes it suitable for real-time operations. Despite being lightweight, EfficientNet-B0 is known for its good performance in a variety of computer vision tasks, including object detection and image classification. EfficientNet-B0 can be fine-tuned for a specific task using a smaller dataset. It has a faster inference time owing to its small size, which is important for real-time applications. Another advantage of EfficientNet-B0 is its adaptive architecture, which allows it to easily adapt to different input sizes and resolutions.

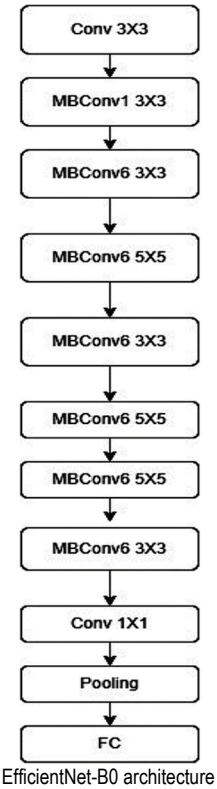


Figure 5 EfficientNet-B0 architecture flowchart

EfficientNet-B0, the foundational model of the EfficientNet family, offers a novel and holistic approach to neural network architecture optimization, balancing both accuracy and computational efficiency. The architecture of

EfficientNet-B0 is shown in Fig. 5. At the heart of its design is the innovative concept of compound scaling, a method that diverges from traditional practices by uniformly scaling the depth, width, and resolution of the network, as defined in Eq. (1). This technique ensures that no single dimension is overoptimized at the expense of the others. As the architecture delves deeper, these blocks, equipped with squeeze-and-excitation operations, manage the intricate task of learning channel-wise dependencies, thereby ensuring a comprehensive understanding of features. Beginning with a convolutional stem that transforms the 3-channel RGB input into a 32-channel feature map, the network sequences through a series of MBConv blocks.

The core architecture involved MBConv blocks equipped with squeeze-and-excitation operations. These blocks manage channel-wise features to understand intricate details in the images. Starting with a convolutional stem that processes RGB inputs into a feature map, the network navigates through these blocks to adaptively recalibrate features, making the model more perceptive of essential information.

EfficientNet-B0 boasts a unique design along with practical techniques such as DropConnect regularization to prevent overfitting during training. This model utilizes global average pooling to compress spatial dimensions and a fully connected layer for final classification while maintaining accuracy across varying computational budgets. Its goal is to provide improved accuracy and faster inference without sacrificing model size or complexity. The inclusion of compound scaling and efficient channel attention allows better performance and adaptability in various scenarios.

$$d = \alpha^{\phi}, w = \beta^{\phi}, r = \gamma^{\phi}. \tag{1}$$

Where EffiecientNetB0 introduced scaling in the depth d, width w, and resolution r. α, β, and γ are scaling coefficients, and ϕ symbolizes the scaling factor that controls the extent to which the depth, width, and resolution of the network should be adjusted.

The essence of Eq. (1) allows EfficientNet to adjust its model complexity effectively by manipulating the depth, width, and resolution through the application of scaling coefficients and a scaling factor. This approach enables the model architecture to be tailored for different computational budgets while striving to preserve high accuracy.

$$F_{\mathrm{sq}}(u_{\mathrm{c}}) = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} u_{\mathrm{c}}(i,j). \tag{2}$$

The following equation Eq. (2) calculates the average value of all the elements in the tensor by adding all its elements and then dividing this sum by the total number of elements, which is $H \times W$. This process compresses the spatial information and reduces the spatial dimensions of the tensor while preserving important information for subsequent operations in the network.

$$S = \sigma\left(\boldsymbol{W}_2 \partial(\boldsymbol{W}_1 z)\right), \tag{3}$$

where S represents the output of the excitation operation, $\boldsymbol{W}_1$ and $\boldsymbol{W}_2$ represent the weight matrices; z is the output of the squeeze block, $g(z, W)$ represents an intermediate computation; ∂ refers to the intermediate computations including ReLU operation and where σ is the sigmoid function.

The above equation Eq. (3) is vital in channel attention mechanisms because it enables the neural network to concentrate on critical channels by dynamically adjusting the significance of each channel in the feature map (z). Weight matrices $\boldsymbol{W}_1$ and $\boldsymbol{W}_2$ are learned during training to emphasize the relevant channels while deemphasizing the less informative ones. The excitation operation facilitates the network's ability to efficiently capture channel-wise dependencies, leading to enhanced feature representation and improved global information access for superior decision making.

3.3 Unified Spatial-Channel Attention

Attention mechanisms play a crucial role in assisting neural networks to concentrate on essential input data, thereby enhancing their learning capabilities and predictive accuracy. They are particularly advantageous in handling variable sequence lengths because they enable the network to focus adaptively on different input segmment.

Attention mechanisms contribute to an improved model performance by capturing intricate patterns and long-range dependencies. Moreover, the transparency provided by attention mechanisms helps clarify the significance of input elements in the decision-making process. Attention mechanisms optimize the computational efficiency and processing speed by directing attention to specific elements.

Spatial attention focuses on spatial relationships within an image, as defined in Eq. (4), by focusing on specific regions or pixels relevant to the task. It helps models highlight critical spatial features such as edges or textures, enabling them to identify key visual patterns.

$$S = \sigma\left(\boldsymbol{W}_2 \delta(\boldsymbol{W}_1 z)\right), \tag{4}$$

where $\boldsymbol{W}_1$ and $\boldsymbol{W}_2$ represent the weight matrices; δ denotes ReLU operation; where σ is the sigmoid function and z is the output of the spatial squeeze block.

By contrast, channel attention operates across channels or feature maps, as defined in Eq. (5), which allows the model to assign different weights to each channel based on its importance. By capturing channel-wise dependencies, it refines feature representations and enhances the model's understanding of the semantic information in the data. When combined with unified spatial-channel attention, these attention mechanisms enable the network to discern both spatial and semantic details, optimizing its ability to extract meaningful information from images.

$$CA = \sigma\left[\boldsymbol{W}_2\delta\left(\boldsymbol{W}_1 \cdot F_{\mathrm{sq}}(u_{\mathrm{c}})\right)\right], \quad (5)$$

$F_{\mathrm{sq}}(u_{\mathrm{c}})$ denotes the spatially squeezed representation of the channel.

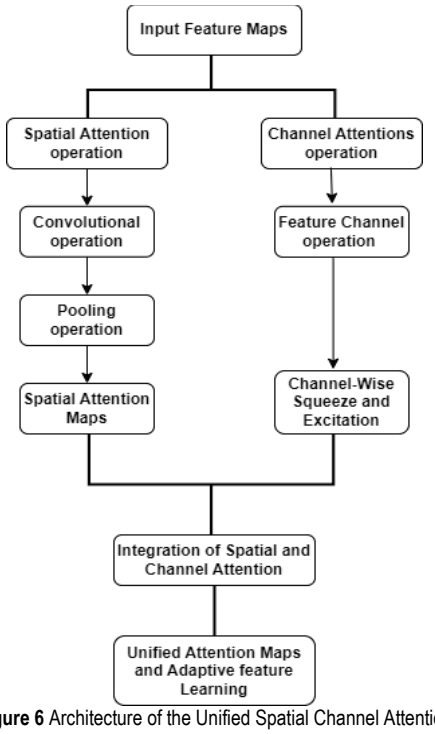


Figure 6 Architecture of the Unified Spatial Channel Attention

The flowchart in Fig. 6 shows the unified spatial-channel attention mechanism, which begins with the input data and proceeds through convolutional operations for feature extraction. The process then branches into spatial and channel attention modules. In the spatial attention module, convolutional layers are followed by global pooling operations, highlighting spatial details. Conversely, the channel attention module utilizes operations such as squeeze-and-excitation, focusing on the feature interdependencies within each channel. These separate pathways converge, allowing the unified attention mechanism to successfully combine the spatial and channel-wise information. The resulting attentional output is reintegrated into the network, enhancing feature representation and aiding classification tasks. Ultimately, this unified spatial channel attention mechanism harmoniously merges spatial and channel-dependent features, improving the model comprehension and classification accuracy, which can increase the accuracy of our road surface classification and generalize better for the task.

3.4 Transfer Learning

The deep learning algorithm addresses the limitations of traditional machine learning methods in feature extraction; however, it requires many images for training to achieve high accuracy. In addition, creating a diverse dataset of rock images can be time consuming. To overcome this issue, transfer learning was used in this study to fine-tune a specific model using the parameters and weights of a pretrained model trained on a large-scale annotated image dataset. By re-training and fine-tuning the specific model, a more accurate classification model can be obtained using fewer rock images and a shorter training time. In road surface classification using transfer learning, the process begins with the selection of a suitable pretrained model for feature extraction. The EfficientNet-B0 architecture was used as the pre-trained model, and customization involved loading the pre-trained EfficientNet-B0 weights and freezing most of the layers to preserve the learned features while fine-tuning only the final layers for the specific task. This methodology captures generalized features from broader image datasets and refines them to cater to the nuances of road-surface classification. By freezing the layers, the model optimizes computational efficiency and reduces the need for extensive training on the new dataset.

After this adaptation, the model proceeds to a training phase with the road surface dataset, allowing it to learn task-specific features while benefiting from the generalizable knowledge initially obtained from the pretrained model. Through this sequential process, the model gained insights into the distinctive characteristics of road surfaces, leveraging the foundational knowledge acquired from its pre-trained state to enhance its classification capabilities.

3.5 Fine Tuning

Fine-tuning is a transfer learning technique that entails further training of a pre-trained model on a new dataset while retaining the knowledge it previously acquired. This approach builds upon the weights learned during the initial training and adjusts them to suit the new task or dataset better.

By fine-tuning the pre-trained EfficientNet-B0 model, which has already gained knowledge about image features and patterns through its previous training, we aimed to refine the model's ability to recognize road surface features. We achieve this by making slight adjustments to the learned features such that they align better with the unique features of our new dataset.

Table 4 Hyperparameters used in road surface classification

Parameters	Value
Optimizer	Adam, RMSprop
Learning rate	1×10^{-5}
Batch Size	10
Dropout	0.3 to 0.5
Early Stopping	Validation loss, Patience = 5
Activation Function	SoftMax function
Loss function	Categorical Cross Entropy

The fine-tuning process involved two strategic steps. In the first step, the core layers of the model remain unaltered, whereas we focus on optimizing the newly integrated classification components, such as global average pooling and dense layers, all hyperparameters details are mentioned in Tab. 4. This phase is critical for adapting the model to discern unique road surface attributes identified during the

transfer learning stage. The second step involves refining the accuracy of the model in road surface classification by fine-tuning specific advanced sections without altering the foundational layers. Utilizing the RMSprop optimizer, these adjustments aim to amplify the model's discernment of crucial spatial nuances that are essential for accurate classification.

Overall, this methodical fine-tuning approach meticulously tailors the EfficientNet-B0 architecture, bolstered by unified spatial channel attention, to excel in discerning the intricate features inherent in road surface images.

4 RESULT AND DISCUSSION

We tested the model for detecting road classification for visually impaired pedestrians and compared it with the basic model EfficientNet-B0 [26] and traditional machine learning algorithms, such as ResNet50 [27] and Random Forest [30], which are shown in Tab. 5. Compared to the models mentioned above, our approach provides promising results. We conducted this experiment using our hand-labelled dataset, which includes three distinct categories: Asphalt, Paved, and Unpaved. We allocated 70% of the data for training, 20% for testing, and the remaining 10% for validation. We calculated the F1-score for each class as false negatives and false positives, which are more important than true negatives and true positives, as in our case, the dataset was not balanced. We ran this test on our manually classified dataset which contains 4015 training images, 557 validation images, and 986 testing images. In the dataset, we included real-world images, including images with other vehicles, while avoiding images that contained transitions between road surfaces and frames that consist of the very strong glare of sun rays causing reflection. Even after including images with complex conditions, our approach can detect the surfaces of vehicles with good accuracy. The Confusion matrix helps us understand the model performance for all classes of the dataset. The matrix compares the actual target with those predicted by our road surface quality classification model.

In both machine learning and statistics, the confusion matrix is a crucial instrument for assessing the performance of the classification models. It offers a detailed comparison of predicted results against true values, shedding light on the model's overall precision and the nature of mistakes it commits. Fundamentally, in tasks involving multiclass classification, there are four primary components: True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN), where TP and TN capture accurate predictions, and FP and FN highlight instances in which the model's prediction contradicts the actual outcome. By diving deeply into these elements, we can pinpoint areas where the model falls short, underscoring the essential role of the confusion matrix in fine-tuning and enhancing classifiers.

The model demonstrates as in Fig. 7 strong discernment between Asphalt and Unpaved classes, with high precision and true positive rates. However, it shows a slightly higher tendency for misclassification within the Paved class, resulting in a few false positives and false negatives. Despite this, the model maintains a high overall accuracy, especially in distinguishing between Asphalt and Unpaved surfaces, and has a lower rate of misclassifications within the Paved class.

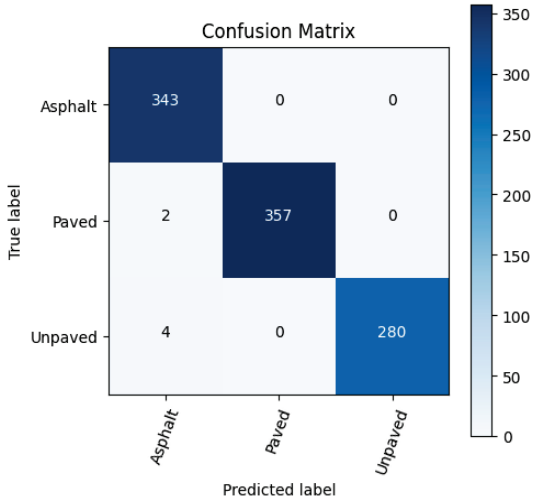


Figure 7 Confusion Matrix for road surface classification

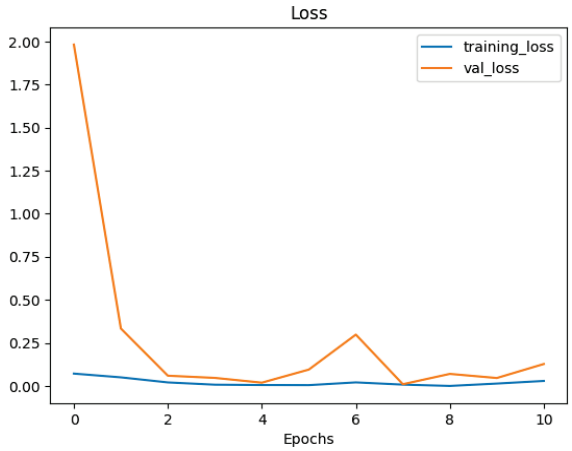


Figure 8 Training loss and validation loss graph for the proposed model

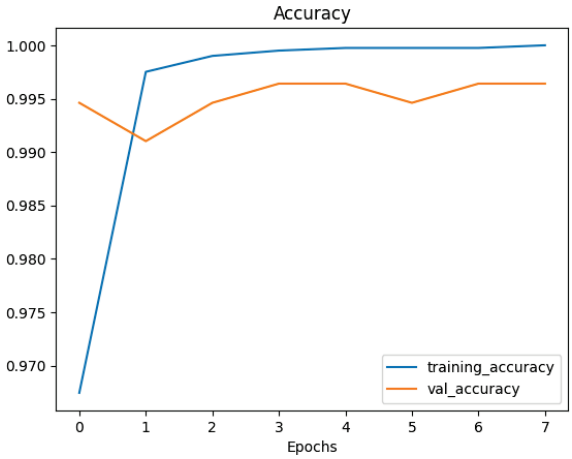


Figure 9 Training accuracy and validation accuracy graph for the proposed model

Fig. 8 and Fig. 9 show the accuracy and loss values during the training and validation phases of our modified EfficientNetB0, respectively. The proposed novel

architecture, based on EfficientNetB0 with the fusion of a unified spatial channel attention mechanism, achieved 99.39% testing accuracy, whereas EfficientNetB0 achieved 96.85% accuracy. The modelis deployed in the form of a web application on the Heroku platform using the Flask application, which provides satisfactory resultsin real-world conditions.

The data presented in Tab. 5 unequivocally demonstrates that the proposed model outperformed the other models by a considerable margin.

Table 5 Comparative analysis of accuracy of different models

Classes	EfficientNet [26] (%)	ResNet [27] (%)	Random Forest [30] (%)	Proposed Model (%)
Asphalt	97.80	92.56	72.34	99.24
Paved	96.23	93.34	82.67	99.68
Unpaved	96.54	89.12	78.89	99.15

Table 6 Classification report of proposed model

Name of the class	Precision (%)	Recall (%)	F1-score (%)	Accuracy
Asphalt	98.28	100	99.12	99.24
Paved	99.44	99.44	99.44	99.68
Unpaved	98.59	98.59	98.59	99.15

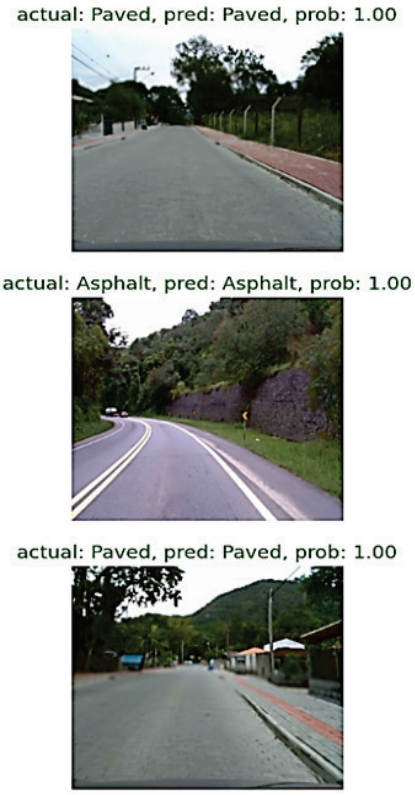


Figure 10 Output of the proposed model based on EfficientNet-B0 and Unified spatial channel attention

A classification report serves as an instrument for machine learning to assess the performance of classification models. It provides an overall assessment of the accuracy of the model as well as class-specific evaluation, which helps identify which classes require improvement. Tab. 6 shows the classification report by which we can further determine the performance of the proposed model over individual classes.

The output of the proposed model is shown in Fig. 10, and it was clearly able to predict the road surface with good accuracy.

5 CONCLUSION AND FUTURE WORK

Previous research on visually impaired pedestrians has largely focused on the detection of obstacles in their paths to help them avoid potential hazards. However, our work has focused on assessing the quality of the road surface upon which these individuals must navigate. By increasing awareness of their surroundings, this approach can aid visually impaired pedestrians in adjusting their walking patterns and speeds. In this study, we present a novel architecture based on EfficientNetB0 with a unified spatial channel attention mechanism that achieves state-of-the-art results and outperforms both individual models and traditional machine-learning algorithms. Our proposed model achieved an accuracy of 99.39%, surpassing the 96.85% achieved by EfficientNetB0. Additionally, our model is well suited for deployment on embedded devices with limited computational power. Experimental results confirm the efficacy of our proposed approach, and we plan to further expand our research by incorporating additional classes while maintaining high accuracy and by identifying various obstacles and potential hazards on the road surface, including stray animals.

6 REFERENCES

[1] World Health Organisation. (2020). Visual impairment and blindness. Retrieved from https://www.who.int/news-room/fact-sheets/detail/blindness-and-visual-impairment.

[2] Helal, A., Moore, S. E. & Ramachandran, B. (2001). Drishti: an integrated navigation system for visually impaired and disabled individuals In *Proceedings of the Fifth International Symposium on Wearable Computers*, 149-156. https://doi.org/10.1109/ISWC.2001.962119

[3] Ramadhan, A. J. (2018). Wearable Smart Systems for Visually Impaired People: *Sensors, 18*(3), 843. https://doi.org/10.3390/s18030843

[4] Porzi, L., Messelodi, S., Modena, C. M. & Ricci, E. (2013). Smart watch-based gesture recognition system for assisting people with visual impairments. In *Proceedings of the 3rd ACM International Workshop on Interactive Multimedia on Mobile and Portable devices*, 19-24. https://doi.org/10.1145/2505483.2505487

[5] Gerber, E. (2003). The Benefits of and Barriers to Computer use for visually impaired individuals *Journal of Visual Impairment & Blindness, 97*(9), 536-550. https://doi.org/10.1177/0145482X0309700905

[6] Krishna, S. & Panchanathan, S. (2010). Assistive technologies as effective mediators in interpersonal social interactions for persons with visual disabilities In: Miesenberger, K., Klaus, J., Zagler, W., Karshmer, A. (eds) *Computers Helping People with Special Needs, ICCHP 2010, Lecture Notes in Computer Science, vol 6180*. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-14100-3_47

[7] Aziz, N., Roseli, N. & Mutalib, A. A. (2011). Visually Impaired Children's acceptance of assistive courseware. *American Journal of Applied Sciences, 8*, 1019-1026. https://doi.org/10.3844/ajassp.2011.1019.1026

[8] Shah, S., Bandariya, J., Jain, G., Ghevariya, M. & Dastoor, S. (2019). CNN based Auto-Assistance system as a boon for directing visually impaired persons. In *The 3rd International Conference on Trends in Electronics and Informatics (ICOEI)*, 235-240. https://doi.org/10.1109/ICOEI.2019.8862699

[9] Fang, Y., Wang, K., Cheng, R., Yang, K. & Bai, J. (2019). Visual place recognition based on multilevel descriptors for visually impaired people. *Proc. SPIE 11158, and Target and Background Signatures V* 1115808. https://doi.org/10.1117/12.2532524

[10] Lee, Y. (2023). Verification of CCTV image data through unsupervised learning model of deep learning. *Tehnički glasnik, 17*(3), 353-358. https://doi.org/10.31803/tg-20221227094126

[11] Kim, D. (2023). Text classification based on neural-network fusion. *Tehnički glasnik, 17*(3), 359-366. https://doi.org/10.31803/tg-20221228154330

[12] Lee, S. (2023). Text classification of mixed models based on deep learning. *Tehnički glasnik, 17*(3), 367-374. https://doi.org/10.31803/tg-20221228180808

[13] Župerl, U. & Kovačič, M. (2023). Artificial Neural Network System for Predicting Cutting Forces in Helical-End Milling of Laser-Deposited Metal Materials. *Tehnički glasnik, 17*(2), 223-230. https://doi.org/10.31803/tg-20230417145110

[14] Alghamdi, S. (2019). Shopping and tourism for blind people using RFID as an IoT application. In *The 2nd International Conference on Computer Applications & Information Security (ICCAIS)*, 1-4. https://doi.org/10.1109/CAIS.2019.8769581

[15] Bai, J., Liu, Z., Lin, Y., Li, Y., Lian, S. & Liu, D. (2019). Wearable travel aids in the environment perception and navigation of visually impaired people. *Electronics, 8*(6), 697. https://doi.org/10.3390/electronics8060697

[16] Kaur, B. & Bhattacharya, J. (2019). A scene perception system for the visually impaired based on object detection and classification using a multimodal DCNN. *Journal of Electronic Imaging, 28*(01), 1. https://doi.org/10.1117/1.JEI.28.1.013031

[17] Afif, M., Ayachi, R., Said, Y., Pissaloux, E. & Atri, M. (2020). An evaluation of RetinaNet for indoor object detection for blind and visually impaired persons assists navigation. *Neural Processing Letters, 51*(3), 2265-2279. https://doi.org/10.1007/s11063-020-10197-9

[18] Veeramsetty, V., Singal, G. & Badal, T. (2020). Coinnet: A platform-independent application to recognize Indian currency notes using deep learning techniques. *Multimed Tools Appl, 79*, 22569-22594. https://doi.org/10.1007/s11042-020-09031-0

[19] Kanna, S. B., Kumar, T. G., Niranjan, C., Prashanth, S., Gini, J. R. & Harikumar, M. E. (2021). Low-Cost Smart Navigation System for Blinds. In *The 7th International Conference on Advanced Computing and Communication Systems (ICACCS), 1*, 466-471. https://doi.org/10.1109/ICACCS51430.2021.9442056

[20] Praveen, R. G. & Paily, R. P. (2013). Blind Navigation Assistance for Visually Impaired based on the Local Depth Hypothesis from a Single Image. *International Conference on Design and Manufacturing (IConDM2013), Procedia Engineering, 64*, 351-360. https://doi.org/10.1016/j.proeng.2013.09.107

[21] Aladrén, A., López-Nicolás, G., Puig, L. & Guerrero, J. J. (2016). Navigation Assistance for the Visually Impaired Using RGB-D Sensor with Range Expansion. *IEEE Systems Journal, 10*(3), 922-932. https://doi.org/10.1109/JSYST.2014.2320639

[22] Ahmetovic, D., Gleason, C., Kitani, K. M., Takagi, H. & Asakawa, C. (2016). NavCog: A turn-by-turn smartphone navigation assistant for people with visual impairment or blindness. In *Proceedings of the 13th International Web for All Conference*, 1-2. https://doi.org/10.1145/2899475.2899509

[23] Chaccour, K. & Badr, G. (2016). Computer vision guidance system for indoor navigation of visually impaired people. At *The 8th IEEE International Conference on Intelligent Systems (IS)*, 449-454. https://doi.org/10.1109/IS.2016.7737460.

[24] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N. & Polosukhin, I. (2017). Therefore, attention is required. *Advances in neural information processing systems*, *30*.

[25] Rateke, T., Justen, K. A. & von Wangenheim, A. (2019). Road surface classification with images captured from low-cost camera-road traversing knowledge (RTK) dataset. *Revista De Informática Teórica E Aplicada, 26*(3), 50-64. https://doi.org/10.22456/2175-2745.91522

[26] Tan, M. & Le, Q. (2019). EfficientNet: Rethinking model scaling for Convolutional Neural Networks. *Proceedings of the 36th International Conference on Machine Learning, 97*, 6105-6114. Retrieved from https://proceedings.mlr.press/v97/tan19a.html.

[27] 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 (CVPR2016)*, Las Vegas, NV, USA, 770-778. https://doi.org/10.1109/CVPR.2016.90

[28] Simonyan, K. & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

[29] Huang, G., Liu, Z., Van Der Maaten, L. & Weinberger, K. Q. (2017). Densely connected convolutional networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 4700-4708.

[30] Louppe, G. (2014). Understanding random forests: From theory to practice. *arXiv preprint arXiv:1407.7502*. https://doi.org/10.48550/arXiv.1407.7502

Author's contacts:

Amit Chaudhary, Research Scholar
(Corresponding Author)
Harcourt Butler Technical University,
Nawabganj, Kanpur, Uttar Pradesh 208002, India
amitchaudhary.gkg@gmail.com

Prabhat Verma, Associate Professor Dr.
Harcourt Butler Technical University,
Nawabganj, Kanpur, Uttar Pradesh 208002, India
pverma@hbtu.ac.in

ISSN 1846-6168 (Print), ISSN 1848-5588 (Online)
https://doi.org/10.31803/tg-20231114190214

Original scientific paper
Received: 2023-11-14, Accepted: 2024-01-02

Renovation Measures for Reduction of Primary Energy Consumption and CO_2 Emissions of Hospital Building

Džana Kadrić*, Rejhana Blažević, Hadis Bajrić, Adna Peco, Edin Kadrić

Abstract: Hospitals, as highly energy-intensive buildings, significantly contribute to a country's energy consumption and CO_2 emissions. The study focuses on a hospital building in Sarajevo, Bosnia and Herzegovina, and investigates the impact of selected energy renovation measures on hospital primary energy consumption and CO_2 emissions. The main goal of this paper is to develop a mathematical model for establishing relationship between primary energy consumption and CO_2 emissions (responses), and the three renovation measures (factors). The research uses dynamic simulation of the building's energy performance in Design Builder, validated with the actual energy consumption. Three energy renovation measures were considered in the study: installation of thermal insulation on external walls and flat roof, and the installation of a photovoltaic power plant. The Analysis of Variance and regression analysis were used to estimate factor effects, and to develop mathematical models. The analysis revealed that the installation of photovoltaic power plant on the roof and thermal insulation on the external walls had the most significant impact on reducing the building's primary energy consumption and CO_2 emissions. Installation of insulation on the roof did not significantly affect these performances compared to the other two measures. Developed models are suitable for evaluation of potential for energy and CO_2 savings through the implementation of energy efficiency measures. Study results can be extrapolated to all buildings within the same category, providing a valuable tool for energy efficiency planning in the healthcare sector.

Keywords: CO_2 emissions; design of experiments; energy renovation; hospital building; primary energy

1 INTRODUCTION

The healthcare sector, particularly hospital buildings, represents a significant portion of countries energy consumption and CO_2 emissions [1]. Hospitals are highly energy-intensive buildings, requiring large amounts of energy for operations including energy consumption in buildings and from medical procedures, such as sterilisation, radiology, laboratories, operating rooms, ventilation, air conditioning, heating, kitchens and laundry rooms [1, 2]. High energy consumption and CO_2 emissions of hospital buildings provide opportunities for implementing solutions that can reduce energy consumption and CO_2 emissions, therefore increase energy efficiency. These solutions can range from energy renovation of building envelope, equipment upgrades and integration of renewable energy systems, to the implementation of energy management and the promotion of user behaviour change towards more efficient energy use. The benefits of increasing energy efficiency in hospital buildings are numerous. Besides reducing energy costs and CO_2 emissions, energy efficiency can also have positive consequences for both human and environmental health and can improve the quality of patient care. Also, hospitals can serve as influential models for other sectors in society, demonstrating the feasibility and benefits of energy efficiency measures and contributing to broader efforts to mitigate climate change. The study [3], conducted in hospital building located in Grece found that the implementation of energy-saving measures, such as replacing all luminaires with LED lights and installing photovoltaic panels, could significantly reduce annual electrical energy consumption. This could potentially bring the consumption close to or even below 45% of the current annual electrical energy consumption. Study [4] offers a comprehensive decision-making approach for implementation of energy-saving renovation strategies in hospital building. This approach provides valuable insights for investors in selecting the most suitable energy-saving renovation plan. The evaluation criteria for this approach includes three important indicators: energy, financial, and thermal comfort. The authors presented results of case study conducted in a selected hospital building to demonstrate applicability of the proposed method. A study [5], focused on various aspects of energy efficiency in healthcare buildings in Spain over a period of 8 years, where it is concluded that significant energy savings can be achieved through implementation of efficient energy-saving techniques.

The general findings from numerous studies indicate that hospitals can achieve considerable reductions in energy consumption by adopting energy-efficient technologies and strategies. These strategies not only reduce energy consumption, but also contribute to economic savings and improved thermal comfort in hospital buildings, which can be incorporated in the national renovation strategies [6, 7].

Healthcare buildings in Bosnia and Herzegovina (B&H) constitute 13% of the total heated area of public buildings and are characterized by high specific energy consumption compared to other public buildings [8]. These buildings, particularly hospital buildings, require careful analysis due to their specific purpose, usage duration, and design parameters related to internal temperature. A hospital building in Sarajevo, constructed in the 1980s, has been selected for analysis based on its typical characteristics common for given construction period, which include poor energy-related characteristics and energy supply system deficient in renewable energy sources. To reduce energy consumption and CO_2 emissions, energy renovation measures of the building envelope and the installation of a photovoltaic power plant are being analysed.

Impact of each measure on energy consumption and CO_2 emission is assessed using Design of Experiments (DOE) methodology. DOE is statistical method widely used in identification of key factors and evaluation of effectiveness of different renovation strategies for reduction of buildings energy consumption. It allows researchers to identify most influential factors, such as energy renovation measures and

to evaluate their individual or combined impact on energy consumption. While numerous studies have applied DOE to analyse residential and public building renovation measures [9, 10], there is limited research using this method to improve energy efficiency, specifically in the healthcare sector. Previous healthcare applications of DOE have focused on enhancing patient flow, scheduling efficiency, and information management [11, 12] rather than energy savings. This represents an opportunity to use DOE for assessing and optimizing the energy performance of healthcare facilities through building renovations.

Therefore, in presented study, a DOE method, Full Factorial Design (FFD) is used to establish relationships between primary energy consumption and CO_2 emissions and energy renovation measures of hospital building. The regression models developed in this study enable quantification of primary energy and CO_2 emissions considering implementation of energy renovation measures. This study aims to model the energy consumption and CO_2 emissions of hospital building, focusing on implementation of energy efficiency measures and the integration of renewable energy sources.

2 METHODOLOGY

2.1 Representative Building within the Hospital Complex

The hospital building analysed in this study is component of a large hospital complex (Fig. 1), constructed during the 1980 with 35800 m^2 of heated area. The analysed hospital building has 4883 m^2 of gross heated area, which represents 13.6% of gross heated area of hospital complex. The representative hospital building, shown in Fig. 1, is a multi-level building with basement, a ground floor, and four additional floors, and flat roof on top. Hospital operates continuously, providing services 24 hours a day, every day of the year.

Figure 1 (a) Satellite view of hospital complex, and (b) visual representation of selected hospital building

The building envelope consists of brick blocks, without thermal insulation and wooden windows, with triple glazing filled with argon. During 2017-2018, the building was renovated and the old windows, with aluminium profiles, were replaced with the new wooden frame windows. The heat transfer coefficient of the new windows is 1.0 W/m^2K. However, heat energy losses are present, due to the low energy characteristics of building external wall. The flat roof is constructed as reinforced concrete slab with 4 cm of thermal insulation.

Table 1 Basic data on building geometry and heat transfer coefficients for key construction elements

Construction element	External wall	Roof	Ground floor	Windows
Total surface area / m^2	2260	923	923	1216
Heat transfer coeff. / W/m^2K	1.698	0.602	0.256	1.0
Net heated surface area / m^2	4698			
Net heated volume / m^3	16787			
Building compactness ratio	0.28			

The heat transfer coefficients for key construction elements, as well as basic data on building geometry, are shown in Tab. 1. According to the provisions of the Rulebook on technical requirements for thermal insulation of buildings and rational use of energy, it is determined that heat transfer coefficient of external wall, ground floor and flat roof do not comply with the requirements set by national regulations [13].

The hospital building is supplied with energy from highly efficient industrial steam boilers, located in the central boiler room, which uses natural gas as fuel. Energy for heating and domestic water heating (DWH) is supplied to the buildings via heating substations (Fig. 2). Two water tanks are installed in heating substation for the preparation of DWH. The overall efficiency of the system, including the boiler room, distribution and internal installations and components, is 93.1%. The hospital building has a radiator heating system for building heating and a central cooling system that serves only a small part of the building.

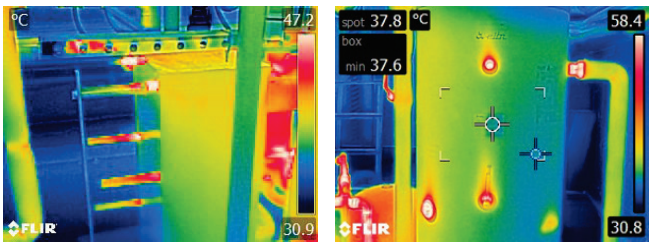


Figure 2 Thermal image of heating substation and DWH storage tank

2.2 Modelling the Energy Performance of the Representative Hospital Building

Design Builder is a software that incorporates the EnergyPlus simulation tool, and it was used for modelling building energy performance in this study. It provides

dynamic simulations of building energy performance, including the analysis of heating, cooling, DWH, ventilation, lighting, and other systems [14, 15].

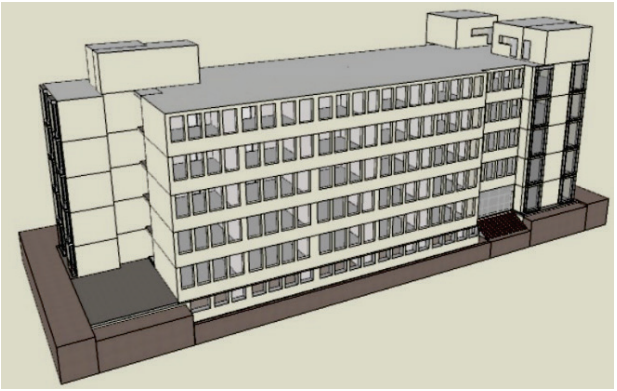


Figure 3 3D model of hospital building designed in Design Builder

The 3D model of hospital building designed in Design Builder is shown in Fig. 3. For building energy performance simulation, EnergyPlus utilizes hourly data to define external conditions in the calculations. These data, monitored by the National Meteorological Service, include parameters such as air temperature, atmospheric conditions, solar radiation, wind speed and direction, and are specific for the building location [14].

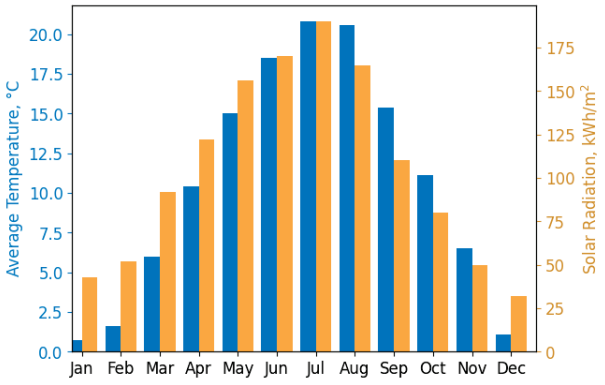


Figure 4 Climatic data for Sarajevo

The hospital building is located in Sarajevo (Capital of B&H), in a northern climate zone, 43.85 latitude and 18.14 longitude. Sarajevo climate is characterized by warm summers, cold and snowy winters, and partial cloudiness throughout the year. The official Heating Degree Days (HDD) for Sarajevo is 3077 [16]. The warmest month is typically July, with an average temperature of 20.8 °C, while the coldest month is January with an average lowest daily temperature of 0.7, as shown in Fig. 4. In terms of solar radiation, July is the brightest month in Sarajevo, while December is the darkest [17].

The zoning of the interior space of the hospital building is performed out according to the hospital layout, where internal design temperature of each zone was specified by the standard BAS EN 12831.

The total primary energy E_{prim}, which includes energy needs for heating, DWH and electrical energy, is calculated using following equation [13]:

$$E_{prim} = \frac{Q_{H,nd} + Q_{DWH}}{\eta_{sist}} f_{p,gas} + E_{el} f_{p,el} \quad (1)$$

where E_{prim} is the annual primary energy consumption in kWh/ann., $Q_{H,nd}$ is annual energy need for heating in kWh/ann., Q_{DWH} is annual energy need for DWH in kWh/ann., η_{sist} is overall system efficiency in %, E_{el} is annual consumption of electricity from the electrical grid in kWh/ann., f_p is fuel primary energy factor (1.1 for natural gas, and 3.0 for electricity) [13].

Annual CO_2 emission, is calculated according to the following equation [13]:

$$m_{CO_2} = \frac{Q_{H,nd} + Q_{DWH}}{\eta_{sist}} c_{p,gas} + E_{el} c_{p,el} \quad (2)$$

where m_{CO2} is annual CO_2 emission in kg/ann., c_p is CO_2 fuel coefficient per unit of energy (0.2 kg/kWh of natural gas, and 0.745 kg/kWh of electricity) [13].

Electricity E_{el} is calculated as total electricity needed, reduced by electricity produced from photovoltaic power plant, which also affect the reduction of CO_2 emissions.

2.3 Energy Renovation Measures

The renovation measures analysed in this study aim to reduce building envelope energy losses and provide electricity from renewable energy sources. Three measures are considered: installation of thermal insulation on the external wall (M1) and on the flat roof (M2), and installation of a photovoltaic power plant on the flat roof (M3). These measures will ensure that energy consumption for heating and electricity is reduced, and according to Eq. (2) and (3), will result in a reduction of primary energy and CO_2 emissions [18-20].

In the current state, the hospital external wall has no thermal insulation. To examine the impact of adding thermal insulation on primary energy consumption and CO_2 emissions, rock wool with a thermal conductivity of 0.033 W/mK is selected as the insulation material. Rock wool offers benefits such as improved acoustics, indoor comfort, and fire safety [21, 22]. Implementing this measure will reduce the external wall heat transfer coefficient U_{wall}, as presented in the next section. The maximum wall thermal insulation thickness analysed in this study is 20 cm. The hospital flat roof has a layer of 4 cm thermal insulation. This helps in reduction of energy losses through this envelope element. To analyse energy and CO_2 saving potential by installing additional thermal insulation on flat roof, installation of perlite board with thermal conductivity of 0.052 W/mK was analysed. Implementing this measure will reduce the roof wall heat transfer coefficient U_{roof}, as presented in the next section. The maximum roof thermal

insulation thickness analysed in this study is 25 cm. Installation of photovoltaic power plant on the flat roof will ensure that a portion of the electricity consumed in the hospital is provided from renewable energy sources. The surface area available for the installation of the photovoltaic power plant is a crucial factor in determining the power output. The total area of the roof available for installation of photovoltaic panels is 566 m^2 that is 60% of the roof surface area. Therefore, the maximum installed power of photovoltaic power plant is estimated to be 80 kW.

Selected measures are represented by factors considered in this study, as shown in the following section.

2.4 Design of Experiments

Design of Experiments (DOE) is a multipurpose methodology used in various fields for identification of important input factors and their influence on the output variables or responses. In DOE, several key steps are involved as follows: determination of input variables (factors) and their levels, selection of output variables (responses), selection of the appropriate experimental design, conducting experiments or simulations, analysis of the obtained results and drawing objective conclusions.

In this study, FFD was used to establish relationships between primary energy consumption and CO_2 emissions, and the three renovation measures (M1, M2, M3), representing three factors. The considered factors are the wall heat transfer coefficient (U_{wall}), heat transfer coefficient of flat roof (U_{roof}) and the power of the photovoltaic power plant installed on the roof (P_{pv}). Factors are considered at two levels, low level (–1) and high level (+1). For the wall heat transfer coefficient (U_{wall}), levels -1 and +1 represent wall without insulation and wall with 20 cm of insulation, respectively. Similarly, for the heat transfer coefficient of a flat roof (U_{roof}), levels -1 and +1 represent roof in its baseline state and roof with 25 cm of thermal insulation, respectively. Finally, for the photovoltaic power plant installation on the roof (P_{pv}), level -1 represents no electricity production from renewable energy sources and a plant power of 0 kW, while level +1 represents plant with installed power of 80 kW. Factors considered in this study and their levels are shown in Tab. 2.

Table 2 Considered factors and their levels

Factor/Related measure	Level	
	–1	1
	Baseline	After renovation
U_{wall} / W/m^{2}K (M1)	1.698	0.272
U_{roof} / W/m^{2}K (M2)	0.602	0.218
P_{pv} / kW (M3)	0	80

For FFD with three factors, each on two levels, appropriate experimental design is 2^3 full factorial design. The experimental matrix contains 8 simulations, with each simulation run representing specific combination of factor levels and corresponding primary energy and CO_2 emissions estimates, as shown in Section 3.2.

The investment costs for each renovation measure are estimated when measures are implemented at their high level (+1). The investment cost for Measure 1 (M1) is 101000 Euros, for Measure 2 (M2) it is 63800 Euros, and for Measure 3 (M3) it is 57900 Euros. However, it is not possible to evaluate the simple payback period for these three renovation measures using DOE methodology at different renovation levels ranging from (–1) to (+1). As highlighted in the study [23], the total costs of renovation measures show sharp increase at level (–0.99) and higher, due to incurring fixed costs associated with the implementation of measures. The total cost of renovation measures is the sum of fixed costs and variable costs, which increase as the level of renovation increases. Consequently, it's not possible to establish a linear mathematical relationship between renovation costs and renovation measures levels ranging from (–1) to (+1). Nevertheless, cost analysis remains crucial and can be utilized in conjunction with modelled energy savings for renovation optimization, as demonstrated in [24]. This aspect is considered for future studies on this topic.

3 RESULTS AND DISCUSSION

After determining the required input parameters, the building's energy performances are calculated using methodology presented in previous section. Also, simulated building energy consumption is validated with the actual annual energy consumption of hospital building. Following, the FFD is used to analyse the impact of selected factors on the building primary energy and CO_2 emissions.

3.1 Building Energy Performance and Validation of the Results

The building model was validated by comparison of the simulated energy consumption obtained by Design Builder with the actual annual energy consumption of the hospital building. The actual energy consumption was obtained based on gas and electricity bills, aggregated for the entire hospital complex. To determine the consumption only for the analysed building, total energy consumption was scaled in relation to the ratio of gross heating surface of analysed building and gross heating surface of the hospital complex. Comparison of actual and modelled annual energy consumptions are shown in Tab. 3.

Table 3 Comparison of actual and modelled annual energy consumption

	Annual electricity consumption / kWh	Annual heating and DWH energy consumption / kWh
Actual data	367359	993213
Model	348936	1056584
Relative error	5.01%	–6.38%

The validation results are very satisfactory and show a relative percentage error ranging from 5.01 to 6.38% of heating, DWH and electricity consumption. This error percentage could be reduced if there were installed heat meters and electricity consumption meters for each building of the hospital complex separately, showing actual consumption of each building in hospital complex. In any case, the obtained results confirm that the model created in

the Design Builder provides reliable simulation of the real building.

Fig. 5 shows hourly diagram of fuel consumption, temperature changes and heat balance for January.

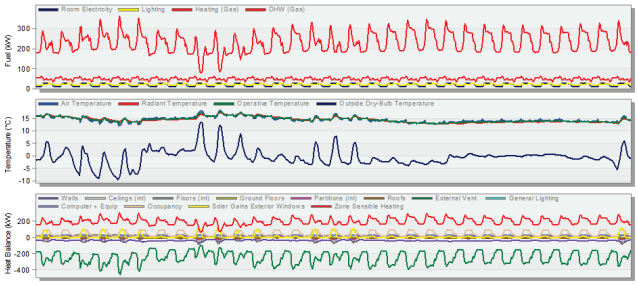


Figure 5 Hourly diagram of fuel consumption, temperature variation and heat balance for January

From Fig. 5, it is noticeable that variations in fuel consumption follow the variations of external temperature, so the highest fuel consumption is in the period of the lowest external temperatures. The heat balance diagram shows that heat losses through building envelope are compensated primarily by thermal energy supplied by the heating system, and by heat gains from the lighting, electrical devices, occupants, and solar gains.

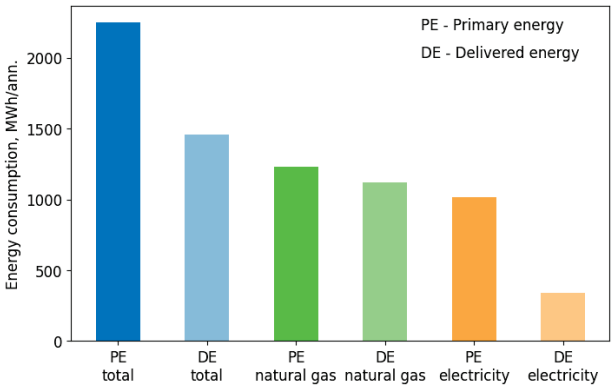


Figure 6 Annual building primary and delivered energy

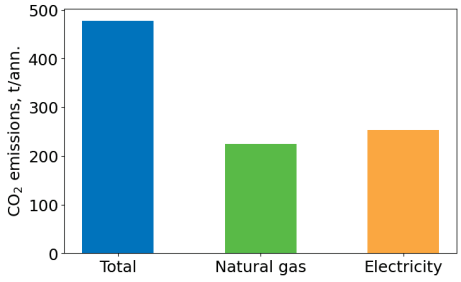


Figure 7 Annual building CO_2 emissions

Fig. 6 shows building annual primary and delivered energy. Natural gas share is 45.3% in total primary energy, while electricity share is 54.7%. Primary energy is greater than delivered energy by 35%, due to losses in the fuel processing, thermal conversion, and transmission and distribution to the hospital.

Fig. 7 presents total annual CO_2 emissions, as well as the share of natural gas and electricity in the total emissions. Natural gas and electricity have a total share in CO_2 emissions of 47.1% and 52.9%, respectively.

3.2 Design of experiments, ANOVA and regression analysis

Analysis of variance (ANOVA) and regression analysis were used for statistical analysis, with the aim of developing of regression models for predicting the building primary energy consumption and CO_2 emissions. The experimental matrix, with coded factor levels and calculated building primary energy consumption and CO_2 emissions, for each simulation, is presented in Tab. 4.

Table 4 Experimental matrix

Exp	M1	M2	M3	m_{CO2} / t/ann.	E_{prim} / MWh/ann.
1	–1	–1	–1	477.8	2250.2
2	1	–1	–1	438.9	2036.5
3	–1	1	–1	466.5	2188.7
4	1	1	–1	432.5	2000.0
5	–1	–1	1	413.2	1991.0
6	1	–1	1	374.3	1777.2
7	–1	1	1	402.0	1929.4
8	1	1	1	367.9	1740.7

ANOVA is used to determine statistical significance of all factors and their influence on the response variables, building primary energy consumption and CO_2 emissions. In this study, ANOVA was performed at a significance level of 0.05. The results of ANOVA, with only statistically significant terms and their corresponding F and p-values, are presented in Tab. 5 for building primary energy, and Tab. 6 for CO_2 emissions.

Table 5 ANOVA of the model for prediction of building primary energy consumption E_{prim}

Source	Degrees of Freedom	Sum of Squares	Mean Square	F-value	p-value
Model	3	220216	73405	933.31	0.000
A (U_{wall})	1	80968	80968	1029.46	0.000
B (U_{roof})	1	4809	4809	61.15	0.001
C (P_{PV})	1	134439	134439	1709.32	0.000
Error	4	315	79		
Total	7				

Table 6 ANOVA of the model for prediction of CO_2 emissions m_{CO2}

Source	Degrees of Freedom	Sum of Squares	Mean Square	F-value	p-value
Model	3	11157.0	3719.00	1298.97	0.000
A (U_{wall})	1	2661.8	2661.84	929.72	0.000
B (U_{roof})	1	156.9	156.85	54.79	0.002
C (P_{PV})	1	8338.3	8338.30	2912.39	0.000
Error	4	11.5	2.86		
Total	7				

The results of ANOVA presented in Tab. 5 and 6 show significantly large F-values and significantly small p-values. In both cases, primary energy and CO_2 emissions, the p-values of all factors are less than 0.05, indicating significant impact of all model terms on primary energy of building and CO_2 emissions. The greatest impacts on the building primary energy and CO_2 emissions result from implementation of

measure M3, the installation of photovoltaic power plant on the roof of the building, and measure M1, the installation of thermal insulation on the external walls of the building. Given that there is already 4 cm of thermal insulation on the flat roof, it is shown that installation of an additional insulation layer of on the roof will not significantly affect primary energy and CO_2 emissions in comparison to the other two measures (M1 and M3). As a result, the implementation of measure M2 according to ANOVA is ranked as the measure with the least impact on primary consumption and CO_2 emissions.

The regression models for predicting building primary energy consumption and CO_2 emissions are given by the following expressions:

$$E_{\text{prim}} = 1989.21 - 100.60U_{\text{wall}} - 24.52U_{\text{roof}} - 129.63P_{\text{pv}} \quad (3)$$

$$m_{CO_2} = 421.64 - 18.24U_{\text{wall}} - 4.43U_{\text{roof}} - 32.28P_{\text{pv}} \quad (4)$$

The high values of the R^2 coefficients for model of building primary energy consumption (adjusted R^2 = 99.75 %, and predicted R^2= 99.43 %), and CO_2 emission (adjusted R^2 = 99.82 %, and predicted R^2= 99.59 %), indicate high accuracy in prediction of responses.

The main effect plot in Fig. 8 and 9 shows the impact of wall heat transfer coefficient (A, U_{wall}), heat transfer coefficient of the flat roof (B, U_{roof}), and photovoltaic plant power (C, P_{pv}) on building primary energy consumption and CO_2 emissions, respectively. It is evident that the installation of a photovoltaic power plant has the most significant effect on the response, followed by installing thermal insulation on external wall. Installation of additional thermal insulation on the flat roof has the least effect on primary energy consumption and CO_2 emissions.

When examining the impact of installing a photovoltaic power plant on the primary energy consumption, it can be observed that primary energy consumption decreases from 2118 MWh annually (when the PV plant is not installed, denoted as P_{pv} = –1) to 1860 MWh annually (when the 80 kW PV plant is installed, denoted as P_{pv} = +1). Simultaneously, CO_2 emissions decreases from 454 tons annually (P_{pv} = –1) to 389 tons annually (P_{pv} = +1). The effect of installing thermal insulation on the external wall on the primary energy consumption and CO_2 emissions is less pronounced, but still significant. Primary energy consumption is reduced from 2089 MWh annually (when the wall is not insulated, denoted as U_{wall} = –1) to 1888 MWh annually (when the wall is insulated with 20 cm of thermal insulation, denoted as U_{wall} = +1). Similarly, CO_2 emissions decreases from 440 tons annually (U_{wall} = –1) to 403 tons annually (U_{wall} = +1). The effect of installing the additional thermal insulation on the flat roof on the primary energy consumption and CO_2 emissions is the lowest among the three measures. With additional roof insulation, primary energy consumption is reduced from 2014 MWh annually (when the roof is insulated with 4 cm of thermal insulation in the current state, denoted as U_{roof} = –1) to 1985 MWh annually (when the roof is insulated with 25 cm of thermal insulation, denoted as U_{roof} = +1). CO_2 emissions decreases from 426 tons annually (U_{roof} = –1) to 417 tons annually (U_{roof} = +1).

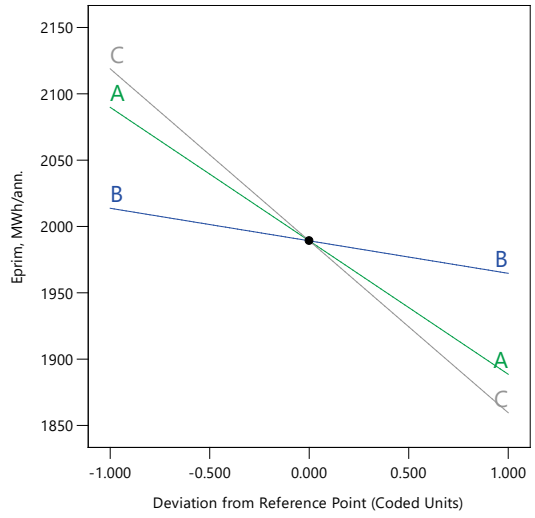


Figure 8 Main effects plot for E_{prim}

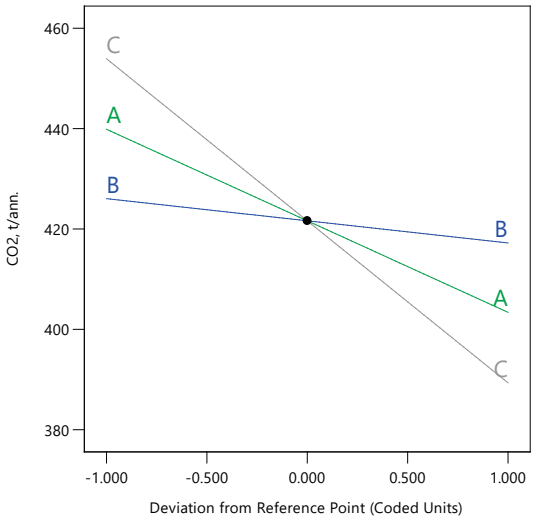


Figure 9 Main effects plot for m_{CO2}

All three energy renovation measures, installing a PV plant, installing thermal insulation on the external wall, and installing the additional thermal insulation on the flat roof contribute to reductions in both primary energy consumption and CO_2 emissions, whereby the installation of the photovoltaic power plant has the most significant impact.

From Fig. 8 and 9, it can be clearly seen that the implementation of energy renovation measures (installation of thermal insulation on the external walls and flat roof, as well as installation of the photovoltaic power plant on the roof of the building) result in reduction of both, primary energy consumption and CO_2 emissions. This is also visible in 3D response surface plots of primary energy and CO_2 emissions, presented in Fig. 10 and 11.

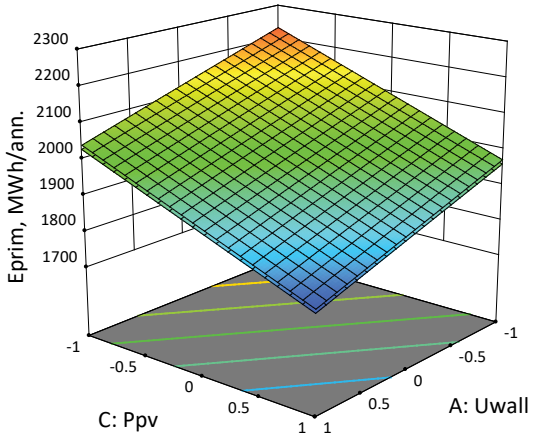


Figure 10 3D response surface plot of E_{prim} as a function of U_{wall} and P_{pv}

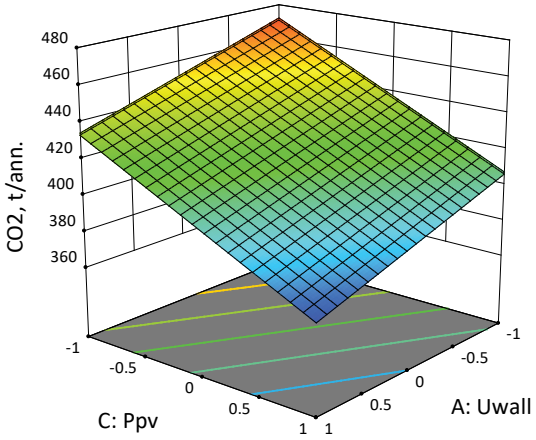


Figure 11 3D response surface plot of m_{CO2} as a function of U_{wall} and P_{pv}

4 CONCLUSIONS

This study demonstrates the benefits of dynamic simulations combined with DOE and regression analysis to model energy performance and identify optimal energy efficiency measures for hospital buildings. The effects of implementing various energy renovation measures on the primary energy consumption and CO_2 emissions of hospital building are analysed.

Hospitals are known to consume significant amount of energy, with energy use intensity being one of the highest among all building types, providing large energy saving potential. In this study, impact of three energy renovation measures on primary energy and CO_2 emission were considered: installing thermal insulation on the external walls and roof and installing photovoltaic power plant.

Hospital building model is designed in Design Builder software, following dynamic simulations of energy performances considering parameters such as climatic conditions, architectural and construction characteristics, installed technical systems, and fuel used. Modelled energy performances of the hospital building, in the current state, show a high agreement with actual energy consumption data, hence validating the accuracy of the building model.

Using Design of Experiments (DOE), followed by ANOVA and regression analysis, mathematical relationships between primary energy consumption and CO_2 emissions (responses) and the renovation measures (factors), were established. The regression models developed in this study enable estimation of primary energy and CO_2 emissions based on renovation measure levels. The analysis revealed that installing photovoltaic power plant and thermal insulation on external walls have the greatest impact on reducing building primary energy and CO_2 emissions. Therefore, study results show that it is possible to develop the mathematical model for prediction of energy performances as a function of various renovation measures, and to use statistical analysis to identify optimal energy renovation measures for hospital buildings.

The research presents a unique dataset of energy characteristics of hospital buildings for various renovation options and levels. Given the lack of studies related to energy renovation in hospital buildings, findings from this study are valuable and may be applicable to other buildings within the healthcare sector. Future research should focus on renovation optimization in healthcare buildings, considering energy and CO_2 emission savings, as well as cost analysis. Presented study results and related analysis will contribute to the broader goal of reducing energy consumption and CO_2 emissions in the built environment, particularly in the healthcare sector.

5 REFERENCES

[1] Tennison, I., Roschnik, S., Ashby, B., Boyd, R., Hamilton, I., Oreszczyn, T., Owen, A., Romanello, M., Ruyssevelt, P., Sherman, J. D., Smith, A. Z. P., Steele, K., Watts, N. & Eckelman, M. J. (2021). Health care's response to climate change: a carbon footprint assessment of the NHS in England. *The Lancet Planetary Health, 5*(2), e84-e92. https://doi.org/10.1016/S2542-5196(20)30271-0

[2] Psillaki, M., Apostolopoulos, N., Makris, I., Liargovas, P., Apostolopoulos, S., Dimitrakopoulos, P. & Sklias, G. (2023). Hospitals' Energy Efficiency in the Perspective of Saving Resources and Providing Quality Services through Technological Options: A Systematic Literature Review. *Energies, 16*(2), 755. https://doi.org/10.3390/en16020755

[3] Bakaimis, B. & Papanikolaou, I. (2017). Electrical energy saving policies, initiatives, results, challenges and lessons learned for the Grevena hospital. *Procedia Environmental Sciences, 38*, 882-889. https://doi.org/10.1016/j.proenv.2017.03.175

[4] Shi, Y., Wang, R. & Chen, P. (2023). Multi-criteria decision-making approach for energy-efficient renovation strategies in hospital wards: Balancing energy, economic, and thermal comfort. *Energy and Buildings, 298*, 113575. https://doi.org/10.1016/j.enbuild.2023.113575

[5] García-Sanz-Calcedo, J., Al-Kassir, A. & Yusaf, T. (2018). Economic and environmental impact of energy saving in healthcare buildings. *Applied Sciences, 8*(3), 440. https://doi.org/10.1016/j.enbuild.2023.113575 https://doi.org/10.1016/j.scitotenv.2020.137446

[6] Balali, A. & Valipour, A. (2021). Prioritization of passive measures for energy optimization designing of sustainable hospitals and health centres. *Journal of Building Engineering, 35*, 101992. https://doi.org/10.1016/j.jobe.2020.101992

[7] Prada, M., Prada, I. F., Cristea, M., Popescu, D. E., Bungău, C., Aleya, L. & Bungău, C. C. (2020). New solutions to reduce

greenhouse gas emissions through energy efficiency of buildings of special importance–Hospitals. *Science of the Total Environment, 718*, 137446.
[8] Nišandžić, M. (2016). Typology of public buildings in Bosnia and Herzegovina. United Nations Development Programme (UNDP).
[9] Li, Q., Zhang, L., Zhang, L. & Wu, X. (2021). Optimizing energy efficiency and thermal comfort in building green retrofit. *Energy, 237*, 121509. https://doi.org/10.1016/j.energy.2021.121509
[10] Baghoolizadeh, M., Rostamzadeh-Renani, R., Rostamzadeh-Renani, M. & Toghraie, D. (2021). A multi-objective optimization of a building's total heating and cooling loads and total costs in various climatic situations using response surface methodology. *Energy Reports, 7*, 7520-7538. https://doi.org/10.1016/j.egyr.2021.10.092
[11] Pan, C., Zhang, D., Kon, A. W. M., Wai, C. S. L. & Ang, W. B. (2015). Patient flow improvement for an ophthalmic specialist outpatient clinic with aid of discrete event simulation and design of experiment. Health Care Management Science, 18(2), 137-155. https://doi.org/10.1007/s10729-014-9291-1
[12] Baril, C., Gascon, V. & Vadeboncoeur, D. (2019). Discrete-event simulation and design of experiments to study ambulatory patient waiting time in an emergency department. Journal of the Operational Research Society, 70(12), 2019–2038. https://doi.org/10.1080/01605682.2018.1510805
[13] Official Gazette of Federation of Bosnia and Herzegovina N 81/19. (2019). Rulebook on minimum requirements for energy performance of buildings. Federal Ministry of Physical Planning.
[14] DesignBuilder Software Ltd. 2019). Design Builder 6.1—User's manual. https://designbuilder.co.uk/download/documents/377-designbuilder-sbem-v6-1-8-manual/file
[15] Nacht, T., Pratter, R., Ganglbauer, J., Schibline, A., Aguayo, A., Fragkos, P. & Zisarou, E. (2023). Modeling Approaches for Residential Energy Consumption: A Literature Review. *Climate, 11*(9), 184. https://doi.org/10.3390/cli11090184
[16] Kulić E. (1990). Heating system design. Savez mašinskih i elektrotehničkih inženjera i tehničara Srbije (SMEITS). (in Serbian)
[17] Diebel, J., Norda, J. & OK. (2022). The Weather Year-Round Anywhere on Earth—Weather Spark.
[18] Paraschiv, S., Paraschiv, L. S. & Serban, A. (2021). Increasing the energy efficiency of a building by thermal insulation to reduce the thermal load of the micro-combined cooling, heating and power system. *Energy Reports, 7*, 286-298. https://doi.org/10.1016/j.egyr.2021.07.122
[19] Yu, D., Tan, X., Liu, Z., Li, D., Wang, Z., Yan, P. & Ni, J. (2023). Energy saving and carbon reduction schemes for hospital with photovoltaic power generation and system upgrading technology. *Heliyon, 9*, e21447. https://doi.org/10.1016/j.heliyon.2023.e21447
[20] Cura, D., Yilmaz, M., Koten, H., Senthilraja, S. & Awad, M. M. (2022). Evaluation of the technical and economic aspects of solar photovoltaic plants under different climate conditions and feed-in tariff. *Sustainable Cities and Society, 80*, 103804. https://doi.org/10.1016/j.scs.2022.103804
[21] Dylewski, R. & Adamczyk, J. (2011). Economic and environmental benefits of thermal insulation of building external walls. *Building and Environment, 46*(12), 2615-2623. https://doi.org/10.1016/j.buildenv.2011.06.023
[22] Lu, G. (2020.). Study on the Properties of Rock Wool for External Thermal Insulation of Buildings under the Soaking and Hot & Humid Conditions. *Journal of Physics: Conference Series, 1622*(1), 012006. https://doi.org/10.1088/1742-6596/1622/1/012006
[23] Kadrić, D., Aganović, A. & Kadrić, E. (2023). Multi-objective optimization of energy-efficient retrofitting strategies for single-family residential homes: Minimizing energy consumption, CO_2 emissions and retrofit costs. *Energy Reports, 10*, 1968-1981. https://doi.org/10.1016/j.egyr.2023.08.086
[24] Costa-Carrapiço, I., Raslan, R. & González, J. (2020). A systematic review of genetic algorithm-based multi-objective optimisation for building retrofitting strategies towards energy efficiency. *Energy and Buildings, 210*, 109690. https://doi.org/10.1016/j.enbuild.2019.109690

Authors' contacts:

Džana Kadrić, Prof.
(Corresponding author)
University of Sarajevo,
Mechanical Engineering Faculty,
Vilsonovo šetalište 9, 71000 Sarajevo, Bosnia and Herzegovina
E-mail: mulahasanovic@mef.unsa.ba

Rejhana Blažević, Prof.
University of Sarajevo,
Mechanical Engineering Faculty,
Vilsonovo šetalište 9, 71000 Sarajevo, Bosnia and Herzegovina
E-mail: muhamedagic@mef.unsa.ba

Hadis Bajrić, Prof.
University of Sarajevo,
Mechanical Engineering Faculty,
Vilsonovo šetalište 9, 71000 Sarajevo, Bosnia and Herzegovina
E-mail: bajric@mef.unsa.ba

Adna Peco, student
University of Sarajevo,
Mechanical Engineering Faculty,
Vilsonovo šetalište 9, 71000 Sarajevo, Bosnia and Herzegovina
E-mail: adna.p.peco@gmail.com

Edin Kadrić, Prof.
University of Sarajevo,
Mechanical Engineering Faculty,
Vilsonovo šetalište 9, 71000 Sarajevo, Bosnia and Herzegovina
E-mail: kadric@mef.unsa.ba

ISSN 1846-6168 (Print), ISSN 1848-5588 (Online)
https://doi.org/10.31803/tg-20240220090703

Original scientific paper
Received: 2024-02-20, Accepted: 2024-04-15

The Effects of 30% Oxygen Concentration Inhalation on Driving Fatigue and Heart Rate Variability

Byung Chan Min*, Kazuyuki Mito, Sang Kon Lee, Seoung Chul Kim, Jeong Han Kim, Seung Hee Hong*

Abstract: This study aimed to evaluate the correlation between heart rate variability and driving fatigue through a physiological approach, examining the effects of oxygen concentrations ranging from normal to high. Driver fatigue, a factor in fatal accidents, has complex causes and varied symptoms. The ambiguity in reporting physiological changes due to mixed terminology such as fatigue and drowsiness prompted this study, which is based on the premise that physiological fatigue originates from the same mechanisms as stress. A driving scenario that induces fatigue was simulated, involving fifteen university student drivers. Experiments were conducted at oxygen concentrations of 21%, 30%, and 40%. Data were collected via an electrocardiogram system and analyzed statistically. Our findings reveal that at a 30% oxygen concentration, drivers showed a significant increase in SDNN and a decrease in LF/HF ratio, indicating enhanced autonomic nervous system stability and reduced sympathetic dominance. This intervention was found to effectively reduce and delay driving fatigue, demonstrating that oxygen supplementation at 30% concentration could notably improve traffic safety by mitigating driver fatigue.

Keywords: driver fatigue; driving stress; Electrocardiogram (ECG); Heart Rate Variation (HRV); 30% Oxygen stimulation

1 INTRODUCTION

Driver fatigue is a principal cause of traffic accidents which, in turn, have a major effect on driver behavior, driving conditions, and driving time in general [1]. Driver fatigue has been reported to cause drowsiness, poor judgment, delayed reaction time, tunnel vision, and occurrences of microsleep. Fatigue while driving is expressed in multiple terms, including drowsiness, distraction, and negligence. Because of the ramifications of the adopted operational definition of driver fatigue, it is necessary to examine the clinical definition and mechanism of fatigue.

Prior to the present study, fatigue had been considered from the perspective of drowsiness; consequently, the reported experimental results encapsulate the incidence of drowsiness rather than the physiological responses to fatigue. Therefore, in this study, the emphasis was to investigate the physiological responses associated with driver fatigue. In addition, we examined the viability of a method for reducing fatigue through oxygen injection.

2 LITERATURE REVIEW

2.1 Definition of Biofatigue

Fatigue may be defined as a form of physical and mental exhaustion resulting on strenuous activity [2], associated with difficulty in initiating and sustaining voluntary activity [3]. Clinically, the manifestation of fatigue is divided into three groups expressed as mental/physical/neural fatigue, central/peripheral fatigue, and acute/chronic fatigue. Mental fatigue occurs because of neurological fatigue caused by psychological factors, whereas physical fatigue occurs because of the accumulation of waste products and insufficient energy supply resulting from continuous muscle use. Neural fatigue occurs because of the processing of large volumes of information collected by sensory organs. An accumulation of fatigue leads to decreasing information processing ability [4]. Peripheral fatigue is a form of muscle fatigue involving neuromuscular junctions, whereas central fatigue is a subjective form of fatigue perceived in the central nervous system [5] which is closely related to socio-psychological factors [6, 7].

2.2 Fatigue and Physiological Indicators

Silverman's research provides a nuanced perspective on fatigue, grounding its physiological basis in mechanisms traditionally associated with maintaining homeostasis under challenging conditions. This framework suggests that fatigue can be understood through the lens of the body's intricate system interactions, involving the nervous, endocrine, and immune systems. Such a perspective implicitly parallels the physiological responses observed in stress-related processes, yet it emphasizes a clear focus on fatigue as a distinct phenomenon.

In exploring the physiological underpinnings of fatigue, Silverman introduces noninvasive evaluation methods aimed at understanding how the body copes with this condition. These methods include the measurement of salivary cortisol, salivary alpha-amylase, and heart rate variability, each offering insights into different aspects of the body's response to fatigue [8]:

1) Salivary Cortisol: This method focuses on measuring the free cortisol present in saliva, which reflects the amount of cortisol circulating freely in the bloodstream. The approach, involving multiple continuous samplings and dynamic testing, is particularly sensitive to changes in the body's physiological state, providing a direct link to fatigue and its effects on the body.
2) Salivary Amylase: The analysis of salivary α-amylase (sAA) serves as a biomarker for the autonomic nervous system's (ANS) activity, particularly the sympathetic branch. Silverman's findings suggest that sAA levels increase in response to various types of stressors, pointing to its utility in understanding how the body's stress response mechanisms are engaged during physical exertion or psychological challenges [11].
3) Heart Rate Variability (HRV): By examining HRV, this approach assesses the balance between the sympathetic and parasythetic branches of the ANS. The variability in

heart rate, especially as influenced by the slow transmission of sympathetic nerve signals, offers a window into the body's adaptive responses, marking a significant index for evaluating physiological states related to fatigue [12].

2.3 Driving Fatigue and Physiological Index

Driver fatigue can be studied in terms of a physiological approach, driver behavior evaluation, vehicle movement, and complex evaluation. According to a database survey spanning fifteen years from 2007 to 2021 [13], 129 studies of driver fatigue were conducted during that period. Forty-three of these studies considered a physiological approach, seventy a behavioral approach, four a vehicular approach, and twelve followed a mixed approach. These studies reflected social concerns and interests about the risk and impact of driver fatigue. Taguchi (1998) studied the relationship between vehicle vibration and driving fatigue, considering HR (Heart Rate), cardiac blood pressure fluctuations (in the form of MWSA: Mayer Wave Related Sinus Arrhythmia), cardiac respiration fluctuations (in the form of RSA: Respiratory Sinus Arrhythmia), and the incidence rate and selection reaction of adrenaline (Adr), Noradrenaline (Nor), salivary cortisol (Cor), and EEG (alpha) in urine. In that study, it was found that adrenaline responded most significantly to the driver's vibration fatigue. Persson et al [14] conducted a set of experiments consisting of driving sessions lasting either 90 min or 135 min, conducted on either motorways or rural roads, and performed at varying stages of alertness of the drivers. With respect to the relationship between HRV and driving fatigue, studies have revealed that HR decreased, whereas SDNN(the standard deviation of NN intervals), RMSSD(the root mean square of successive RR interval differences), pNN50(the percentage of successive RR intervals that differ by more than 50 ms), LF(the rate of low-frequency oscillations), HF(the rate of high-frequency oscillations), and LF/HF(the ratio of low-frequency to high-frequency power) all increased. Zeng et al. [15] reported that the HR decreased, whereas SDNN, RMSSD, pNN50, LF, and HF all increased after driving for 60 min in a highway environment generated by a driving simulator. Buendia et al. [16] reported similar results. However, different HRV results were observed by Jung et al. [17] in their ECG-sensing device development and HRV threshold-setting experiment, which involved driving sessions lasting 2 h on an actual road). The HR, SDNN, RMSSD, and pNN50, as variables in the time domain, and the VLF, LF, HF, and LF/HF, as variables in the frequency domain, were measured and analyzed. The results indicated that a driver's HRV acted as a measure of fatigue in the form of a drowsiness transition (decrease in factors of HR, SDNN, RMSSD and pNN50; increase in factor of LF/HF).

The relationship between oxygen and fatigue has been a subject of numerous studies across various fields, revealing its complex interplay and significant implications for both physical performance and daily activities. Research in sports science and exercise physiology provides a foundational understanding of how oxygen uptake influences fatigue and recovery during physical exertion. For instance, a study highlighted the importance of maximal oxygen uptake ($\dot{V}O_2max$) in determining fatigue and recovery during resistance exercise, underscoring aerobic capacity's role in physical performance [18]. Additionally, gender differences in oxygen saturation and muscle fatigue were explored, suggesting that variations in intramuscular pressures and skeletal muscle perfusion during exercise could lead to differing fatigue onset times between men and women [19].

In the context of team sports like futsal, another study investigated the relationship between heart rate, oxygen consumption, and energy expenditure, aiming to estimate the aerobic and anaerobic contributions to total energy expenditure during a simulated game. This research underscores the complexity of accurately predicting oxygen consumption and its effects on performance in activities characterized by intermittent, high-intensity efforts [20]

The relevance of these physiological insights extends beyond the realm of sports, touching upon everyday activities such as driving. Sung et al. noted the critical impact of stress and fatigue on driving, identifying them as leading causes of traffic accidents. They proposed oxygen injection as a potential solution to alleviate drivers' fatigue and stress, conducting an experiment where oxygen at concentrations of 18%, 21%, and 30% was administered to drivers. The study found that 30% oxygen concentration was the most effective in reducing fatigue and improving driving reaction times, based on subjective evaluations and kinematic measurements. However, this investigation highlighted a gap in research, as it did not assess the drivers' physiological responses to oxygen supplementation.

In light of evidence demonstrating the profound impact of oxygen concentration on physical and cognitive performance, the aim of this study is defined as the examination of how varying oxygen levels influence heart rate variability (HRV) in drivers. HRV, an indicator of autonomic nervous system function, is to be assessed to understand the physiological responses to oxygen variations in a simulated driving environment. Participants are to be exposed to oxygen concentrations of 18%, 21%, and 30%, with the hypothesis being that increased oxygen levels will correlate with improved HRV profiles, indicative of reduced stress and enhanced cognitive and physical resilience. This research is intended to contribute to the understanding of mechanisms by which oxygen can improve driving safety and performance, potentially informing strategies to mitigate driver fatigue-related accidents. The methodology involves a blend of subjective fatigue evaluations and objective physiological measures, aiming to offer insights into optimal driving conditions for safety enhancements.

3 EXPERIMENTAL METHOD

Fifteen male university students with driver's licenses were recruited openly, excluding those with health conditions affecting the experiment. Participants were briefed on precautions and compensated for their time. The study employed a controlled laboratory setup with a graphic

simulator, ECG measurement system, and oxygen injection system. Driving simulations induced mental fatigue on a highway scenario, with three 20-minute driving sessions per participant. HRV parameters were analyzed using electrodes attached via the chest induction method, focusing on time and frequency domains to assess autonomic nervous system balance and vagus nerve activity. Data analysis was conducted using MATLAB and SPSS.

3.1 Participants

In the pursuit of foundational research on Heart Rate Variability (HRV), the present study adopted a controlled sample demographic, limiting participants to male university students within a similar age range. This decision stems from the acknowledgment that HRV can significantly vary across different genders and age groups, necessitating a focused initial investigation to ensure clarity and reliability in the obtained results. Consequently, our experimental group comprised 15 male university students, all of whom possessed valid driver's licenses, symbolizing a homogeneous group in terms of both gender and age, as well as basic driving experience.

Participants were recruited through an open call within the university student community, ensuring that the selection process was transparent and inclusive. The final cohort was carefully chosen based on their health status, explicitly excluding individuals with cardiovascular or respiratory diseases, among other conditions, to minimize external variables that could potentially influence HRV. Additionally, potential participants were briefed about the importance of avoiding alcohol, tobacco, and drug use—factors known to affect the central and autonomic nervous systems—both on the eve of and on the day of the experiment, to further ensure the purity and accuracy of the study's findings.

Upon arrival at the laboratory, participants were thoroughly briefed by the experiment operator on the experiment's objectives, procedures, and safety precautions. It was emphasized that participation was entirely voluntary, with the freedom to withdraw from the study at any moment without any consequences. Following the acquisition of informed consent, physiological signal measuring electrodes were affixed to the participants, and a driving simulation exercise was conducted to standardize the experimental conditions across all subjects. To acknowledge their contribution and time, participants were compensated at a rate of approximately $15 per hour.

3.2 Equipment

The GDS-300S graphic simulator, provided by Grid Space, plays a crucial role in replicating real-world visual stimuli within a controlled setting, allowing for precise manipulation and monitoring of visual inputs. Its sophisticated rendering capabilities and high fidelity make it an indispensable tool for experiments requiring accurate visual simulation, ensuring that the visual stimuli presented to participants are consistent and of the highest quality.

Simultaneously, the MP30 system from BIOPAC was employed for its unparalleled precision in capturing electrocardiogram (ECG) data. This system is designed to facilitate the non-invasive monitoring of electrical activities of the heart, providing real-time insights into the cardiac health and stress responses of participants. Its reliability and accuracy in data acquisition make it an optimal choice for studies focusing on physiological responses to varying conditions.

These components were selected for their proven reliability, precision, and compatibility with the objectives of our study, ensuring that all data collected were of the highest fidelity and relevance. Data extraction and analysis were performed using MATLAB (MATLAB R2019a, MathWorks) for its robust data processing and visualization capabilities. For statistical analysis, SPSS ver.18 was utilized, offering comprehensive tools for in-depth statistical testing and interpretation, thus ensuring the integrity and validity of the experimental findings.

3.3 Environment

To guarantee the precision of our study results, we meticulously implemented control measures for external variables that could potentially influence HRV. These measures were designed to maintain a consistent ambient temperature within the simulation environment, rigorously kept at a constant 23 °C (±1 °C), to negate the effects of temperature fluctuations on participant stress levels and HRV outcomes. Noise control was achieved through the use of sound-insulated headphones that delivered only simulation-related audio cues, thereby eliminating auditory distractions and ensuring that external sound variations did not affect the data.

Our study employed a double-blind design, ensuring that neither participants nor experimenters were privy to the allocation of oxygen conditions throughout the experiment. Participants were informed about exposure to varying oxygen levels in general terms, without specific details, while the sequence of conditions remained undisclosed to experimenters, overseen by an independent team member. This meticulous approach mitigated potential biases, safeguarding the impartiality of experimental implementation and participant response.

3.4 Scenarios

In our driving simulation, scenarios were set on a virtual highway with surrounding vehicles to impose mental fatigue, necessitating the overtaking of preceding vehicles as illustrated in Fig 2. Participants began their journey in the highway's lower lane, shifting to the second lane upon receiving specific instructions. If a preceding vehicle emerged randomly while the participant was in the second lane, they were directed to overtake by moving to the first lane. To moderate the graphic simulation's driving intensity, this procedure was conducted three times, with each driving session spanning 20 minutes. The comprehensive duration of the experiment, as shown in Fig. 1, totaled 150 minutes, including breaks before and after driving sessions, summing up to approximately 3 hours when factoring in preparation and simulator training time. The ordering of oxygen

concentration levels during these sessions was randomized and concealed from participants to prevent order effects. The primary objective of this experiment was to evaluate the influence of HRV and oxygen intake alterations on mental fatigue in conditions of induced driving fatigue. For a thorough assessment of central and acute fatigue, we recorded Electrocardiograms (ECGs) showcasing the activity of the parasympathetic nervous system, particularly the vagus nerve.

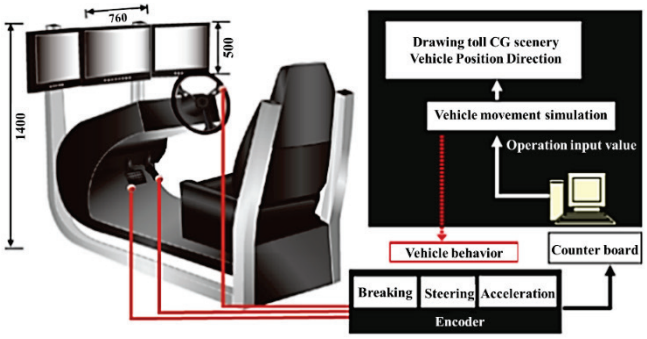


Figure 1 Graphic Simulator Configuration

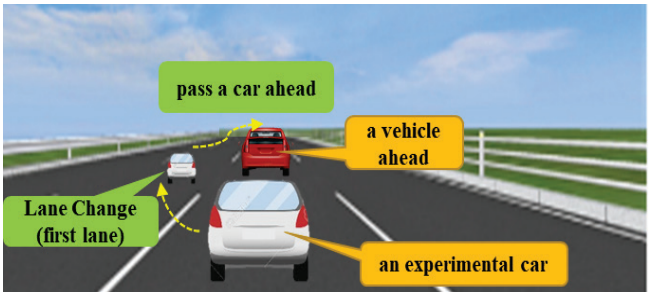


Figure 2 Experimental Driving Scenario

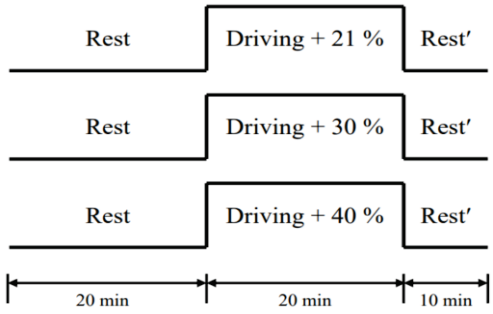


Figure 3 Experiment process and time required

Table 1 HRV Parameters

Parameters	Unit	Definition
		Analysis of 5 min (short-term) recordings
SDNN	ms	Standard deviation of NN intervals
RMSSD	ms	Square root of the mean squared differences between successive RR intervals
PNN50	%	NN50 divided by the total number of RR intervals
LF norm	nu	LF power in normalized units LF/(total power-VLF)×100 *LF Frequency range: 0.04–0.15 Hz
HF norm	nu	HF power in normalized units HF/(total power-VLF)×100 *HF Frequency range: 0.15–0.4 Hz
LF/HF	-	Ratio of LF power to HF power

The order of oxygen concentration released during driving was not notified to the participants in the experiment. In order to exclude order effects, the order of oxygen release was determined by randomization.

The experiment was conducted to evaluate the effect of a change in HRV and oxygen intake on mental fatigue in a situation where mental driving fatigue was imposed. In order to evaluate central and acute fatigue, ECGs reflecting the behavior of the parasympathetic nerve (vagus nerve) were recorded.

3.5 HRV Parameters

Electrocardiogram electrodes were attached using the chest induction (tripole) method depicted in Fig. 4. HRV measured from an electrocardiogram can be analyzed in the time domain and frequency domain. In this study, autonomic nervous system balance and vagus nerve evaluation variables among time and frequency domain parameters were utilized [19].

The variables used for evaluating fatigue were the normalized HF power (HFnorm) and normalized LF power (LFnorm), and those used for evaluating ANS balance were LF/HF and SDNN. Key time domain parameters were used for estimating changes in the vagus nerve. The RMSSD constitutes the HF of HRV. The proportion of NN50 among the entire beat interval (PNN50) was extracted and analyzed.

Continuously measured ECG data were analyzed after removing noise with a 60 Hz NF (Notch. Filter) and an IIR filter (Infinite Impulse Response Filter). HRV parameters were extracted in standard 5-minute units (short-term recordings) [20, 21]

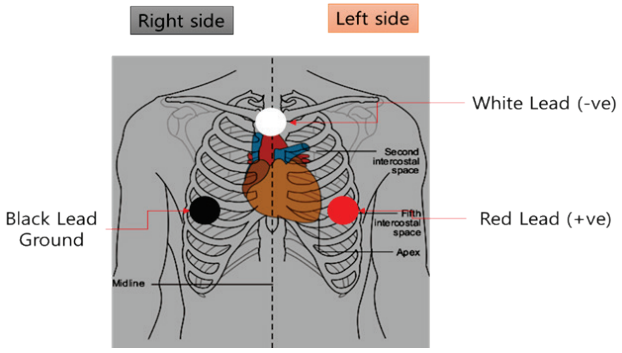


Figure 4 ECG Acquisition system and Configuration

4 EXPERIMENTAL RESULTS

Physiological fatigue responses were compared and evaluated in various ways based on the measurements of HRV in the form of time-lapse and oxygen concentration while resting and driving.

4.1 Changes in HRV over Time

First, we compared the change in heart rate variability with time for resting and driving conditions and the effect of driving on mental stress as a function of time.

In the resting condition, the SDNN maintained a higher value for the total time than in the driving condition, and the SDNN, RMSSD, PNN50, LFnorm, HFnorm, and LF/HF all showed an increase over time. However, in the driving condition, RMSSD. PNN50 and HF showed a decreasing trend over time, whereas LFnorm and LF/HF showed a slightly increase.

The multivariate analysis of HRV with respect to time generated a statistically significant difference between the values obtained for resting and driving conditions, respectively. The time-dependent differences of the SDNN were P5~10 min = 0.08, P10~15 min = 0.01, and P 15~20 min = 0.000, whereas that of PNN50 was P15~20 min = 0.021, and that of the RMSSD was P 15~20 min = 0.013, reflecting the activity of the parasympathetic and vagus nerves.

Table 2 HRV variation with time expressed as *P*-value

(min)	SDNN	RMSSD	PNN50	LFnorm.	HFnorm	LF/HF
0~5	.370	.243	.151	.308	.029	.382
5~10	.008	.091	.947	.744	.166	.392
10~15	.001	.061	.024	.183	.553	.462
15~20	.000	.013	.021	.620	.490	.073

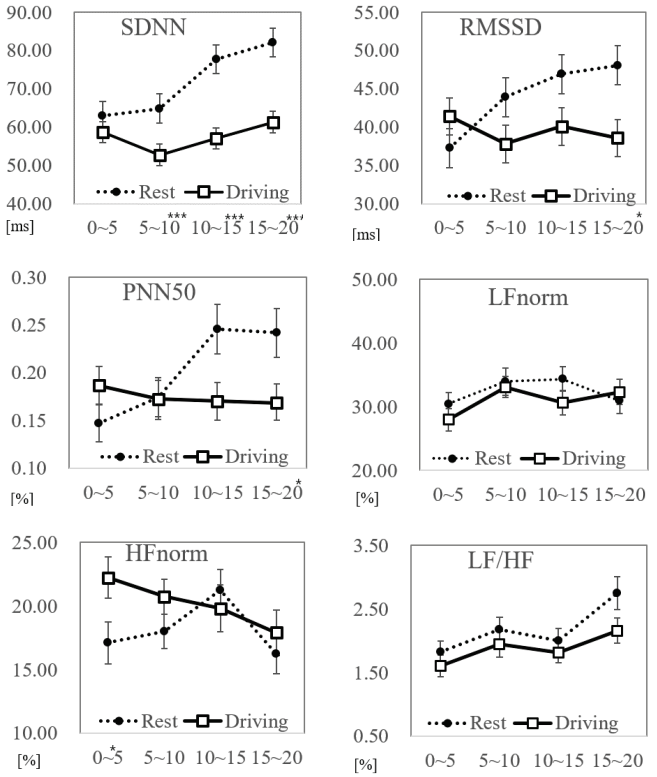


Figure 5 HRV over time while resting and driving

4.2 HRV According to Oxygen Concentration While Resting and Driving

In order to demonstrate the difference in heart rate variability measured by oxygen concentration before driving (that is, while in the resting condition) and after driving (that is, in the driving condition), the values of the parameters in the resting condition and in the driving condition are graphically displayed in Fig. 5. In addition, the statistically significant differences were verified by performing a one-way ANOVA.

According to the results, the SDNN, RMSSD, and PNN50, which are time-dependent variables, all decreased after driving, relative to their values before driving. The SDNN exhibited a statistically significantly decrease (P21 % = 0.039, P30 % = 0.000, and P40 %. = 0.001). The values of the LFnorm, HFnorm, and LF/HF, which are frequency-dependent, exhibited contrasting trends when measured in terms of oxygen concentration. Values of LFnorm decreased at both 21 % and 30 % oxygen concentrations and significantly decreased at 30 % (P30 % = 0.047), increasing only at 40 % concentration. Values of HFnorm exhibited a very small difference at 21 % and 40 %, whereas the value increased significantly at 30 % (P30 % = 0.000). For LF/HF, the difference in values was also small at 21 % and 40 %, whereas the value decreased significantly at 30 % (P30 % = 0.001).

Table 3 The methodology processes and their descriptions.

%	SDNN	RMSSD	PNN50	LFnorm	HFnorm	LF/HF
21	.039	.391	.272	.423	.842	.707
30	.000	.102	.272	.047	.000	.001
40	.001	.072	.338	.358	.987	.854
* Results with a P-value of 0.05 or less are indicated in bold.						

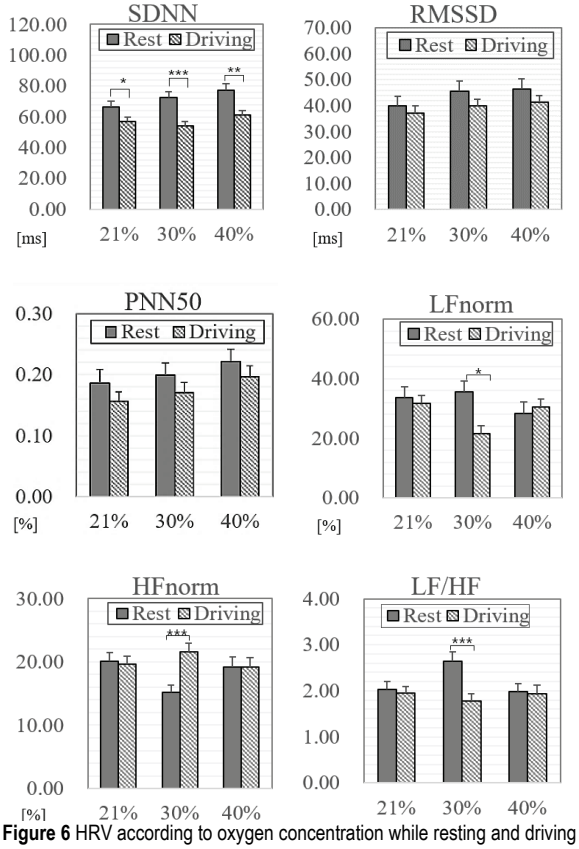


Figure 6 HRV according to oxygen concentration while resting and driving

4.3 HRV According to Oxygen Concentration and Time While Driving

This study was assessed the effect of oxygen on the fatigue and HRV of participants while driving after providing 21%, 30%, and 40% oxygen, respectively. The results were

displayed in Fig 6. At 21% oxygen concentration (the normal concentration in the atmosphere), the SDNN, RMSSD, PNN50, LFnorm, and LF/HF were observed a tendency to increase over time, whereas HFnorm was observed a tendency to decrease. When 30% oxygen was inhaled, the SDNN, RMSSD, PNN50, and HFnorm first decreased and then increased, whereas LFnorm and LF/HF showed a consistently increasing trend. At the highest oxygen concentration of 40%, the SDNN and PNN50 maintained constant levels from the start to the end of driving, whereas the RMSSD and HFnorm decreased sharply, and, accordingly, LF/HF displayed an increasing trend. A multivariate analysis according to the time interval and oxygen concentration while driving indicated that there was no significant difference in any of the variables with respect to both oxygen concentration and elapsed time.

Table 4 HRV by time interval expressed as *P*-value

(min)	SDNN	RMSSD	PNN50	LFnorm.	HFnorm	LF/HF
0-5	.409	.100	.561	.285	.365	.107
5-10	.219	.518	.412	.935	.359	.445
10-15	.733	.733	.604	.580	.848	.528
15-20	.643	.810	.819	.102	.309	.795

4.4 Comparison of Variance in HRV by Oxygen Concentration

To verify whether the variance in HRV before driving and after driving differed with respect to the oxygen concentration, the variance of (HRV while driving - HRV while resting) was calculated using a one-way ANOVA. A significant difference in the variance by oxygen concentration was found in HFnorm ($P = 0.005$). Application of post hoc analysis (Scheffe's) indicated significantly differences between O_2 21% and O_2 30% ($P = 0.011$) and between O_2 30% and O_2 40% ($P = 0.025$).

Table 5 Variance in HRV by Oxygen Concentration expressed as *P*-value

SDNN	RMSSD	PNN50	LFnorm.	HFnorm	LF/HF
.703	.733	.974	.128	.006	.152

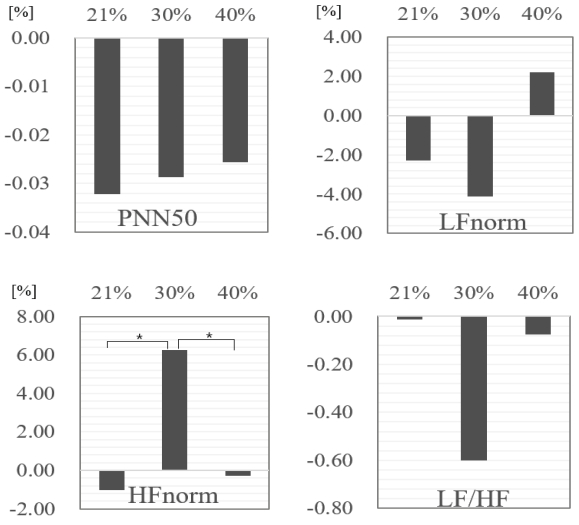


Figure 7 HRV according to oxygen concentration

4.5 A Pearson's Correlation Analysis of HRV

A Pearson's correlation analysis of HRV revealed that the SDNN exhibited a positive correlation with the RMSSD and PNN50 (r = 0.66 and r = 0.56, respectively) and a negative correlation with HFnorm (r = –0.20). Also, RMSSD exhibited a positive correlation with the SDNN, PNN50, and HFnorm (r = 0.66, r = 0.78, and r = 0.21, respectively), whereas PNN50 showed a positive correlation with the SDNN, PNN50, and HFnorm (r = 0.56, r = 0.78, and r = 0.20, respectively). However, it is also consistent with previous studies that LFnorm exhibited no correlation with other variables except LF/HF, HFnorm exhibited positive correlations with the RMSSD and PNN50 in the time domain (r = 0.21 and r = 0.20, respectively), and a negative correlation with the SDNN (r = –0.20), and the parameters of heart rate variability showed correlations with other variables. HF, which represents the activity of the parasympathetic nerve, exhibited a positive correlation with the RMSSD, which represents the vagus nerve activity of the parasympathetic nerve.

Table 6 The methodology processes and their descriptions

SDNN	RMSSD	PNN50	LFnorm.	HFnorm	LF/HF
.703	.733	.974	.128	.006	.152

		SDNN	RMSSD	PNN50	LFnorm	HFnorm	LF/HF
SDNN	Pearson Correlation	1	.656	.563	-.046	-.193	.186
	Sig. (2-tailed)		.000	.000	.397	.000	.001
RMSSD	Pearson Correlation		1	.783	.065	.205	-.059
	Sig. (2-tailed)			.000	.232	.000	.294
PNN50	Pearson Correlation			1	.075	.195	-.060
	Sig. (2-tailed)				.156	.000	.273
LFnorm	Pearson Correlation				1	.059	.432
	Sig. (2-tailed)					.267	.000
HFnorm	Pearson Correlation					1	-.699
	Sig. (2-tailed)						.000

5 DISCUSSION AND CONCLUSION

This study contributes to the expanding research on the non-invasive assessment of driver fatigue through heart rate variability (HRV), aligning with the pioneering insights of Silverman (2010). Our findings underscore the utility of HRV as a reliable indicator of the autonomic nervous system's (ANS) regulatory dynamics, especially in relation to fatigue induced by driving activities. By analyzing both time-domain and frequency-domain parameters of HRV, we have demonstrated the sensitive interplay between physiological responses and environmental factors such as oxygen concentration.

Recent literature reaffirms the potential of HRV analysis in the context of driving fatigue. For instance, the systematic review by Lu et al. (2022) on driver fatigue detection systems using HRV highlights the promising yet variable performance of these systems, indicating a need for

standardization in future research to consolidate HRV's role as a fatigue marker [21]. Similarly, the study applied mutual information for feature selection in ECG-based fatigue detection, represents an advancement in the analytical methodologies, achieving significant accuracy in detecting fatigue states [22]. These studies not only validate the relevance of HRV in fatigue assessment but also illustrate the ongoing efforts to enhance the precision and applicability of HRV-based detection systems.

Our investigation into the effects of different oxygen concentrations on HRV parameters adds a novel dimension to understanding how environmental factors can modulate ANS activity, potentially influencing driver fatigue levels. This inquiry complements existing research, such as the work, which employed wavelet transform and ensemble logistic regression for driving fatigue detection, achieving an accuracy of 92.5% [23]. Moreover, the integration of vehicle data with ECG signals in the study by Halomoan et al. (2023) exemplifies the potential for multimodal approaches to enrich the detection framework.

Acknowledging the limitations inherent in our study, such as the confined sample size and specific driving conditions, we echo the call for broader research to solidify HRV's predictive utility across diverse scenarios. Future endeavors could benefit from exploring the synergy between HRV and other physiological or behavioral indicators of fatigue, offering a more comprehensive framework for real-time fatigue monitoring and intervention.

In sum, our research affirms the critical role of HRV in monitoring driver fatigue, paving the way for further exploration and technological development in this domain. As we move forward, the continued refinement of detection methodologies, bolstered by a deeper understanding of HRV's physiological underpinnings, will be key to realizing the full potential of HRV analysis in enhancing road safety and driver well-being.

Acknowledgement

This research was supported by the 3NF Co., LTD, Republic of Korea.

6 REFERENCES

[1] Hong, S. H., Park, S. J., Min, B. C., Suzuki, K. & Doi, S. (2017). In-Vehicle System Design - Considering Cognitive Characteristics of Elderly Drivers. *International Conference on Platform Technology and Service (PlatCon2017)*, 1-4. https://doi.org/10.1109/PlatCon.2017.7883713

[2] Evans, W. J. & Lambert, C. P. (2007). Physiological basis of fatigue. *Am J Phys Med Rehabil, 86*, 29-46. https://doi.org/10.1097/PHM.0b013e31802ba53c

[3] Chaudhuri, A. & Behan, P. O. (2004). Multiple sclerosis is not an autoimmune disease. *Arch Neurol, 61*, 1610-1612. https://doi.org/10.1001/archneur.61.10.1610

[4] Thayer, J. F. & Lane, R. D. (2009). Claude Bernard and the heart-brain connection: further elaboration of a model of neurovisceral integration. *Neuroscience and Biobehavioral Reviews, 3*(2), 81-88. https://doi.org/10.1016/j.neubiorev.2008.08.004

[5] Davis, M. P. & Walsh, D. (2010). Mechanisms of fatigue. *J Support Oncol, 8*(4), 164-174.

[6] Bower, J. E. (2008). Behavioral symptoms in patients with breast cancer and survivors. *J Clin Oncol, 26*, 768-777. https://doi.org/10.1200/JCO.2007.14.3248

[7] Silverman, M. N., Heim, C. M., Nater, U. M., Marques, A. H. & Sternberg E. M. (2010). Neuroendocrine and immune contributors to fatigue. *PM&R, 2*(5), 338-346. https://doi.org/10.1016/j.pmrj.2010.04.008

[8] Chrousos, G. P. (2009). Stress and disorders of the stress system. *Nat Rev Endocrinol, 5*, 374-381. https://doi.org/10.1038/nrendo.2009.106

[9] Clow. A., Thorn, L., Evans, P. & Hucklebridge, F. (2004). The awakening cortisol response: methodological issues and significance. *Stress, 7*, 29-37. https://doi.org/10.1080/10253890410001667205

[10] Kirschbaum, C. & Hellhammer, D. H. (2000). Salivary cortisol. In: Fink, G., editor. *Encyclopedia of Stress, Vol 3*. Academic Press, San Diego, CA, 379-383.

[11] Nater, U. M. & Rohleder, N. (2009). Salivary alpha-amylase as a non-invasive biomarker for the sympathetic nervous system: current state of research. *Psychoneuroendocrinology, 34*, 486-496. https://doi.org/10.1016/j.psyneuen.2009.01.014

[12] Thayer, J. F. & Sternberg, E. M. (2006). Beyond heart rate variability: Vagal regulation of allostatic systems. *Ann NY Acad Sci, 1088*, 361-372. https://doi.org/10.1196/annals.1366.014

[13] Kamti, M. K. & Iqbal, R. (2022). Evolution of Driver Fatigue Detection Techniques—A Review from 2007 to 2021. *Transportation Research Record, 2676*(12), 485-507. https://doi.org/10.1177/03611981221096118

[14] Persson, A., Jonasson, H., Fredriksson, I., Wiklund, U. & Ahlstrom, C. (2021). Heart Rate Variability for Classification of Alert versus Sleep Deprived Drivers in Real Road Driving Conditions. *IEEE Transactions on Intelligent Transportation Systems, 22*(6), 3316-3325. https://doi.org/10.1109/TITS.2020.2981941

[15] Zeng, C., Wang, W., Chen, C., Zhang, C. & Cheng, B. (2020). Sex differences in time-domain and frequency-domain heart rate variability measures of fatigued drivers. *International Journal of Environmental Research and Public Health, 17*(22), 1-17. https://doi.org/10.3390/ijerph17228499

[16] Buendia, R., Forcolin, F., Karlsson, J., Sjöqvist, B. A., Anund, A. & Candefjord, S. (2019). Deriving heart rate variability indices from cardiac monitoring—An indicator of driver sleepiness. *Traffic Injury Prevention, 20*(3), 249-254. https://doi.org/10.1080/15389588.2018.1548766

[17] Jung, S., Shin, H. & Chung, W. (2014). Driver fatigue and drowsiness monitoring system with embedded electrocardiogram sensor on steering wheel. *IET Intelligent Transport Systems, 8*(1), 43-50. https://doi.org/10.1049/iet-its.2012.0032

[18] Lundberg, T. R., Larsson, G., Alstermark, R. et al. (2024). Relationship between maximal oxygen uptake, within-set fatigue and between-set recovery during resistance exercise in resistance-trained men and women. *BMC Sports Sci Med Rehabil., 16*, 45. https://doi.org/10.1186/s13102-024-00830-8

[19] Curiel, A. (2019). The Relationship between Oxygen Saturation and Muscle Fatigue in Men and Women. *PhD Theses*. University of Oklahoma. https://shareok.org/handle/11244/321125

[20] Santos da Silva, H., Nakamura, F. Y., Papoti, M., Santos da Silva, A. & Dos-Santos, J. W. (2021). Relationship between Heart Rate, Oxygen Consumption, and Energy Expenditure in Futsal. *Frontiers in Psychology, 12*. https://doi.org/10.3389/fpsyg.2021.698622

[21] Lu, K., Dahlman, A. S., Karlsson, J. & Candefjord, S. (2022). Detecting driver fatigue using heart rate variability: A

systematic review. *Accident Analysis & Prevention, 178*, 106830. https://doi.org/10.1016/j.aap.2022.106830
[22] Sudiana, D., Gunawan, T. S. & Salman, M. (2023). ECG-Based Driving Fatigue Detection Using Heart Rate Variability Analysis with Mutual Information. *Information, 14*(10). https://doi.org/10.3390/info14100539
[23] Halomoan, J., Ramli, K., Sudiana, D., Gunawan, T. S. & Salman, M. (2023). A New ECG Data Processing Approach to Developing an Accurate Driving Fatigue Detection Framework with Heart Rate Variability Analysis and Ensemble Learning. *Information, 14*, 210. https://doi.org/10.3390/info14040210

Authors' contacts:

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

Kazuyuki Mito
Department of Informatics, The University of Electro-Communications,
1-5-1 Chofugaoka, Chofu, Tokyo 182-8585, Japan
k.mito@uec.ac.jp

Sang Kon Lee
NF Co., LTD,
24, Sinsojaesandan 1-ro, Jangan-eup, Gijang-gun, Busan, Republic of Korea
sklee@nfeco.co.kr

Seoung Chul Kim
Affiliated Research Institute, NF Co., LTD,
24, Sinsojaesandan 1-ro, Jangan-eup, Gijang-gun, Busan, Republic of Korea
sckim@nfeco.co.kr

Jeong Han Kim
Affiliated Research Institute, NF Co., LTD,
24, Sinsojaesandan 1-ro, Jangan-eup, Gijang-gun, Busan, Republic of Korea
jhkim3@nfeco.co.kr

Seung Hee Hong
(Corresponding author)
Corporate Support Team, Healthcare & Spa Industry Promotion Agency,
84, Yeomchisandan 1-gil, Eumbong-myeon, Asan-si, Chungcheongnam-do,
Republic of Korea
zeele82622@gmail.com

ISSN 1846-6168 (Print), ISSN 1848-5588 (Online)
https://doi.org/10.31803/tg-20240220092303

Original scientific paper
Received: 2024-02-20, Accepted: 2024-04-15

Detection and Predictive Analysis of Drowsiness Using Non-contact Doppler Sensor

Chung Kyo In, Byung Chan Min*

Abstract: The demand for continuous monitoring of vital signs is steadily increasing. Drowsiness occurs when individuals are tired or engaged in repetitive tasks, and driving or working in this state can lead to serious accidents. Various methods for detecting heartbeats based on Doppler sensors have been proposed due to their non-contact nature. Previous research involved developing Doppler radar sensors and verifying their reliability, with over 95 % accuracy compared to traditional ECG devices for heart rate measurement. This study proposes a method utilizing existing Doppler radar sensors to detect and predict drowsiness. To verify the test subjects' drowsy states, their faces were recorded with a camera, and the moments when their eyes were closed were validated as instances of drowsiness. Analytical methods were employed, including cross-method analysis, logistic regression analysis, and panel logistic regression analysis. The analysis revealed a p-value for drowsiness detection lower than 0.001, indicating statistical significance. Moreover, the significance of drowsiness states and stages was confirmed with an accuracy of over 95 %. Particularly, panel logistic regression analysis suggested its suitability as an indicator for predicting drowsiness states. In terms of predicting drowsiness stages and actual drowsiness states, it was observed that a time error of approximately 20-30 seconds exists. The study aimed to detect drowsiness and predict drowsiness based on data acquired through a non-contact Doppler radar sensor.

Keywords: cross analysis; Doppler radar sensor; drowsiness; logistics regression; RRI

1 INTRODUCTION

Drowsiness is akin to falling asleep and is triggered when the body is physically or mentally tired in response to the need for rest. The primary factor that causes drowsiness is the secretion of melatonin, a sleep hormone. Drowsiness can also be induced when there is a high concentration of carbon dioxide in the air or when the environment is warm. Drowsiness is particularly likely to occur during monotonous, repetitive, and demanding tasks, a notable example being driving. Drowsiness can unconsciously creep when handling a steering wheel in a fatigued state, especially during extended high-speed highway driving. On highways, where driving at high speeds is common, driving accidents related to drowsiness can lead to a significant loss of life. Therefore, many studies have focused on detecting drowsiness in advance while driving or working to prevent drowsy driving and industrial accidents.

The method of determining drowsiness based on bio-signals uses electrocardiograms and brain waves, and many studies have been conducted relatively recently. Drowsiness detection technology based on biological signals is accurate, but it has the disadvantage of being uncomfortable for the driver and making conscious judgments because electrodes are attached directly to the driver's body. Therefore, in this study, we develop a program that detects drowsiness based on bio-signals in a non-contact state and analyze the accuracy of drowsiness prediction through experiments.

In a previous study, a Doppler radar sensor was developed and heart rate was measured and analyzed in a non-contact manner. As a result of the measurement, more than 95% accurate reliability was secured compared to existing ECG equipment.

In this study, we present drowsiness and prediction accuracy through a Doppler radar sensor system that remotely detects the breathing and heart rate signals of a stationary target located at a distance of 1 m or less. The proposed radar sensor system is implemented with a 24 GHz band high-sensitivity sensor module to transmit and receive unique operating frequency signals and a signal processing technique to detect biological signals in real time and monitor changes. As a result of measuring nine subjects using a Doppler radar sensor, it was shown that the proposed radar sensor can monitor breathing rate and heart rate per minute with a detection accuracy of over 95%.

In Chapter 2, a program that can detect drowsiness is developed using data acquired from a Doppler radar sensor, and three analysis methods, including cross-tabulation, logistic regression, and panel logistic regression, are used to predict drowsiness. A prediction of drowsiness was presented through this study. Chapter 3 presents the time difference results for whether or not drowsiness is detected and prediction based on the detection results of bio-signals measured from the subject. Chapter 4 discusses various application fields and mentions future research development directions.

In future study, we plan to conduct research that can not only detect drowsiness but also test bio-signals in a non-contact manner by replacing the existing electrocardiogram measurement method in the medical field.

2 METHODS

2.1 RRI Algorithm Analysis

This study aimed to improve peak and drowsiness detection accuracy through Doppler-sensor-based *RRI* (R-R interval) analysis. Fig. 1 presents a flowchart of the proposed *RRI*, which consists of three stages. The first stage involves preprocessing, the second involves spectrum extraction, and the final involves estimating the *RRI* (R-R intervals).

In the pre-processing stage, a Bandpass Filter (BPF) was used to remove frequency components other than the heartbeat, such as respiration and minor movements. Subsequently, short-term Fourier transform (STFT) was applied to calculate the spectrum. To extract the spectrum related to heartbeats, the time window used in STFT should be shorter than that of *RRI* and encompass only one heartbeat.

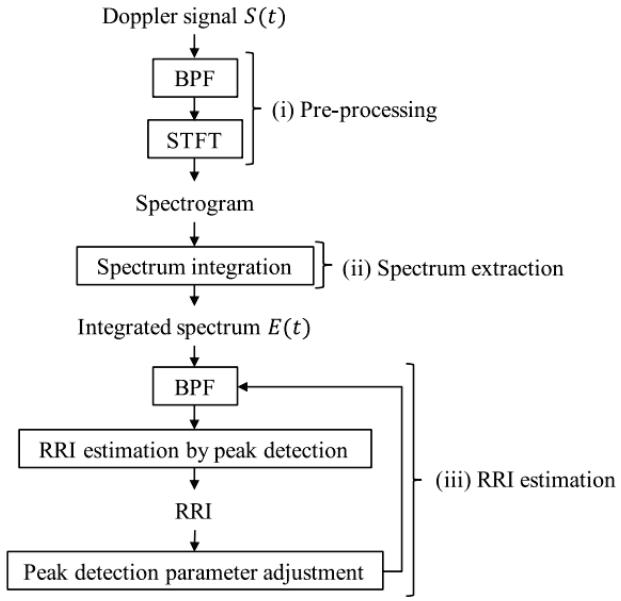


Figure 1 The Flowchart of the RRI method

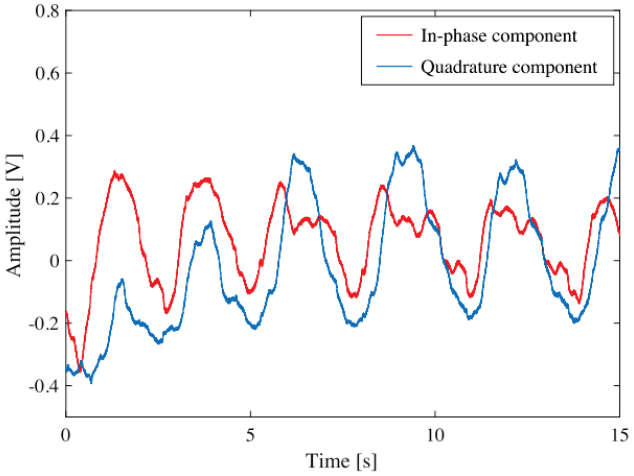


Figure 2 An Examples of in-phase and quadrature components

Fig. 2 provides examples of the interface and quadrature components. Fig. 3 shows eight peaks observed within the window, some related to heartbeats, whereas others are due to noise. Therefore, only the peaks associated with the heartbeats should be selected. Among the 1, 2, 3, ..., 8 peaks, three peaks, p_1, p_2, p_3 ($p_1 < p_2 < p_3$), are selected so that the differences between the previously predicted *RRI* (RRI_{prev}) and the two pairs of *RRI*s are minimized. This was because the adjacent *RRI*s did not vary significantly. Assuming that RRI_{prev} is accurately estimated in Fig. 3, peak 1 is selected as p_1. Subsequently, pairs $RRI_{1,k}$ ($2 < k < 8$) and $RRI_{k,m}$ ($k < m < 8$) were generated. Peaks 1, 3, and 5 are chosen as the final selection for p_1, p_2, and p_3, satisfying the condition that the difference between RRI_{prev}, $RRI_{1,k}$, and $RRI_{k,m}$ is the smallest. Because *RRI*s do not vary significantly between adjacent *RRI*s, a time window is set using the previously estimated *RRI* (RRI_{prev}). If we set the maximum difference between the current *RRI* and the previous *RRI* as 1, when *RRI* increases consecutively twice, the current *RRI* (RRI_{curr}) will equal $RRI_{\text{prev}} + 1$, and the next *RRI* will be $RRI_{\text{curr}} + 1$, which is $RRI_{\text{prev}} + 21$. Therefore, the length of time window W was set according to Eq. (8). This ensured that the window included exactly 3 heartbeats.

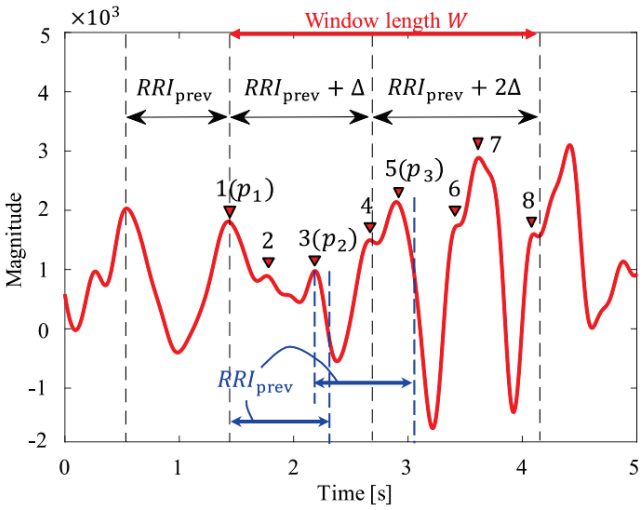


Figure 3 A Concept of the Proposed Peak Detection Algorithm

2.2 Development of Drowsiness Detection Program

This program extracts three main types of data: heart rate, *HRV* (Heart Rate Variability), and drowsiness levels. After specifying the duration of the experiment, data were automatically saved when the set time expired.

The data were stored under folder names, as shown in Tab. 1.

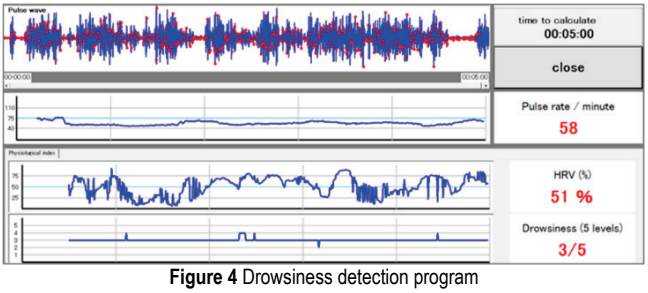


Figure 4 Drowsiness detection program

Table 1 Data file savings

File Name	Example	Description
*****.dat	No_191015155525.dat	Copy of raw data files used for offline analysis. The filename remains the same as the imported file.
No_Processing time.csv ※Saved for offline processing	No_191104131358.csv	Heart Rate, HRV
No_Processing time RR.csv ※Saved for offline processing	No_191104131358_RR.csv	Time (HH:mm: ss.msec) of waveform correction detection and the numerical value of RR interval (msec) for the correction interval concerning input data.

2.2.1 Driving Fatigue and Physiological Index

The drowsiness stage was derived by analyzing the standard deviation of the RR interval, and the RR interval data from the pulse wave peak were resampled and corrected, as shown in Fig. 5.

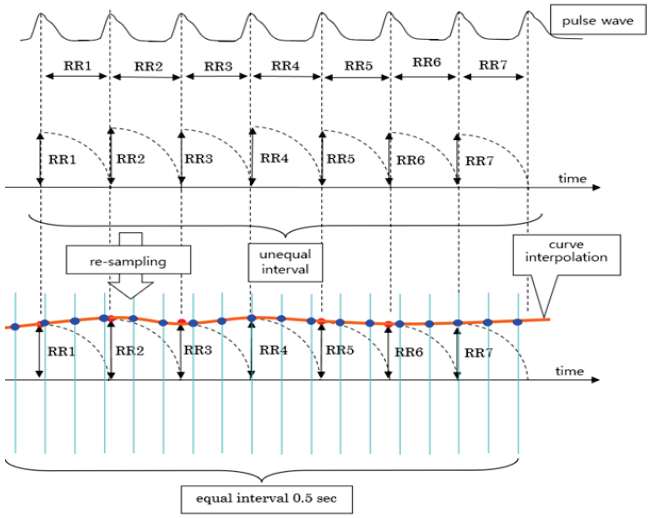


Figure 5 Standard deviation analysis of RR interval

Calculate the power spectrum density of the equally spaced RR interval data, calculating the power between 0.04 and 0.15 Hz as LF and the power between 0.15 and 0.4 Hz as HF.

2.3 Subjects

Nine healthy male and female university students with no cardiovascular or respiratory disorders participated in the experiment. The experiment involved distinguishing between the wakefulness and drowsiness states of each participant. Wakefulness state measurements were taken in the morning, while drowsiness state measurements were conducted after participants stayed awake as late as possible the previous day, and drowsiness was induced by eating a heavy lunch the following day.

2.4 Test Procedure

The participants were instructed to achieve a stable state for 10 min and keep their eyes closed if drowsiness was induced. A camera was placed in front of the participants, and drowsiness was measured using a Doppler sensor. The camera recorded the time displayed on the program, and the moment the participants' eyes closed was used to indicate drowsiness. Drowsiness is detected using this criterion (Tab. 2).

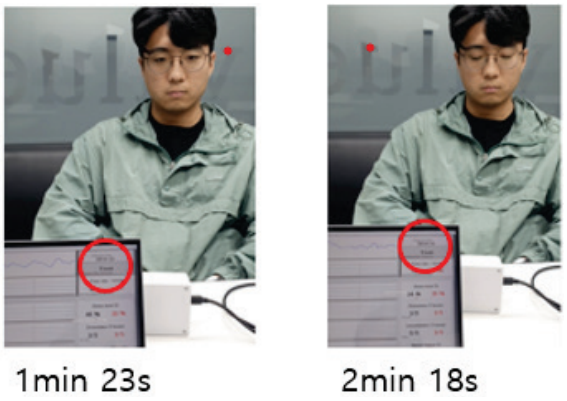


Table 2 Drowsiness detection experiment

Time	Pulse rate	drowsiness	drowsiness (auxiliary number)	Note
0:00:00	78.2354	0	0	
	78.0948	0	0	
	76.0561	0	0	
	76.3104	0	0	
	77.9916	0	0	
	80.3826	3	256.4692	
	78.066	3	244.5294	
	73.8956	3	249.8995	
0:02:18	70.819	3	257.4451	Subject's actual time of drowsiness
	70.819	3	257.4451	
	69.4248	3	254.3162	
	71.1686	3	255.33	
	72.909	3	247.123	
0:02:54	49.7231	4	145.2678	Doppler sensor drowsiness detection time
	49.7231	4	145.2678	
	49.4186	4	144.8847	
	46.4945	4	124.442	
	49.5174	4	148.3674	
	49.6452	4	146.8294	
	49.9585	4	138.2954	
	46.5547	4	117.7181	
	46.5547	4	117.7181	
	46.4956	4	120.5528	
	46.6686	4	115.6547	
	46.4956	4	120.5528	
	45.1074	4	83.6849	
	45.1074	4	83.6849	

2.5 Experimental Analysis Method

This study aims to determine the accuracy of drowsiness detection and prediction using a Doppler sensor. For this experiment, the independent variable data (heart rate) were collected from nine participants, and the program screen was recorded to document the onset of drowsiness. Specifically, by recording the process leading to drowsiness in the wakefulness state and changes in the independent variable, data were collected from nine participants, creating a panel dataset with time information. The independent variable was a continuous variable ranging from 0 to 5, and the dependent variable was coded as 0 for the wakeful state and 1 for the drowsy state. In this study, the following analyses were conducted to verify the hypothesis that as the drowsiness stage increased, the participants were more likely to be in a drowsy state.

First, to examine the descriptive statistics for drowsiness stages and states without differentiation by time, cross-analysis and decision trees were used to assess prediction accuracy.

Second, logistic regression analysis was conducted to determine whether the likelihood of being in a drowsy state increased as the drowsiness stage increased without differentiation over time.

Third, panel logistic regression analysis was conducted using the panel dataset to determine whether the likelihood of being in a drowsy state increased as the drowsiness stage increased. Additionally, a lag of 10-second intervals was used to estimate the degree of temporal error in predicting the drowsy state based on the drowsiness stage.

2.5.1 Cross-analysis

A cross-analysis (chi-square test) was used to compare the frequencies of the two different qualitative variables to identify and analyze their associations. The equation for the cross-analysis is as follows:

$$\chi^2(chi-sqaure) = \sum\left(\frac{(O-E)^2}{E}\right). \quad (1)$$

O is the observed frequency (the actual observed frequency in the contingency table after cross-analysis), and E is the expected frequency (when there is no statistical association between the two variables). The value of χ^2 increases when the observed frequency is greater than the expected frequency, and conversely, the p-value decreases, indicating a statistically significant difference. Based on this, the following hypotheses were formulated: The hypothesis for this study, which examines the relationship between drowsiness and drowsiness states, is as follows: Null Hypothesis (H0): There is no association between the response categories of drowsiness and drowsiness states. Alternative Hypothesis (H1): There is an association between the response categories of drowsiness and drowsiness state. In accordance with this hypothesis, the cross-analysis result for the drowsiness stage and drowsiness state yielded a χ^2 value of 1936.208, and the p-value was less than 0.001. The statistical significance level was set at $p < 0.05$. Therefore, the null hypothesis, which suggests no significant association between the two variables, was rejected, and the alternative hypothesis, indicating an association between the response categories of the two variables, was supported. Based on these results, it can be concluded that there is a statistically significant difference in the ratio of drowsiness and wakefulness states based on the drowsiness stage. Specifically, when examining these values, it was observed that lower drowsiness stage values corresponded to a higher proportion of wakefulness states, and as the drowsiness stage values increased, the proportion of drowsiness states also increased.

Table 3 Analysis of wakefulness and drowsiness states using cross-analysis

Drowsiness Status					
Drowsiness Stage	Drowsiness wakefulness State	Drowsiness State	Total	χ^2	p
0	308 (100%)	0 (0%)	308	1936.208***	<0.001
2	20 (100%)	153 (88.439%)	173		
3	996 (100%)	6815 (87.249%)	7811		
4	0 (100%)	682 (100%)	682		
5	0 (100%)	35 (100%)	35		
total	1324	7685	9009		
* $p < 0.05$					

A Decision Tree algorithm was used to analyze the data and determine combinations of rules that could predict patterns. This was used to assess the prediction accuracy of the drowsiness stage and drowsiness state. When predicting the actual drowsiness state based on the drowsiness stages in Fig. 7 and Fig. 8, it was observed that there were samples with high prediction accuracy (=.92) and samples with relatively low accuracy (=.68). To quantify this further, a logistic regression analysis was performed.

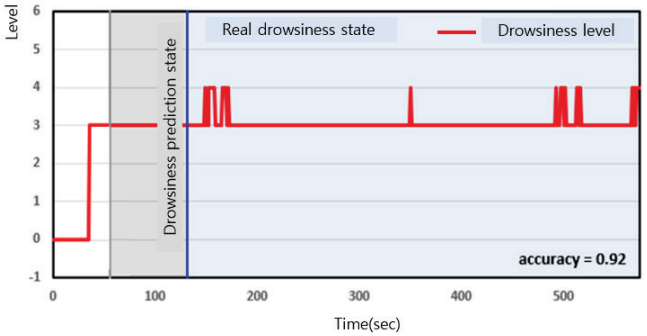


Figure 6 Drowsiness and drowsiness state (Accuracy 92%)

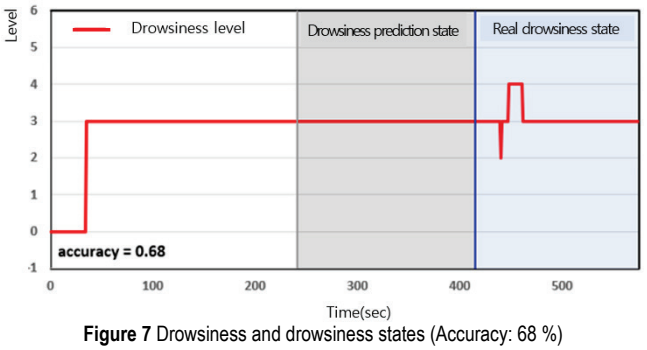


Figure 7 Drowsiness and drowsiness states (Accuracy: 68 %)

2.5.2 Logistic Regression Analysis

Due to its focus on the differences in response frequencies within each cell, cross-analysis has the limitation of being unable to precisely determine how the likelihood of being in a drowsy state increases as the drowsiness stage increases. To address this, logistic regression analysis was employed to provide insight into the probability of being in a drowsy state as the independent variable, the drowsiness stage, increased. The Eq. (2) for the logistic regression analysis is as follows:

$$\text{logit} = \ln\left(\frac{p(x)}{1-p(x)}\right) = \beta_0 + \beta_1 x_1 + \varepsilon. \quad (2)$$

Using a logistic regression model, the analysis aimed to determine the extent to which the likelihood of being in a drowsy state increased as the value of the drowsiness stage increased. The logistic regression analysis results, which predicted drowsiness state as the drowsiness stage increased, showed that the drowsiness stage had a significance level of less than 0.001, making it statistically significant. The Odds Ratio (OR), which indicated how many times the likelihood

of being in a drowsy state increased as the drowsiness stage increased by one level, was calculated as 5.554. Thus, it can be concluded that as the drowsiness stage increases by one level, the likelihood of being in a drowsy state increases significantly by approximately 5.554 times.

Table 4 Logistic Regression Analysis Results

	OR	Significance Level	95% C.I. of EXP(B)	
			Lower	Upper
Drowsiness Stage	5.554***	<0.001	4.765	6.474
Constant	0.0427***	<0.001		
* $p < 0.05$				

2.5.3 Panel Logistic Regression Analysis

While logistic regression analysis can reveal the impact of the drowsiness stage on the drowsiness state, this study further examined the relationship between the drowsiness stage and the drowsiness state by conducting a panel logistic regression analysis using panel data from all nine participants. Moreover, considering the potential time lag in the effect of the drowsiness stage on predicting the drowsiness state, logistic regression analysis was performed using 10-second lags for the drowsiness stage. The analysis showed that as the drowsiness stage increased, the likelihood of drowsiness increased significantly by approximately 2.082 times ($OR = 2.082$, $p < 0.001$). Furthermore, the results for each 10-second lag are as follows:

With a 10-second lag, as the drowsiness stage increased, the likelihood of being in a drowsy state significantly increased by approximately 2.098 times ($OR = 2.098$, $p < 0.001$).

With a 20-second lag, as the drowsiness stage increased, the likelihood of being in a drowsy state significantly increased by approximately 2.105 times ($OR = 2.105$, $p < 0.001$).

With a 30-second lag, as the drowsiness stage increased, the likelihood of being in a drowsy state significantly increased by approximately 2.015 times ($OR = 2.015$, $p < 0.001$).

With a 40-second lag, as the drowsiness stage increased, the likelihood of being in a drowsy state significantly increased by approximately 1.857 times ($OR = 1.857$, $p < 0.001$).

With a 50-second lag, as the drowsiness stage increased, the likelihood of drowsiness significantly increased by approximately 1.874 times ($OR = 1.874$, $p < 0.001$).

With a 60-second lag, as the drowsiness stage increased, the likelihood of being in a drowsy state significantly increased by approximately 1.792 times ($OR = 1.792$, $p < 0.001$).

With a 70-second lag, as the drowsiness stage increased, the likelihood of being in a drowsy state significantly increased by approximately 1.749 times ($OR = 1.749$, $p < 0.001$).

With an 80-second lag, as the drowsiness stage increased, the likelihood of being in a drowsy state significantly increased by approximately 1.764 times ($OR = 1.764$, $p < 0.001$).

With a 90-second lag, as the drowsiness stage increased, the likelihood of being in a drowsy state significantly increased by approximately 1.862 times ($OR = 1.862$, $p < 0.001$).

By analyzing these results, it was observed that the *OR* values decreased to 2 or below after 30 s. Among the lag values, the *OR* values were highest between 20 s and 30 s, indicating a strong and statistically significant relationship between drowsiness and drowsiness states. A time lag of approximately 20-30 s ($OR = 2.105$, $p < 0.001$) appeared to offer the highest prediction accuracy for the relationship between drowsiness and drowsiness states.

	Drowsiness state									
Drowsiness level	2.082***									
	-23.04									
lag1_(10 sec)		2.098***								
		-23.65								
lag2_(20 sec)			2.105***							
			-23.82							
lag3_(30 sec)				2.015***						
				-21.96						
lag4_(40 sec)					1.857***					
					-19.64					
lag5_(50 sec)						1.874***				
						-18.14				
lag6_(60 sec)							1.792***			
							-16.83			
lag7_(70 sec)								1.749***		
								-16.03		
Lag8_(80 sec)									1.764***	
									-15.54	
Lag9_(90 sec)										1.862***
										-15.56
_cons	-4.281***	-4.043***	-3.643***	-2.867**	-1.715	-1.054	1.266	4.654***	3.216***	7.128***
lnsig2u	1.207*	1.384**	1.635***	1.921***	2.467***	2.719***	3.479***	4.138***	3.888***	4.598***
t	-2.54	-2.9	-3.42	-4	-4.58	-5.1	-5.45	-40.84	-5.96	-82.35
N	9009	8919	8829	8739	8649	8559	8469	8379	8289	8199

* p<0.05 ** p<0.01 *** p<0.001

Figure 8 Panel logistic regression analysis results

3 CONCLUSIONS

The results of this study were an experiment on non-contact continuous monitoring using Doppler radar technology, and it was confirmed that drowsiness detection and drowsiness prediction were possible by developing a drowsiness detection program.

Summarizing the analysis results, a significant relationship between the drowsiness stage and drowsiness was confirmed through cross-and decision tree analyses. Among stages 1–5, drowsiness began at stage 3 or higher and was confirmed in more than 95 % of stages 4 and 5.

Next, the intensity of the drowsiness state was confirmed as the drowsiness level increased through a logistic regression analysis. The sample in this study was 9 people, and by panelizing each sample and time, it was confirmed that the drowsiness stage increased the likelih drowsiness.

Finally, to verify the degree of temporal error when the drowsiness stage predicted the drowsiness state, a lag of 10 s was used for each drowsiness stage to identify the section that best explained the drowsiness state.

The drowsiness prediction results of this study can be of great significance in providing a basis for accurately determining the timing of drowsiness. Previous research has simply been to detect drowsiness, but the core of this study is to accurately predict the timing of drowsiness. Therefore,

based on the results of this study, it can be hoped that traffic accidents can be prevented in advance by applying an accurate prediction method for drowsiness to drowsy driving.

4 DISCUSSION

In previous research, a Doppler sensor was developed to ensure the reliability of heart rate measurements. This study was conducted on drowsiness detection using a Doppler sensor. As part of our research, we developed a dedicated program for drowsiness detection and analyzed its reliability of drowsiness detection using three analysis methods.

During the study, a slight lag was found between the participants' actual onset of drowsiness and the moment when the Doppler sensor detected it. Because accidents caused by drowsy driving generally occur after a driver becomes aware of drowsiness, it is important to predict drowsiness before it is consciously recognized.

Although it will take some time for this technology to be applied in real life, it could overcome the shortcomings of traditional monitoring methods using body contact in many applications.

In future research, we believe it will be necessary to further increase the number of experimenters and analyze more accurate data using artificial intelligence-based data correction.

5 REFERENCES

[1] Mu, Z., Hu, J. & Min, J. (2017). Driver fatigue detection system using electroencephalography signals based on combined entropy features. *Applied Sciences, 7*(2), 150. https://doi.org/10.3390/app7020150

[2] Hu, W., Zhao, Z., Wang, Y., Zhang, H. & Lin, F. (2013). Non-contact accurate measurement of cardiopulmonary activity using a compact quadrature Doppler radar sensor. *IEEE Transactions on Biomedical Engineering, 61*(3), 725-735. https://doi.org/10.1109/TBME.2013.2288319

[3] Leem, S. K., Khan, F. & Cho, S. H. (2017). Vital sign monitoring and mobile phone usage detection using IR-UWB radar for intended use in car crash prevention. *Sensors, 17*(6), 1240. https://doi.org/10.3390/s17061240

[4] Mogi, E. & Ohtsuki, T. (2017). Heartbeat detection with Doppler radar based on spectrogram. *IEEE International Conference on Communications (ICC2017)*, 1-6. https://doi.org/10.1109/ICC.2017.7996378

[5] Droitcour, A. D., Boric-Lubecke, O., Lubecke, V. M., Lin, J. & Kovacs, G. T. (2004). Range correlation and I/Q performance benefits in single-chip silicon Doppler radars for non-contact cardiopulmonary monitoring. *IEEE Transactions on Microwave Theory and Techniques, 52*(3), 838-848. https://doi.org/10.1109/TMTT.2004.823552

[6] Li, C., Xiao, Y. & Lin, J. (2006). Experiment and spectral analysis of a low-power Ka-band heartbeat detector measuring from four sides of a human body. *IEEE Transactions on Microwave Theory and Techniques, 54*(12), 4464-4471. https://doi.org/10.1109/TMTT.2006.884652

[7] European Transport Safety Council. (2001). The role of driver fatigue in commercial road transport crashes. Retrieved from https://etsc.eu/wp-content/uploads/The-role-of-driver-fatigue-in-commercial-road-transport-crashes.pdf

[8] Wang, D., Yoo, S. & Cho, S. H. (2020). Experimental comparison of IR-UWB radar and FMCW radar for vital signs. *Sensors, 20*(22), 6695. https://doi.org/10.3390/s20226695

[9] Alshaqaqi, B., Baquhaizel, A. S., Ouis, M. E. A., Boumehed, M., Ouamri, A. & Keche, M. (2013). Driver drowsiness detection system. In *The 8th IEEE international workshop on systems, signal processing and their applications (WoSSPA2013)*, 151-155. https://doi.org/10.1109/WoSSPA.2013.6602353

[10] Rajendra Acharya, U., Paul Joseph, K., Kannathal, N., Lim, C. M. & Suri, J. S. (2006). Heart rate variability: a review. *Medical and Biological Engineering and Computing, 44*, 1031-1051. https://doi.org/10.1007/s11517-006-0119-0

[11] Cheng, J. H., Yeh, J. F., Yang, H. Y., Tsai, J. H., Lin, J. & Huang, T. W. (2012). 40-GHz vital sign detection of heartbeat using synchronized motion technique for respiration signal suppression. In T*he 42nd IEEE European Microwave Conference*, 21-24. https://doi.org/10.23919/EuMC.2012.6459289

[12] Mateo, J. & Laguna, P. (2003). Analysis of heart rate variability in the presence of ectopic beats using the heart timing signal. *IEEE Transactions on biomedical engineering, 50*(3), 334-343. https://doi.org/10.1109/TBME.2003.808831

[13] Zhou, Q., Liu, J., Host-Madsen, A., Boric-Lubecke, O. & Lubecke, V. (2006). Detection of multiple heartbeats using Doppler radar. In *2006 IEEE International Conference on Acoustics Speech and Signal Processing Proceedings, 2*, II-II. https://doi.org/10.1109/ICASSP.2006.1660554

[14] Shinar, Z., Akselrod, S., Dagan, Y. & Baharav, A. (2006). Autonomic changes during wake–sleep transition: A heart rate variability based approach. *Autonomic Neuroscience, 130*(1-2), 17-27. https://doi.org/10.1016/j.autneu.2006.04.006

[15] Horne, J. & Reyner, L. (1999). Vehicle accidents related to sleep: a review. *Occupational and environmental medicine, 56*(5), 289-294. https://doi.org/10.1136/oem.56.5.289

[16] Lu, G., Yang, F., Jing, X. & Wang, J. (2010). Contact-free measurement of heartbeat signal via a Doppler radar using adaptive filtering. In *2010 IEEE International Conference on Image Analysis and Signal Processing*, 89-92.

[17] National Heart, Lung, and Blood Institute. (1998). Drowsy driving and automobile crashes. Retrieved from https://rosap.ntl.bts.gov/view/dot/1661

[18] Sahayadhas, A., Sundaraj, K. & Murugappan, M. (2012). Detecting driver drowsiness based on sensors: a review. *Sensors, 12*(12), 16937-16953. https://doi.org/10.3390/s121216937

[19] Yang, G., Lin, Y. & Bhattacharya, P. (2010). A driver fatigue recognition model based on information fusion and dynamic Bayesian network. *Information Sciences, 180*(10), 1942-1954. https://doi.org/10.1016/j.ins.2010.01.011

[20] Lin, C. T., Wu, R. C., Liang, S. F., Chao, W. H., Chen, Y. J. & Jung, T. P. (2005). EEG-based drowsiness estimation for safety driving using independent component analysis. *IEEE Transactions on Circuits and Systems I: Regular Papers, 52*(12), 2726-2738. https://doi.org/10.1109/TCSI.2005.857555

[21] Vicente, J., Laguna, P., Bartra, A. & Bailón, R. (2011). Detection of driver's drowsiness by means of HRV analysis. In *2011 IEEE Computing in Cardiology*, 89-92.

[22] Hong, T. & Qin, H. (2007, December). Drivers drowsiness detection in embedded system. In *2007 IEEE International Conference on Vehicular Electronics and Safety*, 1-5.

[23] Zhang, W., Cheng, B. & Lin, Y. (2012). Driver drowsiness recognition based on computer vision technology. *Tsinghua Science and Technology, 17*(3), 354-362. https://doi.org/10.1109/TST.2012.6216768

[24] Sakamoto, T., Imasaka, R., Taki, H., Sato, T., Yoshioka, M., Inoue, K., Fukuda, T. & Sakai, H. (2015). Feature-based correlation and topological similarity for interbeat interval estimation using ultrawideband radar. *IEEE Transactions on Biomedical Engineering, 63*(4), 747-757. https://doi.org/10.1109/TBME.2015.2470077

[25] Tariq, A. & Shiraz, H. G. (2010). Doppler radar vital signs monitoring using wavelet transform. In *2010 IEEE Loughborough Antennas & Propagation Conference*, 293-296. https://doi.org/10.1109/LAPC.2010.5666002

[26] Bechet, P., Mitran, R. & Munteanu, M. (2013). A non-contact method based on multiple signal classification algorithm to reduce the measurement time for accurately heart rate detection. *Review of Scientific Instruments, 84*(8). https://doi.org/10.1063/1.4818974

[27] Park, J., Ham, J. W., Park, S., Kim, D. H., Park, S. J., Kang, H. & Park, S. O. (2017). Polyphase-basis discrete cosine transform for real-time measurement of heart rate with CW Doppler radar. *IEEE Transactions on Microwave Theory and Techniques, 66*(3), 1644-1659. https://doi.org/10.1109/TMTT.2017.2772782

[28] Tu, J. & Lin, J. (2015). Fast acquisition of heart rate in non-contact vital sign radar measurement using time-window-variation technique. *IEEE Transactions on Instrumentation and Measurement, 65*(1), 112-122. https://doi.org/10.1109/TIM.2015.2479103

Authors' contacts:

Chung Kyo In
Department of Industrial Management Engineering, Hanbat National University,
125, Dongseo-daero, Yuseong-gu, Daejeon, Republic of Korea
inboss@gmail.com

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-20240319120727

Original scientific paper
Received: 2024-03-19, Accepted: 2024-04-15

A Deep Learning Approaches for Enhancing Clinical Solutions to Cardiovascular Prediction Using EHR

Mounika Valasapalli*, Nallagatla Raghavendra Sai

Abstract: The prediction of cardiovascular disease gained immense significance in the medical field with the alignment of increasing focus on promoting healthier lifestyle. Current methods for cardiovascular disease prediction is leading to so many miss classifications, urging the need of modern automated Deep learning approaches. The main purpose of these approaches is to detect the occurrence of cardiovascular disease (CVD) using patient information from comprehensive electronic health records (HER). Moreover, it is a complex task to choose appropriate features from Electronic Health Records data, and it is a huge confronts to attain robust and accurate results because of the incomplete data entry errors, incomplete record of the patient and patient self-reporting and data integration issues. In this paper we propose an efficient end-to-end framework known as Risk prediction with Deep Residual Neural Network (DRNN), which not only acquires the most influencing features; but also considers the time-based medical data and temporal data to help the patient disease progression, treatment effectiveness, and to check how other diseases are affecting the state of patient. The experimentation is done with the online available Kaggle dataset for cardiovascular disease (CVD) prediction. The result of DRNN demonstrate that the anticipated model significantly enhances the prediction accuracy and F-Measure, Sensitivity compared to various existing approaches. The anticipated model establishes superior trade-off among other approaches.

Keywords: cardiovascular disease; deep residual neural network; high-risk prediction; medical data; prediction accuracy

1 INTRODUCTION

One of the important reasons for death is cardiovascular diseases such as CVD in today's world [1]. Because of cardiovascular diseases, there are 17.9 million people affected out of 57.5 million people, which has been proven as a reported death in 2015 across the world. In addition, the non-negligible burden in economic which is brought by patients with cardiovascular diseases and create critical disability throughout their lifetime. However, this is estimated that around 90 % of cardiovascular diseases are prevented using relevant precautions [2]. The onset of cardiovascular diseases needs to be predicted for the person in the medical domain. There are few acknowledged metrics of pathology for detecting biomarkers for cardiovascular diseases, like angiography and electrocardiogram (ECG). The Authority method used for diagnosis is called angiography in the domain of medicine for cardiovascular diseases, and this obtains greater accuracy in the prediction and diagnosis of cardiovascular diseases. Moreover, the invasive and expensive method is angiography. Another general method for predicting and diagnosing cardiovascular diseases is called the ECG. This accuracy is highly dependent on the knowledge and experience of the experts or the medical staff in the medical domain. Then the guaranteed and considerable area of study is called computer-assisted higher prediction of risk in cardiovascular diseases. Added to the point, the conventional method for higher risk prediction depends on machine learning that focuses on achieving the automatic computer system that needs to be turned with crucial and potential features extracted from the patients' historical EHR, such as electronic health records. It is a manual vulnerable, lower-cost, and non-invasive method compared to conventional pathological measures [3].

EHR drives the higher risk prediction task as the critical difficulty in accurately achieving the patient's portrait, called learning the representation of the patient or the feature engineering. In addition, the EHR consists of different information about the patient, which helps to represent the series of timely ordered visits to hospitals. Everyone has a lot of medical variables, like the diagnosis of demographic procedures and medications of vital science and outcomes from laboratory tests. The number of distinct medical variables in EHR systems is generally very high. The previous predictive models help manage it by having a sparse representation of the features using different dimension reduction approaches. Conventional human intervention measures of feature engineering obtained poor generalization and scalability due to the high dependents on the particular EHR system and the authors' experiences. More scalable and uncomplicated methodologies are used with the help of automatic feature representation that is proposed, like as Bag-of-Words as BoW and One-Hot [4]. Moreover, every feature is considered the independent and discrete word in these methods that create the features that cannot obtain the hidden semantic information accurately among the temporal dynamics and features in the data of EHR. Hence, the process of designing an efficient method for handling the representation of the feature for high-dimension heterogeneous EHR data and sequential is the essential problem.

The important contributions are presented below as the summary session in the proposed system:

- DRNN is a robust and end-to-end model that is suggested in the proposed system for predicting the answer to the greater risk of cardiovascular diseases in accurately having diabetes, hypertension, or hyperlipidemia history for patients who are not having any medical experts' assistance.
- Different clinical data is integrated as a whole, having a close relationship in the proposed system. There is no full utilization of the difference between the heterogeneous data of EHR by the DRNN and the relationship between them.
- DRNN has the performance to be evaluated on the real medical data set in the proposed system. It makes the

DeepRisk obtain better performance on the higher dimensional, heterogeneous data and timely ordered EHR data having the integration of model and reduction in dimensionality.

The work is drafted as: section 2 gives a wider analysis on various prevailing researches on CVD prediction; section 3 elaborates methodology with DRNN. Section 4 describes the numerical outcomes. Section 5 projects the research summary with future research directions.

2 RELATED WORKS

There is difficulty even now in obtaining greater performance using the combination of temporal and heterogeneous clinical data, which is more comprehensive even though many authors have obtained the accurate representation of learning for the higher dimension of data with conventional neural networks (CNN) and recurrent neural network (RNN). Relationships and the differences among the various kinds of medical data are eradicated in the feature learning in the previous research [5]. Also, the temporal data processing of EHR is done with the models based on the CNN [6] that are often restricted due to the need for medical codes during the visits to the hospital or the timely ordered episode. The theoretical attention related to the model of RNN called DeepRisk is suggested in the proposed system for the prediction of the onset of the higher risk cardiovascular diseases and also to deal with the earlier mentioned difficulties for the patient to have hypertension, diabetes, or the history of hyperlipidemia. DeepRisk can excavate the essential information automatically between the raw data and also represent features accurately for the patient having very less dimensions when the patients are turned with the heterogeneous temporal and higher dimensional raw data of EHR.

Long short-term memory (LSTM) and attractive technique in the predictive model is used in the proposed system to utilize the sequential data of EHR more comprehensively [7]. An independent module is used in the proposed system to obtain the relevant characteristics to consider the differences among various types of medical data [8]. On the other hand, the combination of these various data is given to another module in the proposed system to achieve the data regarding the relationship that is potentially used [9]. At last, the modules give all the features learned to integrate for obtaining the greater risk prediction task in the proposal system. The outcomes from the experiments depend on the real data set of the clinic of the cardiovascular diseases that establishes the DeepRisk, which is more robust and accurate when compared with the modern methodologies used in the prediction task based on the HER [10].

3 METHODOLOGY

3.1 Proposed Model

The origination of attention technique originated from the perception of human vision. The global images are scanned first when the human perceives an object and then concentrates on the particular region to obtain more detailed data and also tries to suppress other unwanted data. Google machine translation research team proposes the self-attention technique that helps to obtain extensive attraction having the upcoming study of the attention technique due to the learning process of relationship among the other positions and the particular position to obtain the dependence of context. The self-attention network is used to broad the task. The visual attention technique is utilized to enhance the accuracy. Self-attention technique is first suggested in the proposal system for predicting seizure research to attend to the information of the HER globally. These processes will be elaborated in the upcoming session.

Fig. 2 depicts the process of feeding to the conventional layer for producing two feature maps, Y and Z, provided a local feature $X \in R^{C*H*W}$. Here, Y and Z are the dimensions for both R^{C*H*W}. In addition, the process of reshaping is done for R^{C*N}. Here, $N = H \times W$. Thus, the matrix multiplication is performed among the transportation of Y and Z in the proposed system and then achieves the weight matrix having the dimension with the help of the softmax layer.

$$S_{ji} = \frac{e^{Y_i \cdot Z_j}}{\sum_{i=1}^{N} e^{Y_i \cdot Z_j}} \tag{1}$$

Here, j^{th} position influences on i^{th} position is presented by S_{ji}. More same features give representations of the channels higher relevance among them. In addition, a new representation of feature $T \in R^{C*H*W}$ is created using the process of feeding X and is reshaped to R^{C*N} and the convolution layers simultaneously. However, the matrix multiplication of T and S is performed in the proposed system, and outcome is reshaped to R^{C*H*W}. At last, the scaling parameter α is used to multiply the outcome. The element-wise sum operation having X is executed to obtain the last outcome as $E \in R^{C*H*W}$ is presented in Eq. (2). Here, 0 is assigned to α, which is allocated little by little more weight. Hence, E is the last feature, the original feature, and the sum of weighted feature which includes the perspective related to context and gathering the information globally, which depends on the attention map optionally.

$$E_j = \alpha \sum_{i=1}^{N} (S_{ji} * T_i) + X_j \tag{2}$$

The various features of the channel represent various semantics of EHR signals. The interdependence among the mapping of the channel is mined by using the channel attention technique, and various representations of semantics are relevant to one another. In the first step, the process of reshaping X to R^{C*N} is done using the transpose of X and the matrix multiplication among X. At last, the soft wax layer is used in the proposed system for achieving the map of channel attention $P \in RC \times C$ presented in Eq. (3). Here, i^{th} channel influences on the j^{th} channel is measured by P_{ji}.

$$P_{ji} = \frac{e^{X_i \cdot X_j}}{\sum_{i=1}^{C} e^{X_i \cdot X_j}} \tag{3}$$

The matrix multiplication is performed on X and P in the proposed system, and result is reshaped to R^{C*H*W}. The scaling parameter β is used to multiply the result in the proposed system. The element-wise sum operation having X is executed to obtain the last output as $E \in R^{C*H*W}$ is presented in Eq. (4). Hence, 0 is initialized to β. The integration of two attention models using the element-wise sum in the proposed system to completely utilize the spectrum and channel information of context. After fusing original features, last feature map needs to be achieved using the average pooling in the proposed system.

$$E_j = \beta \sum_{i=1}^{C} (P_{ji} * X_i) + X_j \tag{4}$$

Here, k-fold cross-validation (CV) is chosen for every patient in the proposed system to obtain outcomes that are the same as the real conditions. The complete interictal recordings are classified into n parts. Every part includes the t/n hours gathered randomly, having any patient's recordings when the patient has t hours and n recordings. In addition, feature pairs are provided for testing, this is performed for n times. On the other hand, balance $n - 1$ pair or utilized for the training stage. A few researchers commonly randomly classified 20 % testing data and 80 % training data. These sets are utilized as the validation to focus on overfitting. Moreover, the anticipated methodology is relevant for HER data independency. In addition, HER data is time-dependent and samples are chosen in the proposed system at various periods to observe whether the model started to overfit during the training period. The proposed system chooses 25 % samples from the recordings of patients as the set of validation to monitor in the training set, and the balance 75 % samples are utilized as training set. The problem of overfitting is still present to the number of alterations that upsurges the training accuracy in training process and the early stop method is used to solve this issue. The training has to be stopped and storing the network parameters at the low loss of validation is done immediately during the detection of the loss on the set of validation which is started to maximize.

Tab. 2 depicts the parameter of DRNN model. The model has the input, which is $1 \times 22 \times 9 \times 114$. Here, the number of data is presented as 22, and the dimension of the features is represented as 9×114. The batch normalization, ReLu activation function, and dropout are used to follow every conventional layer. The earlier feature map helps feed the matrix to obtain $64 \times 7 \times 28$, which leads to the convolutional layer, followed by the reshaping operation. There are 4 ResBlock layers used in the proposed system for extracting the deep features of the EHR subsequently as in Fig 1. The global features are fused using the dual self-attention, fully connected layers with activation function. Loss function is chosen in the proposed system as the function of cost. However, 32 are considered the size of the batch. Accordingly, 0.0005 and 0.5 are fixed as the rate of learning and rate of dropout. The proposed new model has been accomplished with MATLAB 2020a.

Table 1 Network architecture of the Proposed Model DRNN

Layers	Output size	Structural description
Input	$1 * 22 * 9$	--
Conv_1	$64 * 1 * 7 * 55$	$22 * 3 * 5$ conv Stride $1 * 1 * 2$
Pooling	$64 * 1 * 7 * 28$	$1 * 1 * 2$ max pooling
RB1	$64 * 7 * 28$	$\begin{pmatrix} 3*3 & 64 \\ 3*3 & 64 \end{pmatrix} * 2$
RB2	$128 * 4 * 14$	$\begin{pmatrix} 3*3 & 128 \\ 3*3 & 128 \end{pmatrix} * 2$
RB3	$256 * 2 * 7$	$\begin{pmatrix} 3*3 & 256 \\ 3*3 & 256 \end{pmatrix} * 2$
RB4	$512 * 1 * 4$	$\begin{pmatrix} 3*3 & 512 \\ 3*3 & 512 \end{pmatrix} * 2$
Attention model	$512 * 1 * 4$	Attention
Avg. pooling	$512 * 1$	--
Classification	$2 * 1$	Softmax

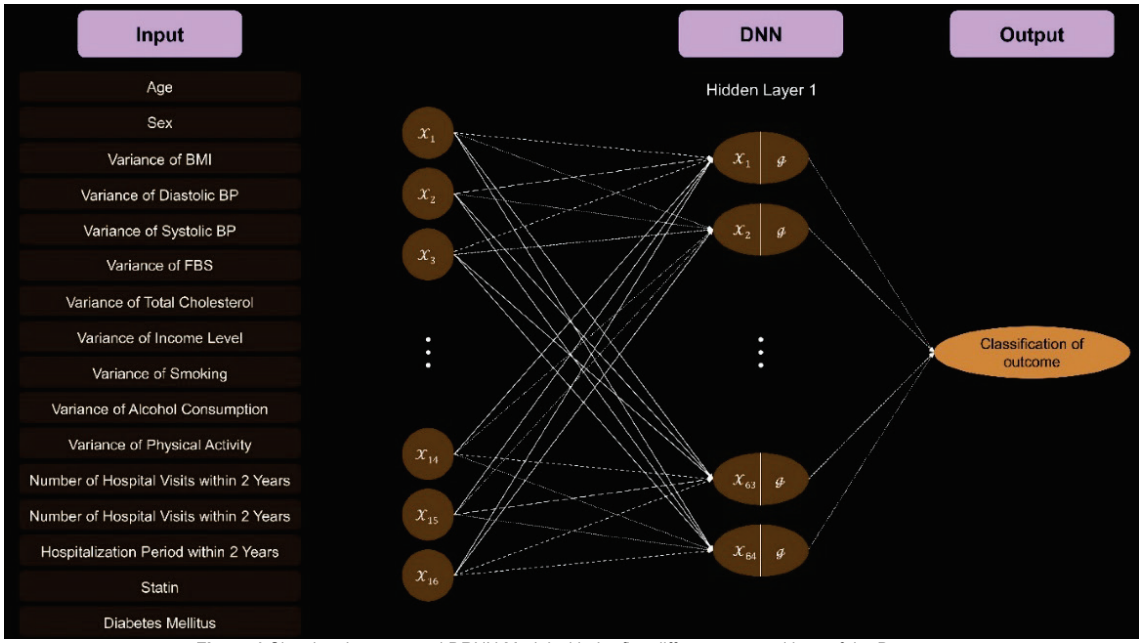


Figure 1 Showing the proposed DRNN Model with the five different compositions of the Datasets

3.2 Model Description

Residual network (ResNet) and dual self-attention techniques are presented in the proposed model. Fig. 2 depicts attraction of the internal module; the EHR data is used in the proposed system as the input to network and heart disease features are hauled out via ResNet potentially. Thus, the input is given to the attention module using the below 3 steps for creating the new global features. An attention matrix is created in the first step to present the special relationship among any features. In the second step, the multiplication matrix on the original features and the attention matrix is executed in the proposed system. In the third step, the final global feature is performed in the proposed system with the help of execution of element-wise sum among original feature and matrix outcome. The channel attention process is the same as the attention model. At last, features (ResNet and attention) are merged and combined with original features to acquire EHR's better characteristics.

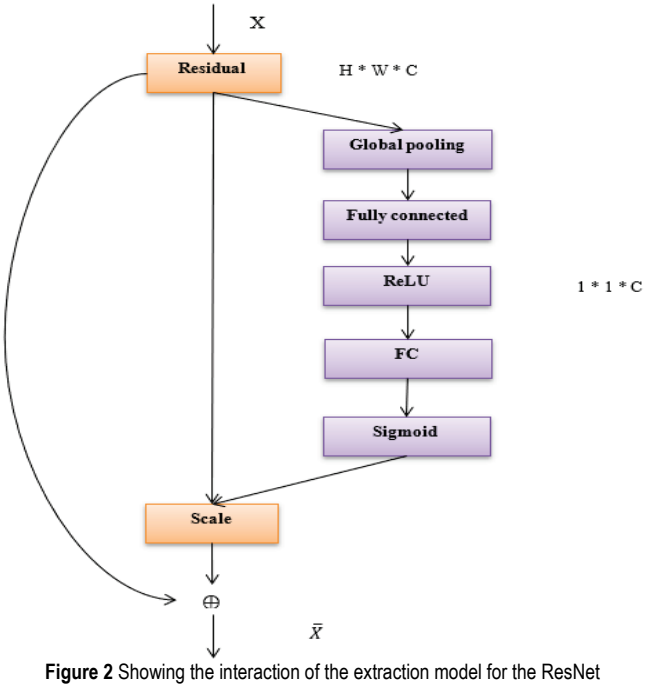


Figure 2 Showing the interaction of the extraction model for the ResNet

CNN is used broadly in natural language processing, computer vision, and so on [30]. The Residual function is used to alleviate the vanishing gradient problem and used to training of very deep neural networks. Features are processed through residual connections, global pooling, fully connected layers with ReLU activations, and finally a sigmoid activation function for binary prediction. It is efficient for depending on total network layers for obtaining the ability of the expression and fitting the mapping relationship better and potentially. There is an issue of the disappearance of radiance, and gradient descent optimization becomes very challenging as the total layers depend on the network model. In recent years, ResNet has been used for combating the mentioned issue when the training has happened for the very deep conventional network consisting of different restaurant blocks and was proposed. Fig 1 presents the addition of identity mapping by the residual block with the network via connections of shortcut that is not based on the calculation and not adding the extra parameters to solve the model degradation. Every residual block consists of 3 × 3 convolution layer, batch normalization and ReLU. The model includes identity map as x and the residual path as $F(x)$. This research has four residual blocks to use that are connected to the proposed system.

The sample data is castoff to calculate the limits of the model. The probability equation is nonlinear for this distribution, therefore analytic solutions are not available. As a result, for getting parameter estimations, we use an iterative approach such as the ResNet Model.

For the ResNet algorithm, the limitations of the given model revised equations are evaluated. The probability of the present model, is

$$Q(\theta,\ \theta^{(0)}) = E_{\theta^{(0)}}\left[\frac{\log L(\theta)}{\bar{x}}\right] \tag{5}$$

This implies

$$\log L(\theta) = \sum_{s=1}^{N} \log\left(\sum_{i=1}^{k} \alpha_i^{(l)} f_i(x_s, \theta^{(l)})\right) \tag{6}$$

The conditional probability region k is

$$P_k(x_s,\ \theta^{(l)}) = \left[\frac{\alpha_k^{(l)} f_k(x_s,\ \theta^{(l)})}{p_i(x_s,\ \theta^{(l)})}\right] \tag{7}$$

$$p_k(x_s,\ \theta^{(l)}) = \left[\frac{\alpha_k^{(l)} f_k(x_s,\ \theta^{(l)})}{\sum_{i=1}^{k} \alpha_i^{(l)} f_i(x_s,\ \theta^{(l)})}\right] \tag{8}$$

Therefore, the ResNet distribution:

$$f_i(x_s,\ \theta^{(l)}) = \frac{\left[\frac{3}{(3p+\pi^2)}\right]\cdot\left[p+\left(\frac{x_s-\mu_i^{(l)}}{\sigma^{(l)}}\right)^2\right]\cdot e^{-\left(\frac{x_s-\mu_i^{(l)}}{\sigma^{(l)}}\right)}}{\sigma_i^{(l)}\left[1+e^{-\left(\frac{x_s-\mu_i^{(l)}}{\sigma^{(l)}}\right)}\right]^2} \tag{9}$$

The likelihood function of example comments must be estimated as the first step of ResNet Algorithm.

E-step: The following expectation (E) step, $\log L(\theta)$ is the expectation value, with respect to the preliminary parameter vector $\theta^{(0)}$ is

$$Q(\theta,\ \theta^{(0)}) = E_{\theta^{(0)}}\left[\frac{\log L(\theta)}{\bar{x}}\right] \tag{10}$$

Given the preliminary parameter $\theta^{(0)}$.

M-step: Parameters estimation,

$$F = \left[E\left(\log L(\theta^{(l)})\right) + \beta\left(1 - \sum_{i=1}^{k} \alpha_i^{(l)}\right)\right] \tag{11}$$

The Updated equations of α_i: The appearance for α_i, we solve the following calculation

$$\frac{\partial F}{\partial \alpha_i} = 0;\ \sum_{i=1}^{N} \frac{1}{\alpha_i} P_i(x_s,\ \theta^{(l)}) + \beta = 0;\ \beta = -N \tag{12}$$

Consequently,

$$\alpha_i = \frac{1}{N} \sum_{s=1}^{k} P_i(x_s,\ \theta^{(l)}) \tag{13}$$

The updated equations of α_i for $(l + 1)^{\text{th}}$ iteration is

$$\alpha_i^{(l+1)} = \frac{1}{N} \sum_{s=1}^{k} P_i(x_s,\ \theta^{(l)}) \tag{14}$$

This implies

$$\alpha_l^{(l+1)} = \frac{1}{N} \sum_{s=1}^{N} \left[\frac{\alpha_l^{(l)} f_l(x_s,\ \theta^{(l)})}{\sum_{i=1}^{k} \alpha_i^{(l)} f_i(x_s,\ \theta^{(l)})} \right] \tag{15}$$

The Rationalized equations of μ_i: For limitation logistic type dispersal - By put on the derived with respect to μ_i, we have

$$\frac{\partial}{\partial \beta_i}\left[\sum_{s=1}^{N} \sum_{i=1}^{k} P_i(x_s, \theta^{(l)}) \log \alpha_i \frac{\left[\frac{3}{(12+\pi^2)}\right] \cdot \left[4 + \left(\frac{x_s - \beta_i}{\sigma_i}\right)^2\right] \cdot \mathrm{e}^{-\left(\frac{x_s - \beta_i}{\sigma_i}\right)}}{\sigma_i \left[1 + \mathrm{e}^{-\left(\frac{x_s - \beta_i}{\sigma_i}\right)}\right]^2} \right] = 0 \tag{16}$$

$$\frac{\partial}{\partial \beta_i}\left[\sum_{s=1}^{N} \sum_{i=1}^{k} P_i(z_s, \theta^{(l)}) \log \alpha_i \frac{\left[\frac{3}{(12+\pi^2)}\right] \cdot \left[4 + \left(\frac{z_s - \beta_i}{\sigma_i}\right)^2\right] \cdot \mathrm{e}^{-\left(\frac{z_s - \beta_i}{\sigma_i}\right)}}{\sigma_i \left[1 + \mathrm{e}^{-\left(\frac{xz - \beta_i}{\sigma_i}\right)}\right]^2} \right] = 0 \tag{17}$$

This implies

$$\sum_{s=1}^{N} P_i(z_s, \theta^{(l)}) \left[\left[\frac{2\left(\frac{z_s - \beta_i}{\sigma_i}\right)\left(-\frac{1}{\sigma_i}\right)}{\left[4 + \left(\frac{z_s - \beta_i}{\sigma_i}\right)^2\right]} \right] + \left[\frac{1}{\sigma_i}\right] - \left[\frac{\mathrm{e}^{-\left(\frac{z_s - \beta_i}{\sigma_i}\right)^2}}{\sigma_i \left(1 + \mathrm{e}^{-\left(\frac{z_s - \beta_i}{\sigma_i}\right)^2}\right)} \right] \right] = 0 \tag{18}$$

$$\mu_i^{(l+1)} = \frac{\sum_{s=1}^{n} \frac{P_i(z_s,\ \theta^{(l)})(2y_s)}{(\sigma_i^{(l)})^2 \left[p + \left(\frac{z_s - \beta_i^{(l)}}{\sigma_i^{(l)}} \right)^2 \right]} \sum_{s=1}^{n} \frac{P_i(z_s,\ \theta^{(l)})}{\sigma_i^{(l)}} + \sum_{s=1}^{n} \frac{2P_i(z_s,\ \theta^{(l)})}{\sigma_i^{(l)} \left[1 + \mathrm{e}^{\left(\frac{z_s - \beta_i^{(l)}}{\sigma_i^{(l)}} \right)} \right]}}{2 \sum_{s=1}^{n} \frac{P_i(z_s,\ \theta^{(l)})}{(\sigma_i^{(l)})^2 \left[p + \left(\frac{z_s - \beta_i^{(l)}}{\sigma_i^{(l)}} \right)^2 \right]}} \tag{19}$$

3.3 Dataset Description

Therefore, the use of a repository gives the data, which are all available online. Data pre-processing is performed, and then different records of patients are gathered. There are 303 records of the patients presented in that data. On the other hand, the missing values are presented from the six records. Missing values for eliminated and remaining data are maintained, such as 297 records. In addition, the introduction of the multi-class variables and the binary classification are done for the dataset attributes. Some variables are employed to measure disease occurrence or not. Values are fixed as 'zero' or 'one', and no symptoms of the disease are represented by 0. Hence, the occurrence of the disease is shown as 137, and the non-occurrence of the disease is shown as 161 accordingly. Dataset description is provided in Tab. 2.

Table 2 Dataset attributes in Kaggle Dataset

Attributes	Descriptions	Type
Age	CVD patients age	N
Sex	CVD Patients' gender	N
CP	CVD type 1. Typical/atypical angina, 3. Non-anginal, 4. Asymptomatic	Nm
Tresbps	BP state	Nm
Chol	Serum cholesterol	N
FBS	Blood sugar level	N
Resting	ECG outcomes	Nm
Thali	Max. heart rate	N
Exchange	Angina owing to exercise	Nm
OldPeak	Exercise owing to depression	N
Slope	peak exercise segment	Nm
Ca	Fluoroscopy coloured vessels	N
Thal	CVD status	N
Num	CVD diagnosis value	N

N - Nominal; Nm – Numerical

4 RESULTS AND DISCUSSIONS

The proposed model needs to compute the performance using statistical analysis like precision, sensitivity, F-measure, recall, Matthews' correlation coefficient (MCC), true negative rate (TNR), false positive rate (FPR) and false negative rate (FNR). Existing techniques have the evaluation such as the Bayesian model (BM), regression model (RM), feed-forward neural network (FFNM), linear support vector machine (L-SVM), random forest (RF), decision tree (C4.5), deep auto-encoder (DAE) and ensemble classifier techniques are used to obtain the significance of the model. Eq. (5) to Eq. (12) expresses the mathematical representation of these measures.

$$Accuracy = \frac{TP + TN}{TP + FN + FP + TN}$$

(20)

$$Precision = \frac{TP}{TP + FP}$$

(21)

$$Recall / Sensitivity / TPR = \frac{TP}{TP + FN}$$

(22)

$$F\text{-}measure = \frac{2 * Precision}{p + r} \quad (23)$$

$$MCC = \frac{(TP * TN) - (FP * FN)}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \quad (24)$$

$$FPR = \frac{FP}{FP + TN} \quad (25)$$

$$FNR = \frac{FN}{FN + TP} \quad (26)$$

$$TNR = \frac{TN}{TN + FP} \quad (27)$$

Deep Auto-encoder (DAE), Ensemble, and DRNN have achieved notably higher performance across various metrics compared to other models (Tab. 3).

The comparison of the anticipated in the simple model is presented in Tab. 3, having the prevailing techniques. In addition, a few statistical metrics include precision, accuracy, F-measure, sensitivity, FPR, MCC, FNR, and TNR. These measures have the results across the anticipated methodology: precision at 98.8%, accuracy at 98%, sensitivity at 97.5%, F-measure at 97.5%, MCC at 95%, FPR at 4.54%, FNR at 3.15% and TNR as 98% accordingly. Then, the existing model shows an accuracy of 8%, 13%, 10%, 26%, 22%, 14%, 1% and 0.8%, which are superior to the anticipated model. The existing model shows precision of 8.8%, 10.8%, 13.8%, 22.8%, 22.8%, 11.8%, 0.8% and 0.6%, which are superior to the prevailing model. The existing model shows a sensitivity of 12.5%, 11.5%, 16.5%, 44.5%, 25.5%, 22.5%, 1.5% and 1.3%, which are superior to the prevailing model. The existing model shows an F-measure of 12.5%, 12.5%, 11.5%, 35.5%, 24.5%, 16.5%, 1.5% and 1.3%, which are better than the proposed model. The existing model shows an MCC of 20%, 20%, 25%, 50%, 43%, 25%, 1% and 0.8%, which are better than the proposed model. The existing model shows FPR of 10.46%, 9.46%, 8.46%, 12.46%, 0.19%, 9.46%, 1.46% and 1% higher than the proposed model. The existing model shows an FNR of 16.85%, 18.85%, 16.85%, 48.85%, 27.85%, 21.85%, 1.2% and 1%, which are higher than the proposed model. The existing model shows TNR of 8%, 8%, 6%, 12%, 20%, 8%, 1% and 0.7%, which are higher than the anticipated model. The complete performance is considerably compared with other approaches using the comparative analysis, and the anticipated ensemble model is proved that the model provides better performance for prediction and the complexity is reduced in the learning techniques.

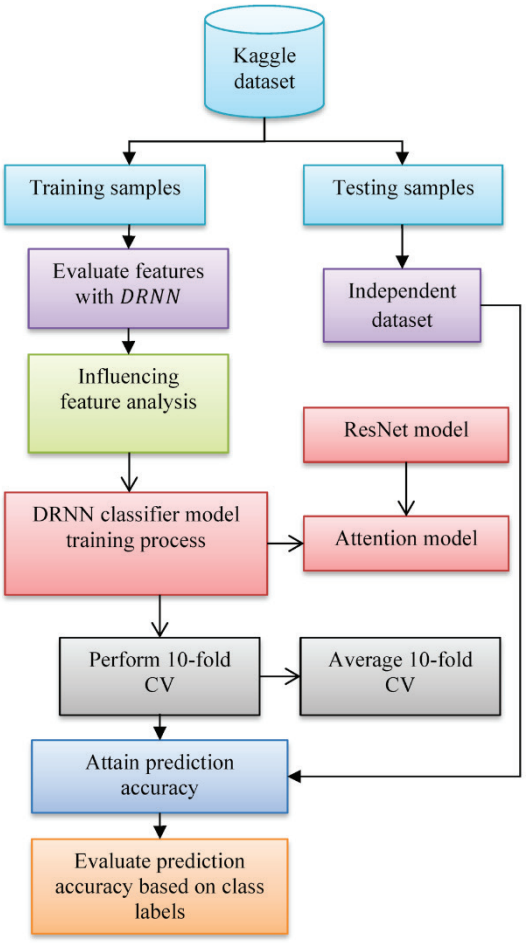


Figure 3 Flow chart of the Proposed Methodology for the different composition of the datasets

Table 3 Overall comparisons of the results with proposed model

Approaches	Accuracy	Precision	Sensitivity	F-measure	MCC	FPR	FNR	TNR
Bayesian model (BM)	90%	90%	85%	85%	75%	15%	20%	90%
Regression model (RM)	85%	88%	86%	85%	75%	14%	22%	90%
Feed Forward network model (FFNM)	88%	85%	81%	86%	70%	13%	20%	92%
Linear Support Vector Machine (L-SVM)	72%	76%	53%	62%	45%	17%	52%	86%
Decision tree (C4.5)	76%	76%	72%	73%	52%	4.35%	31%	78%
Random Forest (RF)	84%	87%	75%	81%	70%	14%	25%	90%
Deep Auto-encoder (DAE)	97%	98%	96%	96%	94%	6%	4.35%	97%
Ensemble	97.2%	98.2%	96.2%	97%	94.2%	5.54%	4.15%	97.3%
DRNN	98%	98.8%	97.5%	97.5%	95%	4.54%	3.15%	98%

The comparison of baseline classified measures is presented in Tab. 3 without a CV and with a 10-fold CV. In addition, the anticipated model's average precision, accuracy, sensitivity and F-measure are 98.5%, 99.5%, 99% and 99%, which is substantially higher than other approaches. The average accuracy of the existing approaches is 6.5%, 11.5%, 8.5%, 24.5%, 20.5%, 12.5%, 1.5%, and 1% higher than BM, RM, FFNM, L-SVM, C4.5, RF, DAE and Ensemble classifier model. The average precision of the existing approaches is 7.5%, 9.5%, 12.5%, 21.5%, 21.5%, 10.5%, 1% and 0.7% higher than other approaches like BM, RM, FFNM, L-SVM, C4.5, RF, DAE and Ensemble classifier model. The average sensitivity of the existing approaches is 12%, 12%, 16%, 44%, 25%, 22%, 2% and 1.5% higher than BM, RM, FFNM, L-SVM, C4.5, RF, DAE and Ensemble classifier model. The average F-measure of the existing approaches are 12%, 12%, 11%, 35%, 24%, 16%, 2% and 1.5% higher than BM, RM, FFNM, L-SVM, C4.5, RF, DAE and Ensemble classifier model (See Fig. 4 and Fig. 5). The re-substitution loss depends on the learning cycle presented in Fig. 7. The consideration of k - fold CV is done in machine learning as in Fig 6. Here, 10 is fixed to the value of k. Hence, the 10 recursions are needed to perform and consider the average accuracy in the prediction state. The average prediction rate is 98.5%, which is presented in Tab. 2 accordingly.

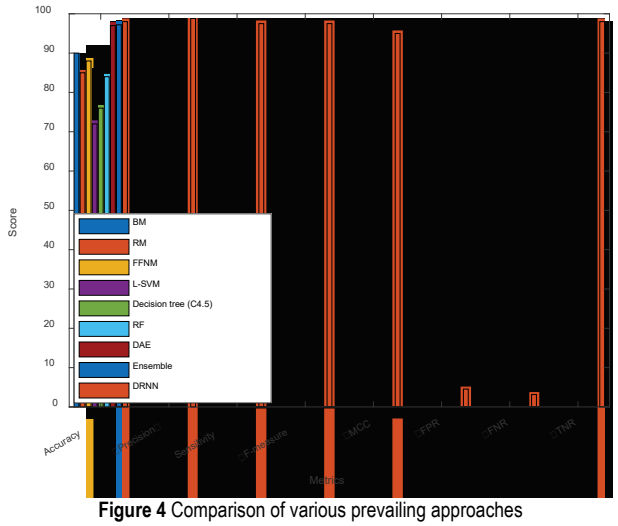


Figure 4 Comparison of various prevailing approaches

Table 4 Average value after 10-fold CV

Model	Accuracy	Precision	Sensitivity	F-measure
Bayesian model (BM)	92%	92%	87%	87%
Regression model (RM)	87%	90%	87%	87%
Feed Forward network model (FFNM)	90%	87%	83%	88%
Linear Support Vector Machine (L-SVM)	74%	78%	55%	64%
Decision tree (C4.5)	78%	78%	74%	75%
Random Forest (RF)	86%	89%	77%	83%
Deep Auto-encoder (DAE)	97%	98.5%	97%	97%
Ensemble	97.5%	98.8%	97.5%	97.5%
DRNN	98.5%	99.5%	99%	99%

The complete performance measure of the classifier is presented in Tab. 4. The baseline DRNN classifier has the complete performance that needs to be determined. The average precision of the anticipated model is 99.5%, the average sensitivity is 99%, and the average F-measure is 99% which is comparatively superior to other techniques. The p-value analysis ranges from 0.001. This analysis helps to prove that the answerable technique provides better accuracy in prediction using the under-fitting and over-fitting problems. The problems are eliminated one after the other and model across DRNN classifier gives better functionality.

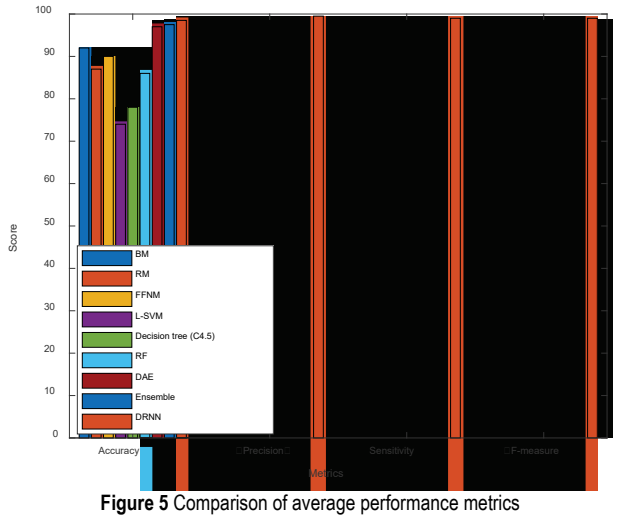


Figure 5 Comparison of average performance metrics

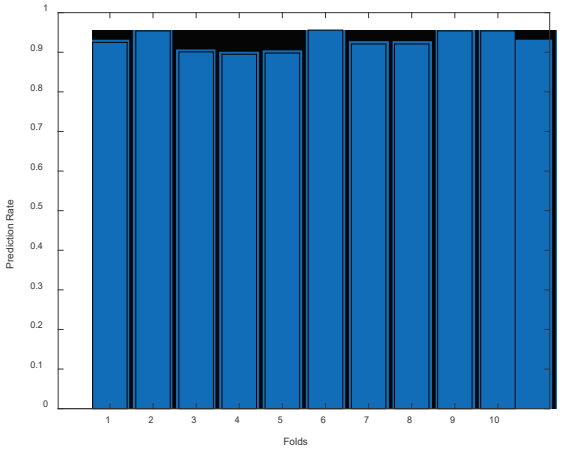


Figure 6 k-fold cross validation (k = 10)

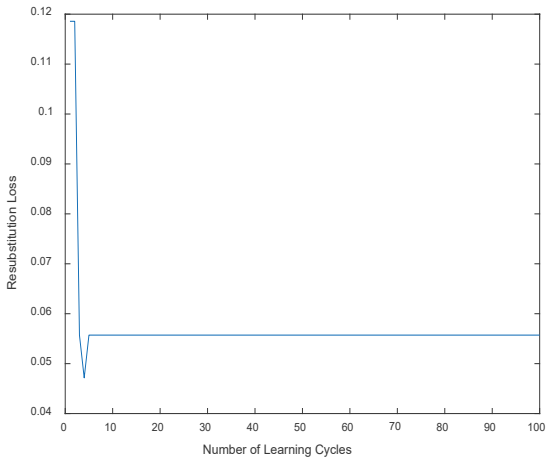


Figure 7 Learning vs. loss analysis

5 CONCLUSION

The process of predicting CVD is a novel approach presented in the studies. The proposed model uses the deep network model for identifying global feature set and is utilized for identifying the optimal configuration. In addition, three main problems are resolved with the help of predicting heart disease using the proposed DRNN. The optimal network configurations need to be identified, and the important issues are underfitting and overfitting. The anticipated model is evaluated with traditional DNN using optimization approaches. However, other network configurations are analyzed to compare the optimized network configuration outcomes. The proposed model performed well than other approaches, and the prevailing techniques are presented to the machine learning using the comparative analysis. The features selection technique improves the performance and obtains good performance over other approaches compared with the different feature selection methods. In addition, the proposed DRNN improves the prediction of heart diseases with prediction accuracy of 94% from comparative and experimental outcomes in the proposed system. Also, the clinical experts helped to generate effective decisions. It is essential for future studies that the complexity of time in the proposed cabinet approach is required to be investigated as the most important feature in the healthcare field. Also, the proposed approach is suggested to test some data sets to evaluate the technique's effectiveness with more instances.

6 REFERENCES

[1] Bharti, R., Khamparia, A., Shabaz, M., Dhiman, G., Pande, S. & Singh, P. (2021). Prediction of heart disease using a combination of machine learning and deep learning. *Comput Intell Neurosci*. https://doi.org/10.1155/2021/8387680

[2] Ramalingam, V. V., Dandapath, A. & Raja, M. K. (2018). Heart disease prediction using machine learning techniques: a survey. *International Journal of Engineering & Technology, 7*(2.8), 684-687. https://doi.org/10.14419/ijet.v7i2.8.10557

[3] Fatima, M. & Pasha, M. (2017). Survey of machine learning algorithms for disease diagnostic. *J Intell Learn Syst Appl., 9*(1), 1-16. https://doi.org/10.4236/jilsa.2017.91001

[4] Garg, A., Sharma, B. & Khan, R. (2021). Heart disease prediction using machine learning techniques. *IOP Conf. Ser.: Mater. Sci. Eng.* 1022 012046. https://doi.org/10.1088/1757-899X/1022/1/012046

[5] Raut, S., Magar, R., Memane, R. & Rupna, V. S. (2020). Heart disease prediction using machine learning. *Journal of Emerging Technologies and Innovative Research, 7*(6), 2081-2085. https://www.jetir.org/papers/JETIR2006301.pdf

[6] Vembandasamy, K., Sasipriya, R. & Deepa, E. (2015). Heart disease detection using Naive Bayes algorithm. *Int J Innov Sci Eng Technol. 2*(9), 441-444.

[7] Louridi, N., Douzi, S. & El Ouahidi, B. (2021). Machine learning-based identification of patients with a cardiovascular defect. *J Big Data, 8*, 133. https://doi.org/10.1186/s40537-021-00524-9

[8] Benjamin, E. J., Muntner, P., Alonso, A., et al. (2019). Heart Disease and Stroke Statistics–2019 Update: A Report from the American Heart Association. *Circulation, 139*(10), e56-e528. https://doi.org/10.1161/CIR.0000000000000659

[9] Ismaeel, S., Miri, A. & Chourishi, D. (2015). Using the Extreme Learning Machine (ELM) technique for heart disease diagnosis. *IEEE Canada International Humanitarian Technology Conference (IHTC 2015*), 1-3. https://doi.org/10.1109/IHTC.2015.7238043

[10] Kanika, P. & Shah, D. R. (2016). Prediction of cardiovascular diseases using support vector machine and Bayesian classification. *International Journal of Computer Applications, 156*(2). https://doi.org/10.5120/ijca2016912368

[11] Siontis, K. C., Noseworthy, P. A., Attia, Z. I., et al. (2021). Artificial intelligence-enhanced electrocardiography in cardiovascular disease management. *Nat Rev Cardiol*. https://doi.org/10.1038/s41569-020-00503-2

[12] Fathima, K. & Vimina, E. R. (2022). Heart Disease Prediction Using Deep Neural Networks: A Novel Approach. In: Raj, J. S., Palanisamy, R., Perikos, I., Shi, Y. (eds) *Intelligent Sustainable Systems. Lecture Notes in Networks and Systems, vol 213*. Springer, Singapore. https://doi.org/10.1007/978-981-16-2422-3_56

[13] Dwivedi, A. K., Imtiaz, S. A. & Villegas, E. R. (2019). Algorithms for automatic analysis and classification of heart sounds - a systematic review. *IEEE Access, 7*. https://doi.org/10.1109/ACCESS.2018.2889437

[14] Kiran, J., Debbarma, N. & Ganjala, S. (2023). Heart Disease Prediction Using Machine Learning. In: Bhateja, V., Yang, X. S., Lin, J. C. W., Das, R. (eds) *Evolution in Computational Intelligence, FICTA 2022. Smart Innovation, Systems and Technologies, vol 326*. Springer, Singapore. https://doi.org/10.1007/978-981-19-7513-4_24

[15] Ansarullah, S. I. & Kumar, P. (2019). A systematic literature review on cardiovascular disorder identification using knowledge mining and machine learning method. *International Journal of Recent Technology and Engineering, 7*(6S), 1009-1015.

[16] Nazir, S., Shahzad, S., Mahfooz, S. & Nazir, M. (2018). Fuzzy logic-based decision support system for component security evaluation. *The International Arab Journal of Information Technology, 15*(2), 224-231.

[17] Amin, M. S., Chiam, Y. K. & Varathan, K. D. (2019). Identification of significant features and data mining techniques in predicting heart disease. *Telematics and Informatics, 36*, 82-93. https://doi.org/10.1016/j.tele.2018.11.007

[18] Mienye, I. D., Sun, Y. & Wang, Z. (2020). An improved ensemble learning approach for the prediction of heart disease risk. *Informatics in Medicine Unlocked, 20*, Article ID 100402. https://doi.org/10.1016/j.imu.2020.100402

[19] Amin, M. S., Chiam, Y. K. & Varathan, K. D. (2019). Identification of significant features and data mining techniques in predicting heart disease. *Telematics and Informatics, 36*, 82-93. https://doi.org/10.1016/j.tele.2018.11.007

[20] Enriko, I. K. A., Suryanegara, M. & Gunawan, D. A. (2016). Heart Disease Prediction System Using K-Nearest Neighbor Algorithm with Simplified Patient's Health Parameters. Springer, Berlin, Germany, 59-65.

[21] Kompella, S. & Boddu, V. (2019). Neural network based intelligent system for predicting heart disease. *International Journal of Innovative Technology and Exploring Engineering, 8*. 484-487.

[22] Tarawneh, M. & Embarak, O. (2019). Hybrid approach for heart disease prediction using data mining techniques. *Acta Scientific Nutritional Health, 3*(7), 147-151. https://doi.org/10.1007/978-3-030-12839-5_41

[23] Jagtap, Malewadkar, P., Baswat, O. & Rambade, H. (2019). Heart disease prediction using machine learning. *International*

Journal of Research in Engineering, Science and Management, 2(2), 352-355.
[24] Kim, J. K. & Kang, S. (2017). Neural network-based coronary heart disease risk prediction using feature correlation analysis. *Journal of Healthcare Engineering*, Article ID 2780501. https://doi.org/10.1155/2017/2780501
[25] Yin, L. W., Gregory, P. A., Amanda, Z. & Margaux, G. (2021). Development of a novel clinical decision support system for exercise prescription among patients with multiple cardiovascular disease risk factors. *Mayo Clinic Proceedings: Innovations, Quality & Outcomes, 5*(1), 193-203. https://doi.org/10.1016/j.mayocpiqo.2020.08.005
[26] Pe, Rubini, Subasini, C., Katharine, A., Kumaresan, V., Gowdhamkumar, S. & Nithya, T. (2021). A Cardiovascular Disease Prediction using Machine Learning Algorithms. *Annals of the Romanian Society for Cell Biology, 25*(2), 904-912.
[27] Sherubha, P., Sasirekha, S. P., Manikandan, V. et al. (2020) Graph based event measurement for analyzing distributed anomalies in sensor networks. *Sādhanā, 45*, 212. https://doi.org/10.1007/s12046-020-01451-w
[28] Sherubha, P. & Mohanasundaram, N. (2019). An Efficient Network Threat Detection and Classification Method using ANP-MVPS Algorithm in Wireless Sensor Networks. *International Journal of Innovative Technology and Exploring Engineering (IJITEE), 8*(11), 1597-1606. https://doi.org/10.35940/ijitee.K3958.0981119
[29] Sherubha, P. & Mohanasundaram, N. (2020). An Efficient Intrusion Detection and Authentication Mechanism for Detecting Clone Attack in Wireless Sensor Networks. *Journal of Advanced Research in Dynamical and Control Systems (JARDCS), 11*(5), 55-68.
[30] Singh, A. & Kumar, R. (2020). *Heart disease prediction using machine learning algorithms. International Conference on Electrical & Electronics Engineering, (ICE3)*, Gorakhpur, India, 452-457. https://doi.org/10.1109/ICE348803.2020.9122958

Authors' contacts:

Mounika Valasapalli, Research Scholar
(Corresponding author)
Department of Computer Science and Engineering,
Koneru Lakshmaiah Education Foundation,
Vaddeswaram, Guntur-522302, India
mounikatkrcse@gmail.com

Nallagatla Raghavendra Sai, Associate Professor
Department of Computer Science and Engineering,
Koneru Lakshmaiah Education Foundation,
Vaddeswaram, Guntur-522302, India
nallagatlaraghavendra@kluniversity.in

ISSN 1846-6168 (Print), ISSN 1848-5588 (Online)
https://doi.org/10.31803/tg-20240510165209

Original scientific paper
Received: 2024-05-10, Accepted: 2024-05-15

Die Casting or Sheet Metal Forming: A Comparison of Car Body Manufacturing in Times of the "Giga Press"

Martin Bednarz

Abstract: Die casting as an alternative for sheet metal structures in car body design has been around for a few years. While the application of die cast body components by legacy automakers is rather limited, especially for mass market products, Tesla bases their mass-produced cars on a few extremely big die cast structure parts. The impact of this strategy makes it necessary to rethink car body manufacturing. In this study, published information and expert interviews are used to evaluate the strengths and weaknesses of the competing philosophies of car body manufacturing.

Keywords: car body manufacturing; large-scale die casting; sheet metal construction

1 INTRODUCTION

The motivation behind this work lies in the continuous development of manufacturing technologies for vehicle bodies in the automotive industry. Regarding mass production, two main technologies play a decisive role: the increasingly important large die-casting technology and the sheet metal shell construction method. As part of this research work, the aim is to make a comprehensive comparison between these two manufacturing processes.

The automotive industry faces constant challenges, ranging from increasing efficiency and cost savings to meeting stricter environmental regulations. Choosing the right manufacturing technology can have a significant impact on the overall performance and competitiveness of an automotive manufacturer. It is therefore crucial to fully understand the advantages and disadvantages of large die casting technology compared to the established sheet metal shell construction method.

2 STATE OF THE ART

2.1 Car Body Manufacturing with Sheet Metal

The sheet metal shell construction or self-supporting shell construction (see Fig. 1) is used in the automotive industry in body construction and represents the current standard for the construction of passenger cars with steel construction [1]. This is a chassis construction method in which steel or aluminium half-shells are connected and there is no longer a distinction between the chassis and the body of the vehicle. Individual areas such as the roof, floor or side walls consist of large sheet metal areas and take on load-bearing tasks as part of the body. With this construction method, it is possible to build complex 3-dimensional structures and to achieve targeted lightweight construction by integrating almost all components. The sheet metal shell design also enables a high degree of automation in the systems and is suitable for the production of high quantities at low cost. Due to the high acquisition costs of the production tools required for forming the sheet metal, customization and a high number of variants are not economically viable. [2]

Figure 1 Body with self-supporting shell construction ([3], p.18)

2.2 Car Body Manufacturing with Die Casting (Mega or Giga Casting)

The term mega-casting is a die casting process that is also referred to as large-scale casting [3], large-scale die casting or more general high performance die casting. The process has initiated a change in the automotive industry. Compared to the established die-casting processes, the main difference is the size of the systems [4].

For vehicle production, these new processes mean the possibility of reducing the complexity of joining and automation technology as well as cycle time and space requirements due to the lower number of components. However, the challenge to reproducible quality, which is due to the lack of knowledge and experience regarding the highly complex die casting process, must be addressed in the future [5]. Large-scale die casting also has an impact on the car body as a product and offers manufacturers the opportunity to further optimize the lightweight construction potential. Accordingly, it means that joints can be reduced and at the same time it is possible to integrate connection points directly into the components. This integration facilitates the connection of add-on parts. Compared to current processes, however, the body can only be adapted or modified to a limited extent. In 2022, the authors stated that problems with component quality, pore formation and the possibility of repairability still need to be solved in car body series [5].

The changes to the body are illustrated in Fig. 2 using a model of a Tesla Model Y body as an example. By using large die casting, the number of individual components can be significantly reduced. In the left-hand image, 171 components are marked, which can be replaced in the right-hand image by the two components marked in colour.

Figure 2 Change of car body parts due to large die casting a Tesla Model Y [6]

2.3 Brownfield versus Greenfield

The term Brownfield refers to an approach that provides for planning based on existing buildings and infrastructure. In this case, the planners of new industrial plants must plan with the existing structures and take these into account when implementing new technologies, production facilities or processes. The given conditions of the existing facilities, both the existing buildings and the processes envisaged in the original planning, limit flexibility and lead to compromises during implementation. Accordingly, the planners are faced with the challenge of realizing the planning objective despite various restrictions. The counterpart to the Brownfield approach is the term Greenfield. This refers to an approach that provides for planning on an open, undeveloped area. The planners of the plants have maximum freedom in the planning and design of the industrial plants, which are only limited by the capacities desired by the client and the financial specifications. This makes it possible to plan and build a factory that meets the latest standards. The freedom in the planning of production processes, material flow but also energy use makes it possible to design the facilities uncompromisingly for efficient processes without being restricted by existing structures. [7]

3 METHODOLOGY

The expert interview was conducted in a quantitative way. As a first step a target group was identified on internet professional platforms using keywords corresponding to die casting. For this sample group a questionnaire was provided. The sample size of 24 complete expert questionnaires was used to compile a quantitative analysis of the current status of large die casting versus sheet metal shell design.

The sample group was set up as shown in Fig. 3.

Most of the experts in the sample group were long time employed in the casting industry, with 8 out of 24 with over 20 years of experience in the field. Fig. 4 displays the duration of relevant employment of the interviewed.

The questionnaire for the sample group compared both car body manufacturing technologies in 14 criteria (Tab. 1).

Figure 3 Expert sample group composition

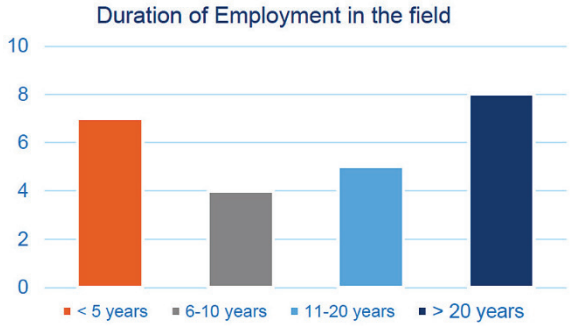


Figure 4 Expert sample group Working Experience.

Table 1 Questionnaire comparison criteria

Product Criteria	Process Criteria
Product Safety	Manual Labour Savings
Leight Weight Potential	Production Time Savings
Energy Consumption	Scalability
CO_2 Emission	Investment Costs
Repairability	Maintenance Savings
Stiffness	Suitable for E-Mobility
Quality and Tolerances	Space Savings

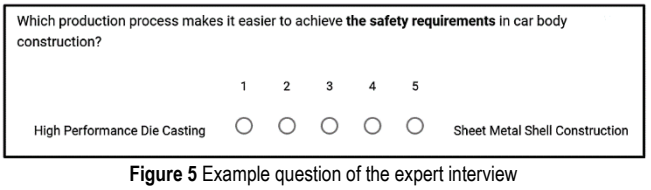


Figure 5 Example question of the expert interview

All questions were displayed as a direct comparison between Sheet metal design and die casting with five different options to answer, as shown in the example in Fig. 5. Option "1" would be strongly in favour of die casting, whereas option "5" is strongly in favour of sheet metal. "3" would be the neutral option. All comparison questions were set up in the same way, so lower numbers favour die casting and high numbers sheet metal design.

All questions were based on the premise, that we have a complete Greenfield approach, so no preliminary conditions for the comparison were defined.

4 RESULTS

The results of the expert interview show a sizeable variation within the answers of the experts, but nevertheless give a consistent picture in most of the categories.

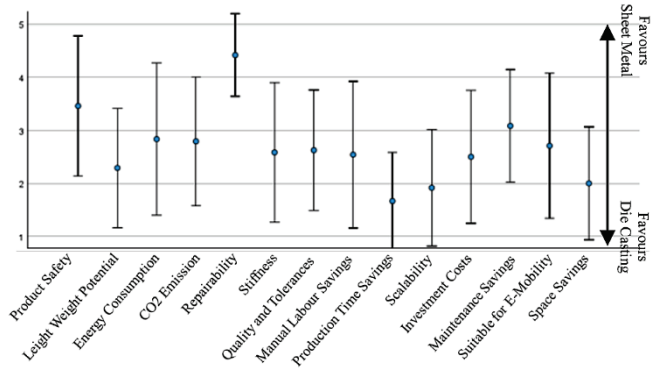


Figure 6 Results of Expert Interview Questionnaire

In 11 out of 14 categories the experts considered large-scale die casting as advantageous. While some of these criteria where almost even, six criteria were evaluated with a very clear outcome. Two indisputable advantages of the sheet metal design, according to the interviewed, were product safety and repairability.

Large-scale die casting was rated superior in light weight potential, production time savings, scalability and space savings.

4.1 Discussion

4.1.1 Product Safety

Even though crash ratings of Tesla Model Y and 3 using large die casting parts for the car body have been very good in comparison, one problem of large-scale die casting processes is the quality of the parts. At the current state of the art lunkers and other casting errors are not completely avoidable for parts with the weight of 50 kg and above.

4.1.2 Light Weight Potential

Although some experts were convinced, that the same or even better results could be achieved by sheet metal forming, most of the experts saw an advantage in die casting, with the two main reasons being the lower specific weight of aluminium compared to steel and the lack of multiple joining or bonding surfaces.

4.1.3 Energy Consumption and CO_2 Emission

Both technologies were rated pretty similar in that regard. When reviewing the specific energy consumption of the base materials sheet metal forming using steel should be the more sustainable solution, with some compensation in the life cycle because of a little higher vehicle weight. In CO_2 emission in most cases steel would also be lower and therefore preferable to aluminium, but this is highly dependable on the energy source used for aluminium production. Overall aluminium production has a path to almost Zero emissions by using exclusively renewable energy, whereas economical Steel production is still relying on coal.

4.1.4 Repairability

Building a car structure out of big casting parts has a great disadvantage when it comes to repairability. If the car structure is damaged by an accident, even just a small local deformation or crack, there is currently no comparable way to repair this, as it can be done with sheet metal, as welding is not possible without altering the material properties. As the parts are pretty much unremovable without taking apart the whole car, a damage of the car structure often time means a total loss of the vehicle.

4.1.5 Stiffness

The flexibility in design of large Casting Parts can give the structure a greater area moment of inertia. Together with less joining points this can make up for the lower Youngs Modulus of Aluminium compared to Steel. Nevertheless, this is highly depending on the design of the structure, so there is no clear advantage with any of the two technologies.

4.1.6 Quality and Tolerances

As mentioned in 4.1.1 big casting parts have issues with lunkers and other casting errors. Additionally, the dimensional accuracy of big castings is not on the same level as sheet metal parts. The reduction of joining points can compensate for that.

4.1.7 Manual Labour Savings

Both technologies offer the opportunity of an almost complete automation. Although more manual quality checks or deburring may be needed for the casting parts, the lower number of parts gives the die casting process a slight edge.

4.1.8 Production Time Savings

According to the experts the major advantage of the large-scale die casting technology is the lower time needed for the production of the car body. Although the big casting parts have a significantly higher cycle time than any sheet metal part, the reduction of parts and joining operations means a significant time saving.

4.1.9 Scalability

Various automobile manufacturers prove that a sheet metal shell design is suitable for lot sizes in the millions, so in principle this technology scales up very well. The better perceived scalability by the experts originates in the lesser parts needed for a die casting structure, so essentially scaling up the lot size is setting up more die casting machines. For a

more complex sheet metal structure changing lot size would result in more complex planning.

4.1.10 Investment Costs

Die casting machines are expensive investment goods, but press lines are as well. The small edge of the large-scale die casting process in this regard may result from the reduction of subsequent process steps for joining and assembly.

4.1.11 Maintenance Savings

In regard of maintenance sheet metal design has a slight advantage. Again, the reduction of parts also reduces the number of tools, that need to be maintained, but deep drawing processes are known to be very reliable at high output with relatively low amount of maintenance needed. Die casting tools on the other hand are not only burdened with high pressures, but also high temperatures and aggressive liquid aluminium. This results in a significantly shorter time of these tools before they need to be serviced.

4.1.11 Suitable for E-Mobility

The biggest difference between an internal combustion engine car (ICE) and a battery electric vehicle (BEV) is the heavy battery pack. The design flexibility of large-scale die casting structures allows a high functional integration of the battery into the car body structure. Mechanical mounting and connections for temperature control of the battery can be directly included in the casting design.

4.1.12 Space Savings

The footprint of the high performance die casting machines is smaller than the footprints of a press line, the necessary tools and all the subsequent assembly operations. According to expert information sheet metal car body manufacturing needs approximately double the footprint of large-scale die casting.

Finally, when asked for the future potential of large-scale die casting the answers were almost unanimously in favour of the die casting process (Fig. 7).

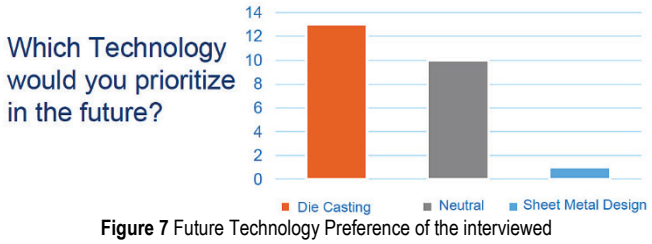


Figure 7 Future Technology Preference of the interviewed

5 SUMMARY

Although the results seem to be very one sided, they have to be taken with caution because of some important caveats. The first is the target group of the expert interview, which was consisting of experts in the field of die casting, so a bias towards die casting technology is very likely in the sample group. The second caveat is, that the two criteria product safety and repairability are very important to the customers and cannot easily be outweighed by other advantages. The last and probably most important caveat is, that all experts were viewing the comparison based on a Greenfield approach without any staff or infrastructure present. Based on that premise large-scale die casting might be the best solution at this point in time for electric vehicles. Legacy automakers rarely have the opportunity of setting up a plant on a green field and have to adapt their existing infrastructure for future projects. In this light a dismissal of conventional sheet metal shell design is absolutely premature. Concluding, it can be said, that large-scale die casting is a powerful new alternative to car body manufacturing and offers some interesting potential especially for Battery Electric Vehicles, but is not a best fit solution for all automakers or all future projects.

6 REFERENCES

[1] Horst, E. F. (Hrsg.) (2017). *Leichtbau in der Fahrzeugtechnik*, 2. Auflage. Wies-baden/Heidelberg: Springer Vieweg. (in German)

[2] Fang, X. (2023). *Karosserieentwicklung und -Leichtbau: Eine ganzheitliche Betrachtung von Design über Konzept- und Materialauswahlprinzipien bis zur Auslegung und Fertigung*. Berlin/Heidelberg: Springer Vieweg. (in German)

[3] Kühling, S. (2022). *In der Gießereiwelt einmalig: VW-Mitarbeiter entwickeln neues Elektro-Auto-Teil*. https://www.hna.de/lokales/kreis-kassel/baunatal-ort312516/ent-wickelt-baunataler-vw-mitarbeiter-haben-teil-fuer-neues-auto-91550655.html (in German) (Accessed on 2024-01-08)

[4] Siebel, T. (2022). Wie sich das Mega-Casting auf die Zulieferer auswirkt. https://www.springerprofessional.de/urformen/automobilproduktion/wie-sich-das-mega-casting-auf-die-zulieferer-auswirkt/23776976 (in German) (Accessed on 2023-10-13)

[5] Schuh, G. et al. (2022). Opportunities and Risks of Mega-Casting in Automotive Production – The Aluminum Die-Casted Body in White. *WT Werkstattstechnik, 112*(9). https://doi.org/10.37544/1436-4980-2022-09-52

[6] N. N. (2023). https://www.electrifiedmag.com/news/get-to-know-megacasting-hype-or-the-next-big-thing/

[7] Imgrund, C. (2023). *Neuplanung einer energieeffizienten (Automobil-)Fabrik*. https://www.ingenics.com/de/blog/neuplanung-einer-energieeffizienten-automobil-fabrik/ (in German)

Author's contacts:

Martin Bednarz, Prof.-Dr.-Ing.
Technische Hochschule Ingolstadt,
Esplanade 10, D-85049 Ingolstadt, Germany
Tel +49 (0) 841 / 9348-3508
martin.bednarz@thi.de

ISSN 1846-6168 (Print), ISSN 1848-5588 (Online)
https://doi.org/10.31803/tg-20240514144703

Original scientific paper
Received: 2024-05-14, Accepted: 2024-05-20

A Comparative Evaluation of Augmented Reality Indoor Navigation versus Conventional Approaches

Tomáš Macháč*, Petr Hořejší, Michal Šimon, Andrea Šimerová

Abstract: In the context of Industry 4.0, augmented reality (AR) is becoming a key element for innovations in the field of lifetime usability of products and optimisation of production processes. Its integrated use outdoors via GPS and Simultaneous Location and Mapping (SLAM) indoors with a focus on inertial technologies enables precise positioning and efficient navigation in industrial environments. The aim of this study is to test the user-friendliness of the proposed indoor navigation system using inertial sensors. The AR application provides indoor navigation between defined points of interest representing key locations. ARCore and ARFoundation were used for development. Participants are tested using three different approaches to orientation in the environment. Their task is to evaluate the user-friendliness and compare the use of a smartphone with an ARNAV app, 2DMAP with visual elements, and variants without wayfinding support "BLIND". User acceptance is validated using a standardized SUS questionnaire. The evaluation results show similar perceptions of the system among most respondents, except for the need for technical support and the presence of inconsistencies, where the ARNAV group showed different opinions, indicating the need for optimisation in these specific areas to improve the user experience. In line with other studies the results suggest that despite advances in augmented reality, there is still a lack of added value, especially for inexperienced users. Despite the number of studies available, user testing of the proposed systems remains poorly carried out, with the emphasis more on functional verification. The results of the study suggest that the low variety and simplicity of tasks in the studied environment limits significant differences in the perception of AR navigation. Future research should focus on the use of augmented reality in complex environments and in solving more complex tasks.

Keywords: Augmented Reality; Indoor Navigation; Industry 4.0; Logistics; User Acceptance

1 INTRODUCTION

Industry 4.0 uses cyber-physical system (CPS) technology to create a platform of cyber-physical production systems (CPPS). This platform is also used to connect the virtual space with the real world in order to increase the intelligence of the equipment in a smart factory [1]. To achieve the fundamental design principles of digital transformation in Industry 4.0, the implementation and integration of diverse information, digital and operational technologies (IDOT) is essential. This integration ranges from simple to advanced technological elements such as industrial sensors, controllers and automated guided vehicles [2]. Within the Industry 4.0 concept, there is an opportunity to interact with technology paradigms that include the Internet of Things (IoT), artificial intelligence (AI), robotics, big data processing, and cloud storage. In addition to these aspects, other innovative technologies, namely virtual and augmented reality, are also being used [3]. Fraga-Lamas et al. [4] describe areas of industry that are suitable for the application of augmented reality systems - manufacturing, service or training.

Augmented Reality (AR) is a technology that provides the user with a means to present a polysynthetic world. Azuma [5] also presents specific application perspectives of augmented reality in the context of situations where it can be effectively used, for example, through navigation maps or visualisation of the current and planned environment. Augmented reality technology relies on different methodological approaches. One of them is markerless AR, which is a form of augmented reality that allows the placement of virtual objects without the need for a reference node. Specifically, this is accomplished by simultaneously displaying a virtual object associated with a precise location obtained through technologies such as GPS and a digital compass. Marker-based AR relies on real-time recognition of images, eyes and other real-world objects to provide digital content to the user [6, 7].

Research oriented towards navigation in the physical environment can be broadly classified into three main groups: external navigation, internal navigation and their integration [8, 9].

Navigation solutions, such as the Google Maps application, rely on information from GPS, aerial imagery and satellite imagery, but with a primary focus on the external environment [10]. Alternatively, an integrated solution can be implemented that combines technologies such as GNSS and LiDAR or SLAM [11, 12]. In outdoor navigation, we exclusively encounter the application of GPS technology [13-15].

For today's large-scale structures such as organisations, universities, business centres and logistics warehouses, navigation is an increasingly complex challenge. An effective solution to this challenge lies in an integrated approach that harmonises external and internal navigation. This integration brings significant benefits in optimising time and facilitating navigation in complex environments [16, 17].

1. 1 Related Work in Augmented Reality Indoor Navigation

De Souza Cardoso et al. [18] identified the prevalent augmented reality applications in industrial environments that are primarily focused on manual assembly processes. These applications provide users with work instructions to perform specific tasks. At the other end of the spectrum is the logistics domain. Rejeb et al. [19] identified limitations to the benefits of augmented reality technology in the area of logistics and supply chain management. They emphasise the

need to support human-centred designs for effective and successful user experiences with AR. Trebuňa et al. [20] describe the issue of truck logistics analysis using GPS location technology.

For indoor navigation solutions, there is a wide range of technologies that are used to pinpoint the user's location in confined spaces. These technologies include Bluetooth, radio frequency identification (RFID), Wi-Fi, magnetic navigation and even LED light technology [21-24]. One frequently used approach is the use of Inertial Measurement Unit (IMU) sensors. These use a magnetometer, gyroscope, and accelerometer, which together enable tracking of the user's movement and orientation in space [13, 17].

Bajpai and Amir-Mohammadian [26] introduced an indoor navigation system that operates without the need for physical signs and uses ARKit technology for the iOS platform. This system works with a server that stores and retrieves 3D map data. An open-source visual SLAM framework called OpenVSLAM is used to implement the 3D map reconstruction.

Romli et al. [27] focused on the development of an AR application with an emphasis on the urbanisation domain. A library was chosen as the demonstration environment for the AR prototype. The Vuforia SDK technology was used for specific scene identification and navigation in the space, which allows setting image markers. The marker system was also introduced by Sato [28], but his solution was extended to include BLE transmitters in order to minimise the energy burden of the mobile device.

The idea of creating augmented reality-enabled 2D and 3D indoor maps has been addressed by Cankiri et al. [29]. They presented a system design that incorporates 3D humanoid agents for realistic navigation scenarios created in an AR environment and integrates IMU and SLAM sensors. Other research is focused on the implementation of a 3D point cloud through Google Glass devices. This study concludes that computer vision technologies are an effective alternative to other methods of simultaneous localisation and mapping [30].

Yao et al. [31] developed a system that relies on the A* algorithm for optimised route planning based on a 2D map. The optimal route is then passed to the navigation module in the form of a sequence of nodes. This navigation module uses AR based on location-based services (LBS). For comparison, they also conducted experiments with a Wi-Fi system, where this approach was found to be significantly less accurate with respect to the real location, in the range of 3 to 8 metres.

Wang [32] presents the concept of an iBeacon positioning mechanism that features the ability to flexibly adapt appropriate detection values for different iBeacon devices at variable distances. The goal of this approach is to provide stable indoor localisation. At the same time, it is intertwined with the AR display functionality without the use of markers.

Extended support in robotics is addressed by Kapinus et al. [33], who present a framework called ARCOR2. This framework focuses on the efficient management of robotic workstations using AR. The authors approach this problem with a methodological foundation, which is paralleled in our paper where we investigate user-friendliness.

1. 2 Hypothesis Setting and Research Questions

We formulated the questions based on the standardised System Usability Scale (SUS) questionnaire [34], which is used later in the paper. For each question we established a null hypothesis (H0) and an alternative hypothesis (H1). H0 assumes that there are no significant differences between respondents in their perception of a particular aspect of the system, and H1 assumes that there is a significant difference between at least two groups of respondents.

Q1: There are significant differences between groups in perceptions of increased use of the system.
Q2: There are significant differences between groups in perceptions of unnecessary complexity of the system.
Q3: There are significant differences between groups in perceptions of the ease of use of the system.
Q4: There are significant differences between groups in the perception of the help needed from a technical support person to use the system.
Q5: There are significant differences between the groups in the perception of good integration of the different functions of the system.
Q6: There are significant differences between groups in perceptions of inconsistencies in the system.
Q7: There are significant differences between groups in perceptions of very quick learning to use by most people.
Q8: There are significant differences between groups in perceptions of the great difficulty of using the system.
Q9: There are significant differences between groups in perceptions of great confidence in using the system.
Q10: There are significant differences between groups in perceptions of the need to learn new knowledge before using the system.

Here we set the research question RQ1: Which of the proposed options is the most favourable in terms of SUS score?

In the introduction of the article, we delve into the elements of Industry 4.0 and their integration with augmented reality, specifically focusing on navigation in logistical processes. Stemming from the main AR panel, we formulate hypotheses and research questions. Subsequently, we describe the methodology, proposing three navigation variants, including an AR application with a graphical user interface and system functionality. We present the results through processed SUS questionnaires, focusing on key aspects of these questionnaires. In the discussion, we compare our findings with similar experimental results from other authors.

2 METHODOLOGY

The aim of this study is to conduct an experimental investigation aimed at comparing the user-friendliness of different forms of indoor navigation support. Specifically, this is in a laboratory environment where a fully illuminated corridor serves as a representation of the space. This area contains individual offices that represent points of interest

involved in a simulated logistics process. A secondary output of this study is also the evaluation of the proposed algorithm operating at the augmented reality level, whereby the algorithm simultaneously operates in a markerless environment. The experiment uses standardised AR packages (described later), which were selected based on a search in the internal environment. We deliberately dropped the use of GPS, which is not relevant in indoor environments. Among the frequently discussed applications are technologies working with inertial measurement units. One of the advantages of this technology is the possibility to operate without the need for additional external tools, moreover, the required technologies are present in common smartphones.

A smartphone serves as the primary device using AR technology in this study. This decision is based on the fact that despite the widespread use of head-mounted display AR glasses, the deployment of AR glasses is still limited, especially due to the high cost of the device (for example, Microsoft HoloLens 2 costs $3,500 - March 2024) [35]. In parallel, we still observe a lack of progress in software capabilities for these devices. We experimentally compare variants using 2D maps, where we are inspired by the concept of visual management [36]. Another source of inspiration for our experiment is the 5S system, which works with colour variants to facilitate spatial understanding and visual management [37]. The purpose of this comparison is to evaluate how technologically advanced orientation support will contribute to user-friendliness. To achieve this idea, it is necessary to explore technological differences, with the most logical approach being to analyse the 2D map aspect of the traditional view.

The application testing was conducted on a Xiaomi Redmi Note 11 Pro mobile device, which has ARCore Depth API support. The characteristics of this test device can be classified as representative of lower-end mobile phones. The operating system of the mobile phone used for testing is Android 11, with a display refresh rate of 120 Hz. The main camera has a resolution of 108 megapixels, and the battery capacity reaches 5000 mAh.

Three variants are analysed in the framework of the indoor navigation analysis:

1) Absence of navigation support to assess the natural orientation in space - BLIND.
2) Navigation using a two-dimensional (2D) map as part of a visual management concept - 2DMAP.
3) Navigation through an augmented reality (AR) application on a smartphone - ARNAV.

The user experience was retrospectively evaluated using the System Usability Scale (SUS) questionnaire [34], which is the most widely used standardised instrument for evaluating the perceived usability of the system being tested. In the context of its psychometric properties, it can be concluded that the SUS questionnaire proved to be effective in terms of reliability, validity, and sensitivity of the outputs. In this questionnaire, it should be considered that users who experience a higher level of satisfaction in the initial use of the system are likely to show a tendency to use the product more often, which will enable them to achieve a high level of skill with the product. The current trend favours the use of short scales because of their efficiency and ease of administration. The standard version of the questionnaire contains a mix of positive and negative items, with odd items worded positively and even items negatively. Respondents rate their level of agreement with each item using a five-point scale from 1 (strongly disagree) to 5 (strongly agree). The SUS total score is calculated by converting each item to a 0-4 scale, where higher values correspond to higher perceived usability. We then sum the converted scores. Finally, we multiply the sum by ∗2.5. This procedure generates scores that can range from 0 to 100, with higher scores indicating a higher level of perceived usability of the system [38]. The analysis was performed in Excel with RealStatistic using Excel add-in version 2210 installed [39].

2. 1 Description of the AR Application for the ARNAV Variant

There are specific requirements that must be met by the application. The primary purpose of the application is for the user to easily navigate using AR, which in our case involves displaying virtual navigation lines leading to points of interest. The system includes a simulation of the processes of loading and unloading objects, and these processes are displayed in the form of a task list that visualises the sequential steps. The application is optimised to work on specified frameworks. A limitation of the application is the specific environment for which it was designed. The application uses the standardised ARCore and ARFoundation libraries, which are designed for Android devices.

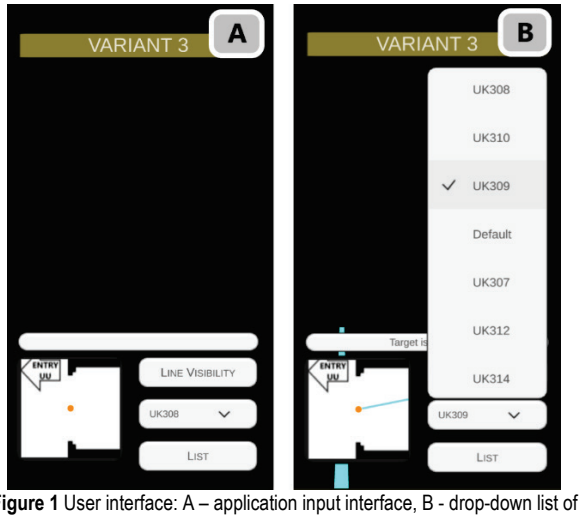


Figure 1 User interface: A – application input interface, B - drop-down list of positions.

The experimental application, the interface seen in Fig. 1, features a miniature floorplan of the defined environment. The control of the application is implemented through buttons in the graphical interface. The "LINE VISIBILITY" button displays a virtual line that navigates in space according to the generated NavMesh. To generate a line to the exact point of interest, a dropdown list is implemented, in

which the individual locations are listed. From this list we can select our preferred location. To minimise the need for additional cues during the experiment, a "LIST" button was created. Pressing it displays a list of tasks that the proband must complete during the experiment. During the implementation of the user interface, we emphasised the criteria of simplicity, clarity, and maximum user-friendliness.

2.2 Methodology for Testing Variants

Participants are thoroughly briefed on the testing procedure at the beginning of the experimental process. For the purposes of the first phase of the experiment, we chose the environment of the University of West Bohemia, considering the possible influence of the complexity of the environment and unstable lighting conditions on the results of the experiment. The experiment itself is initiated in a fairly sterile environment consisting of a flat and symmetrically lit corridor, which allows better control of any environmental effects. In this primary part of the experiment, special emphasis is placed on evaluating the results from the questionnaire that the probands complete.

This experiment represents a logistical process where the proband assumes the role of a new worker in a warehouse. The aim of the experiment is to use one of the available spatial orientation aids to facilitate and optimise the navigation process.

The proband's task is to determine the direction of the location of objects in the warehouse and then, according to a predetermined procedure, to distribute these objects around the production area. Each specific location in this context represents the location of a production machine and one central warehouse.

2.2.1 Variant 1 (Without support tools - BLIND)

This experimental variant represents a basic and direct approach to interacting with the surrounding space, but at the same time presents a more challenging spatial orientation task. The proband is provided with a list of tasks to complete. Coloured labels representing different objects are placed on a door symbolising a warehouse. The proband follows the task list, collecting the required number of labels of different colours from the door. They then place these labels on the doors that represent the production machines. The colours of the labels must match the task list. When all the labels have been distributed, the experiment is terminated. Immediately after completion, the proband conducts an evaluation using the SUS questionnaire. To simplify and speed up the response process, a QR code is placed on the wall that links to a Google Form questionnaire specifically designed for this experiment.

2.2.2 Variant 2 (Using 2D Maps - 2DMAP)

The second variant of the experiment operates with a background in the form of a 2D map (Fig. 2), which is given to the proband simultaneously with the task list. Emphasis is placed on the fact that the use of the attached map is an integrated part of the experiment and the proband is actively encouraged to use it to improve their orientation in space. Subsequently, the proband completes the same procedure as in the first variant. After completing all tasks from the list, the proband again completes the SUS questionnaire.

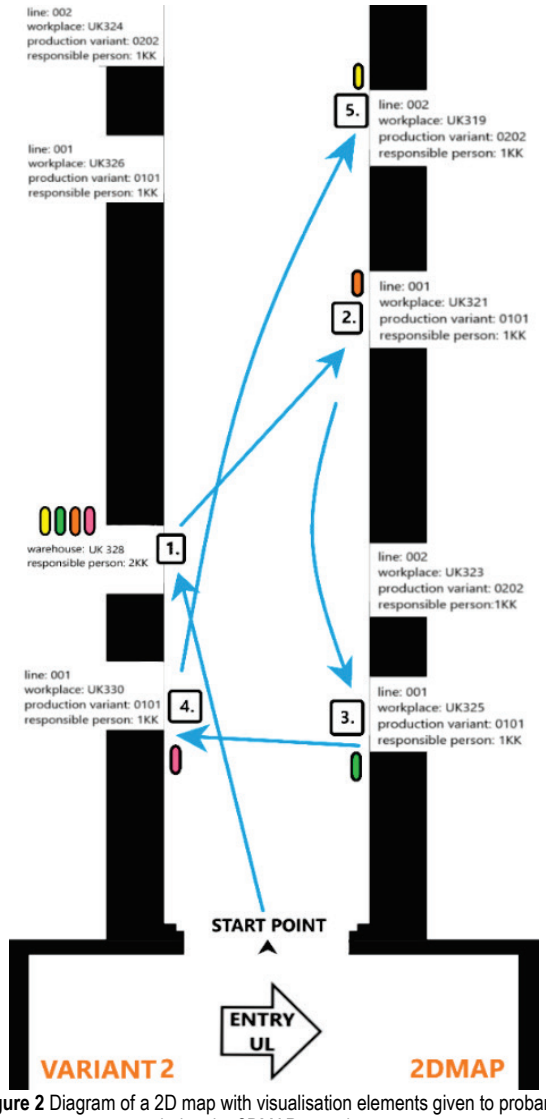


Figure 2 Diagram of a 2D map with visualisation elements given to probands during the 2DMAP experiment.

2.2.3 Variant 3 (Use of AR - ARNAV)

The experiment starts with the proband receiving a mobile phone with an app installed which is capable of visualising the surrounding space through augmented reality. The aim of the app is to display a virtual route leading to points of interest, represented by each door in the corridor. These doors represent locations in a warehouse or production area and are followed in the same way as in the previous versions of the experiment. Before running the experiment, the functions of the app are explained to the proband, and a brief description is provided of what to expect in the app and how to interact with it. As mentioned earlier, we tried to configure the augmented reality system in such a way that the

proband does not need any additional information. The information about the tasks to be performed by the proband is available directly in the augmented reality application.

Before starting the application, the proband stands in the zero position and starts the application. After quick familiarisation with the list of tasks found under the "LIST" button, they select the desired position from the drop-down list. They then click on the "LINE VISIBILITY" button and a virtual line in space appears on the mobile device's display to navigate to the selected location. Following the on-screen instructions, the proband performs an action, such as picking up or sticking a piece of paper on a door. They then check the list of tasks again and repeat the process until all the tasks are completed. An example of the process can be seen in Fig. 3. The experiment ends with the proband completing a questionnaire evaluating the system and its features.

Figure 3 Visualisation of the experimental procedure in the ARNAV variant; A - Sheet with the list of tasks, B - Representation of the navigation line, C - Placement of a coloured label at a specific position.

3 RESULTS

The evaluation works with three different variants of support systems for orientation in an unfamiliar environment, namely:

1) Variant 1: Navigation without support "BLIND" with the participation of 18 respondents.
2) Variant 2: Navigation with 2D map "2DMAP" with 25 respondents.
3) Variant 3: Navigation with augmented reality application "ARNAV" with 22 respondents.

The Kruskal-Wallis test was used for a targeted comparison of the means of the three different variants. This is a non-parametric statistical method, applicable when comparing the means of several independent groups, especially in cases where the assumptions for parametric tests are not met. These assumptions include, for example, the independence of the data (meaning different numbers of respondents coming from different groups), failure to meet the assumption of a normal distribution of the data, or the use of unmeasured metrics where only rankings and comparisons are available. Its undeniable advantage is its relative ease of interpretation and implementation, which is significant for statistical analysis without extensive assumptions [40, 41]. As part of the evaluation process, ten questions (Q1 ... Q10) were formulated that directly correspond to the standardised questions in the SUS questionnaire. The alpha level of significance was set at 0.05 when the Kruskal-Wallis test was performed. This level determines whether we can reject the null hypothesis. The result of this statistic is obtained through the p-value, which is the result of the test performed, and can be seen in Tab. 1. Furthermore, the comparison of the median values for each variant is shown in Fig. 4.

Table 1 Summary of the values of the Kruskal-Wallis test for the stated hypotheses supplemented by the evaluation of the hypotheses.

Question	H-ties	p-value	Result
1	2.519	0.284	There are no significant differences between the variants
2	0.528	0.768	There are no significant differences between the variants
3	2.119	0.347	There are no significant differences between the variants
4	8.225	0.016	Significant differences in perception were shown by the DUNN method between the 2DMAP and ARNAV groups
5	1.417	0.493	There are no significant differences between the variants
6	5.273	0.037	Significant differences in perception were shown by the DUNN method between the 2DMAP and ARNAV groups
7	2.628	0.269	There are no significant differences between the variants
8	1.869	0.393	There are no significant differences between the variants
9	0.199	0.886	There are no significant differences between the variants
10	3.577	0.031	There are significant differences between the variants

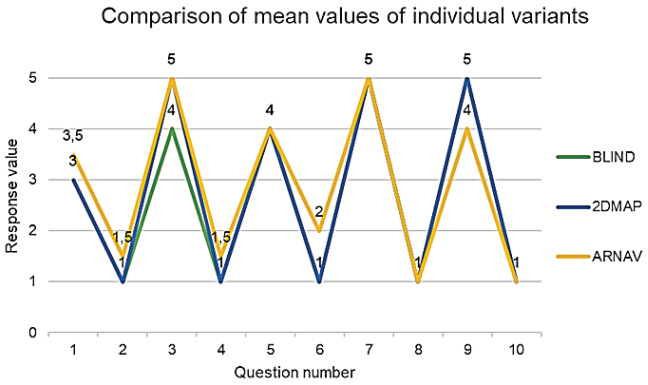


Figure 4 Comparison of mean results for each question of the SUS questionnaire.

For Q1, Q2, Q3, Q5, Q7, Q8 and Q9, no significant differences were found for any respondent groups and the **null hypotheses are not rejected**. Therefore, only Q4, Q6 and Q10, where **significant differences were found between at least two groups of respondents**, need to be examined in further detail.

Q4:
H0: *There are no significant differences between the groups in perceptions of assistance from a technical support person.*
H1: *There are significant differences between groups in perceptions of the help of a technical support person.*

The *p*-value of the test was found to be $p = 0.016$, leading to the rejection of the null hypothesis. However, the median values of the BLIND and 2DMAP groups are the same and equal to 1, while the ARNAV group is equal to 1.5. Half of the respondents from the BLIND and 2DMAP groups responded with a value of 1. That is, they strongly disagree that they would need the support of a technical person to use the system. Respondents in the ARNAV group reported responses ranging from 1 to 2, or that they did not need the support of a technical person. Significant differences in perceptions were found between the 2DMAP and ARNAV groups, with ARNAV respondents showing higher levels of agreement with the need for technical resource support than 2DMAP group respondents.

After rejecting the null hypothesis in the test of mean equality, a post hoc analysis was conducted to identify groups with statistically significant differences. The Dunnett's test was used for this analysis, which is suitable when compared groups have different sample sizes. The analysis revealed a statistically significant difference between the 2DMAP and ARNAV groups ($p = 0.004 < p = 0.016$), see Fig. 5.

Q4	DUNNETT´s TEST	alpha = 0.05
group 1	*group 2*	*p-value*
BLIND	2DMAP	0.095
BLIND	ARNAV	0.324
2DMAP	ARNAV	**0.004**

Figure 5 Comparison of Dunnett test variants for question 4

Q6:
H0: *There are no significant differences between the groups in their perception of good integration of different system functions.*
H1: *There are significant differences between groups in perceptions of good integration of different system functions.*

The *p*-value of the test was found to be $p = 0.037$, again leading to the rejection of the null hypothesis. The medians of the BLIND and 2DMAP groups are equal to 1. The median of the ARNAV group is equal to 2. Half of the respondents in the BLIND and 2DMAP groups answered 1. So they do not perceive any irregularities in the system. The third group ARNAV answered 2 or less. Significant differences in perception were found between the 2DMAP and ARNAV groups, with ARNAV respondents showing a higher level of agreement with too many irregularities in the system than 2DMAP group respondents.

Q6	DUNNETT´s TEST	alpha = 0.05
group 1	*group 2*	*p-value*
BLIND	2DMAP	0.719
BLIND	ARNAV	0.058
2DMAP	ARNAV	**0.014**

Figure 6 Comparison of Dunnett test variants for question 6

We conducted the same analysis, the Dunnett's test, after rejecting the null hypothesis for question 6 as well. We found a statistically significant difference between the 2DMAP and ARNAV groups ($p = 0.014 < p = 0.037$), as shown in Fig. 6.

Q10:
H0: *There are no significant differences between the groups in their perceptions of gaining new knowledge before using the system.*
H1: *There are significant differences between groups in perceptions of new knowledge gain before using the system.*

Despite the rejection of the null hypothesis, the same median values were found in all groups, namely the value of 1. Thus, it can be concluded that respondents do not perceive the necessity of gaining new knowledge before using the system. In each group, one half of the respondents clearly chose the answer 1. For this reason, a post-hoc analysis was performed, but no pair with a significant difference was identified.

Table 2 Summary of SUS questionnaire scores for individual variants.

	SUS score		
Respondent	BLIND	2DMAP	ARNAV
1	52.5	97.5	42.5
2	90	100	82.5
3	85	77.5	82.5
4	97.5	82.5	80
5	80	87.5	82.5
6	97.5	95	62.5
7	45	65	87.5
8	72.5	92.5	80
9	92.5	82.5	77.5
10	100	97.5	80
11	77.5	92.5	95
12	55	70	100
13	90	97.5	90
14	77.5	97.5	92.5
15	92.5	90	87.5
16	100	87.5	97.5
17	37.5	90	77.5
18	80	100	72.5
19		27.5	82.5
20		62.5	65
21		95	85
22		95	80
23		52.5	
24		82.5	
25		85	
MEAN VALUE	**79.03**	**84.10**	**81.02**
MEDIAN	**82.5**	**90**	**82.5**

Parallel to this, an evaluation of the SUS questionnaire was carried out in a standardised form, which is described in the introduction of the methodology. These outputs are summarised in Tab. 2. The higher the SUS score, the higher the level of user acceptance of a given system. This score answers our question: which of the proposed options is the most favourable in terms of SUS score? The output shows 2DMAP as the most preferred option with a value of exactly 90 for the median score. The secondary output of the evaluation of the AR navigation algorithm demonstrated the successful completion of the functionality of the whole system.

A final Kruskal-Wallis test was performed for the SUS score. Here the hypothesis was posed: There are significant differences in median SUS scores between groups. After performing the Kruskal-Wallis test, the p-value was calculated as $p = 0.301$ with an alpha value of 0.05. The hypothesis was rejected.

4 DISCUSSION

The analysis and evaluation of mobile navigation compared to traditional maps has been studied in the past. With our experimental variants, we achieved identical designs - mobile app, a paper 2D map and direct experience. The research confirmed that the directional errors of the participants in the different groups did not show significant differences from each other [42]. A similar study is designed to assess the AR experience in comparison to the map-based experience. However, it is conducted within a VR context. This analysis mainly focuses on the effect of AR on driving performance and route learning ability in the context of a virtual environment [43]. In a similar vein, they conducted an analysis of the forklift navigation process in VR within a logistics process [44]. In many aspects, we can observe the substitution of augmented reality for virtual reality environments. As VR currently has more advanced technology, it provides more significant support for the application of AR elements and their research [45].

In an outdoor environment, Mulloni et al. [46] performed an experimental comparison of an AR application with a traditional map. A significant finding of the research is that users predominantly use augmented reality near road intersections. This location appears to be crucial to effectively support accurate tracking. This result shows that despite the progress made in AR, there is still a lack of sufficient added value, especially for inexperienced users.

Conversely, Dünser et al. [47] identified that while using map backgrounds integrated into an AR application, users focused their gaze on the screen for the longest time, especially when walking. The authors state that this can be explained by the different design of the study conducted. The results of this analysis also show that users generally look at the screen less frequently with the AR interface. The authors interpret this fact to mean that this interface is primarily used for quick verification at decision points and for confirming directions. Sekhavat and Parsons [48] discuss the design and comparison of two tracking methods in outdoor augmented reality applications: location-based augmented reality (LAR) and marker-based augmented reality (MAR). The results show that the characteristics of the location-based tracking method bring the need for less time to locate points of interest; reducing the number of errors in locating specific POIs; inducing a higher perception of the quality of the user experience and promoting a higher acceptance rate to using AR technology.

QR codes are an important integration into AR navigation systems, serving not only as markers for location definition, but increasingly as key elements for identifying points of interest (PoI). In parallel with this functionality, applications are being developed to optimise cloud connectivity. The main impetus for this implementation is to minimise the energy consumption of hardware devices - smartphones or AR glasses [49-52].

Tadepalli et al. [53] presented a solution with some similarities using the A* algorithm. However, it should be noted that this is a rather simplistic approach that has not been developed, for example, in the form of experimental testing.

5 CONCLUSION

Industry 4.0 opens up the possibility of interacting with technological paradigms such as the Internet of Things, artificial intelligence, robotics, big data processing and augmented reality. Augmented reality is a technology with various applications, including navigation and environmental visualisation. In the area of navigation in physical environments, the integration of external and internal navigation is key to effectively addressing the challenges of navigating complex environments, bringing significant benefits in optimising time and facilitating navigation.

In industrial environments, augmented reality is widely applied in manual assembly processes, while in logistics it faces constraints requiring human-centric designs for optimal user experiences. In indoor navigation a variety of technologies from Bluetooth to IMU sensors enable accurate user location in confined spaces. Research in AR navigation includes innovative approaches such as the use of standardised frameworks for iOS and Android platforms or the combination of 3D map data with visual SLAM. The broad spectrum of research also includes optimised route planning using the A* algorithm and the development of positioning mechanisms such as iBeacon and RFID for stable indoor localisation and efficient integration with AR displays.

Our experiment conducted on indoor AR navigation is structured to systematically evaluate the factors affecting the effectiveness of AR technologies and the user experience. The creation of an augmented reality-enabled application based on IMU sensors for a smartphone will enable the assessment of the potential of AR navigation and the optimisation of the research results. The proposed algorithm uses standardised packages to support Android devices, namely ARCore and ARFoundation.

The experiment systematically presents the process of testing AR navigation in indoor environments, starting in a test environment for the first phase and gradually moving to an industrial setting. Participants are introduced to the experimental process and detailed tasks, with each variant of the experiment involving specific means of navigation. The experimental conditions are carefully designed to maintain symmetrical results and obtain quality feedback from participants. Each variant of the experiment simulates a logistics process with the proband in the role of a warehouse worker. The experimental variants include different approaches to interacting with the environment, including the use of a 2D map and a mobile app with AR navigation, as well as a variant without any orientation support.

During the evaluation of the system, no statistically significant differences were found between the three groups

of respondents on most of the questions investigated. Questions related to frequent use of the system, its complexity, ease of use, correct integration, presence of inconsistencies, speed of learning and confidence in use did not show statistically significant differences between the different variants. It can be concluded that the respondents from the total sample had similar views on different aspects of the system.

However, significant differences in the perception of different aspects of the system emerged on the issue of the need for technical support. The 2DMAP group, which required a lower level of technical support, was statistically different from the ARNAV group, which required a higher level of technical support. Respondents in the ARNAV group showed greater agreement that they would need the support of a technical person. Similarly, there were significant differences for the question regarding the presence of inconsistencies in the system, where respondents in the ARNAV group were more likely to agree that there were too many inconsistencies in the system than respondents in the 2DMAP group.

Overall, while most respondents' perceptions of the system were similar, it is important to focus on specific areas where significant differences emerged between the groups. These findings may be key to further optimising the system and improving the user experience. It is significant to note that the initially expected negative evaluation of the system by the BLIND and 2DMAP groups was not confirmed. Nevertheless, the ARNAV group was not statistically different in its perception of the system.

The cause of these results is attributed to the low level of task variety and complexity, as well as the specifics of the environment. This fact can be considered a limitation of this study. Within a low-diversity and non-complex environment, the chosen technology is not a decisive factor, and the user experience is very similar. A key question for further research is to what extent the use of augmented reality support is significant in complex environments and when dealing with complex tasks. This step is the subject of future research.

Some studies suggest that experiments with mobile apps, paper maps and direct experience have confirmed that directional errors between these methods do not show significant differences. Other studies have compared the use of an AR app with a traditional map in outdoor environments, with proximity to road intersections being a key area for effective support. The results also suggest that despite the successes in AR, users still prefer quick verification at decision points and confirmation of directions.

Integrating QR codes into AR navigation systems and optimising cloud connectivity are also identified as key elements for the effective use of these technologies. Although advanced algorithms such as A* exist, the need for further experimental testing to fully evaluate their effectiveness in real navigation conditions should be emphasised. Overall, there is a dynamic and diverse approach to the development of navigation technologies, especially in conjunction with augmented reality.

Acknowledgements

This work was supported by the Internal Science Foundation of the University of West Bohemia under Grant SGS-2023-025 ‘Environmentally sustainable production’.

Institutional Review Board Statement

The study was approved by the internal university ethics committee (ZCU 002459/2024).

6 REFERENCES

[1] Zhou, K., Liu, T. & Zhou, L. (2015). Industry 4.0: Towards future industrial opportunities and challenges. In *The 12th IEEE International Conference on Fuzzy Systems and Knowledge Discovery (FSKD)*, Zhangjiajie, China, 2147-2152. https://doi.org/10.1109/FSKD.2015.7382284

[2] Ghobakhloo, M. (2020). Industry 4.0, digitization, and opportunities for sustainability. *Journal of Cleaner Production, 252*, p. 119869, https://doi.org/10.1016/j.jclepro.2019.119869

[3] Hofmann, E. & Rüsch, M. (2017). Industry 4.0 and the current status as well as future prospects on logistics. *Computers in Industry, 89*, 23-34. https://doi.org/10.1016/j.compind.2017.04.002

[4] Fraga-Lamas, P., Fernandez-Carames, T. M., Blanco-Novoa, O. & Vilar-Montesinos, M. A. (2018). A Review on Industrial Augmented Reality Systems for the Industry 4.0 Shipyard. *IEEE Access, 6*, 13358-13375. https://doi.org/10.1109/ACCESS.2018.2808326

[5] Azuma, R. T. (1997). A Survey of Augmented Reality. *Presence: Teleoperators & Virtual Environments*, 6(4), 355-385. https://doi.org/10.1162/pres.1997.6.4.355

[6] Jomsri, P. (2018). Implementing Virtual 3D Model and Augmented Reality Navigation for Library in University. *International Journal of Modeling and Optimization (IJMO), 8*(6), 315-317. https://doi.org/10.7763/IJMO.2018.V8.671

[7] Fleck, S., Hachet, M. & Bastien, J. M. C. (2015). Marker-based augmented reality: instructional-design to improve children interactions with astronomical concepts. In *Proceedings of the 14th International Conference on Interaction Design and Children*, Boston Massachusetts: ACM, 21-28. https://doi.org/10.1145/2771839.2771842

[8] Iftikhar, H., Asghar, S. & Luximon, Y. (2020). The efficacy of campus wayfinding signage: a comparative study from Hong Kong and Pakistan. *Facilities, 38*(11/12), 871-892. https://doi.org/10.1108/F-04-2020-0035

[9] Munion, A. K., Stefanucci, J. K., Rovira, E., Squire, P. & Hendricks, M. (2019). Gender differences in spatial navigation: Characterizing wayfinding behaviors. *Psychon Bull Rev, 26*(6), 1933-1940. https://doi.org/10.3758/s13423-019-01659-w

[10] Martinez-Cruz, S., Morales-Hernandez, L. A., Perez-Soto, G. I., Benitez-Rangel, J. P. & Camarillo-Gomez, K. A. (2021). An Outdoor Navigation Assistance System for Visually Impaired People in Public Transportation. *IEEE Access, 9*, 130767-130777. https://doi.org/10.1109/ACCESS.2021.3111544

[11] Ren, G., Ai, C., Xu, Q., Wang, Z., Wang, Z. & Geng, D. (2020). Research on Indoor and Outdoor Navigation Technology Based on the Combination of Differential GNSS and Lidar SLAM. In *The IEEE International Conference on Real-time Computing and Robotics (RCAR)*, Asahikawa, Japan, 134-139. https://doi.org/10.1109/RCAR49640.2020.9303252

[12] Aoki, R., Tanaka, H., Izumi, K. & Tsujimura, T. (2018). Self-Position Estimation based on Road Sign using Augmented

Reality Technology. *The 12th IEEE France-Japan and the 10th Europe-Asia Congress on Mechatronics*, 39-42. https://doi.org/10.1109/MECATRONICS.2018.8495815

[13] Chung, C. O., He, Y. & Jung, H. K. (2016). Augmented Reality Navigation System on Android. *International Journal of Electrical and Computer Engineering (IJECE), 6*(1), p. 406. https://doi.org/10.11591/ijece.v6i1.pp406-412

[14] Rao, J., Qiao, Y., Ren, F., Wang, J. & Du, Q. (2017). A Mobile Outdoor Augmented Reality Method Combining Deep Learning Object Detection and Spatial Relationships for Geovisualization. *Sensors*, 17(9), p. 1951. https://doi.org/10.3390/s17091951

[15] Lin, C.-H., Chung, Y., Chou, B.-Y., Chen, H.-Y. & Tsai, C.-Y. (2018). A novel campus navigation APP with augmented reality and deep learning. *The IEEE International Conference on Applied System Invention (ICASI)*, Chiba, 1075-1077. https://doi.org/10.1109/ICASI.2018.8394464

[16] Vanclooster, A. & De Maeyer, P. (2012). Combining Indoor and Outdoor Navigation: The Current Approach of Route Planners. In *Advances in Location-Based Services*, Gartner, G. & Ortag, F. Eds., in Lecture Notes in Geoinformation and Cartography. Berlin, Heidelberg: Springer Berlin Heidelberg, 283-303. https://doi.org/10.1007/978-3-642-24198-7_19

[17] Yang, J., Kang, Z., Zeng, L., Hope Akwensi, P. & Sester, M. (2021). Semantics-guided reconstruction of indoor navigation elements from 3D colorized points. *ISPRS Journal of Photogrammetry and Remote Sensing, 173*, 238-261. https://doi.org/10.1016/j.isprsjprs.2021.01.013

[18] De Souza Cardoso, L. F., Mariano, F. C. M. Q. & Zorzal, E. R. (2020). A survey of industrial augmented reality. *Computers & Industrial Engineering, 139*, p. 106159. https://doi.org/10.1016/j.cie.2019.106159

[19] Rejeb, A., Keogh, J. G., Leong, G. K. & Treiblmaier, H. (2021). Potentials and challenges of augmented reality smart glasses in logistics and supply chain management: a systematic literature review. *International Journal of Production Research, 59*(12), 3747-3776. https://doi.org/10.1080/00207543.2021.1876942

[20] Trebuňa, P., Mizerák, M., Pekarčíková, M., Kliment, M. & Matiscsák, M. (2023). Tracking of Trucks Using the GPS System for the Purpose of Logistics Analysis. *Advances in Design, Simulation and Manufacturing VI*, Ivanov, V., Trojanowska, J., Pavlenko, I., Rauch, E. & Piteľ, J., Eds., in *Lecture Notes in Mechanical Engineering*. Cham: Springer Nature Switzerland, 162-171. https://doi.org/10.1007/978-3-031-32767-4_16

[21] Torres-Sospedra, J. et al. (2015). Enhancing integrated indoor/outdoor mobility in a smart campus. *International Journal of Geographical Information Science, 29*(11), 1955-1968. https://doi.org/10.1080/13658816.2015.1049541

[22] Kunhoth, J., Karkar, A., Al-Maadeed, S. & Al-Attiyah, A. (2019). Comparative analysis of computer-vision and BLE technology based indoor navigation systems for people with visual impairments. *Int J Health Geogr, 18*(1), p. 29. https://doi.org/10.1186/s12942-019-0193-9

[23] Zhang, H. et al. (2022). Visual Indoor Navigation Using Mobile Augmented Reality. *Advances in Computer Graphics, vol. 13443*, Magnenat-Thalmann, N., Zhang, J., Kim, J., Papagiannakis, G., Sheng, B., Thalmann, D. & Gavrilova, M., Eds., in *Lecture Notes in Computer Science, vol. 13443*. Cham: Springer Nature Switzerland, 145-156. https://doi.org/10.1007/978-3-031-23473-6_12

[24] Gualda, D. et al. (2021). LOCATE-US: Indoor Positioning for Mobile Devices Using Encoded Ultrasonic Signals, Inertial Sensors and Graph-Matching. *Sensors*, 21(6), p. 1950. https://doi.org/10.3390/s21061950

[25] Mulloni, A., Seichter, H. & Schmalstieg, D. (2011). Handheld augmented reality indoor navigation with activity-based instructions. *Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services*, Stockholm Sweden: ACM, 211-220. https://doi.org/10.1145/2037373.2037406

[26] Bajpai, A. & Amir-Mohammadian, S. (2021). Towards an Indoor Navigation System Using Monocular Visual SLAM. In *The 45th IEEE Annual Computers, Software, and Applications Conference (COMPSAC)*, Madrid, Spain, 520-525. https://doi.org/10.1109/COMPSAC51774.2021.00077

[27] Romli, R., Razali, A. F., Ghazali, N. H., Hanin, N. A. & Ibrahim, S. Z. (2020). Mobile Augmented Reality (AR) Marker-based for Indoor Library Navigation. *IOP Conf. Ser.: Mater. Sci. Eng.*, 767(1), p. 012062. https://doi.org/10.1088/1757-899X/767/1/012062

[28] Sato, F. (2018). Indoor Navigation System Based on Augmented Reality Markers. *Innovative Mobile and Internet Services in Ubiquitous Computing, vol. 612*, Barolli, L. & Enokido, T., Eds., in *Advances in Intelligent Systems and Computing, vol. 612*. Cham: Springer International Publishing, 266-274. https://doi.org/10.1007/978-3-319-61542-4_25

[29] Cankiri, Z. T., Marasli, E. E., Akturk, S., Sonlu, S. & Gudukbay, U. (2020). Guido: Augmented Reality for Indoor Navigation Using Commodity Hardware. *The 24th IEEE International Conference Information Visualisation (IV)*, Melbourne, Australia, 708-713. https://doi.org/10.1109/IV51561.2020.00123

[30] Rehman. U. & Cao, S. (2015). Augmented Reality-Based Indoor Navigation Using Google Glass as a Wearable Head-Mounted Display. *The IEEE International Conference on Systems, Man, and Cybernetics*, Kowloon, 1452-1457. https://doi.org/10.1109/SMC.2015.257

[31] Yao, D., Park, D.-W., An, S.-O. & Kim, S. K. (2019). Development of Augmented Reality Indoor Navigation System based on Enhanced A* Algorithm. *KSII TIIS, 13*(9). https://doi.org/10.3837/tiis.2019.09.016

[32] Wang, C.-S. (2019). An AR mobile navigation system integrating indoor positioning and content recommendation services. *World Wide Web*, 22(3), 1241-1262. https://doi.org/10.1007/s11280-018-0580-3

[33] Kapinus, M., Materna, Z., Bambušek, D., Beran, V. & Smrž, P. (2023). ARCOR2: Framework for Collaborative End-User Management of Industrial Robotic Workplaces using Augmented Reality. https://doi.org/10.48550/arXiv.2306.08464

[34] Lewis, J. R. & Sauro, J. (2009). The Factor Structure of the System Usability Scale. *Human Centered Design*, Kurosu, M., Ed., in *Lecture Notes in Computer Science*. Berlin, Heidelberg: Springer, 94-103. https://doi.org/10.1007/978-3-642-02806-9_12

[35] HoloLens 2—Pricing and Options | Microsoft HoloLens. Accessed: Feb. 22, 2024. Available: https://www.microsoft.com/en-us/hololens/buy

[36] El Manti, S. and El Abbadi, L. (2022). Integration of Visual Management in the Industry 4.0: Case Study. *The 2nd IEEE International Conference on Innovative Research in Applied Science, Engineering and Technology (IRASET)*, Meknes, Morocco, 1-4. https://doi.org/10.1109/IRASET52964.2022.9738102

[37] Senthil Kumar, K. M., Akila, K., Arun, K. K., Prabhu, S. & Selvakumar, C. (2022). Implementation of 5S practices in a small scale manufacturing industries. *Materials Today: Proceedings, 62*, 1913-1916. https://doi.org/10.1016/j.matpr.2022.01.402

[38] Borsci, S., Federici, S., Bacci, S., Gnaldi, M. & Bartolucci, F. (2015). Assessing User Satisfaction in the Era of User

Experience: Comparison of the SUS, UMUX, and UMUX-LITE as a Function of Product Experience. *International Journal of Human-Computer Interaction, 31*(8), 484-495. https://doi.org/10.1080/10447318.2015.1064648
[39] Home Page (Welcome) | Real Statistics Using Excel. Accessed: Mar. 05, 2024. Available: https://real-statistics.com/
[40] Svoboda, M., Gangur, M. & Mičudová, K. (2019). *Statistické zpracování dat*, 1. vydání. Plzeň: Západočeská univerzita v Plzni. (in Poland)
[41] Rosenstein, L. D. (2019). *Research design and analysis: a primer for the non-statistician*. Hoboken, NJ: John Wiley & Sons, Inc. https://doi.org/10.1002/9781119563600
[42] Ishikawa, T., Fujiwara, H., Imai, O. & Okabe, A. (2008). Wayfinding with a GPS-based mobile navigation system: A comparison with maps and direct experience. *Journal of Environmental Psychology, 28*(1), 74-82. https://doi.org/10.1016/j.jenvp.2007.09.002
[43] Yount, Z. F., Kass, S. J. & Arruda, J. E. (2022). Route learning with augmented reality navigation aids. *Transportation Research Part F: Traffic Psychology and Behaviour, 88*, 132-140. https://doi.org/10.1016/j.trf.2022.05.019
[44] Pereira, A., Lee, G. A., Almeida, E. & Billinghurst, M. (2016). A Study in Virtual Navigation Cues for Forklift Operators. *The XVIII IEEE Symposium on Virtual and Augmented Reality (SVR)*, Gramado, Brazil, 95-99. https://doi.org/10.1109/SVR.2016.25
[45] Medeiros, D., Wilson, G., Mcgill, M. & Brewster, S. A. (2023). The Benefits of Passive Haptics and Perceptual Manipulation for Extended Reality Interactions in Constrained Passenger Spaces. *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, Hamburg Germany: ACM, 1-19. https://doi.org/10.1145/3544548.3581079
[46] Mulloni, A., Seichter, H. & Schmalstieg, D. (2011). User experiences with augmented reality aided navigation on phones. *The 10th IEEE International Symposium on Mixed and Augmented Reality*, Basel, 229-230. https://doi.org/10.1109/ISMAR.2011.6092390
[47] Dünser, A., Billinghurst, M., Wen, J., Lehtinen, V. & Nurminen, A. (2012). Exploring the use of handheld AR for outdoor navigation. *Computers & Graphics, 36*(8), 1084-1095. https://doi.org/10.1016/j.cag.2012.10.001
[48] Sekhavat, Y. A. & Parsons, J. (2018). The effect of tracking technique on the quality of user experience for augmented reality mobile navigation. *Multimed Tools Appl, 77*(10), 11635-11668. https://doi.org/10.1007/s11042-017-4810-y
[49] Yeh, H.-T., Chen, B.-C., Yang, C.-T. & Weng, P.-L. (2018). New Navigation System Combining QR-Code and Augmented Reality. *J. Internet Technol., 19*(2), 565-571. https://doi.org/10.3966/160792642018031902024
[50] Fajrianti, E. D. et al. (2023). INSUS: Indoor Navigation System Using Unity and Smartphone for User Ambulation Assistance', *Information, 14*(7), p. 359. https://doi.org/10.3390/info14070359
[51] Szymczyk, T., Montusiewicz, J. & Gutek, D. (2016). Navigation in Large-Format Buildings Based on RFID Sensors and QR and AR Markers. *Adv. Sci. Technol. Res. J.*, 10(31), 263-273. https://doi.org/10.12913/22998624/64059
[52] Verma, P., Agrawal, K. & Sarasvathi, V. (2020). Indoor Navigation Using Augmented Reality. *Proceedings of the 4th International Conference on Virtual and Augmented Reality Simulations*, Sydney NSW Australia: ACM, 58-63. https://doi.org/10.1145/3385378.3385387
[53] Tadepalli, S. K., Ega, P. A. & Inugurthi, P. K. (2021). Indoor Navigation Using Augmented Reality. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), 7*(4), 588-592. https://doi.org/10.32628/CSEIT2174134

Authors' contacts:

Tomáš Macháč, Ing.
(Corresponding author)
University of West Bohemia,
Univerzitni 2732/8, 301 00 Plzen, Czech Republic
machact@fst.zcu.cz

Petr Hořejší, Doc. Ing. PhD
University of West Bohemia,
Univerzitni 2732/8, 301 00 Plzen, Czech Republic
tucnak@fst.zcu.cz

Michal Šimon, Doc. Ing. PhD
University of West Bohemia,
Univerzitni 2732/8, 301 00 Plzen, Czech Republic
simon@fst.zcu.cz

Andrea Šimerová, Ing.
University of West Bohemia,
Univerzitni 2732/8, 301 00 Plzen, Czech Republic
simerova@fst.zcu.cz

ISSN 1846-6168 (Print), ISSN 1848-5588 (Online)
https://doi.org/10.31803/tg-20241231153205

Original scientific paper
Received: 2024-12-31, Accepted: 2025-01-17

Optimizing Carbonation Hardening for Lightweight Concrete

Olexander Gara, Anatoliy Gara, Andrii Kolesnykov, Khrystyna Moskalova*

Abstract: This study investigates the accelerated carbonation hardening (ACH) of lightweight aggregate concrete. It evaluates the effects of binder content, ground limestone, lightweight aggregates, plasticizers, carbonation pressure, and duration on compressive strength at 1 hour, 28 days, and 180 days. Statistical models and optimization using desirability functions were employed to identify optimal recipe and technological parameters. ACH enhances concrete strength and promotes sustainable carbon sequestration, providing an alternative to conventional curing methods.

Keywords: accelerated carbonation; lightweight concrete; optimization; recipe factors; response surface

1 INTRODUCTION

Accelerated carbonation hardening (ACH) has become an effective approach to enhancing the mechanical properties and environmental performance of concrete. In contrast to traditional methods like steam curing, which demand significant energy and resources, ACH provides a sustainable alternative by utilizing CO_2 to improve strength and durability while capturing atmospheric carbon dioxide [1-3]. The process accelerates carbonation reactions in cement-based materials, forming calcium carbonate, which densifies the concrete matrix and improves its properties [4-5]. This method is particularly advantageous for applications requiring rapid curing, such as bridges and high-rise buildings in challenging environments, where durability and efficiency are critical [6].

Optimizing ACH involves controlling factors such as binder composition, CO_2 pressure, and curing time. This study investigates these parameters using statistical methods to enhance the effectiveness of ACH, ensuring improved material performance and reduced environmental impact.

2 REVIEW OF CURRENT LITERATURE

Investigations demonstrate that accelerated carbonation hardening (ACH) improves concrete properties such as compressive strength, durability, and resistance to environmental factors. Compressive strength can be improved by 10 % to 50 %, depending on the mix design and carbonation conditions [7, 8]. ACH also reduces permeability and increases resistance to chloride penetration, sulfate attack, and alkali-silica reactions, contributing to the durability of concrete structures [9].

The mechanisms of ACH involve CO_2 diffusion into the moist concrete matrix, where it reacts with calcium hydroxide to form calcium carbonate, enhancing material strength by filling voids [10]. Additionally, ACH densifies the microstructure, improving resistance to environmental and chemical degradation [11-14].

However, carbonation reduces alkalinity, potentially increasing the susceptibility of steel reinforcement to corrosion. Protective calcium carbonate layers formed during ACH may mitigate this risk [15]. The long-term effectiveness of ACH depends on exposure conditions and material composition, highlighting the need for further performance studies [16].

3 PURPOSE AND METHODOLOGY

The primary objective of this study is to analyze the influence of recipe and technological factors—such as binder content, plasticizer concentration, ground limestone proportions, carbonation pressure, and duration—on the compressive strength of lightweight aggregate concrete at various curing stages (1 hour, 28 days, and 180 days). The goal is to develop experimental-statistical (ES) models that enable the optimization of these factors, ensuring maximum strength while balancing practical and economic constraints.

The research employs experimental design principles, including response surface methodology and desirability functions, to model and optimize material properties. Historical experimental data were analyzed using Design Expert software, allowing for a detailed evaluation of factor interactions and their contributions to strength development. This methodology ensures a comprehensive understanding of the processes governing accelerated carbonation hardening and provides guidance for achieving optimal material performance under varying conditions.

4 RESEARCH RESULTS

Accelerated carbonation hardening (ACH) of lightweight aggregate concrete enhances the carbonation of hydrated cement phases, resulting in the precipitation of calcium carbonates that improve the mechanical properties of the concrete matrix. The ACH process consists of several key stages: surface preparation by cleaning the concrete to ensure uniform CO_2 penetration; installation of specialized carbonation equipment, such as chambers or injection systems; and control of environmental parameters such as temperature and humidity to maximize the reaction kinetics. CO_2 is injected into the concrete through surface holes or internal channels, with continuous monitoring and adjustment of parameters such as CO_2 pressure, temperature, and exposure duration. After the process, the results are

evaluated based on mechanical properties, structure, and carbonation degree.

The rapid development of high early strength within 30 to 60 minutes necessitates carbonation regimes employing elevated pressure. In accordance with the theory of heat and mass transfer, vacuuming freshly molded concrete creates a porous system under reduced pressure. Introducing CO_2 under pressure during this phase accelerates the carbonation process by promoting self-absorption and stress relaxation within the capillary structure. The dissolution rate of minerals increases proportionally with CO_2 pressure, facilitating controlled structure formation.

Earlier investigations identified [2] the optimal aggregate packing for lightweight concrete, with a binder dosage of 300 kg/m³. Concreare strength, density, and other characteristics are inflused by the the proportion of fine aggregate particles with a size up to 5 mm. The structure and strength characteristics of carbonized concrete are influenced by three primary factors: the aggregate structure, determined by the distribution of porous aggregate; the binder concentration, defined by cement type and properties; and the mixing water content, adjusted for carbonation technology.

Due to the interaction of recipe and technological factors, the properties of concrete can vary significantly during carbonation. Structure formation is influenced by both constructive and destructive factors, which impact the speed and completeness of reactions. Therefore, optimizing ACH requires analyzing the influence of these factors on early and long-term strength [17]. The optimization criterion focuses on minimizing cement consumption while achieving the required properties, as cement is an expensive and limited resource.

The key recipe-technological factors affecting the physical and mechanical properties of ACH lightweight aggregate concrete include binder consumption, which varies from 300 to 500 kg/m³, with ground limestone content ranging from 0 to 30 %; the plasticizing additive (SYM) content, adjusted between 0 and 0.4 % by binder weight; carbonation pressure, ranging from 0.6 to 1.2 MPa; and the duration of carbonation, set between 30 and 60 minutes.

To develop mathematical models for compressive strength at 1 hour (R_{1h}), 28 days (R_{28}), and 180 days (R_{180}), a second-order five-factor experimental design ("Hartley-5") with six center points was used [18]. The factor levels and variations are presented in Tab. 1.

Table 1 Experimental factors and their ranges of variation

Factors		Quantity unit	Code	Variation levels		
				–1	0	1
A	Binder	kg/m³	*Knitt*	300	400	500
B	SYM additive	%	*SYM*	0	0.2	0.4
C	Ground limestone	%	*GL*	0	15	30
D	Applied carbonation pressure	MPa	P_{CO2}	0.6	0.9	1.2
E	The duration of the carbonization process	min	t_{CO2}	30	45	60

The study was conducted on cubic samples with an edge length of 10 cm. The molded specimens underwent carbonation under specified conditions. Concrete density, CO_2 uptake (determined by the weight change of the samples before and after carbonation), and compressive strength were evaluated at specified intervals during the testing process.

The compressive strength of the concrete increased over time, ranging from 2.1 to 16.5 MPa at 1 hour, 4.6 to 21.8 MPa at 28 days, and 5.9 to 26.1 MPa at 180 days. The collected data enabled the creation of experimental-statistical (EC) models for the composite material properties using the Design Expert software [19]. This software allows importing previously obtained experimental data, referred to as "historical" data. The response surface methodology was applied for data analysis. The resulting models and their statistical parameters are presented in Tab. 2. Statistically significant components of the models were selected through backward elimination.

Table 2 EC strength models of material characteristics

Characteristics	Experimental-statistical models for *R*	Regression Statistics
R_{1h}, MPa	$R_{1h} = +8.80 + 3.70A + 0.90B - 0.50C + 1.30D + 0.50E + 0.40AB + 0.20AD + 0.20AE + 0.40BC - 0.10BD + 0.10BE - 0.30CD - 0.30CE + 0.20DE + 0.98A^2 - 0.52C^2 - 0.52D^2 - 0.32E^2$	$R^2 = 0.9983$ *Adj* $R^2 = 0.9960$ *Preq* $R^2 = 0.9979$ *Adeq Precision* = 93.028
R_{28}, MPa	$R_{28} = +12.69 + 5.88A + 1.12B - 1.22C - 0.22D + 0.12E + 0.47AB - 0.28AC - 0.18AD + 0.47BC - 0.13BD + 0.27DE + 1.36A^2 - 0.44B^2 - 0.64C^2 - 0.54E^2$	$R^2 = 0.9984$ *Adj* $R^2 = 0.9967$ *Preq* $R^2 = 0.9928$ *Adeq Precision* = 90.276
R_{180}, MPa	$R_{180} = +13.87 + 6.80A + 1.10B - 1.30C - 0.40D + 0.10E + 0.30AB - 0.50AC - 0.20AD - 0.30AE + 0.30BC + 0.30BD + 2.13A^2 - 0.97B^2$	$R^2 = 0.9985$ *Adj* $R^2 = 0.9973$ *Preq* $R^2 = 0.9922$ *Adeq Precision* = 106.701

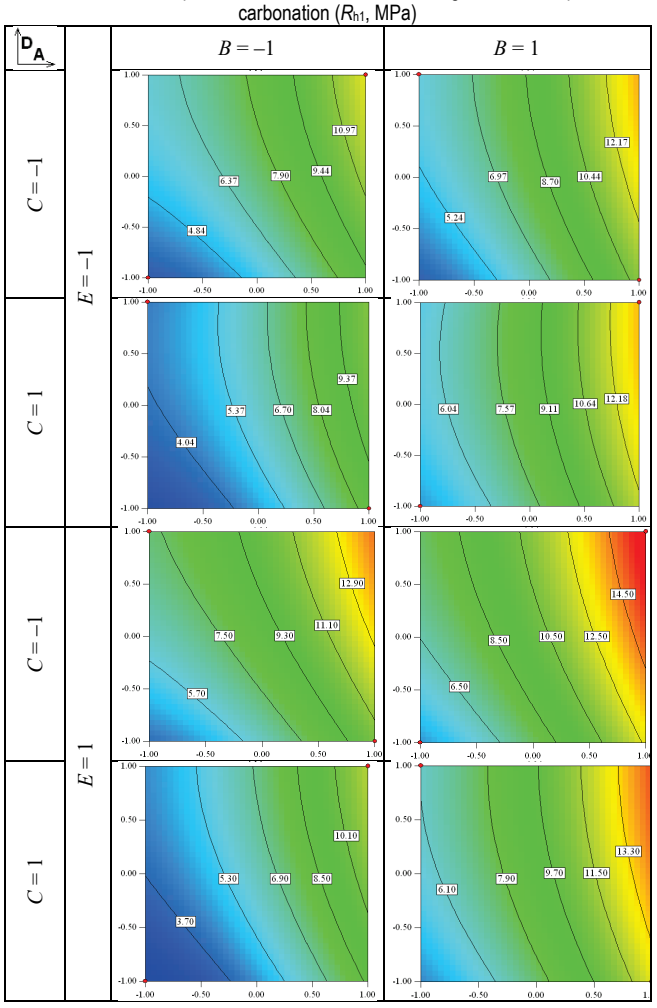
All models exhibited statistical significance. The predicted R^2 values closely matched the adjusted R^2 values, demonstrating consistency. Additionally, the signal-to-noise ratio exceeds the threshold of 4, confirming the reliability of the models. High R^2 values highlight the models' ability to accurately represent the data.

Visualizing the material property field directly is challenging due to the model's multidimensional nature. However, its geometric characteristics can be inferred through line diagrams illustrating the target property (R_{1h} strength in MPa). These diagrams are plotted using two primary factors, *A* and *D*, while keeping the other factors (*B*, *C*, and *E*) constant. Since the model shows low sensitivity to changes in *B*, *C*, and *E*, it is practical to analyze isoline maps at the corners of the factorial subspace defined by these three factors. This approach results in eight sections, as detailed in Tab. 3.

As the strength values constitute a time series, a systematic interpretation necessitates a comprehensive analysis of all data, encompassing both analytical and graphical representations. Limestone content (*GL*, *C*) follows with a notable negative slope. Factors *B*, *D*, and *E* show less pronounced effects on strength. The model's graphical representation uses isoline maps, where *A* and *C* are the primary axes due to their significant influence. Factors *B*, *D*, and *E*, with minimal impact on R_{28}, are presented as discrete

variables. Similar to the analysis for R_{1h}, these factors are positioned at the vertices of the factorial subspace *BDE*. The corresponding isoline diagrams are provided in Tab. 4.

Table 3 Visual representation of the material's strength field 1hour post-carbonation (R_{h1}, MPa)



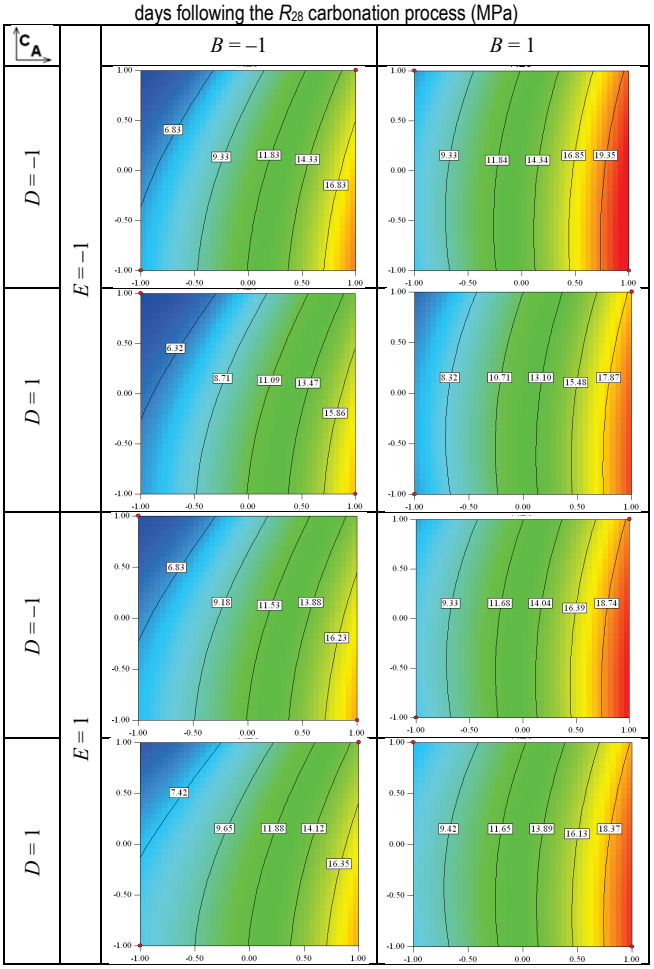
Limestone content (*GL*, *C*) stands out due to its negative contribution, as shown by the slope at the central point of the plan. Factors *B*, *D*, and *E* are treated as discrete variables, while the primary axes for the graphical analysis remain *A* and *C*.

The diagram structure and interpretation approach follow a pattern similar to that in Tab. 3.

The presented mathematical models can be interpreted both from formal mathematical positions and based on the basic concepts of materials science.

There are two main patterns of change in regression models describing the properties over time, provided that they are statistically correctly reduced. During the first stage of development of material structure formation, some of the model components can be split off as insignificant, and at the next stage such components reappear in the models. The second variant observed in the study is associated with stepwise simplification. The terms split off at the first stage do not appear at the subsequent ones. The two variants under consideration are closely related to the existence of "hidden" parameters of the structure that do not directly affect the properties and are not reflected in the ES-models. Since the prerequisite for the formation of properties and, in particular, strength, is the structure of the material, the difference in the case of two variants of changes in regression models can be interpreted structurally using parameterization methods and the theory of morphogenetic rearrangements of the material structure. According to the corresponding theory [20], the structure of the material under the influence of physicochemical factors undergoes a sequence of rearrangements associated with the implementation of the structural sensitivity of the material. If there is one stage of structural sensitivity, in the material after the stage of primary structural differentiation with the formation of areas with high and, accordingly, low density of particles and bonds (clusters and interfaces), the relaxation process prevails (approaching a stable structural equilibrium).

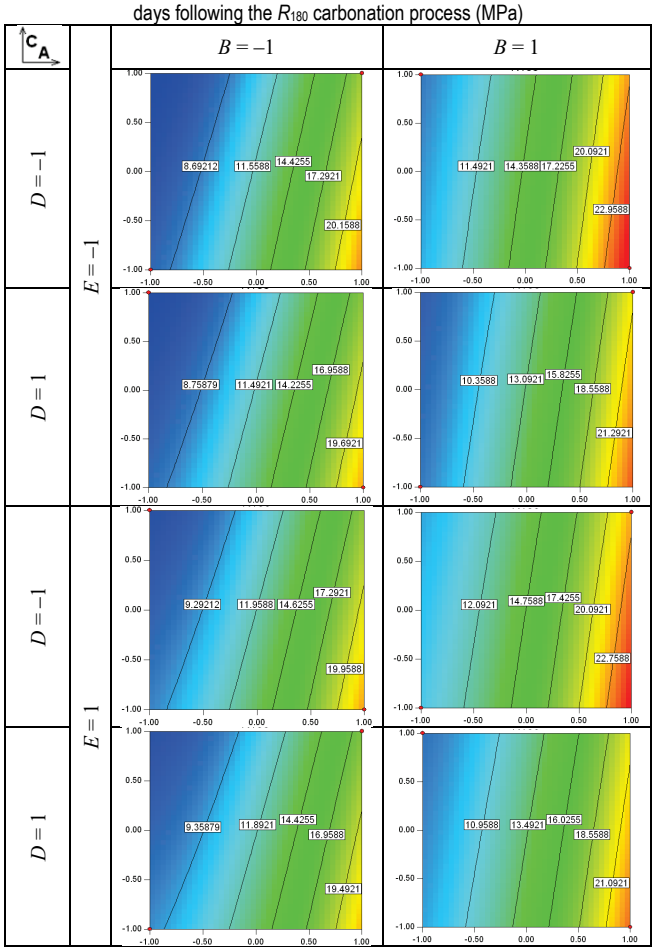
Table 4 Visual representation of the strength distribution within the material 28 days following the R_{28} carbonation process (MPa)



Deviations from this equilibrium decrease, which is associated with a decrease in the influence of a number of structural parameters until they become "hidden". If, however, as a result of the implementation of physicochemical hydration mechanisms, a second phase of sensitivity occurs, associated with dynamic instability, the relaxed characteristics of the structure can again manifest

themselves, affect the properties and, accordingly, enter ES-models. Thus, of the two extreme scenarios, in the material under consideration as a system, the second, relaxation one is realized, with the absence of a repeated stage of sensitivity.

Table 5 Visual representation of the strength distribution within the material 180 days following the R_{180} carbonation process (MPa)



Regression models for strength evolution simplify as hardening progresses. Factors influencing early strength (R_{1h}) diminish in importance over time, with many statistically insignificant terms eliminated by 180 days (R_{180}). This reflects a relaxation process, where structural potential minimizes and linear factor interactions dominate. Thus, the nature of the step-by-step change in regression models can have a phenomenological interpretation from the standpoint of the theory of structure formation.

The influence of binder consumption (A) on strength changes significantly across carbonation stages. At 1 hour, strength increases by 3.2 MPa as levels of *SYM* additive (B), carbonation pressure (D), and duration (E) rise. Maximum strength is observed at high *SYM* levels and optimal carbonation parameters. Conversely, at 28 and 180 days, strength growth peaks with high *SYM* content, no ground limestone (GL, C), and lower pressure and duration values. The relationship between binder consumption and strength follows a second-order function: increasing binder from 300 to 400 kg/m³ boosts strength by 2.7 MPa at 1 hour, 4.2 MPa at 28 days, and 4.5 MPa at 180 days. Ground limestone content exhibits a dual influence on 1-hour strength. In the absence of *SYM*, higher contents are detrimental (3.0 MPa reduction), whereas they prove beneficial (1.0 MPa increase) when *SYM* is incorporated. Over time, limestone's impact diminishes, with binder content and *SYM* additive becoming dominant. At higher binder levels (500 kg/m³) without *SYM*, increased limestone reduces strength by 4.0 MPa at 28 days and 4.1 MPa at 180 days. However, with lower binder levels (300 kg/m³) and *SYM* at 0.4%, replacing cement with limestone up to 30% has negligible effects.

5 OPTIMIZATION OF MATERIAL PROPERTIES

The optimization of recipe-technological factors in this study focuses on three key strength metrics: R_{1h}, R_{28}, and R_{180}. This approach involves multi-criteria optimization, where solutions are derived through compromise to balance competing objectives. A common strategy in such cases is to combine multiple criteria into a single objective function using a convolution method. This method is implemented through desirability functions [21], a tool that facilitates interaction with decision-makers.

The desirability function works by transforming each response or output variable into a dimensionless desirability value, ranging from 0 (completely undesirable) to 1 (fully desirable). These transformed values represent how well each response meets its target or falls within acceptable limits. Once each response is assigned a desirability value, a combined overall desirability score D is calculated as the geometric mean of the individual desirability values (1):

$$D = \left(d_1^{r_1}, d_2^{r_2}, \ldots, d_n^{r_n} \right)^{\frac{1}{\sum_i r_i}} \quad (1)$$

where d_i expresses the desirability of each partial criterion, $0 \le d_i \le 1$ and r_i expresses its importance.

The Design Expert software was employed for effective multi-criteria optimization and the identification of optimal solutions. Each criterion's importance is rated on a scale from 1 (least important) to 5 (most critical), ensuring a structured and transparent decision-making process.

For the constructed ES-models (Tab. 2), numerous optimization scenarios can be defined, each varying by the assigned importance and weight of criteria. Two scenarios of direct engineering relevance are analyzed further.

The primary objective in developing the composite material is to maximize strength throughout the 180-day hardening period. However, achieving sufficient strength by the standard control point of 28 days is equally critical. To focus on these goals, all additional constraints, aside from those related to fixed intervals, are removed. Under these conditions, partial optimization criteria are established, enabling the formulation of a desirability function that reflects the relative importance of each strength metric. This structured approach ensures the prioritization of long-term performance while maintaining compliance with intermediate benchmarks (Tab. 6).

Table 6 Desirability function parameters for long-term hardening optimization

Factors/ properties	Goals	Lower boundary	Upper boundary	Lower weight	Upper weight	Significance
Knitt	within the interval	–1	1	1	1	1
SYM		–1	1	1	1	1
GL		–1	1	1	1	1
P_{CO2}		–1	1	1	1	1
t_{CO2}		–1	1	1	1	1
R_{1h}	maximum	2.1	16.5	1	1	2
R_{28}		3.8	21.8	1	1	2
R_{180}		5.9	26.1	1	1	5

As a result of the implementation of the desirability function method, the following solutions were obtained (Tab. 7).

Table 7 Long-term hardening optimization results

No	*Knitt*	*SYM*	*GL*	P_{CO2}	t_{CO2}	R_{1h}	R_{28}	R_{180}	*Desirability*
1	1	0.89	–1	0.08	0.52	14.8 8	21.2 1	24.6 3	0.9268
2	1	0.93	–0.99	0.06	0.51	14.87	21,23	24.60	0.9265
3	1	0.99	–0.98	0	0.54	14.84	21.25	24.60	0.9262
4	1	0.98	–1	–0.17	0.23	14.25	21.48	24.84	0.9257

The selected composition is the optimal solution to the problem under consideration. Its composition and properties are displayed graphically (Fig. 1).

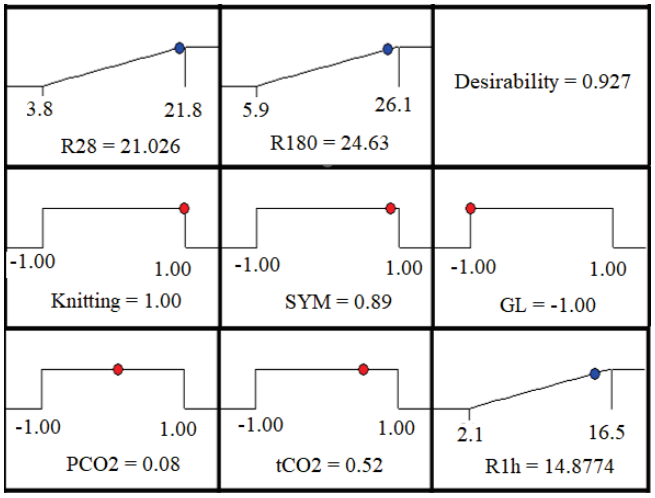


Figure 1 The primary outcome of the multi-criteria optimization of long-term hardening

The composite material optimization focused on minimizing cement consumption and using low-pressure, short-duration carbonation to reduce costs and improve safety. While ensuring the material meets standard strength requirements at 28 days, the primary objective remained maximizing long-term strength. This optimization involved setting partial criteria and assigning importance levels for the desirability function, as shown in Tab. 8.

The following solutions were obtained (Tab. 9).

The corresponding basic solution is reflected in the arch diagram (Fig. 2).

The comparison of Tabs. 8 and 9 reveals that incorporating additional constraints allowed for a more efficient carbonation process and reduced cement usage, albeit with a 6 MPa decrease in long-term strength. The lower desirability values in Tab. 9 highlight the impact of these new limitations and the need to balance conflicting optimization criteria.

Table 8 Desirability function parameters for long-term hardening optimization, with additional conditions

Factors/ properties	Goals	Lower boundary	Upper boundary	Lower weight	Upper weight	Significance
Knitt	within the interval	–1	1	1	1	4
SYM		–1	1	1	1	3
GL		–1	1	1	1	3
P_{CO2}		–1	1	1	1	2
t_{CO2}		–1	1	1	1	2
R_{1h}	maximum	2.1	16.5	1	1	4
R_{28}		3.8	21.8	1	1	5
R_{180}		5.9	26.1	1	1	5

Table 9 Optimization results for long-term hardening under constraints

No	*Knitt*	*SYM*	*GL*	P_{CO2}	t_{CO2}	R_{1h}	R_{28}	R_{180}	*Desirability*
1	0.57	1	–0.49	–1	–1	9.29	17.89	20,23	0.5838
2	0.57	1	–0.48	–1	–1	9.32	17.91	20,24	0.5837
3	0.57	1	–0.48	–1	–0.98	9.33	17.91	20.2 3	0.5836
4	0.57	1	–0.52	–1	–1	9.28	17.92	20,29	0.5835

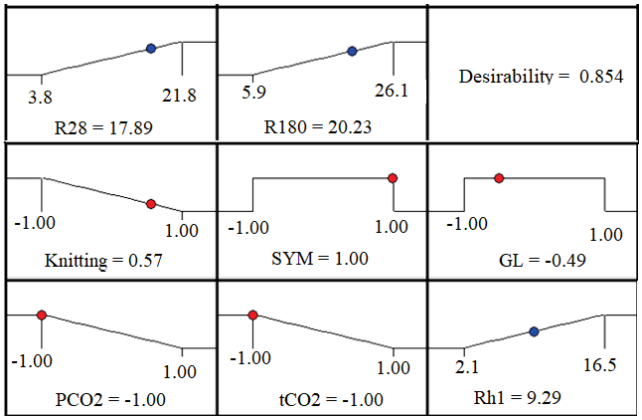


Figure 2 The primary outcome of the multi-criteria optimization of long-term hardening with constraints

6 CONCLUSION

The analysis of strength data for carbonized lightweight aggregate concrete allowed for the development of experimental-statistical models (ES-models) to predict material properties. Compressive strength ranged from 2.1 to 16.5 MPa one hour after carbonation, 4.6 to 21.8 MPa at 28 days, and 5.9 to 26.1 MPa at 180 days. Graphical representations and interpretations of these models support the hypothesis that, as the concrete structure forms, ES-models tend to simplify by removing interactions between structural factors. Two optimization tasks were formulated: the first focused on maximizing long-term strength without additional constraints, while the second incorporated limits on cement usage and carbonation parameters. The desirability function approach determined optimal factor sets. For the second task, the optimal mix included *Knitt* = 0.57, *SYM* = 1, *GL* = –0.49, and P_{CO_2} = –1, yielding strengths of 9.29 MPa (1 hour), 17.89 MPa (28 days), and 20.23 MPa (180 days). These constraints achieved cement savings and optimized carbonation conditions, though long-term strength decreased by about 6 MPa.

Acknowledgments

This work has been supported by Metal Centre Čakovec under the project KK.01.1.1.02.0023

7 REFERENCES

[1] Teir, S., Eloneva, S. & Zevenhoven, R. (2005). Production of precipitated calcium carbonate from calcium silicates and carbon dioxide. *Energy Conversion and Management, 46*, 2954-2979. https://doi.org/10.1016/j.enconman.2005.02.009

[2] Groves, G., Rodway, D. & Richardson, I. (1990). The carbonation of hardened cement pastes. *Advances in Cement Research, 3*(11), 117-125. https://doi.org/10.1680/adcr.1990.3.11.117

[3] Gara, O. A., Kolesnykov, A. V., Semenova, S. V. & Oliinyk T. P. (2024) Optimization of accelerated carbonation hardering effects of expanded clay concrete, *Modern construction and architecture, 8*, 50-64. https://doi.org/10.31650/2786-6696-2024-8-50-64

[4] Kashef-Haghighi, S. & Ghoshal, S. (2013). Physico-chemical processes limiting CO_2 uptake in concrete during accelerated carbonation curing. *Industrial & Engineering Chemistry Research, 52*(16), 5529-5537. https://doi.org/10.1021/ie303372h

[5] Suescum-Morales, D., Kalinowska-Wichrowska, K., Fernández, J. M. & Jimenez, J. R. (2021). Accelerated carbonation of fresh cement-based products containing recycled masonry aggregates for CO_2 sequestration. *Journal of CO_2 Utilization, 46*, 101461. https://doi.org/10.1016/j.jcou.2021.101461

[6] Gara, O. A. (2016). Features of Carbonation Hardening of Lightweight Aggregate Concrete. *Visnjik ODABA, 62*, 22-27. (in Ukrainian)

[7] Rostami, V., Shao, Y., Boyd, A. J. & He, Z. (2012). Microstructure of cement paste subject to early carbonation curing. *Cement and Concrete Research, 42*(1), 186-193. https://doi.org/10.1016/j.cemconres.2011.09.003

[8] Lagerblad, B. (2005). Carbon dioxide uptake during concrete life cycle - State of the art. Swedish Cement and Concrete Research Institute.

[8] Junior, A., Filho, R., Fairbairn, E. & Dweck, J. (2014). A study of the carbonation profile of cement pastes by thermogravimetry and its effect on the compressive strength. *Journal of Thermal Analysis and Calorimetry, 116*, 69-76. https://doi.org/10.1007/s10973-014-3922-9

[10] Šavija, B. & Luković, M. (2016). Carbonation of cement paste: Understanding, challenges, and opportunities. *Construction and Building Materials, 117*, 285-301. https://doi.org/10.1016/j.conbuildmat.2016.04.138

[11] Parrott, L. (1992). Carbonation, moisture and empty pores. *Advances in Cement Research, 4*(15), 111-118. https://doi.org/10.1680/adcr.1992.4.15.111

[12] Shi, C., Tu, Z., Guo, M. Z. & Wang, D. (2017). Accelerated carbonation as a fast curing technology for concrete blocks. In *Sustainable and Nonconventional Construction Materials Using Inorganic Bonded Fiber Composites* (pp. 313-341). Woodhead Publishing. https://doi.org/10.1016/B978-0-08-102001-2.00014-0

[13] Thiery, M., Villain, G., Dangla, P. & Platret, G. (2007). Investigation of the carbonation front shape on cementitious materials: Effects of the chemical kinetics. *Cement and Concrete Research, 37*(7), 1047-1058. https://doi.org/10.1016/j.cemconres.2007.04.002

[14] Morandeau, A., Thiery, M. & Dangla, P. (2014). Investigation of the carbonation mechanism of CH and CSH in terms of kinetics, microstructure changes and moisture properties. *Cement and Concrete Research, 56*, 153-170. https://doi.org/10.1016/j.cemconres.2013.11.015

[15] Bertolini, L., Elsener, B., Pedeferri, P., Redaelli, E. & Polder, R. B. (2013). *Corrosion of Steel in Concrete: Prevention, Diagnosis, Repair*. John Wiley & Sons. https://doi.org/10.1002/9783527651696

[16] Matthews, J. (1984). Carbonation of ten-year-old concretes with and without PFA. *Proceedings of AshTech*.

[17] Gara, A. A. (1990). Development of Carbonation Technology for Lightweight Aggregate Concrete Wall Products. *Doctoral dissertation*. Odesskij inzenerno-stroiteljnij institut. (in Ukrainian)

[18] Adler, Y. P., Markova, E. V. & Granovskiy, Y. V. (1976). Experiment Planning for Finding Optimal Conditions. Nauka. (in Rusian)

[19] Grant, F. (2007). *Design Expert 7.1*. Scientific Computing World.

[20] Kolesnikov, A. V., Semenova, S. V. & Makovetska, O. O. (2023). Modeling of Composite Structure Formation Processes Using Catastrophe Theory. *Sucasne budivnictvo ta arhitektura, 6*, 90-98. (in Ukrainian)

[21] Myers, R. H., Montgomery, D. C. & Anderson-Cook, C. M. (2009). *Response Surface Methodology: Process and Product Optimization Using Designed Experiments.* John Wiley & Sons. https://doi.org/10.1002/9780470744581

Authors' contacts:

Olexander Gara, PhD, Assistant Professor
Odessa State Academy of Civil Engineering and Architecture,
Construction and Technological Institute,
Department of Processes and Apparatuses in the Technology of Building Materials,
Didrihsona st., 4, 65029 Odesa, Ukraine
garaogasa@ukr.net

Anatoliy Gara, PhD, Assistant Professor
Odessa State Academy of Civil Engineering and Architecture,
Construction and Technological Institute,
Department of Production of Construction Products and Structures,
Didrihsona st., 4, 65029 Odesa, Ukraine
ggaryc223420@gmail.com

Andrii Kolesnykov, PhD, Assistant Professor
Odessa State Academy of Civil Engineering and Architecture,
Construction and Technological Institute,
Department of Chemistry and Ecology,
Didrihsona st., 4, 65029 Odesa, Ukraine
Kolesnikov_himek@odaba.edu.ua

Khrystyna Moskalova, PhD, Assistant Professor
(Corresponding author)
Development and Training Centre for the Metal Industry—Metal Centre Čakovec,
Department for Research and Development,
Bana Josipa Jelačića, 22D, 40000 Čakovec, Croatia
kristina@metalskajezgra.hr

ISSN 1846-6168 (Print), ISSN 1848-5588 (Online)
https://doi.org/10.31803/tg-20250101212243

Original scientific paper
Received: 2025-01-01, Accepted: 2025-01-16

Visual Perception of Organic and Polygonal Shapes in the Graphic Communication Processes

Robert Geček, Damir Vusić*

Abstract: Many factors influence the visibility and proper reception of visual messages. The forms used in communication play a crucial yet often overlooked role in this process. This paper investigates which forms are more favourably perceived by the public in graphic communications. A test questionnaire was developed to gauge respondents' preferences for different shapes. The study involved 63 participants whose visual attention to organic and polygonal shapes was measured using an eye-tracker over an 8-second interval. Three scripts were utilized: two freely available ones and one specifically written for this research. These scripts were implemented in MATLAB to analyse and present the visual results of the study. The hypotheses were tested using two statistical methods: the t-test for two vectors and Wilcoxon's rank-sum test. Based on the findings, recommendations were outlined for future research on the application of organic and polygonal shapes in graphic communications.

Keywords: graphic communication; organic shapes; polygonal shapes; visual perception

1 INTRODUCTION

Every day, we encounter a variety of visual stimuli in the form of organic and polygonal two-dimensional shapes. While some shapes are more easily perceived than others, the question arises: are organic or polygonal shapes more acceptable to us? This research aims to help enhance the effectiveness and speed of visual perception in graphic communications and the products derived from these interactions (client–designer–consumer). The objective is to identify the parameters and relationships between organic and polygonal flat shapes that influence the visual perception of the final graphic product.

In today's fast-paced world, the flow of information is incredibly rapid, leaving people increasingly overwhelmed and often causing important details to be overlooked. While the lightning-fast progress of technology brings many advantages, it also presents challenges, as human attention is constantly inundated with vast amounts of information. Distinguishing between various forms of visual stimuli is challenging, as they often merge into a unified structure or visual impression that incorporates technical, artistic, and, most importantly, multidisciplinary elements.

Quick judgments about objects in the environment are influenced by their physical properties, although the exact nature of these properties remains unclear. For example, sharp transitions in contours can evoke a sense of threat, potentially triggering negative biases. The type of contour a visual object possesses – whether it features sharp angles or smooth curves - plays a decisive role in shaping people's attitudes toward it. These first impression tendencies are heavily influenced by the perceptual characteristics of the image, particularly when judgments are made quickly. Research has shown that the shape of an object serves as a significant factor in forming preferences [1, 2].

In the context of graphic reproduction, visual psychophysics - or the psychophysics of visual perception—plays a crucial role. This field, defined as a descriptive science aimed at understanding the sensory capabilities of the normal human visual system [3-5], serves as a key element in this paper for validating the hypothesis.

2 THEORETICAL FRAMEWORK

This paper is based on the research conducted as a part of the dissertation "Determination of visual perception of organic and polygonal shapes in graphic communication processes" [6].

Shape is a fundamental property of visual appearance. Eye movement patterns provide valuable insights into shape analysis strategies during visual perception. However, there is still limited understanding of how shape perception differs between tasks involving motion recognition and those requiring action planning.

Leek et al. discovered that high curvature - particularly extremely concave minimums - can predict eye movement patterns during object recognition tasks. Their study investigated how eye movements might reveal differences in shape analysis strategies between object recognition and action planning. In the context of object recognition, shape detection and recognition are crucial, as they enable the system to identify an object within a given image [7].

The selection of visual aids - such as statistical graphs, images, and symbols - serves as a key tool in graphic communication. A carefully designed message elicits a specific reaction from the target audience. On a subconscious level, individuals seek elements that resonate with their identity. The visual components of packaging aim to evoke a sense of familiarity and connection, creating the impression of "this is my kind of product" while also conveying the product's usefulness. This connection represents the most significant impact of graphic communication.

Visual aids enrich communication by adding depth and clarity to the message. Every object used in graphic communication is thoughtfully designed, carries a story, and serves to encourage interaction or convey a message in a simple yet impactful way. From the moment we wake up, everything that surrounds us communicates visually. Graphic

communication has profoundly enhanced every phase of daily human life. It allows for the expression of personality and ideas that go beyond the limits of words alone. Good design is characterized by innovation, honesty, and aesthetic appeal. When a message is conveyed with clarity and sincerity through graphic communication, people tend to respond positively, forging a deeper connection to the message [8-11].

Graphic communication is showing people the things they urgently need to understand, in ways that they can understand. Graphic communication is the process through which we can decode the visual culture of those we need to communicate with. Sometimes, most of the time, this means using tools and processes that have their origins in graphic design [12]. Shape, as one of the graphic design elements, refers to the external outline of the form or anything that has height and width [13].

3 RESEARCH METHODOLOGY

3.1 Goal and the Hypothesis of the Research

The main goal of this paper is to explore the relationship between the visual perception of organic and polygonal flat shapes in graphic communication processes and to evaluate their visual legibility within graphic communications and across various multimedia systems.

The hypothesis that has been put forward and that will be tested is as follows:

H: *Organic shapes exert a stronger influence on the visual perception of observers compared to polygonal shapes in graphic communication processes.*

3.2 Methodology

The research for this paper was conducted in the laboratory of the University North in Varaždin, involving students from the undergraduate programme in Multimedia, Design, and Application and the graduate programme in Multimedia. A total of 43 first-year undergraduate students participated, all of whom possessed basic knowledge of visual perception but had not yet developed expertise to distinguish the nuanced phenomena occurring in this aspect of graphic communications. Additionally, 20 graduate students, who had gained deeper insights into the processes of visual communication and visual psychophysics after achieving learning outcomes at the undergraduate level, were also surveyed. Of the 63 participants, 33 were male (52.4%) and 30 were female (47.6%), with ages ranging from 18 to 30 years.

The research was conducted in several phases. The first phase entailed the creation of visual templates tailored for the study. In the second phase, participants were presented with a visual questionnaire without any prior explanation of the research objectives. The questionnaire consisted of both textual and visual elements (Fig. 1) and was designed in a minimalist style to help the respondents focus more effectively on the provided shapes.

In this phase of the research, samples were created to display organic and polygonal flat shapes side by side in a monochromatic environment.

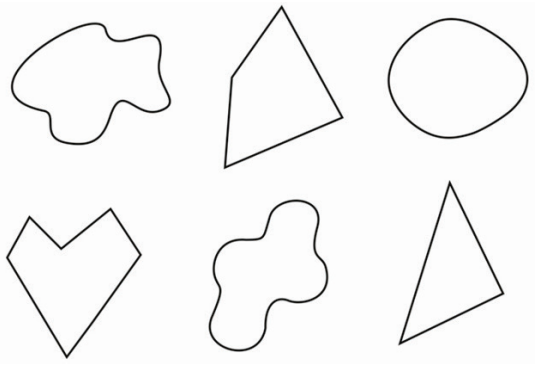


Figure 1 Visual part of the survey questionnaire

The textual portion of the questionnaire included mandatory demographic questions, such as the respondents' gender and age. Below the visual samples, participants were presented with the following questions/tasks/statements:

- *Which shape caught your attention the most and is visually the most acceptable to you?*
- *In a few words, describe why you chose this shape.*
- *Polygonal shapes evoke in me: emotion, fear, threat, happiness, discomfort.*
- *Organic shapes evoke in me: emotion, fear, threat, happiness, discomfort.*

Additionally, a blank field was provided for participants to describe their own impressions or emotions in free text. Once the respondents completed the questionnaire, each was assigned a unique number corresponding to the same number used in the subsequent phase of the research.

The research adhered to ISO 3664:2009 Graphic technology and photography - Viewing conditions. The eye-tracking methodology utilized the Gazepoint GP3 Desktop eye-tracking device in conjunction with a Samsung LCD monitor, model S22A350H (size 21.5", display ratio 16:9, and viewing angle 170°/160°). The Gazepoint Analysis software was used to monitor the eye movements of the participants (Fig. 2), as they viewed the test samples.

Figure 2 Testing place for the Tracking methodology

The test samples presented to the participants featured one organic shape and one polygonal shape, placed side by side. To eliminate any bias regarding the initial viewing direction, the positions of the shapes were alternated between trials (Fig. 3). The samples were separated by a white test shape featuring a cross in the centre to prevent the eyes from lingering on the location of the previous test shape after it changed on the screen.

Each participant had eight seconds to view each test sample, with the sample changing after that time. Eight seconds was chosen based on recent research suggesting that human attention spans have decreased with the rise of new technologies. If a consumer does not find something of interest within this brief time, they tend to dismiss it and move on. As mentioned, the eye-tracking device monitored specific eye pupil movements, capturing data on the first fixation, the longest fixation, and the total amount of time spent focusing on individual points of the test samples.

In the subsequent phases, the results from the visual questionnaire were analysed, along with the statistical processing of the eye-tracking data.

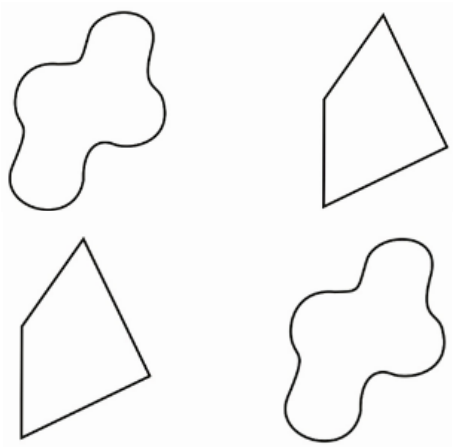


Figure 3 Test samples (visual templates)

4 RESULTS AND DISCUSSION

Fig. 4 shows the interface of the Gazepoint Analysis Professional Edition (v6.0.0.), the software used for testing the participants. Developed by the company Gazepoint, the programme works in conjunction with their eye-tracking device, which monitors eye movements and fixations. While the software is primarily designed for visual presentation of results and is not suitable for advanced statistical analysis or hypothesis testing, the data it collects is highly valuable and can be imported into professional statistical tools for further analysis.

Before testing, each participant undergoes pupil calibration, ensuring that the eye-tracking system accurately follows the designated test points on the screen. The program includes a subprogramme for calibration, and testing cannot begin until all parameters are met.

Participants were investigated with single point fixation thresholds of 30, 50, 70, 90, 110, 130, 150, 170, 190, 210, 230, and 250. Fig. 5 illustrates the density threshold of fixations for 230 observations at a single point. The numerical data shows that 57 fixations were focused on the left organic shape, while 12 fixations were on the right polygonal shape.

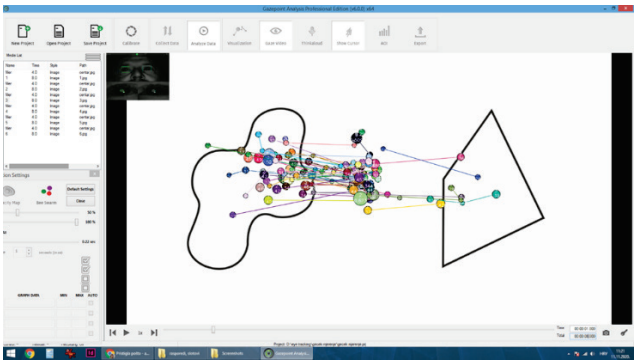


Figure 4 Presentation of fixations of Gazepoint Analysis Professional edition application for the first study sample

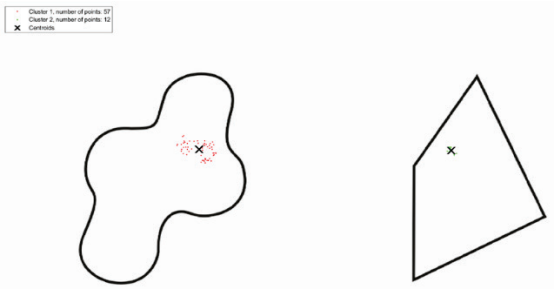


Figure 5 Fixation density threshold for 230 observations at a single point of the first study sample

The statistical analysis of the obtained data was conducted using MATLAB 2020a, with three scripts employed for the visual representation of the results. To test the hypotheses, the two-sample t-test for equal means and the Wilcoxon rank-sum test were applied, based on the results from 63 participants.

Table 1 Results of the t-test with a fixation threshold of 230 observations at a single point of the first study sample

Stupac1	hypothesis_of_equal_mean:	p-value	kurtosis_left_object	kurtosis_right_object	normal_dist_left_object	normal_dist_right_object	Mean_value_first_object	Mean_value_second_object	Statistically:	Stupac2	Stupac3
1_image	reject	5,74E-07	5,313694797		No	No	0,634930635		0 mean_first<mean_second		
2_image	reject	0,004882	18,69866583	10,66029004	No	No	0,174603175	0,650793651	mean_first<mean_second		
3_image	reject	1,65E-05	6,836369164	8,454730814	No	No	0,190476190	0,761904762	mean_first<mean_second		
4_image	reject	0,049386	5,609175073	4,588530827	No	No	0,396825397	0,714285714	mean_first<mean_second		
5_image	reject	2,08E-06		5,656666667	No	No	0	0,571428571	mean_first<mean_second		
6_image	reject	3,51E-05	8,380811184		No	No	0,603174603		0 mean_first<mean_second		

Table 2 Results of Wilcoxon test with a fixation threshold of 230 observations at a single point of the first study sample

Stupac1	hypothesis_of_equal_median:	p-value	Median_value_first_object	Median_value_second_object	Statistically:	Stupac2	Stupac3
1_image	reject	1,48E-08	0	0	median_first>median_second		
2_image	reject	0,001921	0	0	median_first<median_second		
3_image	reject	1,97E-05	0	0	median_first>median_second		
4_image	confirm	0,120982	0	0	no difference		
5_image	reject	1,5E-07	0	0	median_first<median_second		
6_image	reject	3,1E-07	0	0	median_first>median_second		

Tabs. 1 and 2 demonstrate a statistically significant difference in favour of the organic shape, with a fixation threshold of 230 observations at a single point, as shown on the left side of Fig. 5. This is further reflected in the number of fixations on each specific shape. Participants were examined using fixation thresholds of 30, 50, 70, 90, 110, 130, 150, 170, 190, 210, 230, and 250). Fig. 6 shows dot map of fixations for all participants for the first study sample.

Tab. 3 demonstrates a results of fixations for the first study sample, showing fixation thresholds from 30 to 250.

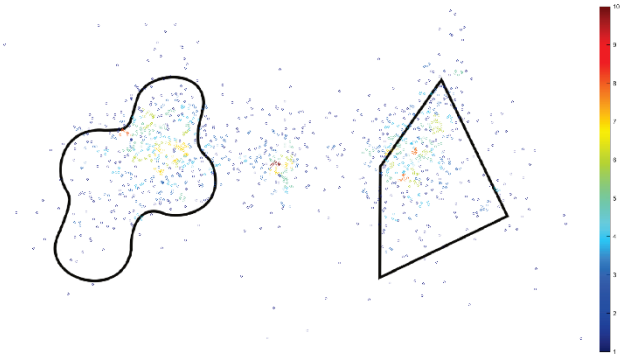


Figure 6 Dot map of fixations for all participants for the first study sample

Table 3 Results of fixations for the first study sample, showing fixation thresholds from 30 to 250

Threshold	N_points1	N_points2	N_points3	N_points3/N_points1*100%	Center1x	Center1y	Center2x	Center2y	Center3x	Center3y
30	507	237	476	93,88560158	448,87953	545,93286	470,63075	917,19375	455,1928	1381,1622
50	491	221	442	90,0203666	447,22504	553,69715	468,58997	922,32038	453,63305	1373,0483
70	459	192	392	85,40305011	445,51925	565,12282	475,28103	918,23181	444,32942	1361,3682
90	416	165	348	83,65384615	440,99616	574,63358	478,55886	903,51296	441,42483	1357,6311
110	389	135	306	78,66323907	438,76988	584,59126	483,86729	911,9445	439,16007	1355,4897
130	343	104	254	74,05247813	431,64386	588,60842	487,11196	922,66331	441,95099	1351,2783
150	285	73	219	76,84210526	426,02581	586,85504	492,36769	942,52175	447,50136	1350,5601
170	235	50	177	75,31914894	427,19752	581,7707	495,96444	944,79798	451,05181	1347,7862
190	183	30	120	65,57377049	430,1646	579,37959	500,28489	939,70236	455,50513	1343,4455
210	112	9	60	53,57142857	437,07345	585,38799	490,4269	934,99303	460,64412	1337,9025
230	57	0	12	21,05263158	448,65212	603,2396			452,49299	1365,1349
250	4	0	0	0	439,17696	630,05341				

Fig. 7 shows a thermal map of fixations of all participants in the first study sample.

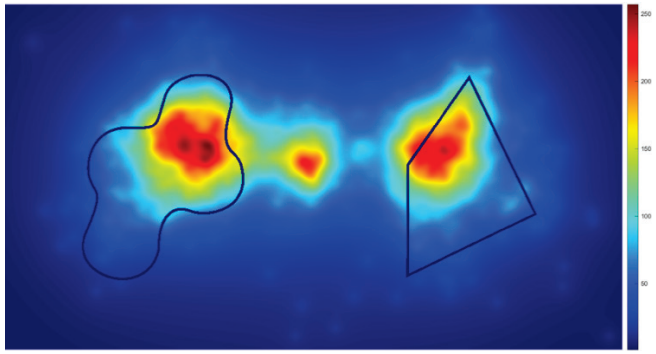


Figure 7 Thermal map of fixations of all participants in the first study sample

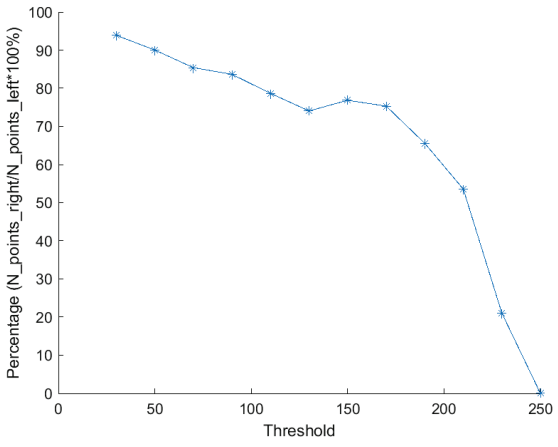


Figure 8 Percentage of fixations at different thresholds for all participants in the first study sample

Based on the results, the t-test (which favours the null hypothesis) demonstrated that the data in vectors x and y come from independent random samples with normal distributions, having equal means and equal but unknown variances. The alternative hypothesis suggests that the data in vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ come from populations with unequal means. The result of the hypothesis test is recorded as 1 if the null hypothesis is rejected at the 5% significance level, using a two-tailed test, and 0 otherwise. In other words, a result of 1 indicates that the means are statistically different, while a result of 0 indicates no significant difference.

Fig. 8 shows the percentage of fixations at different thresholds for all participants in the first study sample.

The Wilcoxon rank-sum test, on the other hand, evaluates the null hypothesis that the data in vectors x and y are samples from continuous distributions with equal medians, against the alternative hypothesis that they are not. The test assumes that the vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ are independent. The result of the hypothesis test is recorded as 1 if the null hypothesis is rejected at the 5% significance level (using a two-tailed test), and 0 otherwise. A result of 1 indicates that the medians of the two groups are statistically different, meaning there is a statistically significant difference in the fixations of subjects between the organic and polygonal forms, with the organic form showing a stronger effect. This outcome confirms hypothesis.

An analogous method was applied to a second test sample, where the positions of the shapes were swapped.

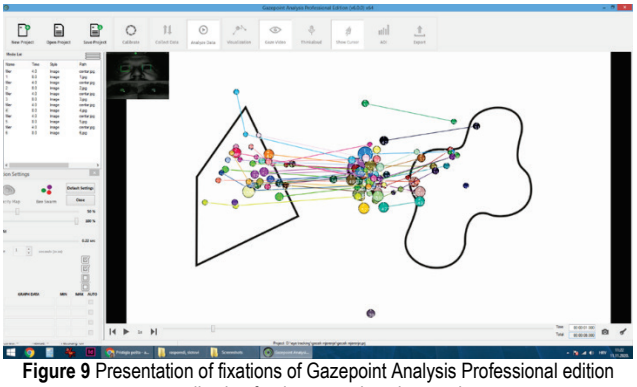


Figure 9 Presentation of fixations of Gazepoint Analysis Professional edition application for the second study sample

Fig. 10 shows dot map of fixations for all participants for the second study sample.

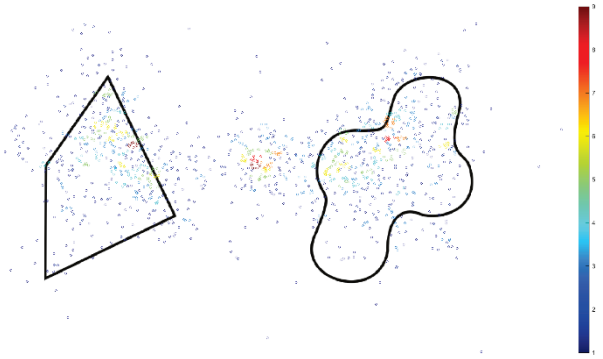


Figure 10 Dot map of fixations for all participants for the second study sample

Tab. 4 demonstrates a results of fixations for the second study sample, showing fixation thresholds from 30 to 250.

Fig. 11 shows a thermal map of fixations of all participants in the second study sample.

Table 4 Results of fixations for the second study sample, showing fixation thresholds from 30 to 250

Threshold	N_points1	N_points2	N_points3	N_points3/N_points1*100%	Center1x	Center1y	Center2x	Center2y	Center3x	Center3y
30	487	292	460	94,45585216	463,58523	491,15218	507,16448	1014,7	445,98749	1373,0032
50	460	282	445	96,73913043	462,86385	498,43141	504,2225	1017,8555	444,54968	1370,5424
70	410	261	411	100,2439024	465,52238	511,48644	496,30799	1012,6232	436,16587	1356,7854
90	364	231	401	110,1648352	454,36499	516,31627	494,18945	1001,1187	437,01091	1343,0967
110	312	207	343	109,9358974	445,01919	513,16115	490,81432	1015,1721	440,77968	1332,9086
130	269	168	294	109,2936803	445,14525	517,95041	490,12969	1000,3645	441,14864	1311,6203
150	235	121	263	111,9148936	443,58689	519,73964	483,34115	977,4239	447,02078	1281,5191
170	184	86	233	126,6304348	440,77868	517,76056	482,93176	960,82299	449,74159	1271,4937
190	144	67	188	130,5555556	443,24125	523,1004	483,65296	949,31124	449,49958	1270,2046
210	87	38	124	142,5287356	429,07059	515,83111	490,9382	944,95368	458,18136	1264,7942
230	25	25	45	180	414,2208	518,39244	492,31386	941,17334	445,37312	1255,172
250	0	0	4 Inf						487,23204	1184,6863

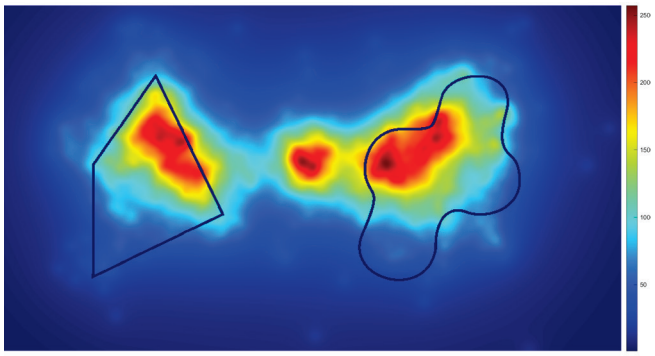


Figure 11 Thermal map of fixations of all participants in the second study sample

Fig. 12 shows the percentage of fixations at different thresholds for all participants in the second study sample.

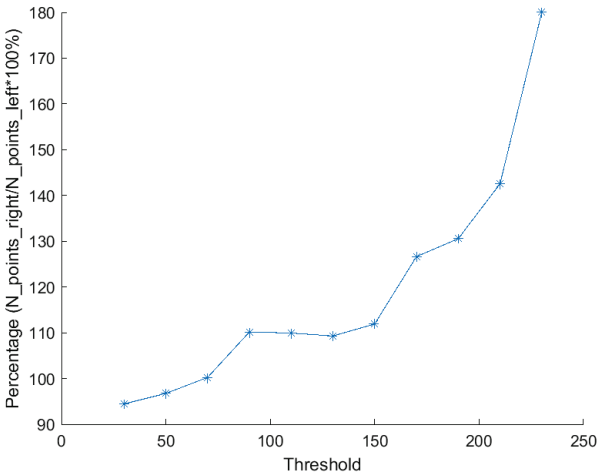


Figure 12 Percentage of fixations at different thresholds for all participants in the second study sample

It can be concluded that there is a statistically significant difference in the fixations of subjects between the organic and polygonal forms even with the replaced positions of shape observation, with the organic form showing a stronger effect. However, the fixation thresholds reveal that the right organic shape received more fixations at all measurement thresholds, except for the 50-fixation threshold at a single point.

5 CONCLUSION

This study paves the way for creating graphic and media communications tailored to the target audience, with scientifically validated parameters to select forms that minimize discomfort and instead evoke emotions such as happiness. Such messages are likely to be perceived more effectively by users.

From the initial analysis of a physical questionnaire, which was designed flexibly to encourage creativity rather than rigid responses, we observed a clear preference for organic two-dimensional shapes (68 %) over polygonal shapes (32 %). Most respondents indicated that organic forms evoked emotions of happiness and calmness, with many adding that these shapes brought them feelings of relaxation and pleasure. In contrast, polygonal shapes were most often associated with emotions such as fear, threat, and discomfort. Interestingly, even among respondents who favoured polygonal shapes, the predominant response was that these shapes instilled a sense of fear.

Statistical analysis of the experimental results confirmed the hypothesis that organic shapes have a stronger impact on visual perception than polygonal shapes in the context of graphic communications. Detailed visual analysis, supported by the fixation images from all respondents, further reinforces the conclusion that organic shapes are highly desirable in graphic and visual communication. These shapes tend to evoke emotions such as happiness and are likely to resonate innately with the target audience. This response may stem from our daily exposure to similar forms in nature, which are deeply embedded in our memory and thus more easily recognized and accepted.

The shapes we use in communication play a vital role in distinguishing products within the mass production landscape. By selecting the right forms, attention can be significantly enhanced, thereby increasing the visual appeal and overall attractiveness of graphic or multimedia products.

As communication becomes increasingly personalized, future research should explore shape preferences based on gender to identify differences in how men and women perceive organic and polygonal two-dimensional shapes. Expanding the scope, future studies could also examine the impact of colour on these shapes, exploring how individuals react to coloured organic and polygonal forms. Additionally, similar investigations should be conducted with 3D organic and polygonal shapes, focusing on their application in user interfaces and video game design.

6 REFERENCES

[1] Bar, M. & Neta, M. (2006). Humans prefer curved visual objects. *Psychological Science*, *17*(8), 645-648. https://doi.org/10.1111/j.1467-9280.2006.01759.x

[2] Leder, H., Tinio, P. L. & Bar, M. (2011). Emotional valence modulates the preference for curved objects. *Perception*, *vol 40*, 649-655. https://doi.org/10.1068/p6845

[3] Vusić, D., Geček, R. & Hajdek, K. (2016). Color appearance of the neon color spreading effect. *Acta graphica 27*(2), 7-13.

[4] Marin, M., Mrvac, N., Vusić, D. (2009). *Vizualna psihofizika i dizajn.* Varaždin: Veleučilište u Varaždinu. (in Croatian)
[5] Johnson, G. M. & Fairchild, M. D. (2003). *Visual psychophysics and color appearance.* CRC Press, Chapter in Digital Color Imaging Handbook. https://doi.org/10.1201/9781420041484.ch2
[6] Geček, R. (2021). *Determination of visual perception of organic and polygonal shapes in graphic communication processes*, Dissertation, Faculty of Graphic Arts, University of Zagreb. https://urn.nsk.hr/urn:nbn:hr:216:217649 (in Croatian)
[7] Leek, E. C., Cristino, F., Conlan, L. I., Patterson, C., Rodriguez, E. & Johnston, S. J. (2012). Eye movement patterns during the recognition of three-dimensional objects: Preferential fixation of concave surface curvature minima. *Journal of Vision*, *12*(1):7, 1-15. https://doi.org/10.1167/12.1.7
[8] Bierut, M., Drenttel, W. & Heller, S. (2002). *Looking Closer Four: Critical Writings on Graphic Design.* First Edition. New York: Allworth Press.
[9] Landow, J., Ritchie, J. & Crooks, R. (2012). *Infographics: The Power of Visual Storytelling.* First Edition. New Jersey: John Wiley & Sons, Inc.
[10] Stoklossa, U. (2007). *Advertising: New Techniques for Visual Deduction.* Second Edition. London: Thames & Hudson Ltd
[11] Smiciklas, M. (2012). *The Power of Infographics: Using pictures to communicate and connect with your audiences*. First Edition. Indianapolis: Que Publishing.
[12] Downs, S. (2011). *The Graphic Communication Handbook*. London: Routlege.
[13] Resnic, E. (2003). *Design for Communication: Conceptual Graphic Design Basics*, Wiley.

Authors' contacts:

Robert Geček, Associate Professor
University North,
J. Križanića 31b, 42000 Varaždin, Croatia
E-mail: rgecek@unin.hr

Damir Vusić, Full Professor
(Corresponding author)
University North,
J. Križanića 31b, 42000 Varaždin, Croatia
E-mail: dvusic@unin.hr

ISSN 1846-6168 (Print), ISSN 1848-5588 (Online)
https://doi.org/10.31803/tg-20241231152941

Original scientific paper
Received: 2024-12-31, Accepted: 2025-01-21

Structure Formation as a Process of Mutual Adaptation of a Product and Material

Valeriy Vyrovoy, Svitlana Semenova, Diana Zenchenko, Aleksej Aniskin*

Abstract: The article analyzes the influence of the product geometry on the adaptation processes occurring during the structure formation of building composite materials. One of the main aspects determining the structure formation processes are the boundary conditions depending on the product geometry. It is shown that changing the product geometry significantly affects the distribution of stresses and strains, the formation of inner surfaces of partitions and cracks. Active structure elements, such as internal surfaces of partitions and microcracks, are key participants in the adaptation processes, providing stress redistribution. The applied graphoanalytical method demonstrates the possibility of visualizing and analyzing the distribution of deformations. The methods of image analysis and damage coefficient assessment allow us to quantitatively evaluate the adaptive capabilities of the material. Thus, the geometric parameters of the product determine the processes of structure formation in materials and its performance characteristics along with the formulation and technological factors.

Keywords: adaptation; cracks; composite; deformations; product shape; structure formation; surfaces of partition

1 INTRODUCTION

Ensuring the functionality of products made of composite building materials, as well as maintaining their integrity, are determined by continuous dynamic processes of structure formation [1, 2]. The material undergoes adaptive changes due to dynamic changes in external conditions and internal states [3]. Active elements of the structure, which include internal inner surfaces of partition, cracks, as well as local and integral residual (initial, technological, genetic) deformations, selectively participate in the adaptation processes [4]. Active elements, representing a single system with other elements of the material structure, can be transformed against the background of a change in the distribution of residual deformations. The transformed elements adapt to each other, to their environment and to a change in the deformed state - unique adaptation processes are realized. Thus, a change in the parameters of some elements provokes a certain reaction of others, which ensures the non-stationarity of the processes of structure formation - unique adaptation phenomena as the ability of a system to develop according to the principle "from what has been achieved" in such a way as to preserve its basic properties through adaptive transformation.

One of the ways to influence the organization of the material structure is to change the geometric shape of the product [5, 6]. To ensure one or more properties of products made of building composites, an unlimited number of structures can be selected. The properties of a material of the same composition under the same curing conditions change depending on the design features of the product. One of the reasons for such a change in properties is residual local and integral deformations.

Residual deformations are present in products, in fiber-reinforced composites, and are formed because of differences in the values of the coefficients of thermal expansion (or shrinkage) of the components of materials in the composite [7]. During the technological period of obtaining materials, the physicochemical processes of hydration of mineral binders, polymerization and polycondensation reactions of polymeric materials, curing of metals and other materials are accompanied by volumetric deformations [8]. Volumetric changes should be considered as summary processes with many components that arise and develop in materials of different nature and purpose during their processing into a product and during the action of operational loads.

The places of occurrence and the nature of development of residual deformation gradients in magnitude and direction depend on the features of the geometry of the products, which can initiate the formation of process cracks and determine the conditions for the formation of the stress-strain state of the finished product [9, 10]. At the same time, the material implements processes of adaptation of the structure to the changes that have occurred to ensure the integrity and functionality of the product. Directed changes in the geometry of the product will reduce deformation gradients to a minimum, which, in turn, will make it possible to manage the technological damage of products and regulate their physical and technical indicators.

The propose of the work is to research the influence of changes in the shape of products on the structure formation of materials as on an adaptation process, which will allow further improvement of the characteristics of products by selecting their geometric configurations.

2 MATERIALS AND METHODS OF RESEARCH

The analysis was carried out on cement samples with different geometric characteristics. The sizes and shapes of the samples are shown in Fig. 1. Fig. 2 shows rectangular samples with cracks sawn into the finished sample (Fig. 2a) and with cracks introduced during the sample molding process (Fig. 2b).

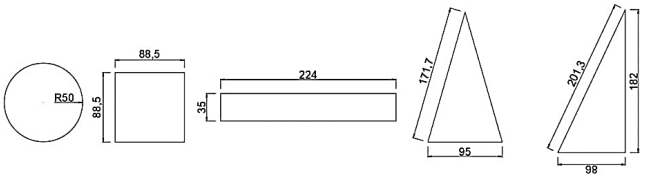


Figure 1 Types and geometric parameters of models

The distribution pattern of deformations was analyzed using a graphoanalytical method, which is used to study the

distribution of shrinkage deformations and shrinkage kinetics when changing the shape of a sample [11]. The use of the graphoanalytical method facilitates the analysis of the deformations that arise due to the visualization of their distribution pattern. In addition, the method is applicable to complex shapes and materials, which allows the design features to be taken into account.

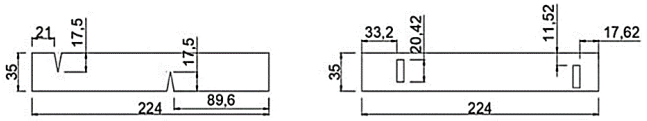


Figure 2 Types and geometric parameters of models: a - sample with a sawn crack; b - sample with a built-in crack

Triangular and round cement samples were also used (Fig. 3), in which the inner surfaces of partition and cracks that arose during the formation of the structure and adaptation of the material to the shape of the product, were revealed using tannin. Then, using image-processing methods [12, 13] in the *Scion Image* software [14], the structural characteristics of the samples, including the damage coefficient, were determined.

3 RESULTS OF THE RESEARCH

The distribution patterns of shrinkage deformations (Figs. 3 and 4) were obtained for samples of different shapes and sizesusing the graphoanalytical method as a result of the analysis of the sample models (Figs. 1 and 2). Cracks formed in the already hardened sample in the sample in Fig. 4c, and cracks formed at the initial stage of product formation in the sample in Fig. 4d.

To construct the deformation diagrams a section was selected passing through the axis of the model and perpendicular to its lateral face (Figs. 3 and 4) in each model.

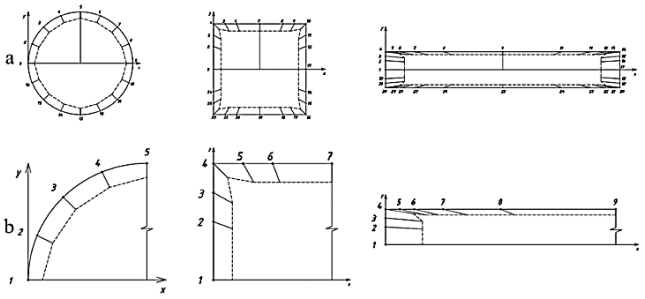


Figure 3 Formation of distribution of technological deformations in models in the process of structure formation: a – first stage; b – fragment of the model

It should be noted that when analyzing using the graphoanalytical method, the material is considered as an isotropic and continuous medium. In this case, each point lying on the interface interacts with all points belonging to its own interface and with all "visible" points on the sample surface and at the interface with voids.

By comparing the shrinkage deformation distribution diagrams in Figs. 3 and 4, it can be noted that the nature of the formation of residual (technological, initial, hereditary) deformations significantly depends on the geometric characteristics of samples and products. The introduction of a coordinate system made it possible to construct an absolute direction of movement of all points of the models, taking into account their geometric features. The movements of all points were brought to the same scale for a quantitative assessment of the obtained results. In each model, the value of the movement of the point related to the *x*-axis was taken as one. The adopted technique allows one to qualitatively assess the change in the nature of the distribution of residual deformations with the allocation of such parameters as the relative magnitude of deformations and the direction of their action through the angle φ.

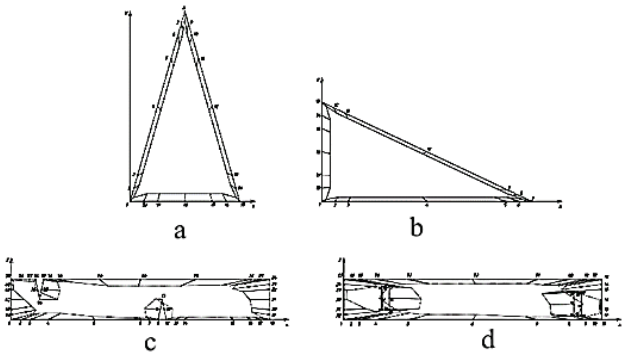


Figure 4 Formation of distribution of technological deformations in models in the process of structure formation

Each model, depending on its shape, develops an individual distribution pattern of technological deformations as can be seen from Figs. 3 and 4, The configurations of the shape of the outer boundary of the models cause the emergence of deformation gradients (Figs. 3a and 4). The curvature of the interface creates conditions for the concentration of multidirectional deformations in places of maximum shape change.

The analysis carried out using the graph-analytical method allows us to conclude that the formation of the distribution pattern of process deformations is a kinetic process that entails constant changes in the configuration of the inner surfaces of partition. Prerequisites for the formation of process cracks and internal inner surfaces of partition arise in areas of the greatest shape change, due to the concentration of multidirectional deformations.

Cement samples (Fig. 5) with a manifested structure were used for visual observation of the process of structure organization. The geometric characteristics of the products were determined according to the models in of a given shape (Figs. 1 and 2).

It is evident from Fig. 5 that the distribution pattern of the internal inner surfaces of partition and cracks significantly depends on the shape of the sample, which confirms the results of the analysis of the distribution of residual deformations in model samples using the graph-analytical method.

It should be noted that one of the aspects of the structure formation of composite materials is the formation of cracks, internal inner surfaces of partition and pores in samples and

products. Since the products are characterized, on the one hand, by the presence of an individual shape and, on the other hand, by the composition and structure of the composite, all these factors in interaction will determine the crack resistance and porosity of the products and, in the case under study, the distribution of internal inner surfaces of partition. The factor that largely predetermines their formation is, as shown above, shrinkage deformations. Thus, for samples of different shapes made of the same material and even for different sections of the same asymmetric sample, completely different geometric characteristics of the active elements of the structure are characteristic.

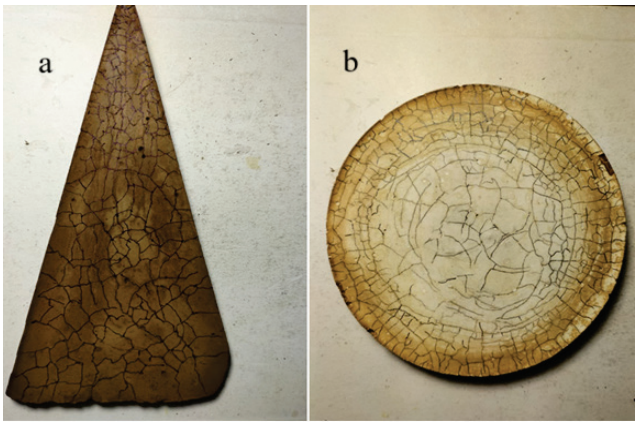


Figure 5 Distribution of internal interfaces and cracks in samples of different shapes: a – triangular; b – round sample

The considered facts can be confirmed by the method of computer processing of images of the studied samples, the first step of which is the "manifestation" of the pattern of internal inner surfaces of partition surrounding each discrete module. The image was transformed to a gray scale, the brightness level was equalized by the rolling sphere method, the boundaries were highlighted, and a median filter was used. As a result, images similar to those shown in Fig. 6 were obtained.

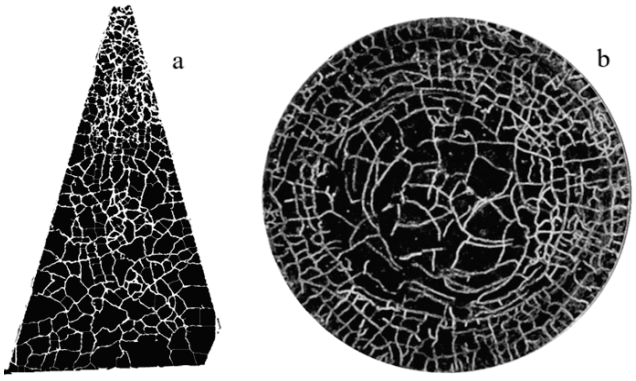


Figure 6 The manifested picture of the distribution of inner surfaces of partition and cracks in samples of different shapes: a – triangular; b – round sample

After threshold division for the samples, the damage coefficients were determined in local geometrically inhomogeneous areas with significantly different shrinkage deformation patterns arising due to the geometric inhomogeneity of the studied areas. Thus, for the triangular sample (Fig. 6a), the parts containing the acute angle (Fig. 7a) and the base (Fig. 7b) separately was studied, in the round sample (Fig. 6b) - the central (Fig. 7c) and peripheral parts (Fig. 7d).

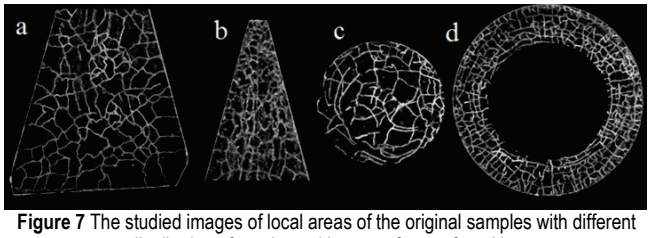


Figure 7 The studied images of local areas of the original samples with different distribution of cracks and inner surfaces of partition

The damage coefficient was estimated using the cumulative polygons method. For the analyzed fragments on a black background (Fig. 7), histograms were constructed, and then, based on them, cumulative polygons were constructed, visually corresponding to some curves. An example of construction for the leftmost fragment (Fig. 7a) is shown in Fig. 8.

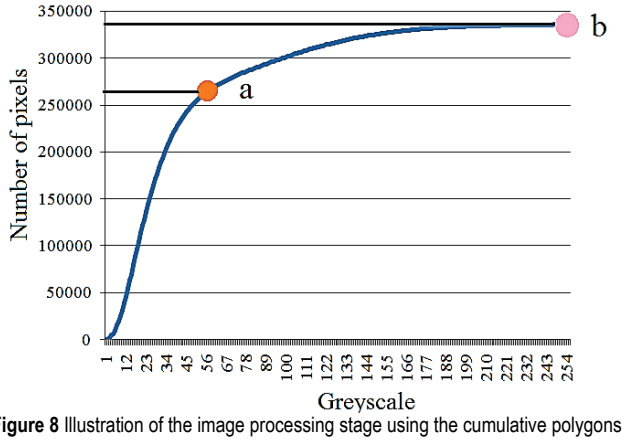


Figure 8 Illustration of the image processing stage using the cumulative polygons method

In the curve corresponding to the polygon, the point of greatest curvature was determined, as well as the maximum number of pixels corresponding to the projection area of the sample under study. In the work under consideration, the definition of inflection points was carried out visually, but if necessary, it can be carried out by constructing a curvature graph. It should be noted that the light lines correspond to the boundaries of the section. Then the damage coefficient can be estimated $\tilde{K}$ using Eq. (1):

$$\tilde{K} = \frac{N_{\max} - N_{\text{borders}}}{N_{\max}}, \quad (1)$$

where $N_{\max}$ is the maximum number of pixels of the object under study; N_{borders} is the number of pixels at the point of maximum curvature.

Because of applying the method under consideration, the following data were obtained (Tab. 1).

Table 1 Assessment of the damage coefficient for different sections of triangular and round-shaped samples

Researchable sample	N_{max}	$N_{borders}$	$\widetilde{K}$
Base of triangle (Fig. 6a)	336117	260867	0,22388
Apex of triangle (Fig. 6b)	102091	60000	0,412289
Center of round sample (Fig. 6c)	198901	136565	0,313402
Periphery of round sample (Fig. 6d)	324858	165940	0,489192

The data (Tab. 1) clearly shows that the damage coefficient differs in sections of products with different shapes. It is particularly high in sections where structure formation occurs under restricted conditions like wall effects and high curvature values. This highlights the importance of considering factors such as material structure, composition, and product geometry systematically. Shrinkage deformations, especially in the area of their unevenness, create a "canvas" for the formation of internal inner surfaces of partition and cracks.

It should also be noted that the nature of the distribution of internal inner surfaces of partition and cracks in the samples in Fig. 5 shows that structural blocks are formed in the continuous medium of the material. Inner surfaces of partition initiate the appearance of discrete structures, and the material adapts to the changes occurring in the system through these surfaces. The shape and configuration of these structural blocks, determined by the network of inner surfaces of partition and cracks formed because of the structure formation processes, varies depending on the location of the material section in the sample.

4 CONCLUSION

The geometric shape of the product determines the distribution of deformation in the material, as well as the configuration of the network of internal inner surfaces of partition and the nature of crack propagation. Changing the geometry of the samples affects the nature of such a network, especially near lines and surfaces of significant curvature. Controlling the structure of this network through changing the geometric shapes of the product is one of the ways to redistribute internal stresses and deformations, and, consequently, reduce the probability of destruction. Since the geometric characteristics of the network under consideration are adapted to the boundary conditions expressed macroscopically in the form of the sample boundaries, the analyzed integral structure-forming process is an adaptive property of the product as a system.

5 REFERENCES

[1] Vyrovoy, V. N., Dorofeev, V. S. & Sukhanov, V. G. (2010). *Composite construction materials and structures: Structure, self-organization, properties*. Odessa: ODABA. (in Russian)

[2] Gusev, B. V. at al. (2006). *Formation of the structure of composite materials and their properties*. Moscow: Nauchny Mir. (in Russian)

[3] Chernyavskiy, V. L. (2002). *Adaptation of concrete:* Monograph. Dnipro: Nova Ideologiya. (in Russian)

[4] Vyrovoy, V. M., Korobko, O. O., Sukhanov, V. G., Kazmirchuk, N. V. & Makarova, S. S. (2020). *Formation and destruction of building composites*: Textbook. Odessa: ODABA. (in Ukrainian)

[5] Vyrovoy, V. M., Sukhanov, V. G. & Korobko, O. O. (2022). *Material structure in the structure of the construction* (2nd ed., revised and expanded). Odessa: ODABA. (in Ukrainian)

[6] Dvorkin, L. Y. (2022). *Theoretical foundations of building materials science*. Rivne: NUVGP. (in Ukrainian)

[7] Solodkyi, S. Y. (2008). *Crack resistance of concretes based on modified cements*: Monograph. Lviv: Lvivska Politechnika. (in Ukrainian)

[8] Vyrovoy, V. N., Dovgan, I. V. & Semenova, S. V. (2004). *Features of structure formation and property development of polymer composite materials*. Odessa: TES. (in Russian)

[9] Lermit, R. (2017). *Problems of concrete technology* (Translated from French, stereotype edition). (in Russian)

[10] Korobko, O., Vyrovoy, V., Zakorchemny, Yu. & Kushnir, O. (2019). Formation of technological damage of building constructions by regulation of deformations. *Actual Problems of Engineering Mechanics*: Materials of the 6th International Conference, Vol. 968, 324-329.

[11] Solomatov, V. I., Vyrovoy, V. N., Dorofeev, V. S. & Sirenko, A. V. (1991). *Composite construction materials and structures with reduced material consumption*. Kyiv: Budivelnyk. (in Russian)

[12] Russ, J. C. (1990). *Computer-assisted microscopy: The measurement and analysis of images* (1st ed.). Springer.

[13] Kolesnykov, A. V., Semenova, S. V., Vyrovoy, V. M. & Kersh, V. Y. (2024). *Image processing methods in building materials science problems*: Monograph. Odessa: ODABA. (in Ukrainian)

[14] https://mesonpiold.cbpf.br/e2002/cursos/NotasAula/ScnImage.pdf

Authors' contacts:

Valeriy Vyrovoy, Doctor of Technical Science, Professor
Odessa State Academy of Civil Engineering and Architecture,
Construction and Technological Institute,
Department of Production of Construction Products and Structures,
4, Didrikhson str., 65029 Odesa, Ukraine
vyrovoy@ukr.net

Svitlana Semenova, PhD, Assistant Professor, Head of the Dep.
Department of Chemistry and Ecology of
Odessa State Academy of Civil Engineering and Architecture,
Construction and Technological Institute,
4, Didrikhson str., Odesa, Ukraine
Tel. +380503366911
semenova@odaba.edu.ua

Diana Zenchenko, Postgraduate Student
Odessa State Academy of Civil Engineering and Architecture,
4, Didrikhson str., Odesa, Ukraine
923dianazenchenko@odaba.edu.ua

Aleksej Aniskin, PhD, Assoc. Professor
(Corresponding author)
University North, Department of Civil Engineering,
Jurja Križanića 31b, 42000 Varaždin, Croatia
aaniskin@unin.hr

ISSN 1846-6168 (Print), ISSN 1848-5588 (Online)
https://doi.org/10.31803/tg-20241113001849

Original scientific paper
Received: 2024-11-13, Accepted: 2025-01-31

Artificial Intelligence vs. Traditional Research Methods: An Empirical Study from Northern Croatia

Katerina Fotova Čiković*, Antonija Mandić, Maja Hoić

Abstract: Due to the emergence and increased development of Artificial Intelligence (AI), research in general has been significantly impacted, particularly in the field of scientific theories and models. The purpose of this study is to analyze the acceptance of both AI tools and traditional methodologies used in research. Moreover, conclusions about the respondents' perception and openness to using AI tools in research regarding gender, age and current academic position are discussed. Another goal is to compare the level of satisfaction from both the AI tools and the traditional research methods. A questionnaire-based survey was carried out between February and March 2024, and it included students and teaching staff at the University North in Croatia. The novelty of this research is mirrored in the scarcity of such empirical studies encompassing the academic community in Croatia.

Keywords: artificial intelligence; Croatia; information retrieval; research methodology

1 INTRODUCTION

Artificial Intelligence (AI) plays an important role in every aspect of people's lives and represents a transformative technology with significant potential for both risk and opportunity in various industries [1]. It is particularly crucial for maximizing the benefits of the Internet of Things and has already revolutionized many industries, including research, manufacturing, finance, and healthcare [2].

The rise and development of AI were mostly influenced by the new media era, with AI technology being optimized and applied in the field of computers and intelligent terminals. AI has thus far revolutionized various, if not all industries. This unprecedented shift is mirrored in the increased integration and application of AI into everyday life, presenting both opportunities and challenges [3]. Generative AI models created "a disruptive impact on teaching and learning, due to their ability to create text, images, and sound, revolutionizing educational content creation and modification" [4].

Desjardins-Proulx discussed the paradigm shift in science and research brought about by AI, especially in the design of algorithms for model-building [5]. Moreover, AI is changing the way people search for information. According to Waly AI is already very much present in the scientific research process [6]. Some of the features and advantages of AI technology in scientific research include speeding up research work, especially literature search and selection, given the emergence of numerous tools that serve as alternatives to traditional search tools. In the last few years, the intensive use of AI tools in higher education is discussed [7]. The use of AI-powered language tools and their impact on the student population is evident. AI tools have been integrated into a wide range of platforms including browsers, social networks, games, and other applications that students and teachers have been using for a long time. However, they were not aware that these applications and tools were based on AI technology. Therefore, it can be established that AI technology has been used in education and science for many years, but the scale of the new tools opens a new niche for researchers to explore. AI tools should be viewed as aids in the information search process rather than as ready-made solutions. They are useful tools that can enhance the creativity of scientists and students and simplify the search process. A crucial aspect of utilizing AI tools to search scientific information is the use of reliable tools connected to pertinent databases. AI presents challenges for scientists and science in general. While AI may have a positive impact, the realisation of benefits relies on the actions and decisions of human users [8]. While using AI tools, one should not forget the importance of traditional information sources.

The integration of AI in academic research tools has significantly increased, particularly following the release of OpenAI's ChatGPT. The proliferation of AI tools and research assistants for scientific publication retrieval over the past three years highlights a trend towards using advanced AI models to keep pace with the growing body of research. This trend is evident in the way these tools are marketed and promoted, emphasizing their ability to handle large volumes of data and provide more relevant search results [9]. AI tools significantly improve the precision and relevance of search results through advanced natural language processing (NLP) and semantic search techniques. By automating the literature review process and providing AI-generated summaries, these tools promise to save researchers considerable time, addressing the ongoing pressure scientists face to publish and manage a constant workload. However, many advanced features in AI tools require subscriptions, indicating a shift towards the commercialization of academic research tools. This trend highlights the increasing reliance on paid services to access advanced AI capabilities in academic research [10]. Globally renowned scientific databases such as Scopus and Web of Science rely on lexical and keyword-based search mechanisms. Their search engines use probabilistic approaches to match search terms with indexed documents, which are heavily dependent on term frequency. This method, while effective to an extent, often provides irrelevant results, necessitating manual filtering by researchers.

Additionally, access to these databases typically requires an institutional subscription, limiting availability to affiliated researchers and academics. In contrast, AI-powered tools like Semantic Scholar and Scispace employ NLP and semantic search techniques. These tools convert text into numerical vectors, enabling a deeper understanding of context and relationships between terms. This approach results in more accurate and relevant search outcomes, addressing the issue of information overload.

Semantic Scholar is one of the oldest popular usages of AI in reference searching. It was used even before the big boom of the new online tools that came to be with the proliferation of the OpenAI AI model. Semantic Scholar, launched by the Allen Institute for AI, utilizes the SPECTER2 model to process and generate word embeddings for document titles and abstracts. The tool's AI-generated summaries (TLDRs) and custom folder organization further enhance its utility. Semantic Scholar sources data from major publishers and databases such as PubMed, Springer Nature, and IEEE, providing comprehensive coverage of scientific literature.

Elicit, developed by Ought in 2021, leverages AI to accelerate literature reviews by extracting key publications from Semantic Scholar [11]. Created in the years following the release of the OpenAI API, Elicit uses a fine-tuned version of the GPT-3 model to perform its tasks. Unlike traditional search tools, Elicit and Scite can identify relevant studies without perfect keyword matches, structuring queries in the form of specific research questions [12]. This capability is particularly beneficial for hypothesis-driven research, where precise data extraction and synthesis are critical. Scispace integrates semantic search with a vast corpus of over 280 million papers, utilizing GPT-3 to offer interactive literature reviews, PDF-based question answering, data extraction, and citation generation. This tool's AI assistant facilitates an in-depth exploration of scientific literature, though access to premium features requires a subscription, reflecting a growing trend towards the commercialization of AI-driven research tools. Sourcely and Minerva cater to specific aspects of reference management. Sourcely analyses longer texts to find relevant references, though it has limitations in language support and filtering options. Minerva focuses on the biomedical field, combining large language models (LLMs) with structured graphs to display interdisciplinary connections and generate summaries from PubMed articles.

In this study, the main objective was to determine how AI tools are associated with various academic positions. For this purpose, an extensive empirical study on a sample of 69 has been done in the period February–March 2024 at the University North in Croatia. The questionnaire was divided into 3 parts. The first part of the questionnaire covered the demographic characteristics of the respondents, the second part of the survey was related to the familiarity and use of AI tools, and the third related to the use of traditional databases in research.

The main scientific contribution of this study is mirrored in the fact that it comprises the first-ever empirical study in the Republic of Croatia, and most probably among the first empirical studies on AI in research globally. The rest of this paper is structured as follows. After the introduction, in Section 2, an extensive literature review of relevant published papers on AI tools, their application in the academic community, their use by different age groups as well as the perception and use of AI versus the traditional research methods is provided. Section 3 includes the methods used and the used sample. Thereafter, in Section 4, the research results are revealed. Section 5 opens a discussion regarding the results and adds concluding remarks.

2 LITERATURE REVIEW

The integration of AI into various aspects of human life, including scientific research, is increasingly widespread. Today, AI tools are assisting scientists in overcoming obstacles related to the growing demand for publishing their work. Scientists face the primary challenge of effectively managing restricted time, limited resources and budget, and finite cognitive capacity. These factors are pivotal when utilizing AI tools to access scientific information. AI offers opportunities for enhanced productivity, enabling scientists to publish more work and fostering greater objectivity in their research endeavours. AI is a branch of computer science that focuses on creating intelligent programs (systems) that largely "imitate" human intelligence [6]. The definition of AI has evolved over the years to include tools that can perform cognitive tasks, particularly learning and problem-solving, with technological innovations such as machine learning and neural networks [13]. According to the citation database Web of Science, 157,947 papers are available on the topic of "AI." The oldest available paper dates to 1960 [14]. The numerical indicators of works show a slight growth in 1984, with increasing interest in the topic from 1990 to 1991. Since 2018, AI has become a "hot" topic, as the number of published papers on this subject has reached high numbers. The countries showing the most interest in this topic are the USA and China. In the field of information librarian science, there are 2,019 available works, and the oldest work is from 1972 by Reilly, KD, titled "Computers and AI."

The scientific community's increasing interest is connected to using AI tools in the scientific information search process, providing a quicker alternative to traditional information search methods.

Analysing the citation database of Web of Science, focusing on AI tools used for the literature search process, it was searched using the key term in the category all fields "AI tools and academic research". The number of papers on this topic is 2276. The oldest paper appears in the early nineties of the 20th century and includes topics dealing with hybrid learning and simulations in computer science.

Most of the papers that are thematically based on AI tools and scientific research are in the field of Computer Science AI and Education Educational research. Using selection criteria on these two fields, 325 papers were extracted. The thematic frameworks of the searched works cover different topics. The main topics are AI tools in academic writing, AI tools like teaching and learning assistants, and ethical components of using AI tools.

Different approaches are described like frameworks for the acceptable use of generative AI in student academic research and writing [15] or evaluating the efficacy of AI content detection tools in differentiating between human and AI-generated text [16].

AI in education can help students develop and improve personalized learning and enhance and improve learning skills [17]. AI-based tools have the potential to customize the learning experience, boost productivity, and increase student engagement [18]. The papers encompass a variety of tools designed to provide learning support. These tools are utilized by teachers to facilitate and enhance the overall learning experience for students. Studying the development of new methods that provide a personalized approach to teaching and learning, including integration of artificial intelligence (AI) tools that can revolutionize education [19].

The ethical aspect is not a new concept in scientific research; therefore, it is logical to maintain interest in this topic, especially regarding AI tools. AI tools can cause both positive and negative consequences. Ethical considerations include data security, safety risks associated with autonomous technologies, and fairness in decision-making processes [6]. In the context of researching the ethical dimensions of using AI tools for scientific purposes, the study focuses on tools that can detect plagiarism in scientific papers [20].

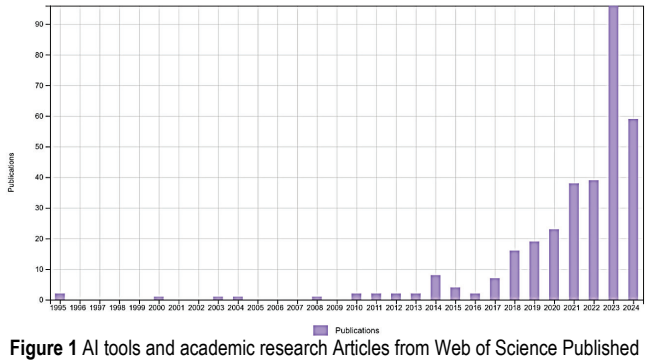


Figure 1 AI tools and academic research Articles from Web of Science Published by Year

Information experts, especially librarians specializing in science, play a vital role in identifying the most effective tools and educating users, such as scientists and students, on how to recognize and utilize these tools correctly. The importance of AI literacy is growing, and this literacy needs to be seamlessly integrated into the information literacy framework. Being AI literate includes knowledge, understanding, use, evaluation, and ethical principles [21]. As information specialists, librarians have the task of familiarizing themselves with the tools that are offered, recognizing their values and shortcomings, and familiarizing their users with their applications. AI literacy has emerged as a new skill set in response to this new era of intelligence.

Finally, as Sağın stated, the academic community is nowadays "polarized, with some embracing AI for its accessibility and efficiency thus advocating it as an indispensable tool, while others cautioning against risks to academic integrity and intellectual development" [4].

3 METHODS AND SAMPLE

To conduct this research, an online anonymous questionnaire was conducted in the period January – March 2024 at the University North in Croatia and included all academic positions (i.e. students and academics). The questionnaire was divided into 3 parts. The first part of the questionnaire covered the demographic characteristics of the respondents, the second part of the survey was related to the familiarity and use of AI tools, and the third related to the use of traditional databases in research. The analysis of data collected based on an anonymous questionnaire is based on statistical methods.

The sample of respondents was described by distribution according to defined characteristics: gender, age groups and academic position at the University North. Selected descriptive statistical indicators were calculated and interpreted for quantitative variables. The distribution of quantitative variables was assessed using the Kolmogorov-Smirnov test. It was found to significantly deviate from a normal distribution. Therefore, the non-parametric Mann-Whitney U test (two-tailed) and the Kruskal-Wallis test were employed to evaluate the statistical significance of the differences among the defined groups of respondents. Differences confirmed at the $p<0.05$ level were considered statistically significant.

The sample consisted of 69 respondents from the academic community of the University North, i.e. respondents who participated and confirmed their familiarity with and use of AI tools. Tab. 1 reveals the distribution of respondents according to gender, age groups and academic position. The sample consisted of 42% male respondents and 58% female respondents. With regard to age, the respondents were divided into three age groups. The first age group consisted of respondents of Generation Z, i.e. respondents aged 18 to 27 years. The second group consisted of millennials, i.e. respondents aged 28 to 43, and the third group respondents aged 44 or older, i.e. Generation X. The first age group was the most numerous. Regarding the academic position at the University North, one-fourth respondents were teaching staff, and the rest were students. The most numerous were undergraduate students.

Table 1 Distribution of respondents according to gender, age and academic position

		Number of respondents	Percentage
Gender	Male	29	42
	Female	40	58
Age groups	Generation Z	32	46
	Millennials	22	32
	Generation X	15	22
Academic position	Teaching staff	17	25
	Undergraduate students	32	46
	Graduate students	20	29

In the second part of the survey, the level of familiarity, use and attitudes towards AI tools for searching and surveying scientific databases and their comparison with traditional databases among members of the academic community of the University North in Croatia.

4 RESULTS

In Fig. 2, the frequency of use of traditional databases to which the University North has a subscription is presented. The results revealed new insights into the practical functions of AI tools in a small sample of the academic community in Croatia. 38% of the respondents answered they sometimes use traditional databases for research, 19% of them claimed they often use traditional databases and 4% very often. Interestingly, 20% of the respondents never use the traditional databases and 19% of them stated they rarely make use of the traditional databases for research. This issue calls for in-detail exploration in future work.

Regarding the question related to the use of AI tools, multiple answers were available to the respondents. More than one-third of the respondents use the Semantic Scholar AI tool when searching scientific databases (Fig. 3). Open Read and Scite tools are used by 19% of them. Next in order of use are PapersGPT (under development), SciSpace, Connected Papers Elicit and Research Rabbi. Other tools were used in less than 10% of cases.

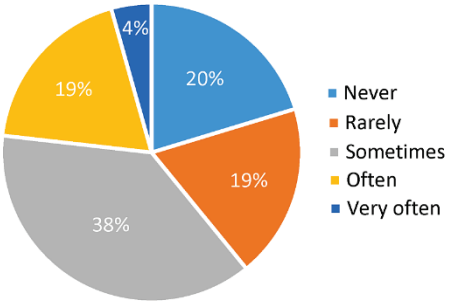


Figure 2 Frequency of use of traditional databases for research

When asked about the academic purpose of their employment of AI tools in research, most of the respondents (more than half of them) answered they use AI tools for initial literature review, whereas the rest of the answers were distributed between "for a more detailed search of relevant works", "for organization and management of references" and "for analysis and synthesis of scientific information". According to the data shown in Fig. 4, more than half of the respondents use AI tools for a more detailed search for relevant works. AI tools for the organization and management of references and the analysis and synthesis of scientific information are used in a slightly lower but similar percentage.

The conducted survey sought to determine the attitudes towards the use and perception of AI tools for browsing scientific databases at the University North in Croatia. Respondents evaluated their views on a five-point Likert scale, ranging from 1 (insufficient rating) to 5 (excellent rating).

Male respondents evaluated their satisfaction with the use of AI with a slightly higher average rating than female respondents (Tab. 2). The median value was the same for both groups of subjects. According to the results of the Mann-Whitney test, no statistically significant difference in attitudes about the use of AI tools was confirmed by gender. Gender does not influence on the use and perception of AI tools for searching scientific databases.

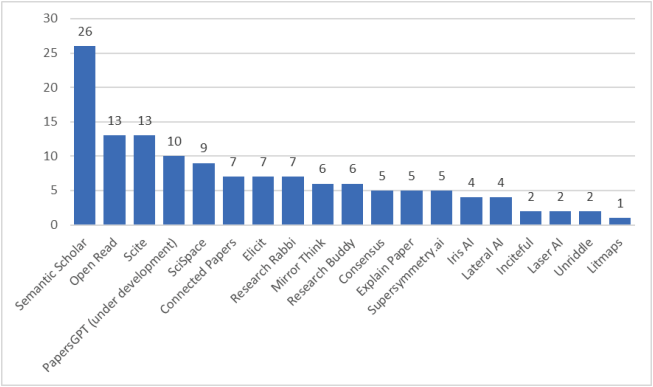


Figure 3 Use of specific AI tools in research

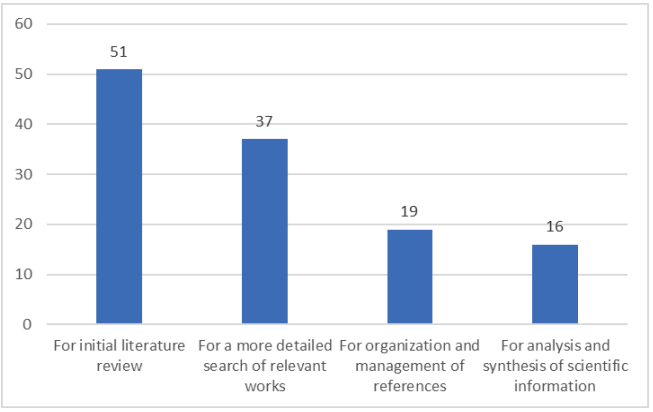


Figure 4 Academic purpose of using AI tools in research

Table 2 Respondents' views on the use and perception of AI tools for searching scientific databases by gender

Gender				Mann-Whitney test	
Male		Female			
Mean	Median	Mean	Median	*Z*	*p*
3.89	4	3.77	4	-0.480	0 .631

Table 3 Respondents' views on the use and perception of AI tools for searching scientific databases by age group.

Age groups						Kruskal-Wallis test	
Generation Z		Millennials		Generation X			
Mean	Median	Mean	Median	Mean	Median	*H*	*p*
3.71	4	4.09	4	3.73	4	3.226	0.199

Table 4 Respondents' views on the use and perception of AI tools for searching scientific databases by academic position.

Academic position						Kruskal-Wallis test	
Undergraduate studies		Graduate studies		Teaching staff			
Mean	Median	Mean	Median	Mean	Median	*H*	*p*
3.79	4	3.90	4	3.8	4	0.089	0.956

Table 5 Respondents' satisfaction with search results obtained through AI tools and traditional databases

Respondents' satisfaction				Mann-Whitney test	
AI tools		Traditional databases			
Mean	Median	Mean	Median	*Z*	*p*
3.83	4	3.52	4	2.253	0.024

With regard to age groups, an average score greater than 4 was calculated only for millennials, while the average scores for Generation Z and X were almost equal (Tab. 3). The median value for all 3 analyzed groups was 4. The results of the Kruskal-Wallis test indicate that there is no statistically

significant difference in attitudes between respondents of the three analyzed age groups.

According to the selected descriptive statistical indicators and the results of the Kruskal-Wallis test (Table 4), it is clear that there is no statistically significant difference in the opinions of respondents with different academic positions about their views on the use and perception of AI tools for searching scientific databases.

To examine whether there is a statistically significant difference in satisfaction with search results obtained through AI tools or obtained through traditional databases, the Mann-Whitney test was applied. The test results are given in Table 5.

The respondents expressed slightly higher average satisfaction with search results obtained through AI tools than those obtained through traditional databases. The median in both groups was 4. The results of the Mann-Whitney test indicate there is a statistically significant difference in the satisfaction of the search results obtained through AI tools and traditional databases among respondents at the 5% significance level.

5 DISCUSSION AND CONCLUSION

Due to its potential advantages over conventional research methodologies, AI is being applied in a growing number of study disciplines. By evaluating student data and suggesting activities, AI can tailor learning experiences in the classroom [22]. AI, in particular neural networks, has been shown to perform more accurately and flexibly than conventional forecasting techniques [23]. With an emphasis on efficacy and effectiveness, AI techniques are frequently employed to anticipate treatment results [24].

Furthermore, numerous studies have shown how much more accurate and versatile some AI techniques—like neural networks—are than more conventional predicting techniques. Recurrent neural networks (RNNs) surpassed conventional econometric techniques in the prediction of conditional volatility, according to Bucci [25]. Forecasting accuracy was increased by Khashei & Bijari's hybrid model, which combined auto-regressive integrated moving average (ARIMA) methods with artificial neural networks [26]. Ghiassi introduced a dynamic neural network model that outperformed traditional neural network and ARIMA models in forecasting time series events [27]. Rasp & Lerch demonstrated that neural networks could significantly outperform benchmark methods in post-processing ensemble weather forecasts [28]. All these papers demonstrate how powerful AI techniques may be to improve research methodologies in a variety of research disciplines. Considering these findings, the main objective of this study is to analyze the acceptance of both AI tools and traditional methodologies used in research. Moreover, the conclusions about the respondents' perception and openness to using AI tools in research regarding gender, age and current academic position were drawn. Another goal was to compare the level of satisfaction from both the AI tools and the traditional research methods.

This empirical study revealed very interesting and a bit surprising result. Namely, the gender, age group and academic position do not influence the use, perception and attitudes toward AI tools for searching scientific databases. Moreover, the findings show that there is a statistically significant difference in respondents' satisfaction with using AI tools and traditional databases. The obtained results call for additional research and education of both students and teaching staff at Croatian universities, regarding the potential benefits of AI tools and their ethical application and use.

The main limitation of this study is the relatively small sample. A bigger sample could contribute to drawing more precise conclusions and would tackle important trends and issues in this research area. In future work, these limitations will be addressed. Namely, the authors plan to investigate the use of AI among the overall Croatian academic community and include members of the other universities in Croatia. Moreover, a comparative study including the Western Balkans is planned for future study. Both the practical and scientific contributions of this paper are relevant. First and foremost, this is among the very few empirical studies conducted in the Republic of Croatia that encompass a fraction of the academic community and tackle the use of AI in research; second, this study is among the very few that incorporate a comparison between the traditional databases and tools and the AI tools in research among the academic members.

6 REFERENCES

[1] Demary, V. & Goecke, H. (2019). Künstliche Intelligenz: Deutsche Unternehmen zwischen Risiko und Chance. *IW-Trends-Vierteljahresschrift zur empirischen Wirtschafts-forschung, 46*(4), 3-18.

[2] Kushwaha, S. (2023). A Futuristic Perspective on Artificial Intelligence. In *IEEE OPJU International Technology Conference on Emerging Technologies for Sustainable Development (OTCON2022)*, 1-6. https://doi.org/10.1109/OTCON56053.2023.10113980

[3] Virvou, M. (2022). The emerging era of human-AI interaction: Keynote address. In *13th IEEE International Conference on Information, Intelligence, Systems & Applications (IISA2022)*, 1-10. https://doi.org/10.1109/IISA56318.2022.9904422

[4] Sağın, F. G., Özkaya, A. B., Tengiz, F., Geyik, Ö. G. & Geyik, C. (2024). Current evaluation and recommendations for the use of artificial intelligence tools in education. *Turkish Journal of Biochemistry, 48*(6), 620-625. https://doi.org/10.1515/tjb-2023-0254

[5] Desjardins-Proulx, P., Poisot, T. & Gravel, D. (2017). Scientific theories and artificial intelligence. *bioRxiv*, 161125.

[6] Waly, M. F. (2024). Artificial intelligence and scientific research. *Sustain. Educ. Globe, 2*, 1-14. https://doi.org/10.21608/seg.2024.269596.1001

[7] Ou, A. W., Stöhr, C. & Malmström, H. (2024). Academic communication with AI-powered language tools in higher education: From a post-humanist perspective. *System, 121*, 103225. https://doi.org/10.1016/j.system.2024.103225

[8] Chubb, J., Cowling, P. & Reed, D. (2022). Speeding up to keep up: exploring the use of AI in the research process. *AI & society, 37*(4), 1439-1457. https://doi.org/10.1007/s00146-021-01259-0

[9] Heidt, A. (2023). Artificial-intelligence search engines wrangle academic literature. *Nature, 620*(7973), 456-457. https://doi.org/10.1038/d41586-023-01907-z
[10] Gabashvili, I. S. (2023). The impact and applications of ChatGPT: a systematic review of literature reviews. arXiv: 2305.18086, Ver. 1. https://doi.org/10.48550/arXiv.2305.18086
[11] Trust in AI: Evaluating Scite, Elicit, Consensus, and Scopus AI for Generating Literature Reviews. (n.d.). Retrieved 3 June 2024, from https://library.hkust.edu.hk/sc/trust-ai-lit-rev/
[12] Rife, S. C., Rosati, D. & Nicholson, J. M. (2021). scite: The next generation of citations. *Learned Publishing, 34*(3), 454-456. https://doi.org/10.1002/leap.1379
[13] Zawacki-Richter, O., Marín, V. I., Bond, M. & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education–where are the educators? *Int. J. Educ. Technol. High. Educ., 16*, 1-27. https://doi.org/10.1186/s41239-019-0171-0
[14] Shubik, M. (1960). Bibliography on simulation, gaming, artificial intelligence and allied topics. *J. Am. Stat. Assoc., 55*, 736-751. https://doi.org/10.1080/01621459.1960.10483374
[15] Rowland, D. R. (2023). Two frameworks to guide discussions around levels of acceptable use of generative AI in student academic research and writing. *J. Acad. Lang. Learn., 17*, T31-T69.
[16] Elkhatat, A. M., Elsaid, K. & Almeer, S. (2023). Evaluating the efficacy of AI content detection tools in differentiating between human and AI-generated text. *Int. J. Educ. Integr., 19*, 17. https://doi.org/10.1007/s40979-023-00140-5
[17] Sekeroglu, B., Dimililer, K. & Tuncal, K. (2019). Student Performance Prediction and Classification Using Machine Learning Algorithms. *ICEIT 2019: Proceedings of the 8th International Conference on Educational and Information Technology*, 7-11. https://doi.org/10.1145/3318396.3318419
[18] Mena-Guacas, A. F., Urueña Rodríguez, J. A., Santana Trujillo, D. M., Gómez-Galán, J. & López-Meneses, E. (2023). Collaborative learning and skill development for educational growth of artificial intelligence: A systematic review. *Contemp. Educ. Technol., 15*. https://doi.org/10.30935/cedtech/13123
[19] Kyrpa, A., Stepanenko, O., Zinchenko, V., Datsiuk, T., Karpan, I. & Tilniak, N. (2024). Artificial Intelligence Tools in Teaching Social and Humanitarian Disciplines. *Inf. Technol. Learn. Tools, 100*, 162-179. https://doi.org/10.33407/itlt.v100i2.5563
[20] Xiao, Y., Chatterjee, S. & Gehringer, E. (2022). A new era of plagiarism the danger of cheating using AI. Presented at *the 20th IEEE International Conference on Information Technology Based Higher Education and Training (ITHET)*, 1-6. https://doi.org/10.1109/ITHET56107.2022.10031827
[21] Ng, D. T. K., Leung, J., Chu, S. & Shen, M. (2021). Conceptualizing AI literacy: An exploratory review. Computers and Education: *Artificial Intelligence, 2*, 100041. https://doi.org/10.1016/j.caeai.2021.100041
[22] Villegas-Ch, W., Palacios-Pacheco, X. & Luján-Mora, S. (2019). Artificial intelligence as a support technique for university learning. In *IEEE World Conference on Engineering Education (EDUNINE2019)*, 1-6. https://doi.org/10.1109/EDUNINE.2019.8875833
[23] Kuo, C. & Reitsch, A. (1995). Neural networks vs. conventional methods of forecasting. *The Journal of Business Forecasting, 14*(4), 17.
[24] Graili, P., Ieraci, L., Hosseinkhah, N. & Argent-Katwala, M. (2021). Artificial intelligence in outcomes research: a systematic scoping review. *Expert Review of Pharmacoeconomics & Outcomes Research, 21*(4), 601-623. https://doi.org/10.1080/14737167.2021.1886083
[25] Bucci, A. (2020). Realized volatility forecasting with neural networks. *Journal of Financial Econometrics, 18*(3), 502-531. https://doi.org/10.1093/jjfinec/nbaa008
[26] Khashei, M. & Bijari, M. (2010). An artificial neural network (p, d, q) model for timeseries forecasting. *Expert Systems with applications, 37*(1), 479-489. https://doi.org/10.1016/j.eswa.2009.05.044
[27] Ghiassi, M., Saidane, H. & Zimbra, D. K. (2005). A dynamic artificial neural network model for forecasting time series events. *International Journal of Forecasting, 21*(2), 341-362. https://doi.org/10.1016/j.ijforecast.2004.10.008
[28] Rasp, S. & Lerch, S. (2018). Neural networks for postprocessing ensemble weather forecasts. *Monthly Weather Review, 146*(11), 3885-3900. https://doi.org/10.1175/MWR-D-18-0187.1

Authors' contacts:

Katerina Fotova Čiković, PhD, Assistant Professor
(Corresponding author)
University North,
Trg dr. Žarka Dolinara 1, 48000 Koprivnica, Croatia
kcikovic@unin.hr

Antonija Mandić, PhD
University North,
Trg dr. Žarka Dolinara 1, 48000 Koprivnica, Croatia
antonija.mandic@unin.hr

Maja Hoić, University North student
University North,
Trg dr. Žarka Dolinara 1, 48000 Koprivnica, Croatia
mahoic@unin.hr

ISSN 1846-6168 (Print), ISSN 1848-5588 (Online)
https://doi.org/10.31803/tg-20250108082640

Original scientific paper
Received: 2025-01-08, Accepted: 2025-01-31

Investigation and Optimization of Adhesive Structure Formation on Shell Limestone Surfaces

Andrii Kolesnykov, Mykola Khlytsov, Svitlana Semenova, Mario Šercer*

Abstract: The article is devoted to the study of the adhesive interaction of restorative compositions with shell limestone as a highly porous material of complex structure. The key physicochemical and mechanical factors determining adhesion are analyzed, including molecular forces, hydrogen bonds, electrostatic interactions, acid-base interactions, as well as the effect of humidity on adhesion. Elements of fractal geometry are used to describe the processes of adaptation of compositions to porous surfaces, a criterion of ideal adhesive contact is introduced. Models of structure formation are considered, including the Smoluchowski equation, diffusion-limited and cluster-cluster aggregation. It is shown how optimization of the restorative composition due to the selection of the necessary fractal parameters can improve its adhesive characteristics. Recommendations are given for the use of various methods for optimizing adhesion processes in restoration. It is also possible to use the obtained results in modern additive technologies.

Keywords: adhesion; fractal dimension; porous structure; restorative compositions; shell limestone; structure formation

1 INTRODUCTION

One of the important tasks of construction materials science, especially relevant in cities - centers of historical, cultural and architectural heritage, is the production of restoration compositions with an optimal set of performance characteristics. An important property of such materials is the ability to form strong adhesive contacts with the surface of the base stone [1, 2]. When forming a contact between the adhesive and the base material (substrate), adhesive structuring occurs − special case of structuring of the adhesive composite, which occurs under conditions of strong geometric constraints specified by the porous structure of the material being restored, for example, shell limestone. In addition, the adhesive processes occurring during the formation of the "adhesive-substrate" contact are of great importance in additive technologies, widely used today in modern construction and restoration [3]. In this case, the key role in such processes is played by the coordination of the geometric characteristics of the material and the adhesive [4]. Such coordination is especially important when applying restoration compounds to complex porous surfaces, for example, fragments of historical buildings made of shell limestone [1].

Adhesion is a complex multifactorial phenomenon based on molecular, physicochemical and mechanical aspects [5, 6]. The mechanical component has a pronounced geometric-topological character associated with the complexity of the hierarchical structure of porous materials [7, 8]. Molecular forces caused by van der Waals interactions, hydrogen bonds, electrostatic and acid-base interactions determine the microscopic nature of adhesion and determine the physicochemical surface properties that form the overall picture of adhesion. At the same time, mechanical interaction, including penetration of the adhesive into the pores and microcracks of the material, forms a significant contribution to the adhesive forces, which determines the importance of studying the geometry and topology of the contact surface.

Geometrical factors play a special role in adhesion optimization and can be described through fractal regularities. The porous structure of the material being restored and the adhesive composition forms a three-dimensional structure of the adhesive contact, which in the ideal case completely fills the surface irregularities and pores. This imposes strict restrictions on the fractal properties of the adhesive material. The correspondence of the fractal dimension of the adhesive and the material being restored can be used as a criterion for the optimality of the contact.

This work is devoted to the consideration of approaches to optimization of adhesive compositions based on geometric, molecular and physicochemical factors using modeling of the structure formation of adhesives and determining their optimal fractal characteristics.

2 FACTORS AFFECTING ADHESION TO POROUS SURFACES

Adhesion (bonding) is a set of intermolecular interactions between the surfaces of different materials, leading to their holding together. For composite materials used for the restoration and reconstruction of shell limestone buildings, adhesion depends on many factors, including the type of restorative composition (e.g., silicate-based) and the physicochemical properties of the shell limestone, especially in the presence of moisture. It should be noted that the factors determining adhesive activity relate primarily to the adhesive surfaces. They can be conditionally divided into molecular, related physicochemical, mechanical and geometric-topological.

2.1 Molecular Factors of Adhesion to the Surface of Shell Limestone

The molecular theory of adhesion explains cohesion through intermolecular interactions such as Van der Waals forces, hydrogen bonds, electrostatic and acid-base interactions. The intermolecular interaction potential $V(r)$ between particles at a distance r is usually expressed as the sum of several contributions and is described by the Eq. (1) [9]:

$$V(r) = V_{\mathrm{VdW}}(r) + V_{ei}(r) + V_{\mathrm{hidrogen}}(r) + V_{\text{acid-based}}(r), \qquad (1)$$

where $V_{\text{VdW}}(r)$ − contribution of Van der Waals interactions; $V_{\text{ei}}(r)$ − electrostatic contribution; $V_{\text{hidrogen}}(r)$ − contribution of hydrogen bonds; $V_{\text{acid-based}}(r)$ − the contribution of acid-base interactions.

Let us consider in more detail the contribution of each type of interaction to the adhesion of restorative compositions that adhere to shell limestone.

Van der Waals forces include dispersion forces, orientation and induction interactions. They are universal and act between all molecules. Dispersion forces play an important role in the adhesion of restorative compositions to inorganic surfaces. These forces depend on the polarizability of the molecules, which in the case of wet shell limestone can be increased due to the presence of water. The Lennard-Jones potential (2) is the most often used to describe Van der Waals forces:

$$V_{\text{vdW}}(r) = 4\varepsilon\left(\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right), \tag{2}$$

where ε − the depth of the potential well, which characterizes the strength of interaction; σ − effective diameter of the molecule.

This potential describes the balance between attraction at large distances (dispersion forces) and repulsion at short distances (Pauli repulsion). In [10], a type of potential for silicates and aluminosilicates is proposed that describes the interaction between oxygen and silicon atoms in the structure of zeolite-like materials, as well as with a correction of the value taking into account ions such as sodium and calcium. For the materials under consideration, typical values are $\varepsilon \approx$ 0.4 - 0.5 kcal/mol and $\sigma \approx 3.0$ Å for interactions between oxygen and silicon atoms. Similar parameters are used for calcite, the main mineral in shell limestone. According to [11], the Lennard-Jones potential for Ca-O interactions is used to model limestone bases, with constant values of $\varepsilon \approx$ 0.32 kcal/mol and $\sigma \approx 3.9$ Å. This potential allows us to consider the interactions of carbonates with silicate and aluminosilicate bonds in the presence of moisture.

Electrostatic interactions can be significant for materials with opposite charges. Aluminosilicate materials can have a surface charge that contributes to their attraction to the carbonate matrix of the shell limestone, especially under high humidity conditions. The electrostatic interaction potential between two charged particles q_i and q_j at a distance r is expressed by the Eq. (3):

$$V_{\text{ei}}(r) = \frac{q_i q_j}{4\pi\varepsilon_0 r}, \tag{3}$$

where ε_0 − dielectric constant.

The electrostatic contribution is significant when taking into account the interaction with silicate restorative compositions, since such materials can have local charges that interact with charged or polarizable groups on the surface of the shell limestone.

Hydrogen bonds are especially important for adhesion in the presence of physically bound water. Water molecules can form bridges between the surface groups of silicate and aluminosilicate composition and the surface of the shell limestone, strengthening the adhesion. This contribution increases under humid conditions, since water is capable of forming hydrogen bonds with –OH groups that may be present on the surface of the adhesive and substrate. Water molecules present in the porous structure create additional hydrogen bonds. Water acts as an intermediate agent that forms hydrogen bond bridges, increasing the adhesive force. In addition, the presence of water reduces the rigidity and increases the mobility of molecules on the surface of the aluminosilicate composition, facilitating their interaction with the shell limestone. Thus, in [12] it is shown that cooperative tridentate hydrogen-bonding interactions significantly enhance underwater adhesion, which can be relevant for understanding the role of water in the adhesion mechanisms of mineral materials. This is especially important when restoring wet shell limestone, since physically bound water not only increases adhesion, but also promotes the diffusion of the restorative composition into the pores of the shell limestone, improving its penetration and fixation.

The hydrogen bond potential can be expressed in terms of empirical potentials, for example using Eq. (4):

$$V_{\text{hydrogen}}(r) = D_{\text{e}}\left(\text{e}^{-a(r-r_0)} - 2\text{e}^{-a(r-r_0)/2}\right), \tag{4}$$

where D_{e} is the bond energy, usually about 5-10 kcal/mol; r_0 is the equilibrium bond distance; a α − rigidity parameter. This potential is used to model water adsorption on the surface of carbonates, which is confirmed by a study of the effect of moisture on adhesive forces.

Acid-base interaction is determined by the ability of surface groups to act as acids or bases. The surface of shell limestone can exhibit weakly acidic properties, and some surface groups of lime-silicate material can be alkaline, which leads to additional adhesion at the level of surface layers [13].

Acid-base interactions are often modeled using the donor-acceptor interaction approach, which can be expressed through the acid-base affinity parameter based on the Lewis donor-acceptor theory method. Acid-base interactions can be modeled using energy contributions to potential functions that account for the donor-acceptor behavior of particles. The Lewis potential is described through the interaction energy E_{AB} (5):

$$E_{\text{AB}} = C \cdot \frac{\alpha_{\text{A}} \cdot \beta_{\text{B}}}{r^2}, \tag{5}$$

where α_{A} − acceptor capacity (acid strength); β_{B} − donor capacity (main force); C − empirical constant depending on the specifics of the materials.

Typical values of the constants depend on the nature of the materials: for aluminosilicates and calcite under humid

conditions, it is proposed to use the value α_A = 0.6 - 0.8 and β_B = 0.4 - 0.6 in approaches based on measurements of surface tension and Hamaker energy.

2.2 Physicochemical Factors and Parameters of Adhesion to Shell Limestone

The interaction potential (1) is the basis for the transition from adhesive molecular interactions to a comprehensive physicochemical description and thermodynamic theory of adhesion. The general expression of surface energy through intermolecular interactions (e.g., Van der Waals forces) can be represented by the integral of the intermolecular potential $\varphi(r)$ over the entire space above the surface (6):

$$\gamma = \int_h^\infty \varphi(r)\mathrm{d}r\,, \tag{6}$$

where h is the distance between molecules on the surface; r is the intermolecular distance.

This formula shows that the surface energy depends on the attractive force of molecules at the phase boundary.

Thermodynamic theory describes adhesion through changes in free energy at the interface between two phases [14]. According to this theory, adhesion is characterized by the work of adhesion W_{adh}, which is defined as the energy required to separate two surfaces in accordance with the Eq. (7):

$$W_{adh} = \gamma_1 + \gamma_2 - \gamma_{12}\,, \tag{7}$$

The work of adhesion is positive for spontaneous adhesion of surfaces, and the higher the W_{adh}, the stronger the adhesion.

Adhesive interactions are closely related to the contact angles formed at the contact of three phases, as well as to the flow and inflow angles. The contact angle θ describes the contact of a liquid with a solid surface and is determined by Young's Eq. (8):

$$\gamma_{SV} = \gamma_{SL} + \gamma_{LV}\cos\theta\,, \tag{8}$$

where γ_{SV}, γ_{SL} and γ_{LV} − surface tension at the interface between a solid, liquid and air.

Angles of runoff and wetting θ_A, θ_R describe the dynamic behavior of a liquid on an inclined surface. The angle of runoff θ_A − this is the maximum angle of inclination at which the drop begins to flow down, and the angle of wetting θ_R − the angle at which the drop stops on the surface. The difference $\Delta\theta = \theta_A - \theta_R$ is called contact hysteresis and is related to adhesion: a large hysteresis indicates high adhesion of the liquid to the surface. A significant hysteresis indicates strong cohesive forces holding the liquid and preventing its movement.

The work of adhesion can also be expressed in terms of the runoff and wetting angles using the relation [15] (9):

$$W_{adh} = \gamma_{LV}(1 + cos\,\theta_{avg})\,, \tag{9}$$

where θ_{avg} − average contact angle, defined as (10):

$$\theta_{avg} = \frac{\theta_A + \theta_R}{2}\,, \tag{10}$$

Thus, the work of adhesion depends on the average value of the runoff and wetting angles, as well as on the surface tension of the liquid $\gamma_{LV.}$ Changing the runoff and wetting angles, for example, by chemically modifying the surface, allows one to regulate the adhesive properties of the system, which is important for practical applications in hydrophobic and hydrophilic coatings.

2.3 Effect of Surface Geometry on Adhesive Interaction. Fractal Analysis of Geometric Parameters

The mechanical theory of adhesion [16] is based on the fact that adhesion is achieved by the penetration of the material into the microroughness and pores of the base surface. For shell rock, with its high porosity and microcracks, the lime-silicate composition penetrates the structure, providing reliable adhesion. The adhesion force can be described by the Eq. (11) [17]:

$$F_{adh} = kA\gamma\left(1 - E^{-\beta P}\right), \tag{11}$$

where F_{adh} is the adhesion force; k is the coefficient of adhesion strength; A is the contact area; γ - surface tension at the interface between the adhesive and the base; β - penetration coefficient; P - pressure when applying the composition.

It is important to note that A is the "real" area of adhesive contact, the accounting of which is complicated by the complex relief of the base surface.

A powerful tool for describing complex surfaces such as shell limestone and other porous materials that are often encountered in adhesive joints is fractal geometry. Traditional models based on Euclidean geometry are often insufficient to adequately describe such structures. The fractal approach allows one to take into account the hierarchical structure and partial ordering characteristic of many natural and artificial surfaces. The main concepts of the fractal theory of adhesion are the fractal dimension D and lacunarity Λ [18, 19]. The fractal dimension characterizes the degree to which the space is filled by the fractal object. For surfaces, D ranges from 2 (smooth surface) to 3 (completely filled space). The higher D, the more developed the surface. Lacunarity Λ describes the "emptiness" or "perforation" of the fractal object. High lacunarity corresponds to a large number of voids or pores. Lacunarity is often associated with the distribution of pore sizes and shape.

One of the most significant in the mechanical theory of adhesion, which is geometric in nature, is the relationship between the fractal dimensions of the base material and the adhesive restorative composition. Let the first surface (shell

limestone) be characterized by fractal dimension D_1, the second surface (adhesive restorative composition) by fractal dimension D_2. Fractal dimension D_1 characterizes the complexity of the shell limestone surface. The greater D_1, the more "rugged" the surface. Fractal dimension D_2 characterizes the complexity of the cluster structure of the adhesive. The greater D_2, the more branched the cluster. Contact occurs in three-dimensional Euclidean space (the embedding dimension $D_E = 3$). Ideal adhesive contact is achieved when the adhesive completely fills all the irregularities and pores on the shell limestone surface. In this case, from the geometric point of view, one can imagine that the adhesive surface is "superimposed" on the shell limestone surface, completely repeating its structure as its negative imprint. In the ideal case, one can assume that the "sum" of their fractal dimensions should be related to the dimension of the space in which the contact occurs.

Let us consider the mutual dimension of two fractal sets. The Minkowski-Buligand dimension [18-20] (or box-counting dimension) is defined by the following Eq. (12):

$$D = \lim_{\varepsilon \to 0} \frac{\log N(\varepsilon)}{\log(1/\varepsilon)}, \tag{12}$$

where $N_A(\varepsilon)$ - the minimum number of cubes (or "boxes") of size ε, required to cover the set.

Next, we will consider two fractal sets A and B with dimensions D_1 and D_2 respectively. Let $N_A(\varepsilon)$ and $N_B(\varepsilon)$ - number of boxes of size ε, required to cover the sets A and B respectively. Then the Eqs. (13) and (14) are fulfilled:

$$N_A(\varepsilon) \sim \varepsilon^{-D_1}, \tag{13}$$

$$N_B(\varepsilon) \sim \varepsilon^{-D_2}, \tag{14}$$

To describe the "mutual arrangement" of sets A and B Let's introduce the concept of Minkowski product $A \oplus B$ (15):

$$A \oplus B = \{a + b \mid a \in A, b \in B\}, \tag{15}$$

Dimension of the Minkowski product $D(A \oplus B)$ related to dimensions D_1 and D_2 the following inequality (16):

$$\max(D_1, D_2) \le D(A \oplus B) \le D_1 + D_2, \tag{16}$$

In the case of ideal contact, when the adhesive completely fills all the irregularities of the shell limestone, it can be assumed that the Minkowski product $A \oplus B$ "fills" all the space near the surface of the shell limestone. In the limiting case, when the adhesive perfectly repeats the structure of the shell limestone, the dimension of the Minkowski product should be related to the dimension of the space D_E. In the case of ideal contact, the dimension of the Minkowski product is approximately equal to $2D_E - 1$. Thus, in the three-dimensional case, if one of the sets is a "plane" ($D_1 = 2$), and the other - a "ragged" surface ($D_2 > 2$), then when they are "superimposed" one can expect that the dimension of the Minkowski product will approach $2D_E - 1 = 5$. This assumption is based on heuristic considerations, acknowledging that when fully filled, the 'interaction' of two fractal sets effectively results in an increase in dimensionality. Thus, for an ideal contact we obtain the approximate relation (17):

$$D_1 + D_2 \approx 2D_E - 1, \tag{17}$$

In three-dimensional space ($D_E = 3$) is performed (18):

$$D_1 + D_2 \approx 2 \cdot 3 - 1 = 5, \tag{18}$$

Expression $5 - D_1 - D_2 \approx 0$ is a criterion for the ideality of adhesive contact within the framework of this theory. The obtained criterion is key for this work.

3 MODEL REPRESENTATIONS OF ADHESIVE STRUCTURE FORMATION IN RESTORATIVE COMPOSITIONS

The structural and geometric properties of the restorative composition are formed in the process of spatially limited structure formation. The process of pore filling and subsequent hardening can be described as aggregation of particles of the restorative composition. In the simplest case, the diffusion-limited aggregation (DLA) model can be used [21]. The cluster-cluster aggregation (CCA) model [21] describes the aggregation of already formed clusters. It leads to the formation of looser and more branched structures than DLA. The fractal dimension of CCA clusters is lower than that of DLA clusters. This model may be relevant if the adhesive consists of already aggregated particles or if the process of its application and hardening promotes aggregation.

To describe more general processes of structure formation, the Smoluchowski equation [22] is applicable, which describes the change in the concentration of clusters of different sizes over time in accordance with Eq. (19):

$$\frac{dc_k}{dt} = \frac{1}{2} \sum_{i+j=k} K(i,j) c_i c_j - c_k \sum_{j=1}^{\infty} K(k,j) c_j, \tag{19}$$

where is c_k - concentration of clusters of size k; $K(i,j)$ - kinetic coefficient describing the rate of adhesion of clusters of sizes i and j.

Eq. (19) allows us to determine how the particle size distribution changes over time as a result of their collisions and associations. The first term of Eq. (19) $\frac{1}{2} \sum_{i+j=k} K(i,j) c_i c_j$ describes the formation of clusters of size k because of collisions and mergers of clusters of sizes i and j, where $i + j = k$. The coefficient $\frac{1}{2}$ takes into account that each collision is counted twice in the sum (for example, the collision of cluster i with cluster j and the collision of cluster

j with cluster i). The second term $-c_k \sum_{j=1}^{\infty} K(k,j)c_j$ describes the loss of clusters of size k as a result of their collisions and mergers with other clusters of any size j. The coagulation coefficient for diffusion-limited aggregation (DLA): In this case, the coefficient depends on the diffusion coefficients of clusters i and j (20):

$$K(i,j) \sim (D_i + D_j)(R_i + R_j), \tag{20}$$

where D_i and D_j – diffusion coefficients; R_i and R_j – radii of gyration of clusters i and j, respectively.

The kinetic coefficient $K(i,j)$ in the general case depends on the physicochemical factors discussed above, such as the forces of intermolecular Van der Waals interaction, electrostatic forces, hydrogen bond forces and acid-base interactions, as well as on the temperature and viscosity of the medium.

The processes of structure formation in a limited pore space can be considered from several other positions, according to which in the considered material, as a result of a number of physical and chemical processes occurring at the micro level, the fractal dimension changes. For cluster systems, to which the considered composition also belongs, the fractal dimension D_f connects the number of particles N in the cluster with its characteristic size R based on the relation (21):

$$N \sim R^{D_f}, \tag{21}$$

For clusters formed in the DLA process in $2D - D_f \approx 1.7$, in $3D - D_f \approx 2.5$ the kinetics of fractal dimension growth is complicated (it involves solving the Smoluchowski equations). However, approximations can be made, for example, by assuming that the average cluster size grows in a power-law manner with time (22):

$$R(t) \sim t^z, \tag{22}$$

where z is a dynamic indicator.

Then, if we assume that the number of particles in a cluster is proportional to time, then the fractal dimension does not change (23):

$$D_f(t) \approx \frac{\ln(t)}{\ln(R(t))} \sim \frac{\ln(t)}{z\ln(t)} = \frac{1}{z}, \tag{23}$$

This is a very simplified argument, but it shows that the growth of fractal dimension is related to the dynamics of cluster growth.

In an idealized case, if the adhesive simply fills the existing pores without changing its internal structure or forming new fractal elements, its fractal dimension does not change. It "copies" the fractal structure of the pore space. However, this is a rather theoretical case. In most real situations, the hardening of the adhesive is accompanied by a change in its structure, which leads to a change in the fractal dimension. Let's consider several possible scenarios:

1) *Polymerization and gelation.* During the solidification process, polymerization may occur, leading to the formation of a three-dimensional network, for example, of an aluminosilicate nature. This network may have a fractal structure. At the initial moment, when the polymer is still liquid, its fractal dimension is close to 3 (volume filling). As polymerization and network formation proceed, the fractal dimension may decrease, approaching values characteristic of percolation clusters (about 2.5) or even lower, depending on the degree of network branching.
2) *Crystallization.* If crystallization of new formations is possible in the adhesive and the resulting crystals can have different shapes and orientations, then the formation of a fractal structure is also possible. The fractal dimension will depend on the crystallization conditions (temperature, cooling rate, presence of impurities).
3) *Particle aggregation.* Since the adhesive is a dispersed system, the particles can aggregate during the hardening process, forming clusters. The considered DLA and CCA models describe this process and lead to the formation of fractal structures with dimensions of about 2.5 (DLA) and lower (CCA). During the hardening process, the mobility of the particles decreases, which can "fix" the formed structure.
4) *Capillary forces and wetting.* Capillary forces and wetting play an important role in the process of filling the pores. If the adhesive wets the surface of the material well, it will tend to fill all the smallest pores and cracks, which can lead to an increase in fractal dimension. On the contrary, poor wetting can lead to the formation of voids and a decrease in fractal dimension.

The fractal dimension of the cured adhesive also depends on the initial pore structure of the material. If the pores are small and numerous, the adhesive will be forced to form a more branched structure to fill them all. This may lead to an increase in the fractal dimension. If the pores are large and few in number, the adhesive can fill them without forming complex fractal structures. In this case, the fractal dimension may be lower. The connectivity of the pores also plays an important role. If the pores are interconnected, the adhesive can move freely and fill them, which contributes to the formation of a more uniform structure. If the pores are isolated, the adhesive will fill them independently, which may lead to the formation of a more heterogeneous and possibly more branched structure.

4 METHODS OF OPTIMIZING GEOMETRICAL PARAMETERS TO IMPROVE THE ADHESION PROPERTIES OF RESTORATION COMPOSITIONS

Based on the above analysis, several methods for optimizing adhesive compositions can be proposed. Due to the multifactorial nature of the structure formation process, determining the fractal dimension of the clusters formed in the adhesive restorative composition requires empirical

methods. Therefore, the initial selection of the composition can be guided by the structural-geometric criterion.

The main optimization problem is characterized by the objective function Eq. (24):

$$F(w_1, w_2,\ldots, w_n) = \left|D_1 + D_A(w_1, w_2,\ldots, w_n) - 5\right| \to \min, \quad (24)$$

where D_1 - fractal dimension of the shell limestone surface (known); $D_A(w_1, w_2,\ldots, w_n)$ - fractal dimension of adhesive clusters, depending on both the granulometric composition (w_1, w_2,…, w_n, w_i – volume fraction of fraction i) and its structure formation processes occurring inside the pore space. Limitations are represented by the conditions (25, 26):

$$\sum_{i=1}^{n} w_i = 1, \quad (25)$$

$$w_i \geq 0, \quad (26)$$

This task is aimed at minimizing the deviation from the ideal contact criterion. $D_1 + D_2 - 5 = 0$. Additional criteria may include minimizing lacunarity in the contact zone (27):

$$F(w_1, w_2,\ldots, w_n) = \Lambda(w_1, w_2,\ldots, w_n) \to \min, \quad (27)$$

It should be noted that minimization $|D_1 + D_2 - 5|$ already indirectly contributes to the minimization of lacunarity, since complete filling implies a reduction in voids. The Richardson curve fit criterion (a plot of $\log(N(\varepsilon)$ versus $\log(1/\varepsilon)$), where ε defines the "scale" of the measurement, [19]) can also be useful, especially if the behavior of the contact at different scales needs to be considered (28):

$$F_{\text{add2}}(w_1, w_2,\ldots, w_n) = \\ = \int_{r_{\min}}^{r_{\max}} \left|L_{\text{shell}}(r) - L_{\text{adh}}(r, w_1, w_2,\ldots, w_n)\right| \mathrm{d}r \to \min, \quad (28)$$

However, this criterion is also partially accounted for by the main criterion, as the fractal dimension is linked to the behavior of the Richardson curves.

4.1 Determination of Geometric Characteristics of the Material and Composition of the Restoration Using Image Processing

The initial data for optimization according to criterion (18) is the fractal dimension of the base material (shell limestone), which can be realized by processing images of the material surface using the box counting method. When implementing it, a grayscale image can be represented as a three-dimensional surface, where x and y are the pixel coordinates, and z is the gray intensity (from 0 to 255), each pixel becomes a point in three-dimensional space. The space in which this surface is located is covered with a grid of cubes of size ε (epsilon). The size ε determines the "scale" of the measurement. The number of cubes $N(\varepsilon)$ that intersect or contain at least one point (pixel) of the surface is counted. The size of the cubes ε is reduced (for example, by 2 times), and the procedure for counting filled cubes is repeated, a new pair of values (ε, $N(\varepsilon)$) is constructed. A graph of the dependence of $\log(N(\varepsilon)$ on $\log(1/\varepsilon)$ is plotted. The slope of the resulting line is the fractal dimension D.

The method was implemented using the *FDim* software tool [23] for different shell limestone samples. The measurement results are given in Tab. 1.

Table 1 Fractal characteristics of typical shell limestone samples

Number	Samples	Fractal dimension D_f	Lacunarity Λ
1		2.557	0.259
2		2.454	0.210
3		2.357	0.282

It should be noted that the shell limestone structure is variable, which is reflected in the fractal characteristics of the corresponding images. As a result, it is proposed to perform the corresponding measurements separately for each restoration project.

Let us calculate fractal dimension of the adhesive composition surface (Tab. 1) according to the relation (18) for ideal adhesive contact. The results are shown in Tab. 2.

The obtained data (fractal dimension $D_2 \approx 2.5$) indicate that the DLA mechanism (and even more so CCA) can ensure the formation of a non-ideal cluster system, and the composition must be designed taking into account the formation of fine-crystalline products in contact with the pore material.

Table 2 Determination of the required fractal dimension of the adhesive composition

Number	1	2	3
Fractal dimension D_1 shell limestone	2.557	2.454	2.357
Required fractal dimension D_2 of the adhesive composition surface	2.443	2.546	2.643

Often the granulometric distribution of dispersed materials is described by a power law, which is an indication of its fractal structure. A special case of a power distribution is the Richardson-Kolmogorov distribution, which is often observed in crushing and grinding processes. The general form of a power distribution (cumulative) (29):

$$N(r) \sim r^{-D_2}, \tag{29}$$

where d - particle size (diameter, linear size, etc.); $N(d)$ - number of particles with size greater than d; D_2 - exponent, which can be interpreted as fractal dimension (Tab. 2).

This ratio allows a rough estimate of the required particle distribution in the adhesive restorative composition. To obtain the size distribution, it is necessary to differentiate expression (29) and obtain the Eq. (30):

$$\frac{\mathrm{d}N(r)}{\mathrm{d}r} - Dr^{-D-1}, \tag{30}$$

To obtain a normalized distribution describing the probability of finding a particle of size r, it is necessary to integrate expression (29) within the range from $r_{\min}$ up to $r_{\max}$ and divide by the obtained result (31):

$$P(r) = \frac{(D+1)r^{-D-1}}{r_{\min}^{-D} - r_{\max}^{-D}}, \tag{31}$$

where $P(r)$ is the probability density of the particle size distribution ($P(r)$ is normalized, $\int_{r_{\min}}^{r_{\max}} P(r)\mathrm{d}r = 1$; $r_{\min}$ - minimum particle size; $r_{\max}$ - maximum particle size; D - fractal dimension.

Let us consider the problem of selecting particle sizes, assuming that there are several fractions with known particle size distributions $P_i(r)$ (e.g., unimodal, such as normal or lognormal distribution) and weighted fractions w_i. Then the overall distribution of the mixture will be (32):

$$P_{\text{mix}}(r) = \sum_{i=1}^{n} w_i P_i(r), \tag{32}$$

where n is the number of fractions, $\sum_{i=1}^{n} w_i = 1$.

To solve the problem of selecting a composition, one can use methods known in mathematical statistics [24, 25]:

1) *Kolmogorov-Smirnov criterion*. This criterion is based on the comparison of the maximum discrepancy between the cumulative distribution functions (33):

$$D_{\text{KS}} = \sup_r \left| F_{\text{mix}}(r) - F_{\text{target}}(r) \right|, \tag{33}$$

where is $f_{\text{mix}}(r)$ and $F_{\text{target}}(r)$ - cumulative distribution functions for the mixture and target distribution, respectively ($F(r) = \int_{r_{\min}}^{r} P(r')\mathrm{d}r'$). Minimizing D_{KS} is a useful way to ensure good distribution fit.

2) *The chi-square test* is represented by Eq. (34):

$$\chi^2 = \sum_{j=1}^{m} \frac{(O_j - E_j)^2}{E_j}, \tag{34}$$

where Oj – observed frequency in the j^{th} interval; Ej - expected frequency in the j^{th} interval, m - number of intervals.

Minimizing χ^2 also allows one to achieve a good fit, especially when the amount of data is large.

3) *Minimizing Kullback-Leibler entropy* (Relative Entropy): This criterion, also known as the Kullback-Leibler divergence, measures the "distance" between two probability distributions according to Eq. (35).

$$D_{\text{KL}} = \int_{r_{\min}}^{r_{\max}} P_{\text{target}}(r) \log \frac{P_{\text{target}}(r)}{P_{\text{mix}}(r)} \mathrm{d}r. \tag{35}$$

Minimizing D_{KL} is aimed at making $P_{\text{mix}}(r)$ approximated P_{target} as accurately as possible, especially in areas where $P_{\text{target}}(r)$ large.

Let's consider an example of calculations. Let's assume that we need to select a granulometric composition with a fractal dimension $D = 2.5$ in the size range from $r_{\min} = 1$ μm to $r_{\max} = 100$ μm. We have two fractions with normal distributions: fraction 1 – average size $\mu_1 = 1.5739$ μm, standard deviation $\sigma_1 = 0.6653$ μm; fraction 2 – average size $\mu_2 = 20$ μm, standard deviation $\sigma_2 = 14.6942$ μm.

The target distribution is given by Eq. (36):

$$P_{\text{target}}(r) = \frac{3.5r^{-3.5}}{1^{-2.5} - 100^{-2.5}}, \tag{36}$$

The probability density of the normal distribution for fractions is given by the Eqs. (37)-(38):

$$P_1(r) = \frac{1}{\sigma_1\sqrt{2\pi}} \mathrm{e}^{-\frac{(r-\mu_1)^2}{2\sigma_1^2}}, \tag{37}$$

$$P_2(r) = \frac{1}{\sigma_2\sqrt{2\pi}} \mathrm{e}^{-\frac{(r-\mu_2)^2}{2\sigma_2^2}}, \tag{38}$$

Now we need to find weight fractions w_1 and w_2 such that the equality (39):

$$P_{\text{mix}}(r) = w_1 P_1(r) + w_2 P_2(r), \tag{39}$$

best matched $P_{\text{target}}(r)$. For example, we use the Kullback-Leibler divergence D_{KL}. Minimum value f Kullback-Leibler divergence: $D_{\text{KL}} = 4.0644$. The corresponding weights are $w_1 = 0.8789$, $w_2 = 0.1211$. The degree of correspondence

between the mixture and target distributions is shown in Fig. 1.

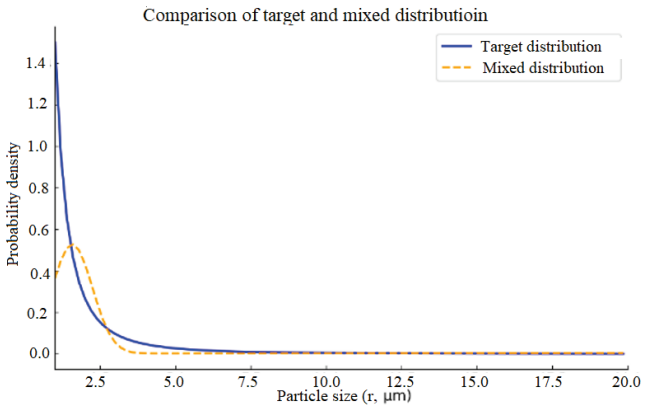


Figure 1 The result of selecting the distribution that best matches the target

Based on the above analysis of adhesion processes and optimization of adhesive interactions using fractal geometry methods, it is possible to evaluate the substrate surface for further selection of the optimal adhesive composition, taking into account the adaptive nature of the ongoing structure formation processes. The restorative material must be able to take the "shape" of pores and the fractal dimension dictated by the relationship (18) for the porous structure during the structure formation process. The adhesive material system adapts to the substrate system, which ensures the integrity and functionality of the resulting adhesive contact system.

Thus, one of the effective concepts for optimizing restorative compositions is the use of fractal patterns that link adhesive interaction with the geometric structure of materials. Fractal dimension D_f is a parameter characterizing the complexity of a surface. In restoration work, she describes how "rugged" the surface of a material, such as stone or limestone, is used in historic buildings. In the case of additive technologies, ideal adhesive contact assumes that the applied composition (adhesive) repeats the surface structure of the restored material as a negative imprint, minimizing voids and improving adhesion. This is only possible by matching the fractal dimension of the adhesive D_2 with a base surface dimension of D_1. The adhesive composition must have the required fluidity and rheological characteristics to ensure uniform distribution over the surface and filling of all voids. Fractal optimization is especially important when using 3D-printing with restoration compositions. Printing allows you to control the micro- and macrostructure of the material, adapting it to complex surfaces. This technology guarantees accurate reproduction of the structural features of a historical building, minimizing the risk of destruction. Moreover, the use of fractal optimization improves the durability of adhesion, reduces the likelihood of delamination and increases resistance to changes in humidity and temperature fluctuations.

5 CONCLUSIONS

The presented work examines theoretical and applied aspects of adhesive structuring of restorative compositions intended for interaction with porous materials such as shell limestone. It is shown that adhesion is a multifactorial phenomenon, including molecular, physicochemical and mechanical components, and the mechanical component has a pronounced geometric-topological character and can be described through fractal parameters. The main approaches to improving adhesive properties are proposed: fractal matching of the surface dimensions of the material and adhesive to achieve ideal contact, optimization of the physicochemical characteristics of the composition by adding nanoparticles and chemical surface treatment. The results of the study showed that successful adhesion is achieved with a comprehensive consideration of molecular, physicochemical and geometric factors. The use of fractal geometry made it possible to quantitatively describe the porosity and texture of the material, as well as to propose criteria for optimizing the composition, including minimizing lacunarity and designing the granulometric distribution of adhesive particles. The practical significance lies in the application of these approaches in additive technologies for restoring historical buildings, where the precise application of compositions, considering the local fractal parameters of the surface, ensures improved adhesion and durability of restoration work. Further research in this area may focus on developing methods to control the structure of adhesives during application and hardening, refining models of structure formation by accounting for temperature and humidity conditions, and utilizing fractal approaches to describe the dynamic properties of adhesive contacts under external influences. Thus, the obtained results emphasize the importance of an integrated approach to the design of adhesive materials based on the coordination of their physicochemical and geometric characteristics with the restored material, which opens up prospects for creating more sustainable and durable technologies in construction and restoration.

Acknowledgments

This work has been supported by Metal Centre Čakovec under the project KK.01.1.1.02.0023

6 REFERENCES

[1] Korobko, O. A. & Lisenko, V. A. (2005). Structure and design of a stone-shell concrete structure. *Bulletin of the Odessa State Academy of Construction and Architecture*, (15), 144-151. (in Russian)

[2] Kinloch, A. J. (1987). *Adhesion and Adhesives: Science and Technology*. Springer. https://doi.org/10.1007/978-94-015-7764-9

[3] Kumar, S. (2020). *Additive Manufacturing Processes*. Springer. https://doi.org/10.1007/978-3-030-45089-2

[4] Israelachvili, J. N. (2011). *Intermolecular and Surface Forces* (3rd ed.). Academic Press. https://doi.org/10.1016/B978-0-12-391927-4.10001-5

[5] Turusov, R. A. (2021). Adhesion interaction and adhesion mechanics. *Polymer Science Series D*, 14, 522–531. https://doi.org/10.1134/S1995421221040250

[6] Mittal, K. L. (2021). *Progress in Adhesion and Adhesives, Volume 6 (Adhesion and Adhesives: Fundamental and Applied Aspects)*. Wiley-Scrivener.

https://doi.org/10.1002/9781119846703
[7] Yang, X. Y., Li, L. H., Li, Y., Rooke, J. C., Sanchez, C. & Su, B.-L. (2017). Hierarchically porous materials: synthesis strategies and structure design. *Chemical Society Reviews*, 45, 481-558. https://doi.org/10.1039/c6cs00829a
[8] Kendall, K. (2012). *Molecular adhesion and its applications: The sticky universe*. Springer Science & Business Media.
[9] Allen, M. P. & Tildesley, D. J. (2017). *Computer Simulation of Liquids* (2nd ed.). Oxford University Press. https://doi.org/10.1093/oso/9780198803195.001.0001
[10] Mishra, R., Kunhi, A., Geissbühler, D., Manzano, H., Jamil, T., et al. (2017). cemff: A force field database for cementitious materials including validations, applications and opportunities. *Cement and Concrete Research*, 102, 68-89. https://doi.org/10.1016/j.cemconres.2017.09.003
[11] Dove, M. T., Winkler, B., Leslie, M., Harris, M. J. & Salje, K. H. (1992). A new interatomic potential model for calcite: Applications to lattice dynamics studies, phase transition, and isotope fractionation. *American Mineralogist, 77*(3-4), 244-250.
[12] Lamberty, Z. D., Tran, N. T., van Engers C. D., Karnal, P, Knorr, D. B. & Frechette, J. (2023). Cooperative Tridentate Hydrogen-Bonding Interactions Enable Strong Underwater Adhesion. *ACS Appl Mater Interfaces, 15*(29), 35720-35731. https://doi.org/10.1021/acsami.3c06545
[13] Akbar, N., Akbar, A., Adlan, H. & Nordin, M. (2021). The Characteristics of Limestone and Anthracite Coal as Filter Media in Treating Pollutants from Groundwater. *International Journal of Environmental Science and Development*, 12, 58-62. https://doi.org/10.18178/ijesd.2021.12.2.1318
[14] Berlin, A. A. & Basin, V. E. (1974). *Fundamentals of Polymer Adhesion*. Moscow: Khimiya (Chemistry). (in Russian)
[15] Neumann, A. W., David, R. & Zuo, Y. (2011). *Applied Surface Thermodynamics*. CRC Press. https://doi.org/10.1201/EBK0849396878
[16] Popov, V., Pohrt, R. & Li, Q. (2017). Strength of adhesive contacts: Influence of contact geometry and material gradients. *Friction*, 5, 308-325. https://doi.org/10.1007/s40544-017-0177-3
[17] Pocius, A.V. & Dillard, D. A. (2002). *Adhesion science and engineering. Volume: v.1 The mechanics of adhesion*, Dekker.
[18] Vicsek, T. (1992). *Fractal Growth Phenomena* (2nd ed.). World Scientific. https://doi.org/10.1063/1.4822864
[19] Falconer, K. (2009). *Fractals: Mathematical Foundations and Applications.* Moscow: Mir. (in Russian)
[20] Feder, J. (1988). *Fractals*. New York: Plenum Press. https://doi.org/10.1007/978-1-4899-2124-6
[21] Meakin, P. (1983). Models for the formation of fractal clusters by irreversible diffusion-limited aggregation. *Physical Review A, 27*(3), 1495-1507. https://doi.org/10.1103/PhysRevLett.51.1119
[22] Mazo, R. M. (2008). The smoluchowski equation', Brownian Motion: Fluctuations, Dynamics, and Applications. *International Series of Monographs on Physics*, 97-110. https://doi.org/10.1093/acprof:oso/9780199556441.003.0008
[23] See https://web.mit.edu/reuter
[24] Lehmann, E. L. & Romano, J. P. (2005). *Testing Statistical Hypotheses* (3rd ed.). Springer.
[25] Mood, A. M., Graybill, F. A. & Boes, D. C. (1974). *Introduction to the Theory of Statistics* (3rd ed.). McGraw-Hill.

Authors' contacts:

Andrii Kolesnykov, PhD, Assistant Professor
Odessa State Academy of Civil Engineering and Architecture,
Construction and Technological Institute,
Department of Chemistry and Ecology,
Didrihsona st., 4, 65029 Odesa, Ukraine
Kolesnikov_himek@odaba.edu.ua

Mykola Khlytsov, PhD, Assistant Professor,
Odessa State Academy of Civil Engineering and Architecture,
Construction and Technological Institute,
Department of Processes and Apparatuses in the Technology of Building Materials,
Didrihsona st., 4, 65029 Odesa, Ukraine
khlytsov@odaba.edu.ua

Svitlana Semenova, PhD, Assistant Professor,
Odessa State Academy of Civil Engineering and Architecture,
Construction and Technological Institute,
Department of Chemistry and Ecology,
Didrihsona st., 4, 65029 Odesa, Ukraine
semenova@odaba.edu.ua

Mario Šercer, PhD
(Corresponding author)
Development and Training Centre for the Metal Industry - Metal Centre Čakovec,
Bana Josipa Jelačića 22 D, 40000 Čakovec, Croatia
ravnatelj@metalskajezgra.hr

ISSN 1846-6168 (Print), ISSN 1848-5588 (Online)
https://doi.org/10.31803/tg-20231008170542

Preliminary communication
Received: 2023-10-08, Accepted: 2025-02-07

Deformation of the Cooling Vessel's Shell

Katarina Pisačić*, Mario Pintarić

Abstract: This paper will present the procedure for calculating the internal volume obtained by deforming the outer shell of the cooling vessel. A nonlinear finite element analysis was conducted using the Ansys software. The material behavior when it transitions into the plastic deformation region was defined, and the deformed shape resulting from the plastic deformation of the outer shell of the cooling vessel was obtained. The volume of the deformed shape was determined with several simplifications: the profile of the outer shell was simplified into a circular arc, and the parameters of the circle were calculated. The formula for the volume of a rotational body was used to calculate the volume.

Keywords: Ansys software; cooling vessel; deformation; shell; volume

1 INTRODUCTION

The Finite Element Method is a technique based on the discretization of a continuum [1]. The continuous domain is divided into a finite number of subdomains called finite elements. The state within each element, such as displacement, deformation, stress, temperature, and other field variables, is described using interpolation functions. These functions must satisfy appropriate conditions to make the discretized model closely approximate the behavior of the continuous system. With the proper formulation of finite elements, the accuracy of the approximation to the exact solution increases as the number of elements increases. During the derivation of equations, we distinguish between the following procedures:

- the process of deriving the finite element equation based on solving a differential equation, i.e., the method of weighted residuals,
- Derivation of the finite element equation based on variational formulations (the principle of virtual displacements, the principle of minimum total potential energy, the principle of virtual forces, and the principle of minimum complementary energy).

If the independent variables are forces or stresses, the Finite Element Method is referred to as the force method. The other method is the displacement method, where the unknowns are the displacements at the nodes of finite elements. If the independent variables are displacements, stresses, and strains, it is a mixed formulation of finite elements. Due to the large number of equations and unknowns, the Finite Element Method uses computers to solve problems. Various software packages are based on the Finite Element Method, and in this work, Ansys software is used for finite element analysis.

We calculate the volume of the cooling vessel so that the vessel does not come under the state inspection. Previous research has not addressed this issue, so due to the lack of prior experience and for the sake of simplicity, this method of volume calculation has been chosen.

2 COOLING VESSEL

For the purpose of cooling food products and alcoholic beverages, vessels with double or triple shells are manufactured, known as duplicators in food technology [2]. Water or glycol flows through the double shell, cooling the contents of the vessel through the wall. In some designs, there is additional thermal insulation on the outside, along with the outer shell of the vessel that protects the insulation and the vessel's fittings. An example of such a vessel for beer is shown in Fig. 1.

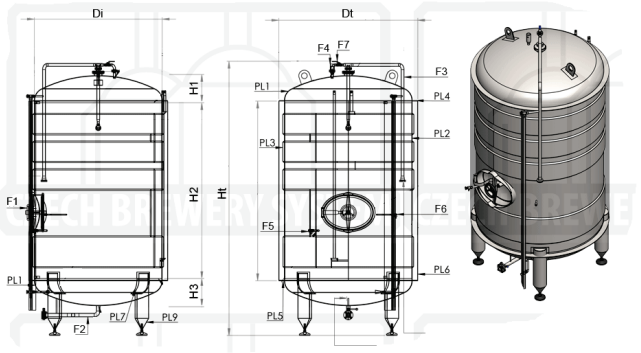


Figure 1 An example of a cooling vessel for beer [3]

One way to implement a double shell is to weld the thinner shell, and a pressure cleaner is connected through the provided opening to create pressure, causing the stretching of the outer, thinner shell. An example of simulating the outer shell of such a vessel is presented in this paper.

The material of such vessels is most commonly stainless steel, such as SS304 or AISI 304 (X5CrNi18-10 acc. to EN 10027-1). The typical behavior of SS304 stainless steel is shown in Fig. 2. The figure depicts the stress-strain relationship in a tensile test. This diagram is obtained from the manufacturer, and a similar characteristic is obtained depending on the production technology and parameters. If the manufacturer does not guarantee the properties, the mechanical properties of samples from a particular batch of material can be tested on a tensile testing machine to obtain the accurate characteristic.

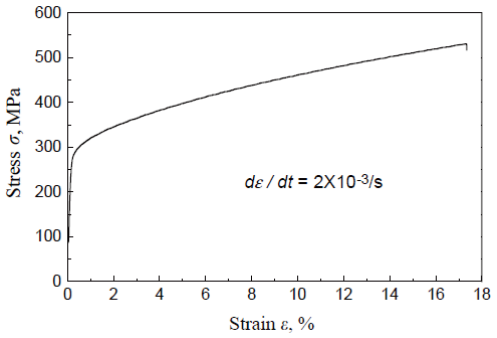


Figure 2 The deformation diagram for stainless steel SS304 [4]

Table 1 Material properties for SS304

Mechanical Properties	
Density (kg/m^3)	7930
Young's Modulus (GPa)	190-200
Poison' Ratio	0,29
Tensile Strength, Ultimate (MPa)	505
Tensile Strength, Yield (MPa)	215
Strain Hardening Coefficient, *n*	0,10-0,50
Chemical Composition	
Carbon, C	≤0,080 %
Chromium, Cr	18-20 %
Iron, Fe	66,345-74 %
Manganese, Mn	≤2,0 %
Nickel, Ni	8,0-10,5 %
Phosphorus, P	≤0,045 %
Silicon, Si	≤1,0 %
Sulfur, S	≤0,030 %

For the purposes of this analysis, and in the absence of an exact model for stainless steel, structural steel and material characteristics from the Ansys material database were used (Fig. 3). Material properties were used for nonlinear analysis, and the values are approximate. The work represents a calculation method rather than an exact solution.

Table 2 Material properties for Structural Steel

Mechanical Properties	
Density (kg/m^3)	7850
Young's Modulus (GPa)	0,3
Poison' Ratio	200
Tensile Strength, Ultimate (MPa)	460
Tensile Strength, Yield (MPa)	250

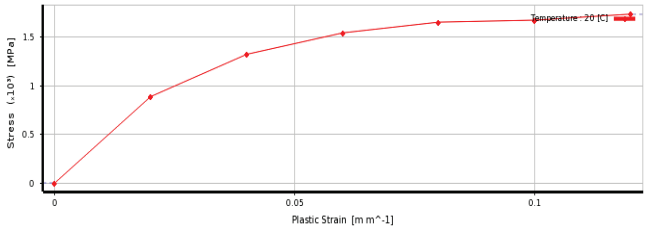


Figure 3 Material characteristics

2.1 The Outer Shell of the Cooling Vessel

The outer shell of the cooling vessel has dimensions: height 300 mm, diameter 1000 mm, and wall thickness of 1 mm (Fig. 4).

It is necessary to calculate the volume between the outer and inner walls of the vessel obtained through the plastic deformation of the outer shell. The shell is subjected to a pressure of 7 MPa and then relieved.

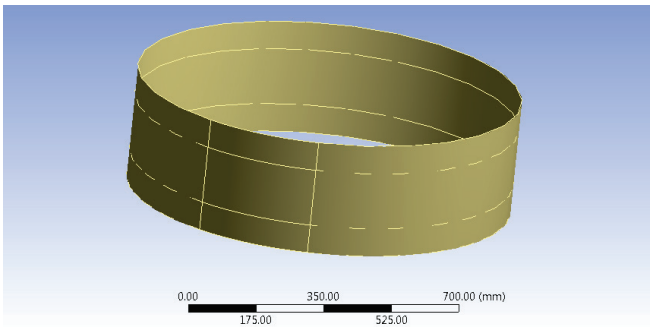


Figure 4 The appearance of the outer shell of the vessel

3 ANALYSIS USING SHELL ELEMENTS

In the first case, we are considering a model designed as a thin-walled structure. We defined a finite element mesh, using SHELL281 elements [5]. The geometry of this element is shown in Fig. 5. The SHELL281 element is suitable for analyzing thin to moderately thin structures. The element has 8 nodes, and each node has 6 degrees of freedom: translation in the *x*, *y*, and *z* directions, and rotation around the *x*, *y*, and *z* axes. When the membrane option is used, the element only has translational degrees of freedom. This element is suitable for linear problems, large rotations, and changes in shell thickness for nonlinear analysis. SHELL281 can be used for layered constructions to model composites or sandwich structures. The element's formulation is based on logarithmic strain and true stress. The element's kinematics allows for finite membrane deformation (stretching).

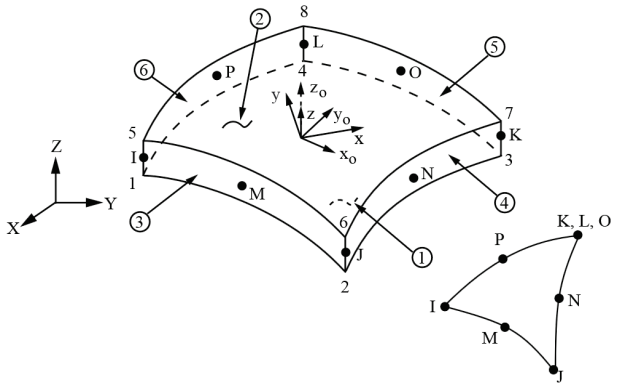


Figure 5 The geometry of the SHELL281 element [5]

The vessel is modeled as a thin-walled structure, and SHELL281 elements have been used. A finite element mesh has been generated (meshing performed), as shown in Fig. 6.

After applying boundary conditions, the outer edge and lower edge are defined as a fixed support, with zero displacements in all axes and zero rotations around all axes (Fig. 7).

The vessel is subjected to a compressive load (Fig. 8), and the load is defined as a time-dependent table (Fig. 9).

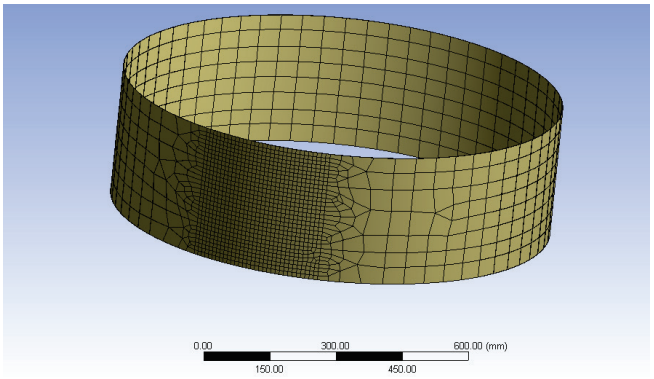


Figure 6 The finite element mesh

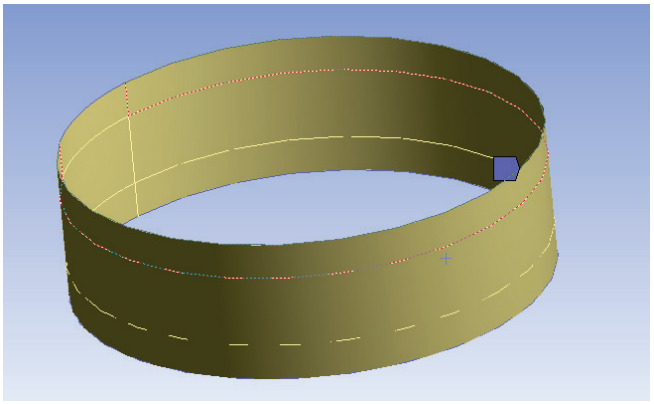


Figure 7 Restraint of the upper and lower edges

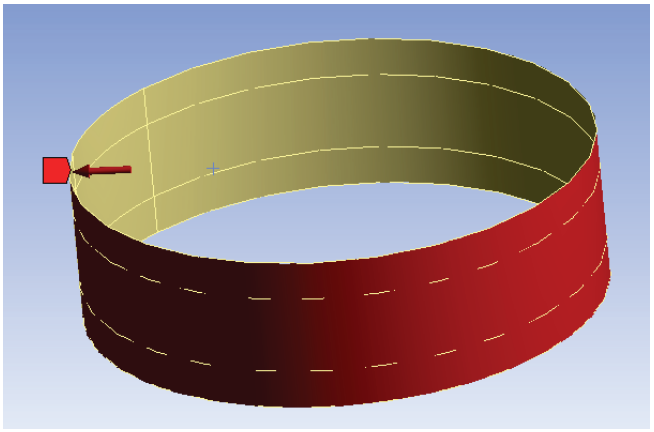


Figure 8 The load on the vessel

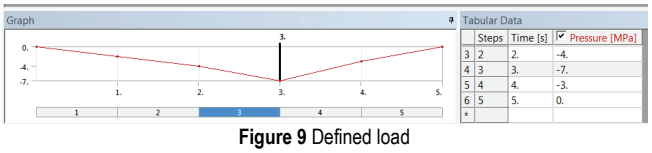


Figure 9 Defined load

Table 3 Defined load in Fig. 9

Steps	Time, s	Pressure, MPa
2	2	–4
3	3	–7
4	4	–3
5	5	0

After all parameters were defined, a finite element analysis was conducted, and results were obtained. Initially, data for the overall deformation throughout the vessel's volume were obtained.

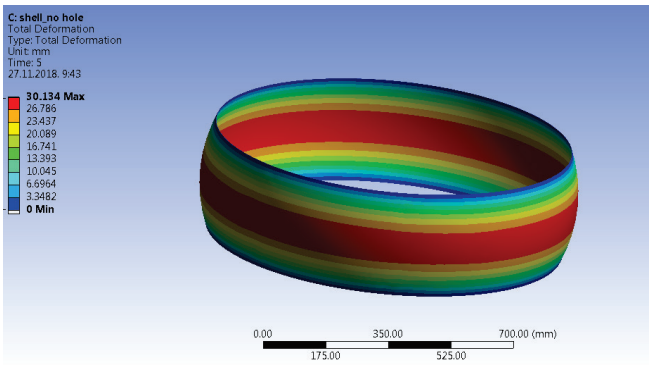


Figure 10 The deformed shape after unloading

From Tab. 1, it is evident that the maximum deflection increases as time increases, corresponding to the increasing pressure. In 3 seconds, the maximum pressure is 7 MPa, after which the pressure decreases, along with the deflection, until it comes to a stop. After unloading, the maximum deformation in the center is 30.134 mm. Deformation at the edges is zero, so the minimum deflection is 0. Tab. 3 displays the results for total deformation over time.

Table 4 Results of maximum total deformation over time for the thin-walled model

	Time, s	Maximum deformation, mm
1	0.2	2.21
2	0.4	4.31
3	0.7	7.17
4	1	9.70
5	1.2	11.26
6	1.4	13.17
7	1.7	15.94
8	2	18.50
9	2.2	21.07
10	2.4	23.86
11	2.7	28.16
12	3	33.11
13	3.2	32.76
14	3.4	32.41
15	3.7	31.87
16	4	31.34
17	4.2	31.08
18	4.4	30.82
19	4.7	30.46
20	5	30.13

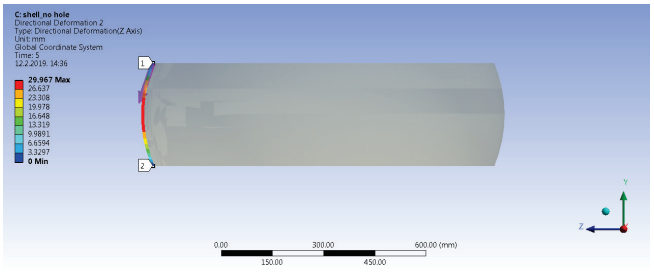


Figure 11 The profile of the deformed shape of the thin-walled shell

To determine the characteristics of the resulting vessel, it is necessary to calculate the internal volume of the vessel obtained in this way. To calculate the desired volume, a path

is defined in the software, and the Finite Element Analysis results are analyzed along that path of cross-section.

Tab. 4 shows displacements in the *z*-axis direction, which in this case represents the horizontal axis. These displacements will be used to determine the profile function for calculating the volume of the rotational body.

Table 5 Results of deformation in the *z*-axis direction for the thin-walled model

	Length, mm	Deformation in the direction of the axis *z*, mm
1	0	0
2	6.25	2.16
3	12.5	4.63
4	18.75	6.81
5	25	8.83
6	31.25	10.81
7	37.5	12.70
8	43.75	14.50
9	50	16.22
10	56.25	17.86
11	62.5	19.40
12	68.75	20.85
13	75	22.19
14	81.25	23.43
15	87.5	24.56
16	93.75	25.58
17	100	26.49
18	106.25	27.30
19	112.5	28.00
20	118.75	28.60
21	125	29.08
22	131.25	29.47
23	137.5	29.75
24	143.75	29.91
25	150	29.97
26	156.25	29.91
27	162.5	29.74
28	168.75	29.47
29	175	29.08
30	181.25	28.59
31	187.5	27.99
32	193.75	27.29
33	200	26.49
34	206.25	25.57
35	212.5	24.55
36	218.75	23.43
37	225	22.19
38	231.25	20.84
39	237.5	19.40
40	243.75	17.86
41	250	16.22
42	256.25	14.50
43	262.5	12.70
44	268.75	10.81
45	275	8.83
46	281.25	6.80
47	287.5	4.63
48	293.75	2.16
49	300	0

The maximum deformation of the shell is on the central cross-sectional line at a height of 150 mm from the bottom. The difference between the total deformation and directed deformation occurs due to material stretching.

4 ANALYSIS USING SOLID ELEMENTS

In the second case, we are considering a model designed as a three-dimensional model. We defined a finite element mesh, using SOLID187 elements [6], as shown in Fig. 12.

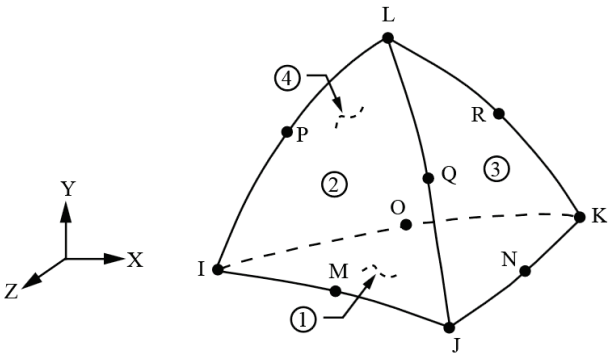


Figure 12 The geometry of the SOLID187 element [6]

The SOLID187 element is a higher-order element with ten nodes. It behaves according to quadratic displacement and is suitable for irregular meshes. Each node on the element has three degrees of freedom: displacement along the *x*, *y*, and *z* axes. The element has the capability to model plasticity, hyperelasticity, creep, strain-hardening, large deflections, and large deformations. It can handle mixed formulations for simulating deformations in nearly incompressible elastoplastic and fully incompressible hyperelastic materials.

The appearance of the finite element mesh is shown in Fig. 13. In the area where the path for the profile is defined, a finer mesh of elements was used. A function was applied to reduce the element size in a specific region, resulting in more accurate results with a smaller step.

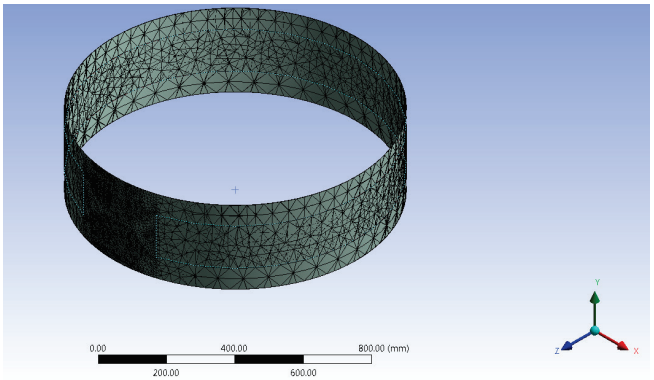


Figure 13 The finite element mesh of the three-dimensional model

Fig. 14 shows the total deformation throughout the volume of the body. The results are very similar to those shown in Fig. 10.

The maximum total deformation is calculated depending on the changing applied pressure. The results of total deformation over time are shown in Tab. 5.

In order to compare the analysis results using SHELL281 elements and SOLID187 elements, data was extracted for the unloaded state at the same cross-section, and a new profile

for calculating the volume of the rotational body was obtained. The results are shown in Fig. 15 and Tab. 6.

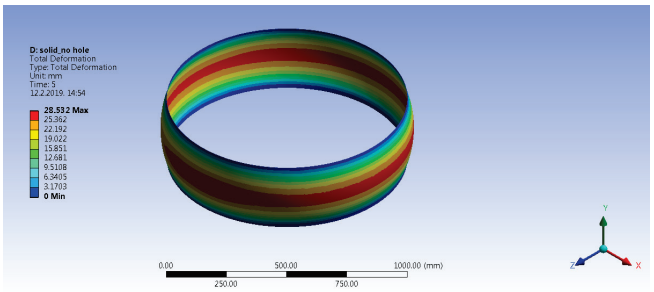


Figure 14 Total deformation of the three-dimensional model

Table 6 The total deformation over time for the three-dimensional model

	Time, s	Maximum deformation, mm
1	0.2	2.59
2	0.4	4.84
3	0.7	7.73
4	1	10.20
5	1.2	11.69
6	1.4	13.34
7	1.7	15.84
8	2	18.23
9	2.2	20.59
10	2.4	23.15
11	2.7	27.12
12	3	31.74
13	3.2	31.39
14	3.4	31.02
15	3.7	30.47
16	4	29.91
17	4.2	29.63
18	4.4	29.35
19	4.7	28.94
20	5	28.53

Table 7 Results of deformation in the *z*-axis direction for the three-dimensional model

	Length, mm	Deformation in the direction of the axis *z*, mm
1	0	–1.91E-14
2	6.25	1.82
3	12.5	3.73
4	18.75	5.60
5	25	7.43
6	31.25	9.22
7	37.5	10.98
8	43.75	12.71
9	50	14.40
10	56.25	16.03
11	62.5	17.61
12	68.75	19.15
13	75	20.63
14	81.25	21.85
15	87.5	22.96
16	93.75	23.96
17	100	24.84
18	106.25	25.62
19	112.5	26.30
20	118.75	26.86
21	125	27.32
22	131.25	27.68
23	137.5	27.93
24	143.75	28.08
25	150	28.14
26	156.25	28.09
27	162.5	27.94
28	168.75	27.70
29	175	27.34
30	181.25	26.90
31	187.5	26.33
32	193.75	25.68
33	200	24.92
34	206.25	24.06
35	212.5	23.08
36	218.75	21.99
37	225	20.79
38	231.25	19.42
39	237.5	17.97
40	243.75	16.44
41	250	14.83
42	256.25	13.14
43	262.5	11.37
44	268.75	9.52
45	275	7.63
46	281.25	5.72
47	287.5	3.79
48	293.75	1.85
49	300	0

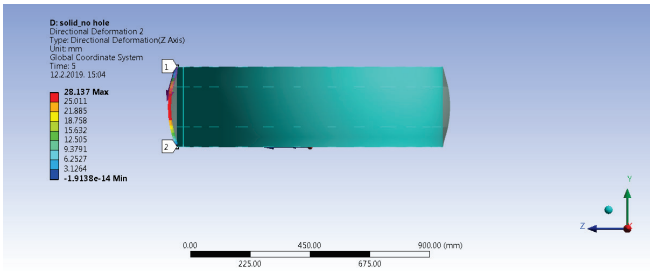


Figure 15 The directed deformation in the *z*-axis direction and the profile of the 3D model

5 THE CALCULATION OF THE VOLUME OF ROTATIONAL BODIES

The volumes of rotational bodies are calculated by taking into account that their cross-section at each level is a known circle, and its area is known [7].

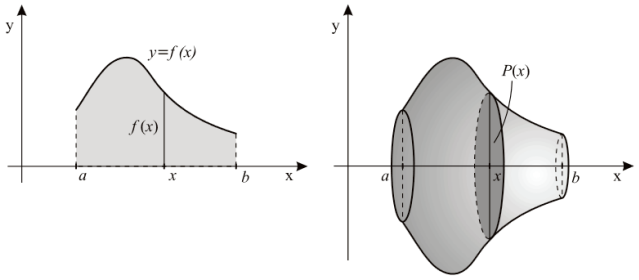


Figure 16 The volume of rotational bodies [7]

The volume of a rotational body when we have a known function $f(x)$:

$$\int_a^b \pi\left(f(x)\right)^2 \mathrm{d}x \tag{1}$$

The volume of a rotational body between two known functions $f(x)$ and $g(x)$:

$$\int_a^b \pi\left(\left(f(x)\right)^2-\left(g(x)\right)^2\right)\mathrm{d}x \tag{2}$$

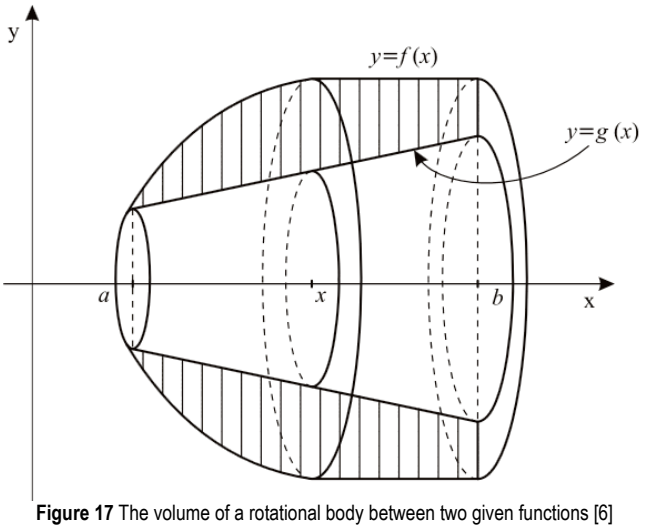


Figure 17 The volume of a rotational body between two given functions [6]

For simplicity, we will assume that the profile line of the deformed shape is a segment of a circle. From the equation of a circle:

$$\left(x-x_0\right)^2+\left(y-y_0\right)^2=R^2 \tag{3}$$

A system of equations is set up to obtain the equation of a circle passing through three points, by which we approximate the deformed profile:

$$\left(x_1-x_0\right)^2+\left(y_1-y_0\right)^2=R^2 \tag{4}$$

$$\left(x_2-x_0\right)^2+\left(y_2-y_0\right)^2=R^2 \tag{5}$$

$$\left(x_3-x_0\right)^2+\left(y_3-y_0\right)^2=R^2 \tag{6}$$

5.1 The Volume of the Thin-Walled Model

The coordinates of three points on the deformed shell of the vessel have been obtained. It is known:

$$x_1=0,\ y_1=500$$
$$x_2=150,\ y_2=530$$
$$x_3=300,\ y_3=500$$

From the system of Eqs. (4), (5), (6), we obtain the parameters of the circle function:

$$x_0=150,\ y_0=140$$
$$R=390\ \text{mm}$$

In this case, the first function describing the upper boundary of the rotational body $f(x)$ is:

$$f\left(x\right)=y_0+\sqrt{R^2-\left(x-x_0\right)^2}$$
$$f\left(x\right)=\sqrt{152100-\left(x-150\right)^2}+140 \tag{7}$$

The second function describing the lower boundary of the rotational body $g(x)$ is:

$$g\left(x\right)=500$$

The volume of the rotating body follows:

$$V=\int_0^{300}\pi\left(\left(f(x)\right)^2-\left(g(x)\right)^2\right)\mathrm{d}x=19457027{,}49\ \text{mm}^3 \tag{8}$$
$$V=19{,}457\ \text{dm}^3$$

The volume of the space inside the vessel obtained by deforming the outer shell calculated using SHELL281 elements amounts to 19,457 dm^3.

6 VOLUME OF THE THREE-DIMENSIONAL MODEL

The coordinates of three points on the deformed shell of the vessel have been obtained. It is known:

$$x_1=0,\ y_1=500$$
$$x_2=150,\ y_2=528$$
$$x_3=300,\ y_3=500$$

From the system of Eqs. (4), (5), (6), we obtain the parameters of the circle function:

$$x_0=150,\ \ y_0=112{,}214$$
$$R=415{,}786\ \text{mm}$$

In this case, the first function describing the upper boundary of the rotational body $f(x)$ is:

$$f\left(x\right)=y_0+\sqrt{R^2-\left(x-x_0\right)^2}$$
$$f\left(x\right)=\sqrt{172877{,}760204-\left(x-150\right)^2}+112{,}214286 \tag{9}$$

The second function describing the lower boundary of the rotational body $g(x)$ is:

$$g\left(x\right)=500$$

The volume of the three-dimensional model body follows:

$$V=\int_0^{300}\pi\left(\left(f\left(x\right)\right)^2-\left(g\left(x\right)\right)^2\right)dx=18406357{,}997\ \text{mm}^3 \tag{10}$$
$$V=18{,}406\ \text{dm}^3$$

The volume of the space inside the vessel obtained by deforming the outer shell calculated using SOLID187 elements amounts to 18,113 dm^3.

7 CONCLUSION

Using the finite element method, it is easy to calculate an approximate volume that encloses the deformed shell of a welded cooling vessel. In this study, the deformed profile is approximated by a segment of a circle, and a more detailed analysis and nonlinear regression of cross-sectional point results could provide a more precise approximation. In that case, it would not be enough to only know the maximum deflection in the central part but also the exact arrangement of elements and the displacement of each element individually. This dependency would be represented as a function of the distance from the coordinate axis, yielding different results.

In this study, two analyses were performed using SHELL281 elements and SOLID187 elements, and the calculation results of the volume are similar, showing discrepancies at the level of 5 %. The results should be confirmed by theoretical calculations and volume measurements after the vessel is manufactured under the same loading conditions as in the simulation.

In the subsequent stages of the research, it will be necessary to create a physical model to validate the theoretical numerical simulation.

8 REFERENCES

[1] Sorić, J. (2004). *Metoda konačnih elemenata*. Golden marketing – Tehnička knjiga, Zagreb. (in Croatian)
[2] Yari, S., Safarzadeh, H., & Bahiraei, M. (2023). Experimental study of storage system of a solar water heater equipped with an innovative absorber spherical double-walled tank immersed in a phase change material. *Journal of Energy Storage, 61*, 106782. https://doi.org/10.1016/j.est.2023.106782
[3] https://www.czechminibreweries.com/production/brewery-components/cold-block/beer-maturation-lager-tanks/
[4] Kang, G., Gao, Q. & Yang, X. (2004). Uniaxial and non-proportionally multiaxial ratcheting of SS304 stainless steel at room temperature: experiments and simulations. *International Journal of Non-Linear Mechanics, 39*, 843-857. https://doi.org/10.1016/S0020-7462(03)00060-X
[5] https://www.sharcnet.ca/Software/Ansys/16.2.3/en us/help/ans_elem/Hlp_E_SHELL281.html
[6] https://www.sharcnet.ca/Software/Ansys/17.0/en_us/help/ ans_elem/Hlp_E_SOLID187.html
[7] Primjene integrala https://www.fsb.unizg.hr/matematika/ download/ZS/dif_racun_II/09_primjene_integrala.pdf (in Croatian)

Authors' contacts:

Katarina Pisačić, mag. ing. mech.
(Corresponding author)
University North, University Center Varaždin,
Jurja Križanića 31b, 42 000 Varaždin, Croatia
kpisacic@unin.hr

Mario Pintarić, mag. ing. mech.
University North, University Center Varaždin,
Jurja Križanića 31b, 42 000 Varaždin, Croatia
mpintaric@unin.hr

ISSN 1846-6168 (Print), ISSN 1848-5588 (Online)
https://doi.org/10.31803/tg-20231118075406

Preliminary communication
Received: 2023-11-18, Accepted: 2024-01-02

Design and Analysis of Nuclear Transportation Cask Using Phase Changing Material

Girish Venkatesh Gudi*, R. G. Nagraj, Abhijit Dandavate

Abstract: In the nuclear sector, secure transportation of radioactive materials is crucial in the nuclear sector, and cask design and analysis are crucial for maintaining fuel integrity. Peak cladding temperature (PCT) is a significant factor in fuel failure likelihood. This research provides a thorough approach to designing and analysing nuclear transportation casks, focusing on PCT. Advanced computer techniques are used in the study, which includes a number of variables including the thermal properties of the cask and external heat sources. In this work, a composite PCM made of polyethylene glycol (PEG) and octadecanol is employed to provide equal heat transfer. The proposed composite PCM is reinforced by (Carbon Nano Tubes) CNTs to ensure homogeneous heat conduction within the PCM particles. The thermal behaviour of the cask is simulated using ANSYS, which takes into consideration conduction, convection, and radiation heat transfer among other aspects. The PCT for spent fuel assemblies for light water reactors operating in helium settings with varying decay heat loads and temperature boundary circumstances was calculated using these simulations. Finally yet importantly, the outcome has been validated by comparison with related studies.

Keywords: ANSYS; Peak Cladding Temperature (PCT); Phase Changing Material (PCM); transportation cask

1 INTRODUCTION

At the conclusion of the nuclear fuel cycle, a conveyance task will be needed to transport spent nuclear fuel (SNF) from temporary containers in utilizing crops (reactors of nuclear power plants or scientific sites) to reprocess along with permanently stored sites [1]. Radioactive compounds are normally transported in a specially constructed large container known as a transportation cask. The container might transported through a number of modes, namely road, rail, sea, and potentially air [2]. The procedure for shipping must fulfill a number of duties to assure the security of people, citizens, along with the environment during all situations, namely nuclear material confinement, mechanical security, thermal safeguarding, along with radiologic shielding. It additionally has to allow for simple loading, drying, tightening up, & handling as a way to maintain operational time as well as expenses [3].

SNF gatherings for simple water reactors are typically composed of fuel rods that are joined in square arrays by headers, footings, and periodical spacer plates. Every rod is a Zircaloy-coated tube filled with highly radioactive fuel pellets with fission gases [4]. Spent components are carried far from nuclear power plant zones in thick-walled vessels. Before the US Nuclear Regulatory Commission (NRC) issuing a Certificate of Compliance for a shipment, Federal Regulations (10CFR71) need the supplier to produce a safety analysis report (SAR) showing that the package will continue to provide its containment, shielding, along with criticality handling abilities under normal transport conditions (NCT) afterwards following a series of complicated cases [5]. Amongst the situations was a 9-meter drop onto an unyielding work layer, a 1-meter drop onto a steel piercing rod, full engulfment in an 800 °C fire for 30 minutes, and water submersion. The fuel cladding contains the extremely radioactive fuel pellets along with fission gases, but its integrity needs to be preserved with the goal to keep the components in the expected configuration throughout regulatory review. Radial hydrides can form within the fuel cladding under NCT, making it brittle if the temperature rises beyond 400 °C. If the cladding temperature rises over 570 °C or 750 °C after a fire, it may undergo creep deformation or burst rupture [6].

A transportation cask's body consists of a hollow cylinder with a minimum of three concentric levels: an outer steel shell that interacts with the outside atmosphere, a thick shielding material coating providing thermal insulation along with neutron absorption operations, along with an interior steel shell that lines the cavity where hazardous substances are maintained [7]. These may include vitrified trash packaged in steel canisters or used fuel rods wrapped in metal and grid-stacked into a basket housing the fuel assembly. The container is kept secure by a tight-fitting lid. Shock absorbers are mounted to both ends of the cask to maintain mechanical integrity in the event of a significant accident [8].

PCMs have generated curiosity regarding their use in thermal security, storing power, thermal regulation, and conserving electricity because of their high quantities of enthalpy and of their transitions from liquid to solid [9]. Since they collect heating/cooling thermal energy in the thermal circuit in addition to safeguarding the back-end equipment/materials from thermal runaway/overcooling, PCMs are frequently used as heat capacitors in thermal shielding. Organic PCMs are suitable for thermal barrier operations due to their poor thermal conductivity along with elevated latent heat [10]. However, organic PCM leakage through the solid-liquid phase transition procedure shows a significant impact on their application effectiveness. To address this problem, stable composite PCMs could be made by mixing PCMs with a variety of porous supporting substances [11]. Because of their significant heat conductivity, porous based on carbon supports including extended graphite, carbon nanotube sponge, along with graphene aerogels are not conducive to increasing thermal safeguarding efficacy [12].

The design of casks is crucial in the field of nuclear transportation for assuring the secure and safe transfer of radioactive materials. Our addition, which builds on the

current cask design, attempts to improve its characteristics for increased security and effectiveness. We may further enhance the current cask to meet changing industry requirements by concentrating on important factors like structural integrity, thermal management, real-time monitoring, modular design, and regulatory compliance. Primarily, it is crucial to reinforce the cask's structural integrity. We can increase its resistance to impact, corrosion, and environmental factors by utilizing cutting-edge materials like high-strength metals or composites. This improvement will reduce the risk of radioactive material escape by strengthening the cask's ability to withstand mishaps or occurrences during transit. Furthermore, the integrity of the radioactive elements depends on efficient thermal control inside the cask. The cask will be able to maintain constant inside temperatures by including better insulation materials and cooling technologies, notably for high-level radioactive waste. We can increase the overall reliability of the cask design by introducing passive cooling techniques that lessen the need for active cooling systems.

Therefore, the present paper does a thermal study of dry storage facilities for spent nuclear fuel at various conditions, relying on designing foundation estimates and inputs to ensure that the peak cladding temperatures do not go over an established limit. Therefore, a distinctive substance for thermal shielding is determined depending on the thermal characteristics of the entire system.

2 LITERATURE SURVEY

Csontos et al. [13] provided the results of a double-blind benchmark created to estimate the accuracy of simulations of dry storage structures and obtain knowledge of the main causes of modelling inaccuracy. The DOE/EPRI High Burn up (HBU) project's experimental results are contrasted with the four entrants from this round robin. The temperature within the fuel is being monitored by thermocouples while it is kept in a bolted cask together with 32 HBU pressurized water reactor components. The results highlighted the need of improving essential uncertainty component assessments in modelling solutions, especially those pertaining to internal gap size changes.

Zinet et al. [14] used finite element modelling to examine the secondary impacts of the vapour transfer mechanisms unique to these endothermic shielding materials. The framework was developed to study heat transport within a standard cask's wall. The framework considers mechanisms including evaporation, diffusion, and re-condensation of the water content that induce a transfer of latent heat within the medium. The materials under consideration include plaster, phenolic foam, and polyester resin compound. This sophisticated model's forecasts are contrasted with the results of a framework that just includes conduction heat transfer. It has been established that heating kinetics are significantly influenced by gas transport phenomena. Following that, the effects of modelling components such as material porosity and condensation coefficient are addressed.

Two thermal simulations of the SNF storing casks CASTOR 440/84 and CASTOR 440/84M were created utilizing the COBRA-SFS code, an internationally recognized and tested code used for dry storage cask licensing and safety evaluation, according to Seveceka et al. [15]. At the Nuclear Power Plant Dukovany, the WWER-440 SNF is being dry stored in the casks. The SUJB examined and utilized both versions for assessing multiple problems in different configurations, primarily for licensing reasons. The Czech Republic's State Office for Nuclear Safety is the regulatory agency in charge of awarding licenses for spent nuclear fuel storage and transportation containers (SCs) (SNF).

Schwartz [16], According to research, the buffer's auxiliary minerals have a big influence on how quickly the copper shield corrodes. According to groundwater chemistry, a buffer containing calcium sulphates (gypsum or anhydrite) produces a Cu° corrosion depth three to eight times bigger than one without calcium sulphates. The existence or absence of iron oxides (including hematite, goethite, or lepidocrocite) has no influence on reactive-transport models. Iron oxides, however, drastically reduce corrosion depth when present, according to mineral equilibrium thermodynamics.

Dong-Won Lim et al. [17] to prevent the intrusion of an inert sodium area into the air-cooling space built a double-cylinder cask. When both the weight and the shielding thickness were considered, the PSO technique produced a shielding thickness of 26 cm. We were able to assess the effectiveness of the shielding by calculating the radiation dosage of wasted fuel retrieved before and after a year of cooling. Two distinct fuel locations discovered throughout transportation are being studied in order to address a functional issue in a cask drive system. As a result of the analysis, it is determined that the current cask design satisfies legal criteria when used ordinarily.

Alsmadi et al. [18] studied the shielding capabilities of the majority of prevalent structurally amorphous metals (SAMs) for potential usage as coating barriers for radiation shielding and corrosion on spent fuel dry cask canister uses. SAM1651 possessed the greatest attenuation coefficients along with smallest rates of exposure at low photon frequencies owing to its substantial density (7.6 g/cm3) along with elemental content in molybdenum (Mo) and yttrium (Y), whereas SAM2X5 possessed the opposite property at elevated photon energies owing to its high density (7.6 g/cm^3) and elemental content in manganese (Mn) and tungsten (Si).

Hyungjin Kim et al. [19] says that depending on the environment they are in, spent fuel assemblies in nuclear power plants may reach different peak cladding temperatures. Manteufel and Todreas calculate peak cladding temperature using three models: the Wooton-Epstein correlation, the current thermal conductivity approach by Bahney Lotz, and the two-region model. Using a two-dimensional CFD simulation, the study measured the peak cladding temperatures of a Babcock and Wilcox (15 × 15 PWR) reactor. The simulation is helpful for determining peak cladding temperature since it overestimates the measured peak cladding temperature in a used fuel dry storage cask. The 17 × 17 array has a more aggressive effective thermal conductivity than the 16 × 16 PWR spent fuel arrangement.

Jie Li et al. [20] says that for spent nuclear fuel assemblies in commercial light water reactors, the peak cladding temperature (PCT) is a significant metric. An effective thermal conductivity model, a computational fluid dynamics model, and a linked active thermal conductivity as well as edge conductance model were utilized to determine the PCT for spent fuel assemblies in storage or transit casks. The outcomes showed that the vacuum environment is more interesting than other gas atmospheres, as the PCT limit is exceeded at lower boundary temperatures for a specified decay heat load. This essay illustrates the PCT computations and contrasts the outcomes produced by various models.

Several issues have been observed regarding the PCT in previous studies:

1) *Inadequate Thermal Insulation:* Higher temperatures within the barrel during shipment may result from inadequate thermal insulation in the cask design. This may lead to higher PCT values that may exceed the safety thresholds established for the fuel rod cladding. Inadequate insulation materials or subpar insulation design may exacerbate this issue.
2) *Insufficient Cooling Mechanisms:* To reduce the heat that the radioactive materials produce while being transported, suitable cooling systems are required. Temperatures may rise and PCT values may increase if the cooling systems inside the cask are improperly constructed or do not operate as intended. Poor design, broken equipment, or shoddy maintenance can all contribute to insufficient cooling.
3) *Inaccurate Temperature Predictions:* For constructing and assessing the performance of nuclear transportation casks, accurate temperature distribution and PCT prediction is essential. However, prior study has demonstrated cases in which the temperature forecasts were wrong, resulting in an underestimating of the PCT. This might jeopardize the fuel rod cladding's transportation safety.
4) *External Environmental Factors:* The external environment may affect the PCT during transit, including the temperature and sunlight exposure. Higher temperatures and higher PCT values can occur because of inadequate consideration of these parameters during barrel design or inadequate preventive measures.

Consequently, it is evident from the study that it is important to construct a nuclear transportation cask with a high temperature resistant capacity.

3 RESEARCH METHODOLOY

This study suggested a unique thermal shielding for the secure transportation of Spent Nuclear Fuel (SNF) and tests its effectiveness at various temperatures using a 3D of a vertical dry cask built for computer simulation utilizing ANSYS/FLUENT. Based on suggested cask, the SNF is initially covered with steel, and next to that is a lead-based radiation shielding. The role of thermal shielding, which is essential for cask safety, is next discussed. The thermal shielding must be strong enough to prevent both the transmission of inside heat and the transmission of outside heat outdoors. Many low thermal conductive materials are in use, but recently PCM materials have seen significant application for greater SNF applicability. Uneven heat transmission caused by large volumetric change is one of the main issues with PCM. In order to provide uniform heat transmission, a composite PCM made of polyethylene glycol (PEG) and octadecanol is utilized. The suggested composite PCM is strengthened with CNTs to provide uniform heat conduction inside the PCM molecules. The CNT aids in the quick and equal transfer of heat energy; as a result, the volumetric change is uniform throughout the body of the cask, extending life.

The next step is to provide a low thermal conductivity encapsulation in order to bind the thermal energy from a free flow convection process. Thus, the suggested system used 60% of Silica aero gel together with 20% fibre glass and 20% TiO2 to greatly reduce convection. Low thermal conductivity silica aerogels are utilized to encapsulate PCMs in polyethylene glycol (PEG) and octadecanol with reinforced CNT to create composite PCMs. The suggested aerogel may exhibit low thermal conductivity, high latent heat, big compressive strength, good hydrophobicity, and higher thermal cycle stability when compared to traditional aerogel. Due to the synergistic effect between the composite PCMs' high latent heat also the low thermal conductivity, the proposed silica aerogel mix-based CNT reinforced composite PCMs have the potential to directly apply to the thermal insulation and protection device. Monolithic mechanically reinforced glass fibre TiO2-silica aerogel composites have been successfully made using two-step sol-gel system. The main silica precursor used was methyl-tri-meth-oxy-silane, which was joint with an environmentally friendly ethanol/water mixture as the solvent and dried at ambient pressure. In this method, the two phases are chemically bound together, and the silica particles take on the elongated shape of fibers, improves the mechanical properties and Young's modulus when compared to the traditional thermal shielding, thereby supplying an impact shield.

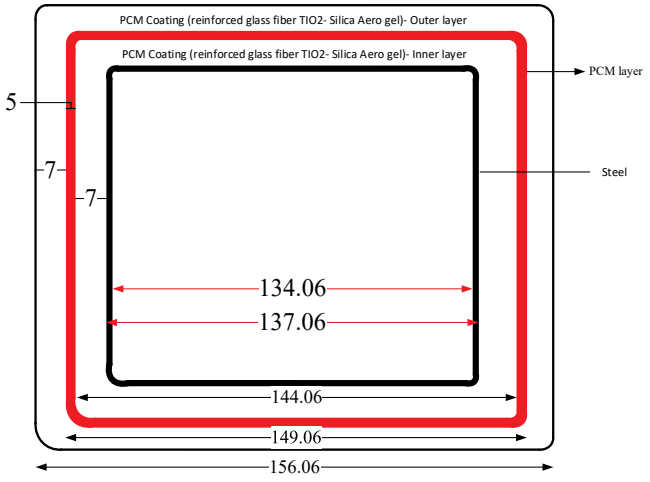


Figure 1 Dimensional parameter of proposed nuclear cask (all the dimension are in mm)

3.1 Modelling of Peak Cladding Temperature

BWR assembly geometries: In Fig. 2, a transverse cross-section of a B&W 9 × 9 fuel assembly, or one-fourth of an assembly, is shown. There are 81 wasted fuel rods in all. Their sheathing has an outside diameter and thickness of 10.77 mm and 0.76 mm, correspondingly. The fuel pellet measures 9.06 mm in diameter. The wasted fuel assembly's active fuel length is 3810 mm. Tab. 1 lists the assembly dimensions that were utilized.

Table 1 Parameter used for modelling spent nuclear cask

Parameter	
Number of rods	79
Rod Pitch	14.53 mm (0.572 in)
UO_2 Pellet Diameter	9.06 mm (0.3565 in)
Cladding Outer Diameter	10.77 mm (0.424 in)
Cladding thickness	0.76 mm (0.030 in)
Number of water Rods	2
Water Rod Outer Diameter	10.82 mm (0.426 in)
Water Rod Thickness	0.81 mm (0.032 in)
Channel Inside Width	134.06 mm (5.278 in)
Channel thickness	2.03, 2.54, or 3.05 mm (0.080, 0.100, or 0.120 in)
Channel Corner Radius	9.65 mm (0.380 in)
Active Fuel Length	3810.0 mm (150 in)

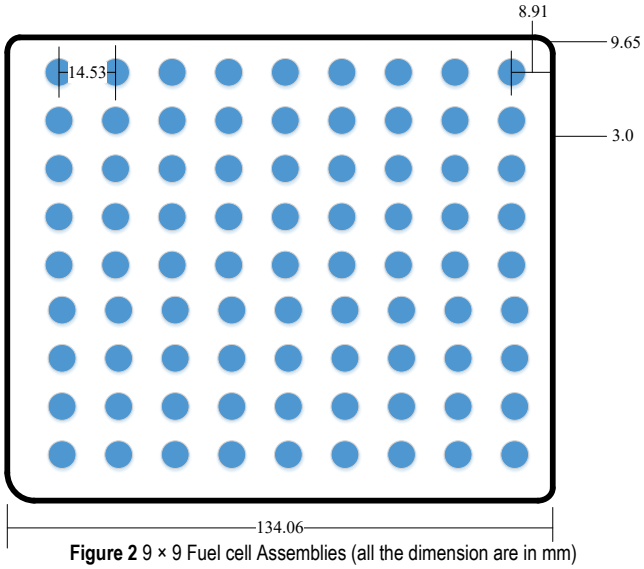


Figure 2 9 × 9 Fuel cell Assemblies (all the dimension are in mm)

3.2 ANSYS Thermal Model

The B&W 9 × 9 BWR assembly's two-dimensional Thermal mesh model is shown in Fig. 4. Using symmetry, this model represents one quarter of an assembly. The method comprises the basket wall with the space between fuel cladding and fuel rods, supposing two planes of symmetry.

A steady-state energy equation was solved consuming a finite volume approach and a second-order upwind discretization methodology, yielding the conduction and radiation temperature findings. The differential equations of the mathematical model are discretized consuming the finite volume technique (FVM), and the resulting algebraic equation system is solved by means of a segregated implicit solver. In order to speed up Gauss-Seidel procedure, calculations are first linearized and then sequentially solved (Hutchinson and Raithby, 1986). Using the Semi-Implicit Method for Pressure-Linked Equations (SIMPLE) technique, the pressure velocity coupling is accomplished (Patankar, 1980).

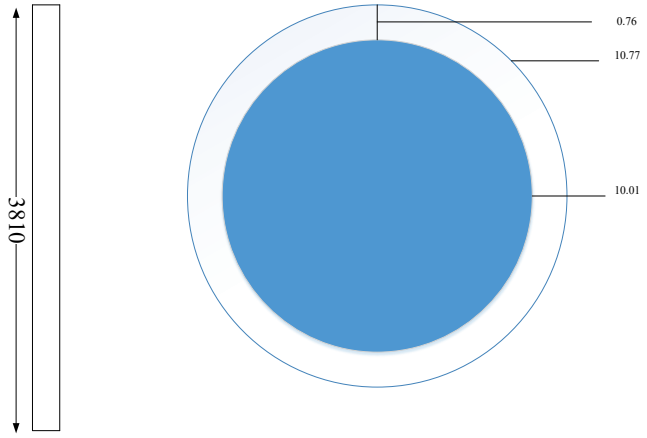


Figure 3 Single fuel rod parameters (all the dimension are in mm)

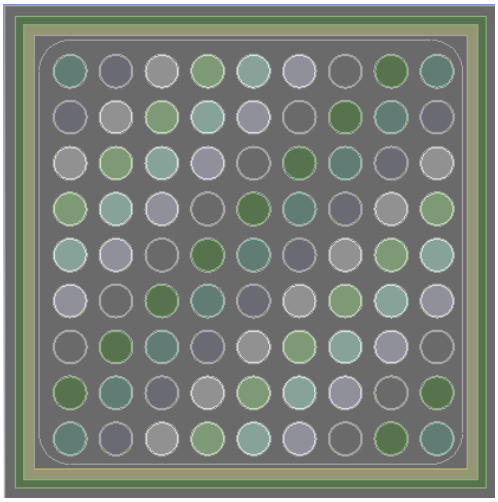


Figure 4 Geometry

The mass and momentum residuals must be 10^{-3} and the energy equation must be 10^{-14} for the convergence criterion to be met. The number 10^{-14} for the energy formula was chosen since the net heat transfer error must be very tiny, signifying that a complete heat balance has been reached. Helium, fuel cladding, and fuel rod's thermal characteristics (specific heat and thermal conductivity) were believed to have a temperature-dependent value (KHNP, 2008).

Radiation model: The radiation of the spent fuel assembly was solved using the discrete ordinates (DO) radiation model. For a finite number of discrete solid angles, each associated with a vector direction fixed in the global Cartesian system (x, y, and z), the discrete ordinates (DO) radiation model solves the radioactive transfer equation (RTE). The user controls the degree of angular discretization that is comparable to selecting the amount of rays in the

discrete transfer radiation model (DTRM). Ray tracing is absent from the DO model. The DO algorithm using a transport calculation instead represents the radiation intensity in the spatial coordinates (x, y, and z). Similar to how the energy and fluid flow equations were solved, so was this one.

The DO model has the benefit of being able to be used for modelling non-gray banded radiation, radiation in semi-transparent material, and all types of optical thickness. It remains computationally expensive and has a restricted amount of radiation directions, resulting in numerical smearing. The discrete ordinates radiation model's angular discretization and pixilation values have a big impact on how temperature behaves. Therefore, it is critical to select an appropriate pixilation and discretization value. In ANSYS (2013) and US NRC (2013), the impact of these values in relation to a pixilation or discretization constant is noted.

3.2.1 Model Creation

The ANSYS models are built for the spent fuel shown in Fig. 4. The PCTs were calculated using the fluent code. The steady-state energy formula was determined by employing the finite volume method and a second-order discretization technique, producing values for radiation as well as conduction temperatures. For obtaining the pressure and velocity fields, the steady-state conservation of momentum formulas was addressed utilizing a second-order upwind technique for natural convection/radiation computations. All conduction/radiation and radiation/natural convection scenarios were conducted to determine the impacts of natural convection on the BWR (9 × 9) spent fuel assembly with helium backfill.

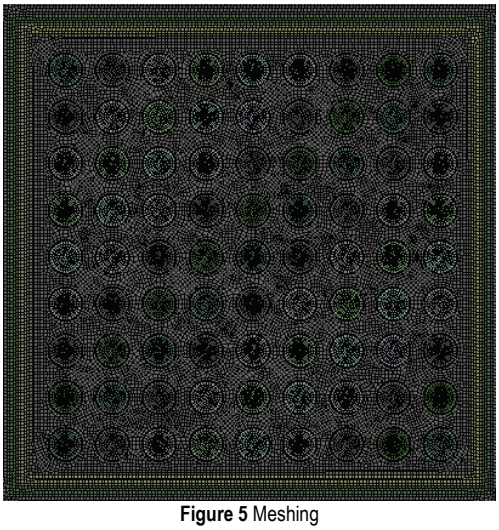


Figure 5 Meshing

3.2.2 Meshing of Geometry

In ANSYS, meshing is the process of turning a geometry into a network of nodes and elements called a finite element mesh. The discretization of the geometry into manageable, small components during the meshing process is a critical stage in the finite element analysis (FEA). A variety of meshing tools and methods are offered by ANSYS to produce high-quality meshes for precise and effective simulations.

Table 2 Meshing size

Meshing size	
Nodes	413971
Elements	76196

3.2.3 Boundary Condition

Boundary conditions in ANSYS are used to specify how a finite element model behaves at its edges. They are necessary for correctly replicating real-world settings and producing insightful findings. You may simulate a variety of physical circumstances by using boundary conditions, which define the limits and loads that are applied to the model.

Table 3 Boundary condition for material

Steel	Density	7850 kg/m^3
	Thermal conductivity	45 W/mK
	Specific heat	0.49 kJ/kgK
PCM with CNT (0.5% of grapheme with erythritol)	Density	1480 kg/m^3
	Thermal conductivity	1.095 W/mK
	Specific heat	1.35 kJ/kgK
Silica Aerogel (Silica + Glassfibre + TiO_2)	Density	0.12 kg/m^3
	Thermal conductivity	0.12 W/mK
	Specific heat	0.023 kJ/kgm^3
Spent Nuclear Fuel (SNF)	Density	2000 kg/m^3
	Thermal conductivity	0.135 W/m^3
	Specific heat	2.640 kJ/kgK
Helium	Density	69.7 kg/m^3
	Thermal conductivity	0.151 W/m^3
	Specific heat	5.1926 kJ/kgK

ANSYS has the tools and capabilities necessary to model time-dependent behaviour and heat generation. Let's examine transient analysis and heat production modelling, two crucial ANSYS concepts relating to time and heat generation.

1) *Transient Analysis:* To examine a system's time-dependent behaviour, transient analysis is performed. It is appropriate for simulations of dynamic events, transient thermal phenomena, or time-varying loads when the model's reaction varies over time. You may carry out transient analysis in ANSYS by taking into account the following:
2) *Time Steps*: The time intervals at which the analysis will be conducted must be specified. You may choose the time step size, overall analysis time, and a number of other time-dependent options using ANSYS.
3) *Loads and Boundary Conditions:* Changing forces or temperature profiles are examples of time-varying loads that may be imposed at various time increments. Similar boundary conditions that rely on time include temperature changes and specified displacements.
4) *Solution Method:* The transient equations are solved numerically by ANSYS. Up until the analysis time is reached, the program use iterative algorithms to calculate the response of the model at each time step.

Modelling of heat generation: Thermal analysis frequently involves heat generation when there are heat sources present, such as electrical components, chemical

processes, or frictional heat. ANSYS offers numerous ways to mimic the production of heat:

1) *Heat Sources:* At particular points in the model, the power density or rate of heat generation can be specified. To do this, either provide the heat generation value or take into consideration internal heat generation by employing the right material attributes.
2) *Joule Heating:* The resistance of conductive materials to electrical current causes joule heating in electrical systems. By defining electrical current and resistive qualities, ANSYS enables you to add the impacts of joule heating.
3) *Radiation and Convection:* Heat transmission by radiation and convection may be modelled using ANSYS. Boundary conditions that take radiation or convective heat transfer coefficients into account for heat exchange with the surroundings can be defined.
4) *Thermal Material Properties:* For a realistic representation of material thermal behaviours, ANSYS offers a wide variety of material models and attributes. These characteristics, including thermal conductivity, specific heat, and thermal expansion, are very important in models of heat generation and transport.

ANSYS enables you to simulate and evaluate time-dependent thermal phenomena, including heat sources, temperature changes, and transient behaviours in diverse engineering applications. This is done by integrating transient analysis with suitable heat generating modelling approaches.

3.2.4 Setup

Configure the thermal solver settings, such as selecting the appropriate solver (e.g., steady-state or transient), specifying convergence criteria, and defining solution methods.

Transient Thermal: Transient thermal analysis in ANSYS Thermal is a powerful tool for studying temperature changes over time in a system. It allows engineers and researchers to simulate and analyse transient heat transfer phenomena, such as how a system responds to sudden temperature changes or how long it takes a system to reach thermal equilibrium. The process begins with creating or importing the system's geometry into ANSYS and generating an appropriate mesh. Next, thermal properties like conductivity, specific heat, and density are defined for the materials involved. Boundary conditions are then specified, including initial temperature distribution, prescribed temperature changes, heat fluxes, and radiation effects. Solver settings are configured, such as the time step size and convergence criteria. The transient analysis is then solved, with ANSYS iterating through time steps to solve the heat transfer equations and update the temperature distribution. Finally, post-processing tools in ANSYS allow for in-depth analysis and visualization of the results, including temperature distribution, heat fluxes, and other relevant quantities at different time points. By following these steps, engineers can gain valuable insights into the transient thermal behaviour of their systems and make informed decisions for design improvements or optimization.

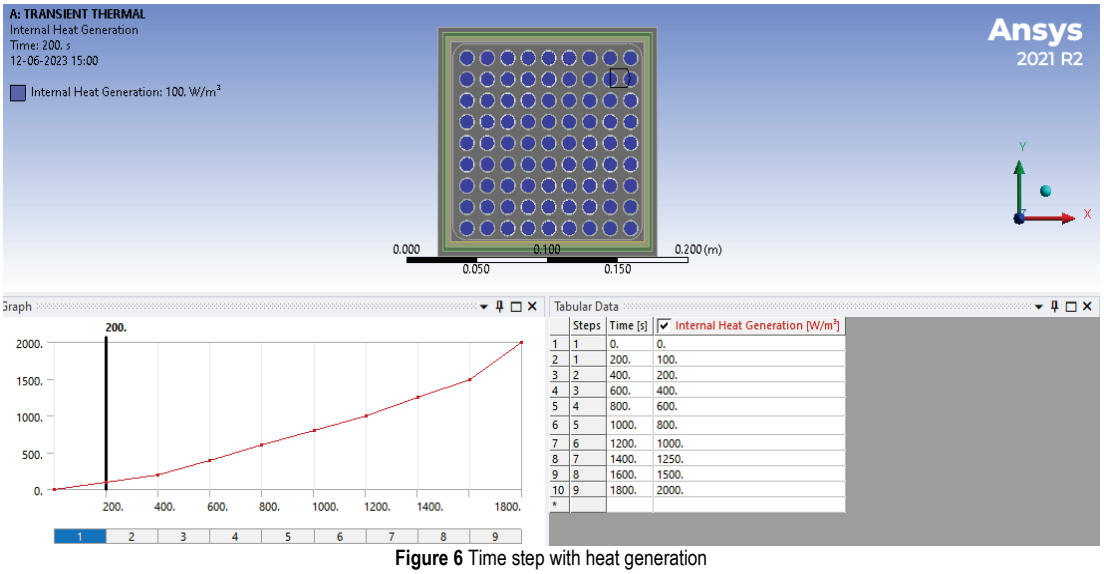


Figure 6 Time step with heat generation

4 RESULT AND DISCUSSIO

Peak cladding temperature (PCT) is the highest temperature that the protective covering of a nuclear fuel rod can reach under regular or unusual operational circumstances. Inside a nuclear reactor, the fuel rods house fuel pellets that produce heat through fission. This protective covering, often crafted from materials such as zirconium alloy, envelops the fuel pellets, acting as a shield to keep them in place and prevent the dispersion of radioactive substances.

4.1 For 25 °C and 100 W Heat Generation

In Fig. 7, the output of the nuclear cask is presented under conditions of a 25 °C ambient temperature and an

internal heat production of 100 W. The graph indicates that the highest temperature reached by the cladding is 23 °C.

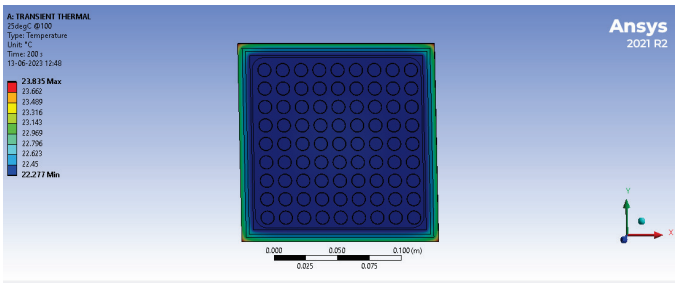


Figure 7 Analysis for 25 °C and 100 W heat generation

4.2 For 25 °C and 800 W Heat Generation

The graph in Fig. 8 displays the performance of the nuclear cask under specific conditions: an ambient temperature of 25°C and an internal heat generation of 800 W. It reveals that the Peak cladding temperature reached 24°C.

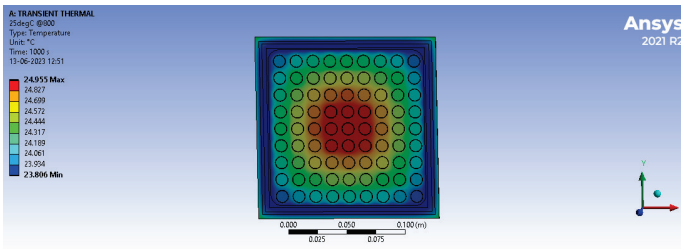


Figure 8 Analysis for 25 °C and 800 W heat generation

4.3 For 25 °C and 2000 W Heat Generation

In Fig. 9, the graph illustrates the output of a nuclear cask when subjected to an ambient temperature of 25 °C and an internal heat production of 2000 W. As indicated by the graph, the highest temperature reached by the cladding is 28 °C.

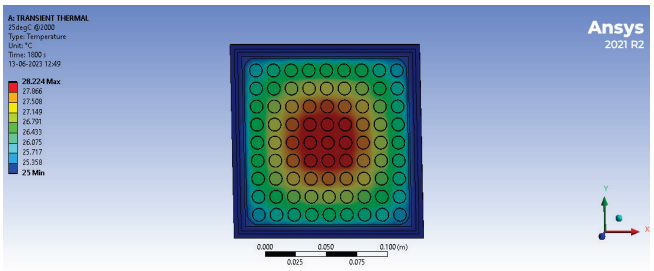


Figure 9 Analysis for 25 °C and 2000 W heat generation

4.4 For 400 °C and 100W Heat Generation

In Fig. 10, the nuclear cask's performance is depicted, indicating an internal heat generation rate of 100 W and an ambient temperature of 400 °C. The graph exhibits a maximum cladding temperature of 253 °C.

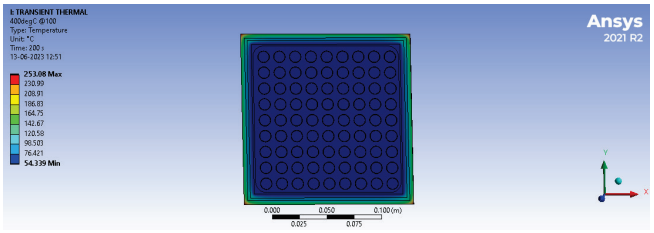


Figure 10 Analysis for 400 °C and 100 W heat generation

4.5 For 400 °C and 800 W Heat Generation

In Fig. 11, the graph illustrates the output of the nuclear cask when subjected to a 400 °C ambient temperature and an internal heat production of 800 watts. Specifically, it shows a high cladding temperature of 320 °C as depicted on the graph.

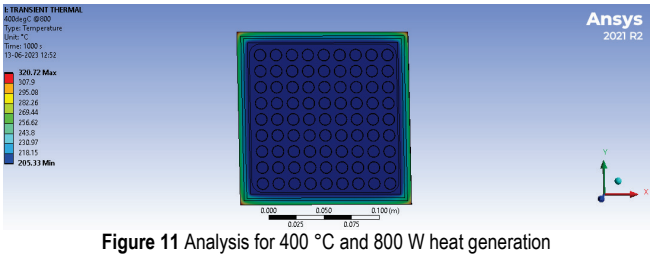


Figure 11 Analysis for 400 °C and 800 W heat generation

4.6 For 400 °C and 2000 W Heat Generation

In Fig. 12, the output of the nuclear cask is depicted under specific conditions: an internal heat generation of 2000 W and an ambient temperature of 400 °C. The graph indicates that the cladding temperature reaches a high of 356 °C.

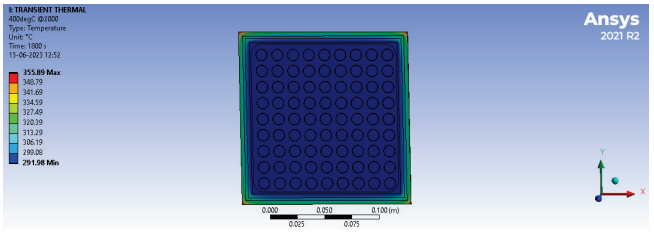


Figure 12 Analysis for 400 °C and 2000 W heat generation

4.7 For 800 °C and 100 W Heat Generation

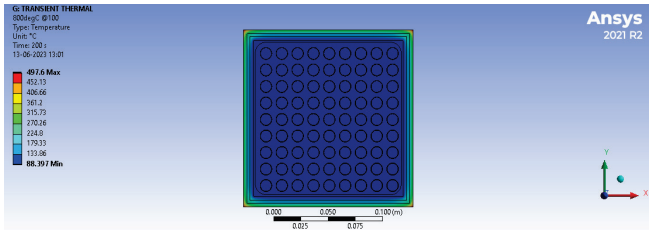


Figure 13 Analysis for 800 °C and 100 W heat generation

In Fig. 13, the graph illustrates the output of the nuclear cask under specific conditions: an ambient temperature of 800 °C and an internal heat production of 100 W. The graph

highlights a notable high cladding temperature, reaching 497 °C.

4.8 For 800 °C and 800 W Heat Generation

In Fig. 14, the data illustrates the output of a nuclear cask under specific conditions. With an internal heat generation of 800 W and an ambient temperature of 800 °C, the graph indicates a peak cladding temperature of 636 °C.

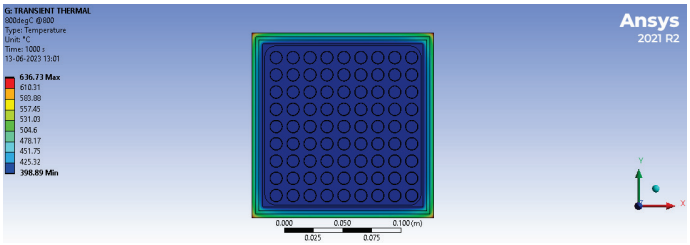


Figure 14 Analysis for 800 °C and 800 W heat generation

4.9 For 800 °C and 2000 W Heat Generation

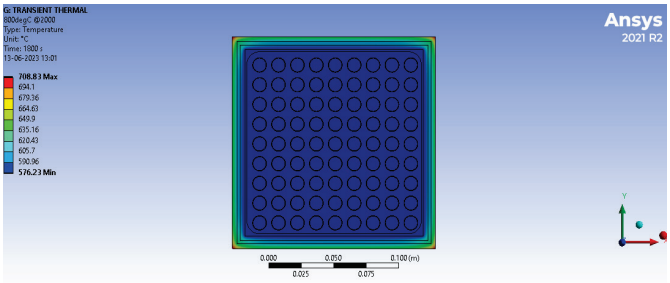


Figure 15 Analysis for 800 °C and 2000 W heat generation

Fig. 15 illustrates the performance of the nuclear barrel, showing an internal heat generation of 2000 W and an ambient temperature of 800 °C. The graph displays a maximum cladding temperature of 708 °C.

The Peak Cladding Temperature (PCT) for a BWR (9 × 9) spent fuel assembly in each environment are shown in Tab. 4 based on the ETC correlations. Again, the vacuum environment produces the greatest PCTs, whereas the helium atmosphere produces the lowest PCTs.

Table 4 Calculated PCTs for a BWR spent fuel assembly

Temperature (°C)/Internal heat (W)	100	200	400	600	800	1000	1250	1500	2000
25	23	24	24	24	24	25	26	27	28
100	69	74	78	81	83	85	87	89	91
200	130	141	150	157	163	168	172	176	179
300	192	210	222	233	242	250	257	263	268
400	253	277	294	308	320	331	340	349	356
500	314	343	365	384	400	413	425	435	444
600	375	411	438	460	478	495	509	522	532
700	436	478	509	535	557	576	593	608	620
800	497	545	581	611	636	658	677	694	708

5 COMPARATIVE ANALYSIS

The suggested study has been evaluated against prior research, as displayed in the Tab. 5. We can conclude from this work that PCM is a suitable material for use in nuclear transportation casks because the table clearly shows that our proposed cask with PCM has higher thermal shielding than other materials like zirconium, which has higher thermal resistance and corrosion resistance but was a bit unstable for higher and longer use.

Table 5 Comparative analysis

Research work	Temperature (°C)/Internal heat (W)	200	400	800	1000	1250	1500
Hyungjin Kim et al. [19]	300	315.7	330.6	344.8	358.5	371.5	384.1
Jie Li et al. [20]	300	303	305	310	313	319	325
Bahney et al. [21]	300	303	306	309	313	316	319
Proposed work	300	210	222	242	250	257	263

6 CONCLUSION

In a 2D ANSYS simulation based on the FLUENT code, the suggested framework was used to estimate the peak cladding temperature along with the efficient thermal conductivity for up to 9 × 9 BWR spent fuel assemblies under helium backfill gas. They were evaluated using a range of assembly heat loads with basket wall temperatures. The ANSYS simulation's peak cladding temperatures were compared to the traditional technique's cladding temperatures.

The peak cladding temperatures as well as the efficient thermal conductivity of spent fuel assemblies utilized in Koodankulam NPPs are being investigated utilizing ANSY Thermal modeling. They were calculated using the 9 × 9 PWR spent fuel assemblies. For this simulation, the assembly heat load ranged from 100 to 2000 W. The simulation findings for peak cladding temperatures are identical, but the temperature drop variance among the three arrays happens very minimally at all basket wall temperatures. The efficient thermal conductivity computed from the 9 × 9 BWR spent fuel assembly data is more conservative than typical shielding found in the literature at a normal operating temperature for spent fuel transportation/storage casks. In the future, we would like to concentrate on the development of improved thermal management technologies, such as passive or active cooling approaches, to effectively disperse heat created during travel.

Acknowledgements

The authors would like to thank the Deanship of Universiti Teknologi Malaysia for supporting this work.

7 REFERENCES

[1] Kurniawan, T. A., Othman, M. H. D., Singh, D., Avtar, R., Hwang, G. H., Setiadi, T. & Lo, W. H. (2022). Technological solutions for long-term storage of partially used nuclear waste: A critical review. *Annals of Nuclear Energy, 166*, 108736.

https://doi.org/10.1016/j.anucene.2021.108736
[2] Alyokhina, S. (2018). Thermal analysis of certain accident conditions of dry spent nuclear fuel storage. *Nuclear Engineering and Technology, 50*(5), 717-723. https://doi.org/10.1016/j.net.2018.03.002
[3] Kim, T., Kim, K., Lee, D., Na, T., Chung, S. & Kim, Y. (2022). Conceptual Design, Development, and Preliminary Safety Evaluation of a PWR Dry Storage Module for Spent Nuclear Fuel. *Applied Sciences, 12*(9), 4587. https://doi.org/10.3390/app12094587
[4] Burns, J. R., Hernandez, R., Terrani, K. A., Nelson, A. T. & Brown, N. R. (2020). Reactor and fuel cycle performance of light water reactor fuel with 235U enrichments above 5%. *Annals of Nuclear Energy, 142*, 107423. https://doi.org/10.1016/j.anucene.2020.107423
[5] Shama, A., Rochman, D., Caruso, S. & Pautz, A. (2022). Validation of spent nuclear fuel decay heat calculations using Polaris, ORIGEN and CASMO5. *Annals of Nuclear Energy, 165*, 108758. https://doi.org/10.1016/j.anucene.2021.108758
[6] Lee, D. & Diaconeasa, M. A. (2022). Preliminary Siting, Operations, and Transportation Considerations for Licensing Fission Batteries in the United States. *Eng., 3*(3), 373-386.
[7] Jeong, G., Kim, S. & Lee, S. (2022). Development of Model to Evaluate Thermal Fluid Flow around a Submerged Transportation Cask of Spent Nuclear Fuel in the Deep Sea. *Journal of Nuclear Fuel Cycle and Waste Technology*, Corpus ID: 256163523. https://doi.org/ https://doi.org/10.7733/jnfcwt.2022.043
[8] Rodríguez-Penalonga, L. & Moratilla Soria, B. Y. (2017). A review of the nuclear fuel cycle strategies and the spent nuclear fuel management technologies. *Energies, 10*(8), 1235. https://doi.org/10.3390/en10081235
[9] Abdelgaied, M., Attia, M. E. H., Kabeel, A. E. & Zayed, M. E. (2022). Improving the thermo-economic performance of hemispherical solar distiller using copper oxide nanofluids and phase change materials: Experimental and theoretical investigation. *Solar Energy Materials and Solar Cells, 238*, 111596. https://doi.org/10.1016/j.solmat.2022.111596
[10] Yousefi, E., Nejad, A. A., & Rezania, A. (2022). Higher power output in thermoelectric generator integrated with phase change material and metal foams under transient boundary condition. *Energy, 256*, 124644. https://doi.org/10.1016/j.energy.2022.124644
[11] Zhang, J., Cao, Z., Huang, S., Huang, X., Liang, K., Yang, Y. & Wen, C. (2022). Improving the melting performance of phase change materials using novel fins and nanoparticles in tubular energy storage systems. *Applied Energy, 322*, 119416. https://doi.org/10.1016/j.apenergy.2022.119416
[12] Li, Y., Huang, X., Wang, J., Lv, F., Jiang, S. & Wang, G. (2022). Enzymolysis-treated wood-derived hierarchical porous carbon for fluorescence-functionalized phase change materials. *Composites Part B: Engineering, 234*, 109735. https://doi.org/10.1016/j.compositesb.2022.109735
[13] Csontos, A. A., Waldrop, K., Durbin, S., Hanson, B. D., Broussard, J. E. & Lenci, G. (2018). Spent Fuel Dry Storage Cask Thermal Modelling Round Robin. https://api.semanticscholar.org/CorpusID:211483369.
[14] Zinet, M., Ghazal, R., Issard, H. & Bardon, O. (2019). Spent fuel transportation cask under accidental fire conditions: Numerical analysis of gas transport phenomena affecting heat transfer in shielding materials. *Progress in Nuclear Energy, 117*, 103045. https://doi.org/10.1016/j.pnucene.2019.103045
[15] Ševečeka, M., Valacha, M. & Leeb, C. H. (2020). Thermal analysis of dry storage and transportation casks castor using COBRA-SFS. *Acta Polytechnica CTU Proceedings, 28*, 32-41. https://doi.org/10.14311/APP.2020.28.0032
[16] Schwartz, M. O. (2021). Corrosion-Enhancing and Corrosion-Reducing Accessories in Bentonite Surrounding Copper-Shielded Containers for Nuclear Waste. *Journal of Hazardous, Toxic, and Radioactive Waste, 25*(4), 04021024. https://doi.org/10.1061/(ASCE)HZ.2153-5515.0000628
[17] Lim, D. W., Lee, C. W., Lim, J. Y. & Hartanto, D. (2019). On the Particle Swarm Optimization of cask shielding design for a prototype Sodium-cooled Fast Reactor. *Nuclear Engineering and Technology, 51*(1), 284-292. https://doi.org/10.1016/j.net.2018.09.007
[18] Alsmadi, Z. Y. & Bourham, M. A. (2022). An assessment of protective coating dry cask canisters with structurally amorphous metals (SAMs) for enhanced radiation shielding. *Nuclear Engineering and Design, 388*, 111647. https://doi.org/10.1016/j.nucengdes.2022.111647
[19] Kim, H., Kwon, O. J., Kang, G. U. & Lee, D. G. (2014). Comparisons of prediction methods for peak cladding temperature and effective thermal conductivity in spent fuel assemblies of transportation/storage casks. *Annals of Nuclear Energy, 71*, 427-435. https://doi.org/10.1016/j.anucene.2014.04.004
[20] Li, J., Murakami, H., Liu, Y., Gomez, P. E. A., Gudipati, M. & Greiner, M. (2007). Peak cladding temperature in a spent fuel storage or transportation cask. *Proceedings of the 15th International Symposium on the Packaging and Transportation of Radioactive Materials*, 21-26.
[21] Bahney, R. H. & Lotz, T. L. (1996). *Spent nuclear fuel effective thermal conductivity report*. Prepared for the US DOE, Yucca Mountain Site Characterization Project Office by TRW Environmental Safety Systems, Inc.

Authors' contacts:

Girish Venkatesh Gudi, Research Scholar
(Corresponding author)
Department of Mechanical Engineering,
Guru Nanak Dev Engineering College,
Mailur Rd, Mailoor, Karnataka, Bidar 585403, India
E-mail: scholar.girishvenkateshgudi@gmail.com
girishgudi@gmail.com

R. G. Nagraj, Dr.
Guru Nanak Dev Engineering College,
Mailur Rd, Mailoor, Karnataka, Bidar 585403, India
dr.abhijitd@dpcoepune.edu.in

Abhijit Dandavate, Dr. Research Scholar
Department of Automobile Engineering,
Dhole Patil College of Engineering,
1284, Near Kharadi EON IT Prak,
Dhole Patil College Road, Wagholi, Pune 412207, India
nagugr@gndecb.ac.in

ISSN 1846-6168 (Print), ISSN 1848-5588 (Online)
https://doi.org/10.31803/tg-20231220092909

Preliminary communication
Received: 2023-12-20, Accepted: 2024-01-23

Safety Work as a Factor in Reducing the Number of Injuries of Tractor Operators

Tomislav Jurić, Mislav Jurić, Ivan Vidaković, Željko Barač*

Abstract: Modern agricultural production requires the use of agricultural machinery. Among all agricultural machines, the tractor is one of the most frequently used agricultural machines. In this paper, we perform a literature review and explore various safety hazards a tractor operator is exposed to. Next, we present the results of a work safety related survey we conducted in Osijek-Baranja and Vukovar-Srijem counties. We conclude the paper with a list of safety related observations we derived from our research.

Keywords: ergonomics; health; operators; safe; tractor

1 INTRODUCTION

Modern agricultural production must ensure the production of quality, quantitatively sufficient and market competitive food. For the most part, manufacturing processes in agricultural production are mechanized, which implies the use of sophisticated machines. By increasing the level of mechanization of a certain process, we reduce the percentage of human labor within them, which contributes to lower injury occurrences in workers who take part in the process.

Agricultural production is specific in that:

- production takes place in the open,
- technological processes are conducted mostly by driving an agricultural aggregate over the production area,
- doing certain technical operations is time constrained due to optimal agrotechnical deadlines,
- agricultural machinery is applied on multitude of terrains (plain, gentle slopes, big slopes),
- the machines used have seasonal character (except tractors), so they remain out of use for a while,
- work is usually done on surfaces which are significantly dislocated from urban environments, and frequently, to reach the production surfaces, the worker must traverse roads that were created by the movement of agricultural machines and are not suitable for the movement of most road vehicles (weighers)
- most frequently the work is being done by only one person (agricultural tractor operator),
- the work is conducted in spite of unfavourable weather conditions (fog, snow) when the visibility is lowered,
- in production they use protective agents and mineral fertilizers that are aggressive and harmful to human health and,
- children participate in agricultural work.

On the production surface usually, there is only the tractor operator who operates the aggregate (common in family farms). Therefore, the problem of the tractor operator injuring themselves is complicated by the fact that the operator has no other choice, but to give themselves first aid, call for medical assistance and wait for their arrival. Furthermore, very often the tractor is being operated by persons who are not acquainted with potential dangers that could occur to the operator (they were not trained to operate the tractor safely).

Agricultural tractor, as a fundamental agricultural machine, is used in nearly all branches of agricultural production (farming, animal husbandry, fruit growing and viticulture, and vegetable growing). It is intended for towing, powering attached machines, carrying attached tools and devices, and for transport. Modern agricultural tractors, technologically perfected, in addition to an educated operator, require well-organized servicing and preventive maintenance, which, if conducted regularly, will avoid the occurrence of breakdowns, ensure high operational reliability of the tractor, extend the service life and reduce the possibility of injury to the operator during operation. Employee injuries while performing certain production processes certainly have economic, health, social and organizational consequences.

The aim of this paper is to provide an overview of the dangers and harms to which agricultural tractor operators and persons working in agriculture are exposed, and to present the results of research into the safe operation of agricultural tractor operators.

1.1 Literature Review

During work, the operator of the agricultural tractor is exposed to the negative influence of numerous factors, which results in a decrease in the performance and quality of the work performed, the appearance of fatigue of the operator, which, along with a decrease in concentration, very often ends in their injury at the workplace.

In parts of the production process where we still have a significant share of human labor (auxiliary workers who fill the containers of mineral fertilizer spreaders with fertilizer, that is, the containers of seed drills with seeds, etc.), there is a risk of injury to participants in the production process during work.

Very often, injuries are the result of the carelessness of the operator (who does not notice another person in the working area around the tractor and thus the tractor runs into people or the tractor passes over certain extremities) or the person who performs the work of an auxiliary worker.

One of the specifics of agricultural production (on family farms) is the participation of children in production processes and their injury during work.

Investigating children's injuries while doing agricultural work in Požega-Slavonia County, authors [1] are found that as many as 30.6 % of respondents (out of 188 children) suffered minor or serious injuries. A special problem regarding the safety of children and adult members of agricultural households is manifested in the fact that monitoring measures (application of regulations, measures and protection rules) of the existing laws cannot be implemented over them, since they are not employed, and in order to reduce the number of injuries to children, it would be beneficial to educate the parents about working in a safe way, note the same authors.

Injuries of persons working in agriculture can occur:

- While coming to work,
- While working with a tractor and other self-propelled machines,
- During work on the supply of agricultural machinery (auxiliary workers),
- During troubleshooting of tractors, self-propelled and attachment machines,
- While participating in road traffic,
- While going to the production area and returning from it,
- During work with protective equipment,
- Due to fatigue,
- Due to recklessness,
- Due to inexperience,
- Due to alcohol and intoxicants,
- While returning from work,
- Inappropriate work speed,
- Bad weather conditions during work,
- Difficult working conditions (working on sloping ground, movement on soil with high humidity, etc.) and
- Lack of expertise (the tractor is driven by people who don't have a driver’s license).

Authors [2] are citing similar causes of injuries, indicate the need for continuous education of tractor operators through courses, regarding the proper use and maintenance of machines. Courses should be organized with the participation of all relevant institutions and associations.

Authors [3] are determine the factors that increase the risk of accidents with a tragic outcome in agriculture (tractor without a safety cabin, older age, people working alone, road, slope of roads and paths, canals and obstacles, tractor manoeuvres, traffic, single-axle tractors, trailer, repairs, too much load, falls, excessive amount of alcohol in the blood). The most dominant factors are the lack of a safety cabin in the event of a rollover, driving on roads and paths, and older age.

Operators and people working in agriculture are mainly exposed to mechanical hazards while working with tractors and other agricultural machines by authors [4]. The high incidence of injuries while working with machines is influenced by their age and correctness, the operator's work experience, fatigue and possible alcohol consumption during work. Therefore, it is necessary to educate all household members regarding the dangers they are exposed to during work, the consequences of their actions, as well as the correct use of machines and pesticides, note the same authors.

The largest number of injuries that occur during work in agricultural households are the result of mechanical hazards when working with agricultural machines and tools, and the main factors for the occurrence of these are the technical malfunction of the machines and the ability and condition of the operator (fatigue, alcoholism, etc.) by author [5].

Investigating injuries in agriculture in the United Kingdom (UK), author [6] are notes that injuries with a tragic outcome most often occur when working with machines, due to falls from a height and due to electric shock. The highest number of accidents during the year occur in the months of July, August and September, and the number of accidents in agriculture is significantly higher than in most other industries, notes the same author.

Analyzing a multi-year period (2004-2013) in which 2.892 traffic accidents involving tractors occurred, by authors [7] are state that the biggest cause of accidents with a tragic outcome was a tractor without a safety cab. Most of the accidents happened on intercity roads. The age of the operator over 45 is a factor that increases the probability of an accident, the authors warn.

Authors [8] are analyzing tractor accidents in agriculture and forestry in the Czech Republic with serious injuries and fatal outcomes, state that in the interval from 2009 to 2018, 89 accidents occurred, of which 72 accidents were with severe (serious injuries) and 17 accidents with tragic consequences. In all accidents, the tractors were driven by men, and the most common cause of accidents was non-compliance with the rules.

Authors [9] are analyzed agricultural accidents in the Republic of Austria (2008-2010) caused by tractors, self-propelled harvesting machines and transport machines and found that 1/3 of the accidents occurred due to overturning of the vehicle, and more than 10 % of the accidents that occurred is due to technical problems, slippery or muddy surfaces or poor vehicle handling.

Investigating the consequences of accidents that occurred while working with a tractor, authors [10] in the period from 1999 to 2003 in the Republic of Macedonia, found that out of a total of 101 tragically injured persons, 56.44 % were killed in traffic, and 43.56 % during work. The older age of the operator results in reduced sensory and psychomotor abilities, which leads to a higher number of injuries with a tragic outcome, according to the same authors.

Regular maintenance measures ensure the technical correctness of the tractor and the machine that is aggregated with the tractor, and since they are performed before the start of work, they should reduce the possibility of injury. Furthermore, these measures ensure the technical correctness of the tractor as a participant in road traffic. The number of traffic accidents involving tractors, the number of fatalities and injured persons are presented in Tab. 1.

The number of traffic accidents involving tractors has a decreasing trend from 360 accidents (2015) to 334 accidents (2019). The number of people killed compared to 2015 (5 people) managed to decrease to 3 people killed in 2017, but,

unfortunately, during 2018 and 2019, it increased to 9 people killed. The number of severely injured and slightly injured persons increased in 2018 and 2019 compared to 2015.

Table 1 Number of traffic accidents involving tractors, number of fatalities, seriously injured and slightly injured persons for the period 2015-2019 [11]

Year	Number of traffic accidents involving tractors	Number of people killed in an accident	Number of seriously injured persons	Number of persons who suffered minor injuries
2019	334	9	28	56
2018	337	9	31	51
2017	334	3	21	48
2016	362	5	29	51
2015	360	5	18	40

Investigating the implementation of maintenance measures on family farms, Authors [12] are indicate that they are not performed well, which will certainly increase the possibility of tractor breakdowns, as well as traffic accidents, since tractors participate in road traffic.

In order to determine the agricultural tractor as a traffic safety factor, authors [13] are conducted research by surveying 107 family farms, where they found a high average age of tractors (17.7 years) and a high percentage of unregistered tractors (52.79 %) on family farms. The fact that even 23.33 % of technically defective tractors participated in public traffic is also worrying, according to the same authors.

By fulfilling the ergonomic requirements, to the greatest extent, the modern constructions of agricultural tractors enable comfortable, health-safe work, reducing the possibility of injury of the operator to a minimum. Unobstructed entry and exit to the workplace (minimizing the possibility of injury) is ensured by the installation of handrails and steps and ladders on tractors and self-propelled agricultural machines. In order to reduce the possibility of injury, the steps must have a high coefficient of friction with the sole of the shoe, the distance between the steps should be uniform, and the first step should be easily visible (even at night) to avoid the possibility of missing it when leaving the cabin [14].

Authors [15] are defined the minimum size of the workspace (Fig. 1) of the operator of the agricultural tractor, which must provide him with unhindered movements of his hands and feet, and at the same time be spacious enough so that the tractor driver is not injured.

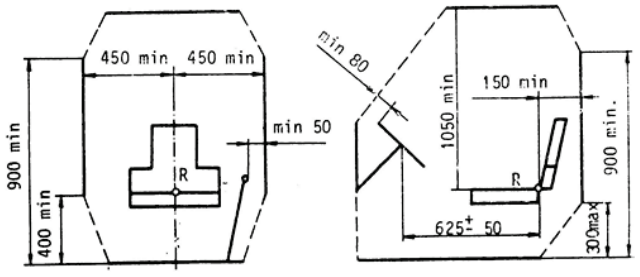


Figure 1 Minimum values of the size of the workspace

Mechanical vibrations at the operator's workplace are the result of the tractor's movement, the operation of individual tractor assemblies (engine, transmission), the operation of the attached machine, etc. The negative impact of mechanical vibrations on the operator is manifested in their traumatic injuries at the place of their impact (direct impact) and indirect impact on nerves and nerve endings, which often results in disorders of the circulatory system, bones, nervous system, bones, joints, muscles, stomach, balance and others by authors [16].

The problem of mechanical vibrations is being solved by active and passive protection measures by author [17]. Active protection measures include: suspension of the front axle of the tractor, suspension of the entire tractor, an increase in the weight of the tractor and newer tire designs. In contrast to active measures, passive protection measures include the suspension of the tractor seat and the tractor cabin.

The limit and warning values of daily exposure during the operator's eight-hour working time to the hand-fist system are 5 ms^{-2} and 2.5 ms^{-2}, respectively, and to the operator's trunk 1.15 ms^{-2} and 0.5 ms^{-2}, [18, 19]. Investigating the level of vibrations of the hull that affects the operator when moving the tractor on different agrotechnical surfaces, seat materials and movement speeds, it was determined that not a single measured value exceeded the permissible limit value of 1.15 ms^{-2} by authors [20]. The level of vibrations that affect the hand-arm system of the operator, when moving the tractor on different agrotechnical surfaces and at different speeds of movement, increases with the greater number of working hours of the tractor by authors [21]. The same authors indicate that all measured values did not exceed the permissible limit value of 5 ms^{-2}.

Noise is any unwanted sound that causes operator fatigue, reduced concentration, orientation problems, reduced ability to receive sound notifications and safety at work by authors [22]. Noise at the operator's workplace is a consequence of the operation of the engine, the operation of the transmission elements, the operation of the air intake system and the exhaust system, the operation of the hydraulic system, and vibrations due to the movement of the tractor on the production surfaces and vibrations of individual parts of the tractor [17]. The noise level at the tractor operator's workplace should not exceed 85 dB, otherwise the operator's hearing may be damaged.

Tractor cabs should be lined with material that absorbs sound well, and if the noise level is still above 85 dB at the operator's place, the operator should use personal protection against noise (ear plugs, ear protectors, etc.). Noise generation is certainly affected by the age of the tractor and the quality of its maintenance.

By measuring the noise inside and outside the operator's cabin at different engine revolutions (1100, 1800 and 2200 rpm^{-1}) and when standing and moving in relation to the number of working hours, it was determined that with the increase in the number of working hours' noise increases, except for external noise at 2200 min^{-1} when standing and moving. None of the measured values exceeded 80 dB, which is favourable in terms of operator protection and safety by authors [23].

Agricultural machines work in an environment where there is a significant amount of dust. Dust in agriculture is

generated during tillage and harvesting by author [24]. In order to prevent dust from entering the tractor cabin (operator's workspace), filters are installed in it, through which the surrounding air passes before entering. If this protection is not sufficient, the operator must use a protective mask or a protective visor.

During work, the operator must be provided with favourable microclimatic conditions. Ensuring these conditions is complicated by the limited volume of the tractor cabin, large glass surfaces (windows and doors), working under a clear sky (exposure to solar radiation), heat radiation from the engine, etc.

If the operator is exposed to high temperatures at the workplace, this may result in heat stroke, and low temperatures may result in the freezing of certain parts of the body. In order for the operator to feel comfortable at the workplace, the air temperature must be 18-20 °C in the winter or 24-27 °C in the summer, the air flow speed 0.1 - 0.5 ms^{-1}, and the relative humidity 30 - 65 %, by author [25].

Modern constructions of tractors are equipped with cooling devices, cabins are lined with thermal insulation materials, which, along with the use of tinted and bevelled glass, as well as blinds and shades, provides the operator with favourable working conditions due to the microclimate.

When applying pesticides and other protective agents and mineral fertilizers, the operator may come into contact with dangerous substances. In addition to harmful substances, there is a risk of exhaust gasses entering the cabin, which contains carbon monoxide among other dangerous and harmful gasses.

During the application of protective agents (spraying, spraying), aerosol droplets are formed, which must never reach the operator's workplace. Therefore, the surrounding air will pass through a filter before entering the cabin. To prevent the entry of dangerous particles into the cabin through the openings through which the control levers, hydraulic hoses, etc. pass, seals must be correct, and there must be a certain overpressure in the cabin. Unfortunately, in practice, operators sometimes know how to perform treatments with the cabin door open and without applying personal protective equipment, which leads to damage to their health. Regular and high-quality measures of service and preventive maintenance are important factors for ensuring favourable working conditions regarding the protection of operators from dangerous substances by authors [15].

During operation, the operator must monitor the measuring and control instruments on the control panel in the tractor, the space in front, on the side and behind the tractor, and control the operation of the attachment device (with which the tractor is aggregated). On modern tractors, the glazed surfaces of the tractor cabin have been increased to the maximum in order to reduce the possibility of injuries during operation and to increase the visibility of the space in front, on the side and behind the tractor. The tractor is equipped with internal and external mirrors, lights and a device for washing and defrosting the windshield in order to work safely in the day and in the night. These structural solutions have reduced the invisible surfaces from the operator's workplace. Further development of agricultural tractor structures with the aim of improving visibility is focused on the application of cameras and sensors for the part of the space that is outside the operator's field of vision (blind spot). These systems detect a person who is in the tractor's trajectory, stop the tractor and thus prevent accidents by authors [26].

During operation, the tractor aggregates (tractor units) move on production surfaces on different reliefs (plains, smaller slopes and larger slopes). The best results of the operation of the tractor unit are achieved during operation on flat terrain, where we have good stability of the unit, less load on the tractor operator, lower fuel consumption, good efficiency of the unit, etc. Working on gentle inclines results in a decrease in the stability of the unit, an increased load on the operator, higher fuel consumption, a decrease in efficiency, etc. The worst results are achieved during the operation of the unit on higher slopes.

During rest and movement of tractor aggregates, the reaction values of the ground on the front and rear axles of the tractor also change, and their values are certainly affected by the ground on which the aggregate moves (flat surface, rise or slope), attachment tool or machine aggregated with the tractor. Although the tractor manufacturers have defined in their instructions the permitted slopes and climbs on which the tractors may move during operation, very often the tractor overturns and injures the operator. The most common reason for this is the neglect of this data due to the self-confidence of the operator ("I can do it, nothing will happen to me"), inexperience, the work of a minor or a person who occasionally drives the tractor, etc.

Due to their construction, tractors have a high center of gravity, which makes them unstable, and tractor overturning is the main factor in accidents by author [27].

In the Republic of Croatia, a safety cabin or safety frame is mandatory on all tractors registered for the first time after January 1, 1983 [28].

Very often, injuries during tractor operation occur:

- Due to the side overturning of the tractor or
- Longitudinal overturning of the tractor (ascent or slope).

The stability of the tractor is influenced by the following factors [15]:

- Linear tractor scale
- Magnitude, direction and direction of the forces acting on the tractor
- The slope of the ground and the direction of movement of the tractor (uphill or downhill).

Lateral overturning of a tractor moving on a slope, Fig. 2, will occur at the moment when the normal reaction of the ground at the point of the wheel that is in a higher position (the right wheel during the slope of the ground to the left side) is equal to zero (N1 = 0). Therefore, the beginning of the tractor overturning with regard to point 2 is at the moment when:

$$Gh_c \sin\alpha = 0{,}5G\,(B+b)\cos\alpha \quad (1)$$

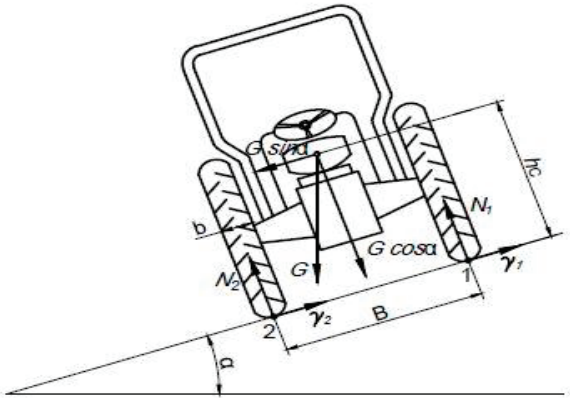


Figure 2 Movement of the tractor on a side-sloping surface

Where: G – tractor weight, N; B – wheel spacing, m; b – pneumatics width, m; h_c – the height of the tractor's center of gravity, m; α – angle of inclination of the surface, °.

The tractor's side tipping can sometimes be preceded by the tractor's side drift. The values that influence the increase or decrease of the risk of lateral overturning are: the height of the tractor's center of gravity (h_c) and the sum (b+b) of the distance between the wheel spacing (B) and the wheel width (b). Smaller values of the height of the center of gravity ensure better stability of the tractor, as a larger sum (B+b).

When analyzing the longitudinal overturning of the tractor during operation, we distinguish between the overturning of the tractor around the axis of the rear wheels (during work on a climb), Fig. 3, and the overturning of the tractor around the axis of the front wheels (during work on a slope). The condition for overturning the tractor in the first case is the complete unloading of the front wheels (N_p =0), that is, for the overturning of the tractor on the slope, the unloading of the rear wheels (N_z = 0).

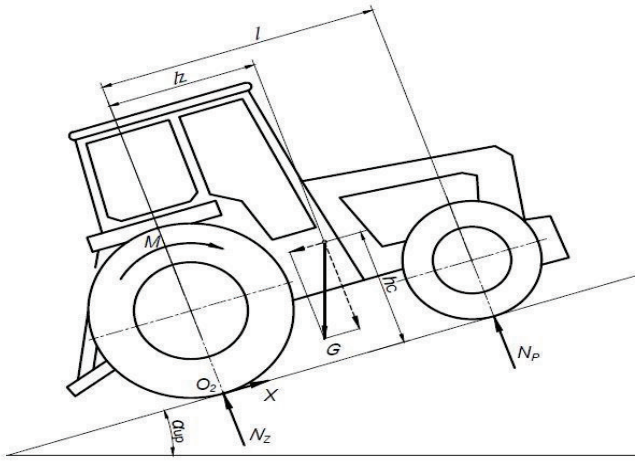


Figure 3 Tractor on a slope

Where: G – tractor weight, N; h_c – the height of the tractor's center of gravity, m; α_{up} – angle of inclination of the surface, °; l – distance between axles; l_z – distance of the center of gravity from the axis of the rear wheel.

The stability of the tractor during standstill depends on the angle of inclination of the ascent or slope. Considering the position of the tractor's center of gravity, the limit angle when the tractor is at rest on the slope (α_p) is greater than the limit angle of ascent (αup). For tractors with wheels, αup = 35-400, and α_p = 600 (it can be higher), according to the authors [15, 29]. Along with overturning, the longitudinal stability can also be hindered by the tractor sliding on the ground.

The loss of longitudinal stability can be the result of the tractor's drive wheels getting stuck, where they do not turn, which will result in the front part of the tractor being lifted off the ground and the tractor rotating around the rear (drive wheels).

Furthermore, the loss of longitudinal stability can occur due to an inadequate coupling height of the trailer and the tractor.

2 MATERIALS AND METHODS

Research regarding the safe work of agricultural tractor operators was carried out 2023 by surveying agricultural tractor operators in Osijek-Baranja and Vukovar-Srijem counties. The questionnaire contained eleven questions (Tab. 2). The results of the research have an indicative character since the answers were obtained by the subjective judgment of the respondents. The research carried out is part of the research that will be carried out in the counties of Eastern Croatia with regard to the safe work of operators.

Table 2 Example of research survey

No	Questions	Answers	
1.	Are you familiar with working procedures that are considered safe?	Yes	No
2.	Do you have technical documentation for all the machines you own?	Yes	No
3.	Do you have a valid first aid kit in the tractor?	Yes	No
4.	Do you follow the instructions for safe operation prescribed by the machine manufacturer?	Yes	No
5.	I perform agrotechnical operations	Mostly alone (yes).	With another person (no).
6.	In the event of an accident during work, are you able to provide professional first aid to the injured person?	Yes	No
7.	During work I wear:	Work clothes and work footwear (yes).	Everyday clothes and everyday footwear (no).
8.	Do you have the correct protection of all PTO shafts?	Yes	No
9.	In the case of removal of plant residues and soil when the attachment machine becomes clogged:	I turn off the drive to the connecting machine (yes).	Turn off the drive machine (no).
10.	Do you have a mobile device with you during work?	Yes	No
11.	When carrying out repairs on carried machines that have been raised by a hydraulic lifting device:	Turn off the tractor and eliminate the fault (yes).	Turn off the tractor and place mats under the machine (no).

The obtained results were processed using the Chi-Square Tests.

3 RESULTS AND DISCUSSION

The obtained results were analyzed and presented in two graphs.

Fig. 4 shows that most of the answers to the questions are positive, except for the answer to question 11 from the survey, where "no" indicates a good answer by looking at the survey. Furthermore, the analysis of the results found that operators in Osijek-Baranja County adhere to the measures necessary for safe work.

The analysis of the results of the Vukovar-Srijem County found that the answers to the questions were mostly positive, except for questions 3, 7 and 8 where the answer was "no" in the range of 55 % to 70 %. Furthermore, by looking at the survey, it was determined that the operators in the mentioned investigated county mostly do not have valid first aid in the tractor, do not wear work clothes and work footwear, and do not have proper protection of all PTO shafts (Fig. 5).

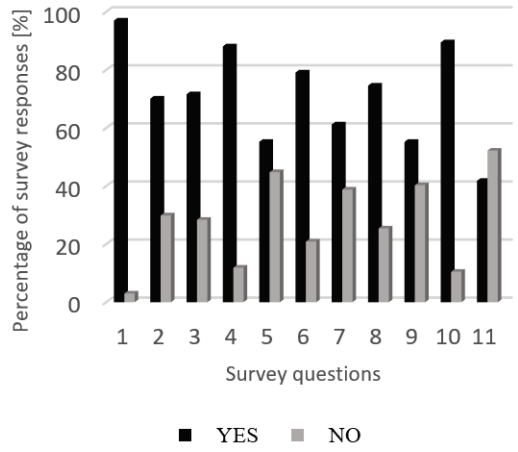


Figure 4 Results of the safe work survey of the investigated Osijek-Baranja County

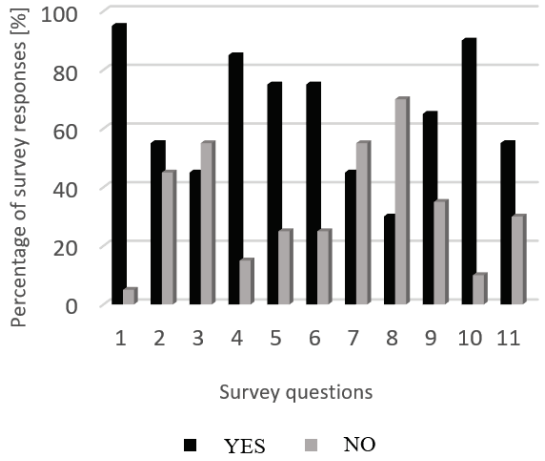


Figure 5 Results of the safe work survey of the investigated Vukovar-Srijem County

By not complying with these measures, they put themselves at risk of minor injuries and serious life-threatening injuries, which is what the authors warn and refer to [1].

Operators should take a course or training on safe work in order to become aware and begin to comply with all measures until it is too late, as stated by the authors [2].

Table 3 County and safety crosstabulation

			Safety		Total
			Positive	Negative	
County	OB	Count	529	208	737
		Expected Count	509	228	737
	VS	Count	132	88	220
		Expected Count	152	68	220
Total		Count	661	296	957
		Expected Count	661	296	957
OB - Osijek-Baranja County; VS - Vukovar-Srijem County					

Tab. 3 shows that in Osijek-Baranja County were 529 positive answers regarding safety, with the corresponding theoretical frequency being 509, while for positive answers in Vukovar-Srijem County, the theoretical value (152) was higher than the actual value (132). Of the total number (957), there were over 69 % positive safety responses.

Table 4 Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	11.001[a]	1	0.001		
Continuity Correction[b]	10.456	1	0.001		
Likelihood Ratio	10.666	1	0.001		
Fisher's Exact Test				0.001	0.001
N of Valid Cases	957				
a. 0 cells (0,0 %) have expected count less than 5. The minimum expected count is 68.05. b. Computed only for a 2×2 table					

Value of the Chi-square test is 11.001 and shows high statistical significance (Tab. 4).

Table 5 Symmetric Measures

		Value	Approx. Sig.
Nominal by Nominal	Phi	0.107	0.001
	Cramer's V	0.107	0.001
	Contingency Coefficient	0.107	0.001
N of Valid Cases		957	

Considering the values of coefficients in Tab. 5, it was concluded that the correlation of the variables is relatively low (0.107), but highly significant. It is possible to observe a trend of a lower number of safety-conscious operators in Vukovar-Srijem County, i.e. more safety-conscious operators in Osijek-Baranja County compared to the expected frequencies.

5 CONCLUSIONS

The specificity of agricultural production is manifested, among other things, by the movement of agricultural aggregates by production bases. Very often, operators of agricultural machinery perform their work alone in locations

far from populated areas. During work, the operator is exposed to numerous factors that negatively affect his safety and health, while the ergonomic solutions of modern agricultural machines strive to ensure the most favourable working conditions.

The conducted research indicates:

- That there is a difference between the investigated counties in terms of not having valid first aid, not wearing work clothes and work footwear, and the absence of PTO shaft protection;
- That not having a valid first aid kit reduces the quality of self-help provided by the operator in the event of an injury;
- That not wearing work clothes and work footwear reduces operator safety during work and increases the possibility of a negative impact on health;
- Due to the lack of protection of the cardan shaft, there is a great risk of minor, serious and life-threatening injuries;
- On the need for continuous education regarding safe work;
- Osijek-Baranja County operators are significantly aware of safety at work. In addition, there is a highly statistically significant relationship between the operators of a Particular County and safety at work.

6 REFERENCES

[1] Janev Holcer, N., Brkić Biloš, I., Jurić, F., Rešić, A. & Vugrinčić, M. (2014). Risk factors and children injuries while performing agricultural work in the Požega-Slavonia County. *Pediatria Croatica*, *3*(58), 190-194. https://doi.org/10.13112/PC.2014.34

[2] Oljača, M. V., Kovačević, D., Radojević, R., Gligorević, K., Pajić, M. & Dimitrovski, Z. (2010): Accidents with tractor drivers in public traffic of the Republic of Serbia. *Poljoprivredna tehnika*, *1*(35), 75-82.

[3] Arana, J., Mangado, J., Arnal, P., Arazuri, S., Alfaro, J. R. & Jaren, C. (2010). Evalution of risck factors in fatal accidents in agriculture. *Spanish Journal of Agricultural Research, 8*(3), 592-598. https://doi.org/10.5424/sjar/2010083-1254

[4] Janev Holcer, N., Deriš, B. & Džakula. A. (2013). Health protection of the farming population in Požeško-slavonska County. *Sigurnost: časopis za sigurnost u radnoj i životnoj okolini, 55*(1), 1-8.

[5] Fabijanić, K. (2010). Occupational safety and health of the farming population. *Sigurnost: časopis za sigurnost u radnoj i životnoj okolini, 52*(4), 367-379.

[6] Solomon, C. (2002). Accidental injuries in agriculture in the UK. *Occupational Medicine, 52*(8), 461-466. https://doi.org/10.1093/occmed/52.8.461

[7] Arnal, P., López-Maestresalas, A., Arazuri, S., Mangado, J. M. & Jarén, C. (2017.): A multi-year analysis of traffic accidents involving agricultural tractors. *Chemical Engineering Transactions, 58*, 109-114. https://doi.org/10.3303/CET1758019

[8] Valeriánova, Z. & Patočka, Z. (2020). Analysis of serious and fatal occupational accidents associated with tractors in agriculture and forestry in the Czech Republic. *Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis, 68*(4), 719-727. https://doi.org/10.11118/actaun202068040719

[9] Mayrhofer, H., Quendler, E. & Boxberger, J. (2014). Narrative text analysis of accident reports with tractors, self-propelled harvesting machinery and materials handling machinery in austrian agriculture from 2008 to 2010 – a comparison. *Annals of Agricultural and Environmental Medicine, 21*(1), 183-188.

[10] Dimitrovski, Z., Gligorević, K., Ružićić, L. & Oljača M. V. (2008). Consequences of tractor accidents involving older farmer population in the agriculture of the repuplic of Macedonia. *Poljoprivredna tehnika, 33*(4), 103-110.

[11] MI (2015-2019). Bulletin on road traffic safety 2015-2019. Retrieved from https://mup.gov.hr/pristup-informacijama-16/statistika-228/statistika-mup-a-i-bilteni-o-sigurnosti-cestovnog-prometa/bilteni-o-sigurnosti-cestovnog-prometa/287330

[12] Jurić, T., Emert, R., Šumanovac, L. & Horvat, D. (2001). The realization of machinery maintenance on family farms. *Proceedings of the 29th International symposium on agricultural engineering "Actual tasks on agricultural engineering"*, Opatija, 43-50.

[13] Jurić, V., Jurić, T., Kiš, D., Emert, R. & Plaščak, I. (2008). Agricultural tractors as factors in traffic safety. *Proceedings of the 36th International symposium on agricultural engineering "Actual tasks on agricultural engineering"*, Opatija, 105-113.

[14] OG 39/2011. Rule book on procedure for the homologation of tractors for agriculture and forestry with regard to the working area, access to the driver's seat and doors and windows tpv 317, *Republic of Croatia*, Zagreb.

[15] Brkić, D., Vujčić, M., Šumanovac, L., Kiš, D., Jurić, T. & Knežević, D. (2005). *Operation of agricultural machinery*. Book of Faculty of Agrobiotechnical Sciences Osijek: Osijek, Croatia.

[16] Aranđelović, M. & Jovanović, J. (2009). *Occupational medicine.* Book of Faculty of Medicine University of Niš, Serbia.

[17] Časniji, F. (1991). *Ergonomic deficiencies of agricultural tractors.* Book of Faculty of technical sciences university of Novi Sad, Serbia.

[18] EC (2002). Council Directive on the minimum health safety requirements regarding the exposure of workers to the risks arising from physical agents (vibration). *J Eur Commun*, EEC 89/391, 2002/44. Retrieved from https://eur-lex.europa.eu/legalcontent/HR/TXT/?uri=CELEX:02002L0044-20081211

[19] OG 155/2008. Rule book on protection from the risk of exposure to vibration at work. Ministry of economy, entrepreneurship and crafts, *Republic of Croatia*, Zagreb.

[20] Barač, Ž., Plaščak, I., Jurić, T., Jurišić, M., Zimmer, D., Vidaković, I., Petrović, D., Duvnjak, V. & Marković, M. (2018). Operator's whole body vibrations dependent of agrotechnical surface, speed of movement and seat upholstery. *Tehnicki glasnik, 12*(2), 68-73. https://doi.org/10.31803/tg-20171207121511

[21] Barač, Ž., Plaščak, I., Jurić, T., Baličević, P., Duvnjak, V., Jurišić, M., Heffer, G. & Marković, M. (2019). Influence of an uneven surface on the vibration occurrence affecting the tractor operator. *Tehnicki glasnik*, *13*(3), 192-196. https://doi.org/10.31803/tg-20190211142404

[22] Kroemer, K. H. E. & Grandjean, E. (2000). *Adjusting of the work to operator.* Book of ergonomics, Zagreb, Croatia.

[23] Barač, Ž., Plaščak, I., Jurišić, M., Tadić, V., Zimmer, D. & Duvnjak, V. (2018). Noise in the cabin of agricultural tractors. *Technical Gazette, 25*(6), 1611-1615. https://doi.org/10.17559/TV-20170223093448

[24] Göhlich, H. (1987). *Lehrbuch der Agrartechnik.* Band 5-Mensch und Maschine, Book, Hamburg und Berlin, Germany.

[25] Renius, K. T. (1987). *Traktoren.* Book, Mönchen, Germany.
[26] Ehlers, S. G. & Field, W. E. (2017). Determing the effectivennes of mirrors and camera system in monitoring the rearwad visibility of self-propelled agricultural machinery. *Journal of Agricultural Safety and Health, 23*(3), 183-201. https://doi.org/10.13031/jash.12034
[27] Jerončić, R. (2008). The main reasons for accidents with the agriculture and forestry tractors. *Poljoprivredna tehnika, 33*(1), 101-110.
[28] OG 51/2010. Rule book on the technical conditions of vehicles in traffic on roads, *Republic of Croatia*, Zagreb.
[29] Petrović, P., Obradović, D., Petrović, Ž. & Petrović, M. (2014). Longitudinal stability of tractor, as an essential para meter security in service. *The 17th Scientific conference current problems and tendencies in agricultural engineering*, Beograd, 133-146.

Authors' contacts:

Tomislav Jurić, PhD, Full Professor
Faculty of Agrobiotechnical Sciences Osijek
Vladimira Preloga 1, 31000 Osijek, Croatia
tjuric@fazos.hr

Mislav Jurić, MSc
Independent researcher
Hrvatskih kraljeva 110, 32100 Vinkovci, Croatia
mislav.juric01@gmail.com

Ivan Vidaković, PhD, Senior assistant
Faculty of Agrobiotechnical Sciences Osijek
Vladimira Preloga 1, 31000 Osijek, Croatia
ividakovic@fazos.hr

Željko Barač, PhD, Assistant Professor
(Corresponding author)
Faculty of Agrobiotechnical Sciences Osijek
Vladimira Preloga 1, 31000 Osijek, Croatia
zbarac@fazos.hr

ISSN 1846-6168 (Print), ISSN 1848-5588 (Online)
https://doi.org/10.31803/tg-20240510171202

Preliminary communication
Received: 2024-05-10, Accepted: 2024-05-20

Carbon Footprint Principles and Challenges in Transport Logistics

Kenn Steger-Jensen, Hans-Henrik Hvolby*, Sven Vestergaard, Mihai Neagoe, Carsten Svensson

Abstract: Logistic operators are under pressure from consumers and governments to reduce the carbon footprint impact. Operators need increased focus on reducing the carbon impact and calculating their actual carbon emissions. EU's Emissions Trading System comes into force from 2027 and works on the 'cap and trade' principle. A cap is a limit set on the total amount of greenhouse gases that logistic operators may emit. Therefore, there is a need to develop a roadmap and action plans to reduce emissions, which will drastically impact the logistic operators. The paper brings forward current standards, frameworks and principles as well as examples of carbon footprint calculations in last-mile transportation.

Keywords: carbon footprint calculation; digitalisation; frameworks; last mile delivery; transport planning

1 INTRODUCTION

Order management and transport planning of last-mile transportation is critical since it is expensive and can account for as much as 75% of the total transportation cost [1]. Freight distribution companies are focusing on improving routing efficiency and reducing the cost of their deliveries. The improvement today also needs to be reductions in externalities such as congestion [2], air pollution [3], and greenhouse gas emissions as well as noise and accidents [4, 5].

The costs of fuel consumption represent over 50% of the total operating costs in last-mile transportation [6], especially when taking load, distance, speed, and time dependencies into consideration [7].

According to a recent literature review on last-mile deliveries [8] there is a growing interest in studying last-mile delivery operations from both academia and industry to identify the underlying challenges and possible solutions. In a recent literature review on last-mile strategies for urban freight delivery [9], 22 last-mile delivery strategies are identified. This is reflected by an increase in the number of publications related to last-mile deliveries [10]. Even though there is a growing body of literature addressing last-mile delivery operations, the authors found a lack of principles and methods for transport planning and calculating CO_2 emissions concerning transportation in last-mile deliveries.

The challenge is that freight distribution companies are faced with new and stronger requirements for documenttation and reporting. These new requirements drive a growing need for innovation and efficiency at freight distribution companies, which has not been seen before. Transportation companies must from 2027 be part of the EU's emissions trading system (ETS). The current EU requirements for the fulfilment and reporting of Greenhouse Gas (GHG), which follows ISO14083 will also be expanded with the new EU requirements (see section 2), which together with the EU's emissions trading system (ETS), will result in freight distribution companies being put in a very difficult situation.

The principle of the ETS is that there is a CO_2 quota allocated to companies in general, which will be continuously reduced to reach the climate targets that have been set.

GHG distinguishes between Scope 1, Scope 2 and Scope, where Scope 3 is the emission of CO_2 (e) from activities carried out by subcontractors. This is where freight distribution companies come into play when transporting goods to or from a company.

Customers who purchase transport services will focus on only "paying" carbon emissions for their share of the transport. This means that freight distribution companies are squeezed on their competitive edge partly because they have to buy CO_2 from the continuously reduced CO_2 quota in the ETS and partly because customers have to document and report on their GHG in scope 3. Transport companies will aim to reduce overall emissions by minimizing idle time and optimizing the utilization of payload as well as consolidating goods on different modes of transport.

The EU is currently working on introducing a digital product passport (DPP), which will help reduce environmental impact and support development towards a circular and sustainable future. The purpose of the DPP is also to help the customer make better decisions when buying, which is why it is expected that the DPP will also contain information about CO_2. DPP not only requires freight distribution companies to provide this data, but the product owner or the financially responsible company (e.g. distributor or retailer) is responsible for this data. A challenge for freight distribution companies is that both their own and the customer's calculation must be reported for the planned DPP.

Overall, there is a need for more data to document and report the calculations used. This includes principles and methods to be able to meet requirements from the EU as well as requirements from customers and for transport companies' own GHG reporting. To our knowledge, no previous studies have investigated this issue, as it is a quite new but challenging planning area, and there is a lack of guidelines to overcome these challenges.

The purpose of this article is to shed light on the challenges of order management and transport planning with a focus on calculating CO_2 emissions regarding transport services. We particularly focus on the differences between pre-calculations, which are often based on offers as well as form the basis for agreements with customers, and the subsequent follow-up on the actual work carried out for the customer, also called post-calculation.

A variation and difference between pre-calculation and post-calculation have always existed, but the question is which of the parties will be exposed to this difference in terms of emissions and what significance it will have concerning the aforementioned documentation and reporting requirements in the future in the area of sustainability. A possible difference between the pre-calculation and post-calculation is expected to become the new and decisive competitive parameter for both the goods owner/customer but also transport companies in the future, which is why planning and optimization of transportation is expected to be an even more important factor in transport companies.

The authors hypothesise that, depending on which planning principles and methods transport in the last-mile is carried out, there will be differences in the amount of environmental pollution. This difference between the offer to the customer and the actual consumption of emissions and thus also the price will be the future challenge that needs to be clarified.

This environmental burden must be expected to be seen by the end customer, customs, public authorities, etc. via the new EU digital product passport. This may result in additional GHG costs for the product and not just the GHG reporting of the customer's business.

We search for answers to the following research questions:

- How will transport planning principles and methods affect the amount of greenhouse gases?
- How will this difference be handled by the supplier of transport services? and whom of the partners (customer vs. supplier) should bear this difference?
- How will competitiveness be affected by current and future reporting and documentation of emissions?

The paper is organized as follows: Section 2 covers the main emission frameworks (Greenhouse Gass Protocol, Global Logistics Emission Council Framework) and ISO's 14083 standard as well as EU's Digital Product Passports and EU's Emissions Trading System. Section 3 covers the elements of Transport Planning and Carbon Footprint and section 4 covers four different principles and methods of calculating the transport work and emission of 3 deliveries. Section 5 is a discussion and finally, section 6 is a conclusion.

2 FRAMEWORKS SUPPORTING EMISSION REDUCTION

In this section, we discuss frameworks supporting emission-reducing frameworks such as the Greenhouse Gass Protocol, Global Logistics Emission Council Framework and ISO's 14083 standard as well as EU's Digital Product Passports and EU's Emissions Trading System

2.1 Emission Frameworks and Standards, EU's Digital Product Passports and EU's Emissions Trading System.

The Greenhouse Gas Protocol (GHG) [11] is the global standard for corporate climate reporting. It distinguishes between scope 1, scope 2 and scope 3. Scope 3 is the emission of CO_2 (e) from activities carried out by subcontractors. For example, when transport of goods to or from the company is carried out by subcontractors for the company.

When it comes to the calculation and declaration of CO_2 (e) emissions for transport, a European standard, EN16258, was adopted in 2011. This standard has today been upgraded to an ISO standard, ISO14083 [12].

An industry initiative, the Global Logistics Emission Council Framework, also called the GLEC framework [13] (GLEC Version 3), is based on the principles described in ISO14083. In some areas, the GLEC framework differs from ISO14083. For example, GLEC only concerns CO_2 (e), i.e. reporting of energy consumption is not included.

The GLEC also contains minimum requirements for the content of the company's GHG reporting to external stakeholders (CSR reporting and ESG data). Here, it is important that a new EU requirement, the Corporate Sustainability Reporting Directive (CSRD), has been introduced. The new standard is called the European Sustainability Reporting Standards (ESRS) and contains all the requirements that must be followed when reporting to CSRD.

ISO14083 is primarily a standard for calculating energy consumption and CO_2 emissions. The standard follows the principles of the GHG protocol. A calculation/declaration according to ISO14083 will thus be following the GHG protocol. However, it is important to know, that reporting CO_2 (e) emissions following the GLEC framework will meet the requirements of ISO14083, whereas reporting according to ISO14083 will not necessarily meet the additional requirements of the GLEC framework.

Science Based Target Initiative, SBTi was established at the initiative of the UN, the Carbon Disclosure Project, the World Resources Institute and the WWF (World Wildlife Fund). The aim was to give companies common guidelines for how they should determine their targets for reducing CO_2 emissions in line with the objectives of the Paris Agreement. SBTi recommends that the GLEC framework forms the basis for the calculation of CO_2 emissions for transport chains.

2.2 Digital Product Passport

The European Digital Product Passport is an important initiative by the European Union towards sustainability and circular economy [14]. Introduced by the European Commission, this passport aims to provide valuable information about the environmental sustainability of products [15]. It is part of the broader efforts to promote

sustainability and achieve the goals of the European Green Deal [5]. The objective of the European Digital Product Passport is to enhance the traceability of products and their components, thereby contributing to a climate-neutral and circular European economy [16].

The passport will enable easy access to information about a product's environmental attributes, such as durability, reparability, recycled content, and availability of spare parts [15]. By providing this information, consumers will be empowered to make more informed purchasing decisions [14]. Additionally, the European Digital Product Passport also supports the goal of reducing CO_2 emissions by 55% by 2030, as part of the Ecodesign for Sustainable Products Regulation [16]

This initiative is part of a comprehensive plan laid out by the European Commission in March 2022 to improve product sustainability within the European Union [11]. The Circular Economy Package released by the European Commission identified the Digital Product Passport as a key solution to enhance product traceability [14]. This passport will not only benefit consumers but also contribute to a more sustainable and circular economy by promoting the design of products with longer lifetimes and better recyclability [11].

By implementing the European Digital Product Passport, the European Union is taking significant steps towards achieving its sustainability goals and promoting a more transparent and eco-friendly marketplace [16]. The passport is expected to play a crucial role in the European Union's transition to a sustainable and circular economy [16, 17].

Additionally, DPPs could be relevant to industries as digital-based supply chain compliance tools for reporting duties such as GHG emissions, life cycle assessments or corporate social responsibility reporting. For example, they could provide detailed and recent information along the value chain to create the conditions for Scope 3 GHG reporting. Scope 3 emissions cover all indirect emissions, both upstream and downstream, that occur in the value chain of a reporting company [11].

At the moment the establishment of the data requirements for the EU digital product passports for each specific product category will depend on a process of industry-wide stakeholder consultation. The entire supply chain has to work together to specify the vital information that could stop a product from going to waste. Some of the data specifications have already been established.

2.3 Emissions Trading System

The Emissions Trading System (ETS) is a central element of the European Union's efforts to combat climate change and reduce greenhouse gas emissions. Implemented in 2005, the EU ETS is the largest carbon market in the world, covering various sectors such as power generation, industry, and aviation [18]. The EU ETS operates on the cap-and-trade principle, wherein a cap is set on the total amount of greenhouse gas emissions allowed by participating entities [19]. These entities are then allocated a certain number of emission allowances, which they can use, either for their internal emissions or trade with other participants [19]. This market-based approach creates incentives for companies to reduce their emissions, as those that exceed their allowances must purchase additional permits, while those with lower emissions can sell their excess permits [19].

Over the years, the EU ETS has demonstrated effectiveness in reducing emissions. According to the European Environment Agency, between 2005 and 2020, the ETS sectors achieved a 42% reduction in emissions, far surpassing the initial reduction target of 21%. [20]. This success has been attributed to the economic incentives created by the carbon market, encouraging innovation and emissions abatement measures within covered industries [21].

Key features of the EU ETS include its flexibility and gradual reduction of allowances over time. The system employs a linear reduction factor, which decreases the number of available allowances year by year, thereby progressively tightening the emission cap [20]. This approach ensures a steady decline in emissions while allowing for adjustments to economic circumstances and policy developments [18].

The environmental integrity of the EU ETS is maintained through strict monitoring, reporting, and verification requirements for participating entities [19]. The European Union's emissions registry, as well as independent third-party verifiers, ensure accurate accounting and transparency within the system [21].

The EU ETS has also taken steps to address potential carbon leakage, where companies may shift production to regions with lax emissions regulations. To prevent this, the EU has introduced the Carbon Leakage List, which identifies sectors at risk and provides additional allowances to mitigate the risk of carbon leakage [22].

It addressed the fact that EU industries had been facing carbon costs from the European Emissions Trading System (ETS), unlike those in non-EU countries. This created a disadvantage for domestic producers. Although not currently expected to be in the current version, an EU CBAM could be used in the future to mandate that importers of regulated products pay a price based on the carbon footprint of their product before it is sold on the EU Single Market.

Going forward, the EU will therefore deploy an alternative policy instrument to avert emissions leakage: the Carbon Border Adjustment Mechanism, CBAM [23]. Supporting the EU ETS, this instrument will apply the carbon price faced by foreign producers to the emissions embedded in imports of some categories of products.

In the future, a DPP could be an important tool to disclose the carbon footprint of products. The need for reliable information throughout the whole product value chain is a key factor in calculating the carbon price through a CBAM

Furthermore, the EU ETS has evolved through time to address emerging challenges and align with the EU's climate ambitions. The system was revised in 2018 to strengthen its functioning and align with the goals of the Paris Agreement, including the introduction of the Market Stability Reserve to address potential surpluses or deficits in allowances [24]. The EU has also proposed expanding the scope of the ETS to

include new sectors such as shipping, bringing further emissions under its jurisdiction. In 2027, the EU ETS will be further extended to cover CO_2 emissions from road transport [25].

3 TRANSPORT PLANNING AND CARBON FOOTPRINT

According to the Eurostat [26], empty backhauls represent about 20% of road transportation activities in Europe. This is 5% lower than reported in 2014 [27] alongside the average use of loaded trucks of 57% of their capacity. In light of the strong focus on reducing CO_2 emissions, improved filling rates of trucks is an area worth researching. Potential benefits of horizontal collaboration for logistics services providers are found in the literature, which reports cost savings from less than 10% up to 50% [28-32]. The cost savings are based on increased efficiency and productivity [33] and are further leading to decreased environmental impact [34].

The transport companies investigated in this paper, all plan transports manually, although domain-specific systems are used for decision-making such as moving transports from one route (truck) to another. Most order requests come from primary or secondary customers but the load are not known until a few hours before the truck starts its route, which makes efficient planning complicated. To utilize the capacity best possible, they make use of "buying or selling" casual transports from/to other transporters to fully utilize their trucks. This is, however, very time-consuming and is mainly based on a huge number of phone calls and, to some degree, emails. As a last option to level capacity, they make use of transport brokers such as Teleroute and Timocom. This is, however, not a preferred option due to limited control of the transport and the limited financial outcome [35-37].

In a previous study [38], we presented a concept to enable a dynamic plan to be used "on route" based on an agent-based approach aiming at supporting selling and buying transports to and from other transport companies. This includes a large number of constraints, as illustrated in Figure 1, to evaluate whether a request qualifies for manual decision-making. The main aim was to increase the truckload while at the same time smoothening the communication and admin processes.

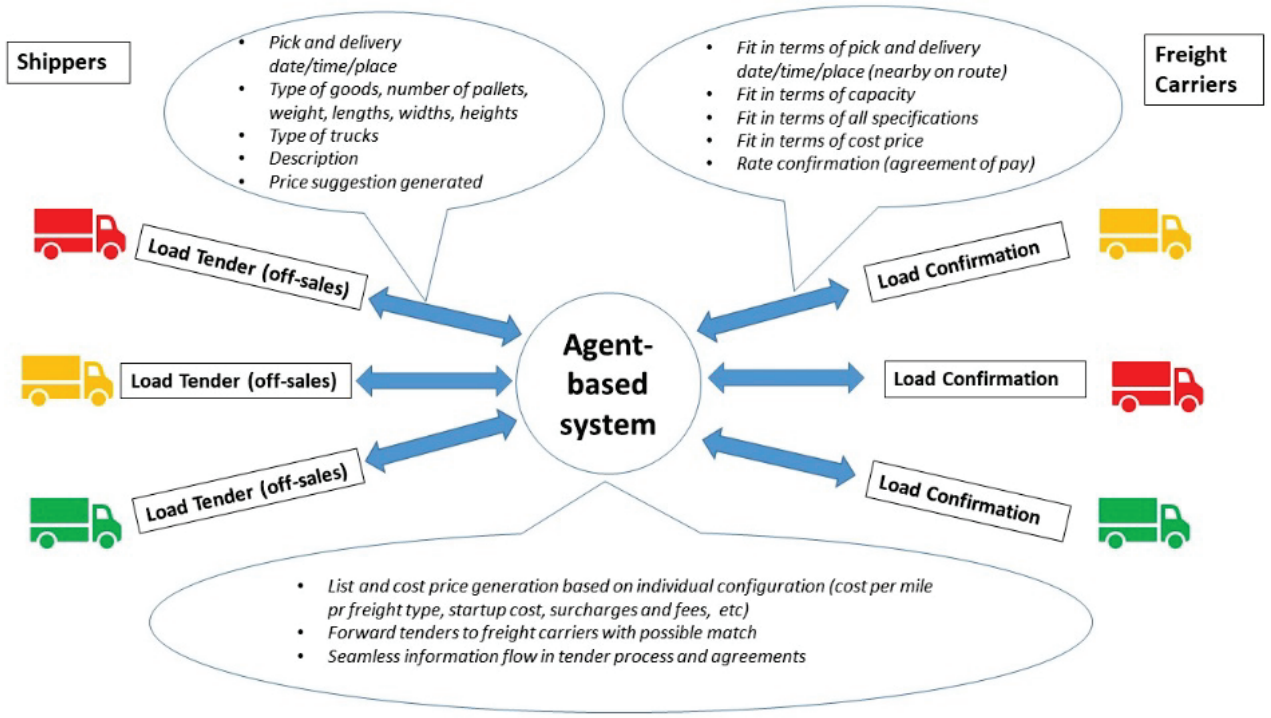


Figure 1 Conceptual illustration of tenders, requests, constraints and business model [38]

In the next section, we will delve into the methods and factors used to determine carbon emissions in the context of last-mile goods transport.

3.1 Carbon Footprint

Due to the increased focus on reducing the Carbon Footprint impact, transport companies experience that more and more customers ask for a declaration of the carbon footprint on the invoice. Further, some customers are asking for a carbon footprint estimate in their request alongside the cost. In the near future, caps on carbon footprint usage will be introduced. This leads to a situation where most or all customers will ask for a carbon footprint quotation and not just an estimate. By then, the carbon footprint will be just as important as the cost quotation.

The challenge for many transporters is, that costs and carbon footprint are difficult to estimate before you have a full truckload, as well as a full picture of which types of energy the customers are willing to pay for such as traditional

fuel, bio-fuel, electro-fuel, electricity etc. A truckload with a variety of customer-specific fuel "conditions" will either be a nightmare or require flexible customers accepting that goods on average are transported according to the fuel customers pay for, meaning that their specific goods may not be transported as environmentally friendly as they pay for. This will also depend on the future rules for calculating the carbon footprint as input for the digital product passport as discussed in section 2.

The starting point for calculating the emissions is the distance, the weight, the road conditions and the mode of transport:

1) **Distance**: The total distance covered during the transport operation is a fundamental factor. This includes each leg of distance travelled by different modes of transport—road, rail, water, or air—including detours or intermediate stops.
2) **Weight**: The weight of the goods being transported significantly impacts the energy required titled payload. The total weight, including the cargo and packaging, should be considered when calculating the transport work. Heavier loads require more energy for transportation.
3) **Road Conditions**: The terrain and road conditions, including inclines or declines, affect the energy needed for transport. The additional work required to overcome slopes or rough surfaces should be factored into the calculations.
4) **Mode of Transport**: Different modes of transport have varying energy efficiencies and associated work requirements. Road transport, for example, may have higher energy demands compared to rail or water transport. Considering the specific mode of transport involved in each leg of the journey is crucial.

By considering these factors, transport work can be calculated using established formulas by estimating emissions associated with last-mile goods transport. Carbon dioxide equivalents, CO_2 (e), are commonly used as a unit of measurement. Calculating emissions involves several extra factors:

1) **Fuel Consumption**: The type and quantity of fuel consumed by each mode of transport play a significant role in emission calculations. Different fuels, such as diesel, gasoline, or aviation fuel, have different carbon intensity values that need to be accounted for when estimating emissions.
2) **Emission Factors**: Emission factors represent the amount of emissions produced per unit of fuel burned. These factors are specific to different modes of transport and fuel types and are usually available from standardized emission databases.
3) **Mode-specific Considerations**: Each mode of transport has specific CO_2 emission factors (as well as methane, nitrous oxide, etc).

By multiplying the transport work with the CO_2 emission factor, emissions for the last-mile transport can be estimated. Carbon footprint calculators and emission inventory methodologies provided by organizations like the International Panel on Climate Change (IPCC) and emission databases by governmental bodies can assist in these calculations as well as companies like EcoTransIT [39] which has specialised in not only developing emission models but also collecting information on types of fuel used globally on ships and trucks.

4 TRANSPORT WORK AND EMISSIONS

In this section, we present 4 ways of calculating emissions for 3 deliveries from a distribution centre (DC) to customers.

- Delivery 1 consists of 8 tonnes and the distance from the DC is 5 km (10 km in total).
- Delivery 2 consists of 4 tonnes and the distance from the DC is 7 km (14 km in total).
- Delivery 3 consists of 6 tonnes and the distance from the DC is 9 km (18 km in total).

The calculations follow current guidelines by the organisation "Danish Freight Forwarders" [40] based on two factors.

- The total driving distance forms the basis for determining energy consumption and emissions.
- The payload is used to calculate transport work for distributing emissions among customers.

Quotations to the transport customer are for one trip at a time. Emissions can be differentiated into 2 types: *well-to-tank* and *tank-to-wheel*. Well-to-tank (WTT) emissions; referred to as energy provision GHG emissions and tank-to-wheel emissions (TTW), are referred to as operational GHG emissions. Jointly, these two add up to the well-to-wheel (WTW) emissions and they build the emissions of an entire transport chain element (TCE). In the following calculations, WTW emissions are used in the case examples.

The values for the emission factors used must be calculated following EU guidelines [19] and guidelines by GLEC [3] and ISO14083 [6]. The trucks drive on average 3.8 km. per litre of diesel (5% biodiesel) and the WTW emission is 3.17 kg CO_2 (e) per litre. The WTW emission is therefore 3.17/3.8 = 0,834 CO_2 (e) per km. The payload is not included in the emission calculations [40].

In the following, a pre-calculation and 3 different ways of conducting post-calculations are presented.

4.1 Pre-Calculation

Each customer inquiry is based on a pre-calculation. At the time of the quotation, it is unknown whether you will manage to get a return trip. Therefore, the pre-calculation is based on an empty return trip and orders are processed one by one. As the calculations entirely are based on transport distance, there is no emission contribution to the transport company, as it will be distributed among the individual shipments in a weighted ratio. Empty driving, repositioning without load and other driving without payload, which

consumes energy and emits emissions, is distributed among the customers.

The CO_2 (e) emission (WTW) pre-calculations of the 3 deliveries are as follows:
(1): 10 km×0,834 kg CO_2 (e) = 8.342 kg CO_2 (e).
(2): 14 km×0,834 kg CO_2 (e) = 11.679 kg CO_2 (e).
(3): 18 km×0,834 kg CO_2 (e) = 15.016 kg CO_2 (e).

Given the above conditions where shipments are carried out one at a time. The truck will drive 42 km in connection with carrying out shipments, of which 21 km without a load. The total emission is 35.04 kg CO_2 (e) as it is the measured fuel consumption that is the basis for calculating the CO_2 (e) emissions post-calculation.

4.2 Post Calculation, Case 1

In this case, orders are processed one by one according to the pre-calculation, but the emission is based on a weighted allocation of the transport work (km × tonnes).

The total emission is similar to the pre-calculation, which is 35.037 kg CO_2 (e) and the total transport work is 8×5 + 4×7 + 6×9 = 122 tonnes-km. The post-calculations of the 3 deliveries' emissions are as follows:
(1): 8×5/122×35,037 kg CO_2 (e) = 11.487 kg CO_2 (e)
(2): 4×7/122×35,037 kg CO_2 (e) = 8.041 kg CO_2 (e)
(3): 6×9/122×35,037 kg CO_2 (e) = 15.508 kg CO_2 (e)

In total 35,037 kg CO_2 (e) is allocated for deliveries, and zero kg for the transport company.

4.3 Post-Calculation, Case 2

In this case, the deliveries are consolidated as one trip instead of 3. This means that the total distance travelled in kilometres is reduced from 42 to 18 and with more cargo. All in all, a better filling rate and less empty driving. Further, both the outbound and inbound emission is distributed among the 3 deliveries according to transport work [40].

The post-calculation of the 3 deliveries' emissions are as follows:
(1): 8×5/122×18×0,834 kg CO_2 (e) = 4,9223 kg CO_2 (e)
(2): 4×7/122×18×0,834 kg CO_2 (e) = 3,3446 kg CO_2 (e)
(3): 6×9/122×18×0,834 kg CO_2 (e) = 6,6460 kg CO_2 (e)

In total 15.02 kg CO_2 (e) is allocated for deliveries, and zero kg for the transport company.

4.4 Post-Calculation, Case 3

In this final case, the deliveries are consolidated as one trip, similar to case 2. However, the emissions are distributed differently, as the outbound emission is distributed among the 3 deliveries according to payload [40] whereas the inbound emission is put on the transport company.

The post-calculation of the 3 deliveries' emissions are as follows:
(1): 8×5/122×9×0,834 kg CO_2 (e) = 2,462 kg CO_2 (e)
(2): 4×7/122×9×0,834 kg CO_2 (e) = 1,723 kg CO_2 (e)
(3): 6×9/122×9×0,834 kg CO_2 (e) = 3,323 kg CO_2 (e)
(R): 9×0,834 kg CO_2 (e) = 7,508 kg CO_2 (e) for the inbound (return) trip

, The results are summarised in Tab. 1, illustrating the huge difference between the different ways of calculating CO_2 emissions and raising a flag in terms of the importance of the chosen quotation strategy for transport companies in the coming years.

Table 1 illustrates 4 ways of calculating CO_2 (e) emissions of 3 deliveries. **Pre-calc**) a pre-calculation based on individual deliveries. **Post-calc 1**) a post-calculation based on individual deliveries. **Post-calc 2**) a post-calculation of consolidated deliveries. **Post-calc 3**) Similar to Post-calc 3, except that the customers are not "claimed" for the return trip

Case	Delivery	Tonnes	Km outbound	Km inbound	Tonnes*Km In truck	Tonnes*Km delivery	Tonnes*Km distribution	CO2 (e) distribution	Emission efficiency
Pre-calc	1	8	5	5	40	40		8,342	0,209
	2	4	7	7	28	28		11,679	0,417
	3	6	9	9	54	54		15,016	0,278
	Sum		**21**	**21**	**122**	**122**		**35,037**	
Post-calc 1	1	8	5	5	40	40	0,328	11,487	0,287
	2	4	7	7	28	28	0,230	8,041	0,287
	3	6	9	9	54	54	0,443	15,508	0,287
	Sum		**21**	**21**	**122**	**122**	**1,000**	**35,037**	
Post-calc 2	1	8	5		90	40	0,656	4,923	0,123
	2	4	2	9	20	28	0,459	3,446	0,123
	3	6	2		12	54	0,885	6,646	0,123
	Sum		**9**	**9**	**122**	**122**	**2,000**	**15,016**	
Post-calc 3	1	8	5	0	90	40	0,328	2,462	0,062
	2	4	2	0	20	28	0,230	1,723	0,062
	3	6	2	0	12	54	0,443	3,323	0,062
	Return	0		9	0	0	0	7,508	na
	Sum		**9**	**9**	**122**	**122**	**1**	**15,016**	

5 DISCUSSION

Order management and last-mile transport planning are challenging. Minimizing the difference between pre-calculation and post-calculation is even more important since CO_2 emission now needs to be documented and reported to the customer and authorities, based on a measurement of the transport service.

Overall, we have found the following answers to the 3 research questions.

It has been shown with the 4 cases, that differences between pre- and post-calculation are likely unless freight quote agreements are more flexible and contain an agreement on how an emission difference is to be distributed between the partners. To minimize this emission difference, it is important to use and maintain good key figures for planning, as they will be decisive for the size of the difference.

Generally, last-mile deliveries are less energy-efficient (with a higher fuel consumption) than long-haul freight, due to frequent stops, a higher share of idle time per trip, multiple delivery points per trip, a higher frequency of deliveries, different drop-off times, and nonoptimal rescheduling, due to the absence of recipients [41, 42].

Emission-driven KPIs for last-mile deliveries will be largely based on empty driving, the utilization rate of payload and the distance travelled versus time unit ratio, as optimization of these will result in the lowest emissions for a transport unit, which will probably also become an important

competitive parameter between freight distribution companies.

Documentation and reporting requirements for freight distribution companies are changing, especially when it comes to the new requirements for emissions and sustainability described in section 2.

Goods owners will have an increasing focus on the emission of their goods, not at least driven by the upcoming DPP. Customers expect minimum emissions related to their delivery. Also, the company's reporting requirements for GHG in scope 3, means that freight last-mile delivery in particular will be challenged as complexity is high and in terms of planning challenging.

Transport quotes will probably develop and adapt but it may have negative consequences for transparency. The partners may, according to the agreement, distribute the emissions between each other, which will be reported on the GHG annual statement, even if it is higher than expected.

The challenge in terms of transparency will probably be that the product owner does not want the actual emission to be disclosed on the planned digital product passport, precisely because it is agreed how the emission is to be distributed. It will be an important competitive parameter for the product owner that the product has the lowest emission printed on the digital product passport, as the customer can see this and will probably use this to decide on product selection. The problem we see is that even if the emission deviation results in a contribution to the transport company's envelope, the real emission will not be disclosed in the digital product passport, as the owner of the goods wants the agreed emission disclosed. Based on this, we believe we have answered the 3 research questions.

6 CONCLUSION

Accurately calculating transport work and emissions for last-mile freight distribution is crucial in understanding the environmental impact of transportation activities. Advancements in data collection, standardized methodologies, and technology-driven solutions can further refine these calculations, leading to better-informed decisions and a more sustainable future.

By considering factors like distance, weight, mode of transport and emission factors, businesses and policymakers can gain insights into the energy requirements and associated emissions of transporting goods. This information can improve decision-making, encourage more sustainable transport choices, and minimize the environmental impact of last-mile freight distribution.

Freight quotes based on a customer's inquiry are by nature a pre-calculation and a challenge in itself as the the remaining transportation tasks are more or less unknown at the time of the quotation. The current guidelines for freight forwarders, where all emissions are distributed to the customers and thus no emissions to the transport company, are under pressure, as this is not fair for the customers. In the future, customers may not accept an actual carbon footprint usage that is way higher than the quote provided by the transporter in advance. This leaves the transporter in a dilemma as the actual carbon footprint usage depends on which other customer orders they manage to get and how they can combine the total load in their truck planning.

The freight quotes agreement is expected to adapt to deal with this issue, which we expect the transport market will do, but the question remains whether the digital product passport will require the actual CO_2 emissions or it will accept the calculated CO_2 emissions related to the freight quotes agreement when entering into the transport agreement.

Acknowledgements

Kenn Steger-Jensen, Hans Henrik Hvolby and Sven Vestergaard would like to express their gratitude to the Interreg OKS for funding the Value4Sea project to which this paper belongs [43]

Disclaimer

This paper is unrelated to one of the author's employment at the International Monetary Fund (IMF) and is undertaken in his capacity. The views expressed in this paper are those of the authors and do not necessarily represent the views of the IMF, its Executive Board, or IMF management.

7 REFERENCES

[1] Gevaers, R., Van de Voorde, E. & Vanelslander, T. (2011). Characteristics and Typology of Last-mile Logistics from an Innovation Perspective in an Urban Context. In *City Distribution and Urban Freight Transport: Multiple Perspectives*. Cheltenham, UK: Edward, Elgar. https://doi.org/10.4337/9780857932754.00009

[2] Jesús, M., Cortés, P., Grosso, R. & Guadix, J. (2012). Selecting the Location of Minihubs for Freight Delivery in Congested Downtown Areas. *Journal of Computational Science, 3*(4), 228-237. https://doi.org/10.1016/j.jocs.2011.12.002

[3] Ranieri, L., Digiesi, S., Silvestri, B. & Roccotelli, M. (2018). A Review of last-mile Logistics Innovations in an Externalities Cost Reduction Vision. *Sustainability, 10*(3), 782. https://doi.org/10.3390/su10030782

[4] Van Loon, P., Deketele, L., Dewaele, J., McKinnon, A. & Rutherford, C. (2015). A Comparative Analysis of Carbon Emissions from Online Retailing of Fast Moving Consumer Goods. *Journal of Cleaner Production, 106*, 478-486. https://doi.org/10.1016/j.jclepro.2014.06.060

[5] Edwards, J. B., McKinnon, A. & Cullinane, S. L. (2010). Comparative Analysis of the Carbon Footprints of Conventional and Online Retailing: A "Last-mile" Perspective. *International Journal of Physical Distribution & Logistics Management, 40*(1/2), 103-123. https://doi.org/10.1108/09600031011018055

[6] Lang, Z., Yao, E., Hu, W. & Pan, Z. (2014). A Vehicle Routing Problem Solution Considering Alternative Stop Points. *Procedia - Social and Behavioral Sciences, 138*, 584-591, https://doi.org/10.1016/j.sbspro.2014.07.242

[7] Franceschetti, A., Honhon, D., Van Woensel, T., Bektaş, T. & Laporte, G. (2013). The time-dependent pollution-routing problem. *Transportation Research Part B: Methodological, 56*, 265-293. https://doi.org/10.1016/j.trb.2013.08.008

[8] Liu, S. & Hassini, E. (2024). Freight last-mile delivery: a literature review. *Transportation Planning and Technology, 47*(3), 323-369. https://doi.org/10.1080/03081060.2023.2268601

[9] Lyons, T. & McDonald, N. E. (2023). Last-mile Strategies for Urban Freight Delivery: A Systematic Review. *Transportation Research Record, 2677*(1), 1141-1156. https://doi.org/10.1177/03611981221103596
[10] Olsson, J., Hellström, D. & Pålsson, H. (2029). Framework of Last-mile Logistics Research: A Systematic Review of the Literature. *Sustainability, 11*(24), 7131. https://doi.org/10.3390/su11247131
[11] Greenhouse Gas Protocol. FAQ. See: ghgprotocol.org/ sites/default/files/standards_supporting/FAQ.pdf. Last visited on March 15, 2024
[12] ISO14083:2023. Greenhouse gases: Quantification and reporting of greenhouse gas emissions arising from transport chain operations, See: www.iso.org/standard/78864.html, last visited on March 22, 2024
[13] GLEC Framework, See: smartfreightcentre.org/en/our-programs/global-logistics-emissions-council/ Last visited on March 21, 2024.
[14] PSQR Software Company(Performance, Service, Quality and Reliability): All You Need To Know About the EU Digital Product Passports. See: psqr.eu/publications-resources/all-about-dpp/. Last visited on March 15, 2024
[15] Ecodesign for Sustainable Products Regulation, EU. See: commission.europa.eu/energy-climate-change-environ ment/standards-tools-and-labels/products-labelling-rules-and-requirements/sustainable-products/ecodesign-sustainable-products-regulation_en, last visited on March 14, 2024
[16] European Circular Economy Stakeholder Platform. See: circulareconomy.europa.eu/platform/en/knowledge/digital-product-passport-ticket-achieving-climate-neutral-and-circular-european-economy, last visited on March 4, 2024
[17] Psarommatis, F. & Gökan, M. (2024). Digital Product Passport: A Pathway to Circularity and Sustainability in Modern Manufacturing. *Sustainability, 16*(1), 396. https://doi.org/10.3390/su16010396
[18] EU Emissions Trading System (EU ETS). See climate.ec.europa.eu/eu-action/eu-emissions-trading-system-eu-ets_en. Last visited on January 12, 2024
[19] UN Framework Convention on Climate Change. Annex I: Parties' greenhouse gas emission reduction targets. See: unfccc.int/sites/default/files/resource/sbi2023_inf07.pdf Last visited 22 March 2024.
[20] ETC CM report 2023/07: Trends and projections in the EU ETS in 2023. The EU Emissions Trading System in numbers. See www.eionet.europa.eu/etcs/etc-cm/products/ etc-cm-report-2023-07-1. Last visited 02 March 2024
[21] https://www.climate-policy-watcher.org/. Last visited 02 March 2024
[22] EU ETS Carbon Leakage List - European Commission. See ec.europa.eu/clima/policies/ets/allowances/ carbon leakage en. Last visited 08 December 2023.
[23] Regulation (EU) 2023/956 of the European Parliament and of the Council of 10 May 2023 establishing a carbon border adjustment mechanism (Text with EEA relevance). See: Regulation - 2023/956 - EN - cbam regulation - EUR-Lex (europa.eu) Last visited 02 March 2024.
[24] EU Market Stability Reserve. See: climate.ec.europa.eu/eu-action/eu-emissions-trading-system-eu-ets/market-stability-reserve_en. Last visited 22 March 2024.
[25] See: cdp.net/en/articles/climate-action/5800-from-targets-to-action-the-eu-emissions-trading-system. Last visited 8 December 2023.
[26] Eurostat: A fifth of road freight kilometres by empty vehicles. See https://ec.europa.eu/eurostat/web/products-eurostat-news/ -/ddn-20211210-1. Last visited on April 30, 2024.
[27] Juan, A. A., Faulin, J., Pérez-Bernabeu, E. & Jozafowiez, N. (2014). Horizontal cooperation in Vehicle Routing Problems with Backhauling and Environmental Criteria. *Procdia, Social & Behavioral Sciences, 111*, 1133-1141. https://doi.org/10.1016/j.sbspro.2014.01.148
[28] Quintero-Araujo, C. L., Gruler, A. & Juan, A. A. (2016)- Quantifying potential benefits of horizontal cooperation in urban transportation under uncertainty: A simheuristic approach. *Conference of the Spanish Association for Artificial Intelligence*, 280-289. https://doi.org/10.1007/978-3-319-44636-3_26
[29] Soysal, M. Bloemhof-Ruwaard, J. M., Haijema, R. & van der Vorst, J. G. (2018). Modeling a green inventory routing problem for perishable products with horizontal collaboration. *Computers & Operations Research, 89*, 168-182. https://doi.org/10.1016/j.cor.2016.02.003
[30] Montoya-Torres, J. R., Muñoz-Villamizar, A. & Vega-Mejia, C. A. (2016). On the impact of collaborative strategies for goods delivery in city logistics. *Production Planning and Control, 27*, 443-455. https://doi.org/10.1080/09537287.2016.1147092
[31] Bailey, E., Unnikrishnan, A. & Lin, D.-Y. (2011): Models for minimizing backhaul costs through freight collaboration. J*ournal of Transport Res. Board. 2224*, 51-60. https://doi.org/10.3141/2224-07
[32] Sanchez, M., Pradenas, L., Deschamps, J.-C. & Parada, V. (2016). Reducing the carbon footprint in a vehicle routing problem by pooling resources from different companies. *NETNOMICS Economic Research and Electronic Networking, 17*, 29-45. https://doi.org/10.1007/s11066-015-9099-2
[33] Gansterer, M. & Hartl, R. F. (2018): Collaborative vehicle routing: A survey. *Eur. J. Oper. Res. 268*, 1-12. https://doi.org/10.1016/j.ejor.2017.10.023
[34] Pérez-Bernabeu, E., Juan, A. A., Faulin, J. & Barrios, B. B. (2015). Horizontal cooperation in road transportation: A case illustrating savings in distances and greenhouse gas emissions. *International Transactions in Operational Research, 22*, 585-606. https://doi.org/10.1111/itor.12130
[35] Timocom, www.timocom.com, last accessed April 12, 2024
[36] Teleroute, www.teleroute.com, last accessed April 12, 2024
[37] Trucker Path, truckerpath.com, last accessed April 12, 2024
[38] Hvolby, H.-H., Steger-Jensen, K., Neagoe, M., Vestergaard, S. & Turner, P. (2019). Collaborative Exchange of Cargo Truck Loads: Approaches to Reducing Empty Trucks in Logistics Chains. *Proceedings of APMS, Advances in Production Management Systems, 2*, 68-74. https://doi.org/10.1007/978-3-030-29996-5_8
[39] EcoTransIT, See: www.ecotransit.org/en/methodology/ last visited on March 4, 2024.
[40] Guidelines for freight forward calculations according to the Danish Association of freight forwarders. See https://dasp.dk/wp-content/uploads/Danske-Speditoerers-Guide-for-CO2-beregning-og-deklaration.pdf. Last visited on April 29, 2024.
[41] Bandeira, R. A. M., D'Agosto, M. A., Ribeiro, S. K., Bandeira, A. P. F. & Goes, G. V. (2018). A Fuzzy Multi-Criteria Model for Evaluating Sustainable Urban Freight Transportation Operations. *Journal of Cleaner Production, 184*, 727-739. https://doi.org/10.1016/j.jclepro.2018.02.234
[42] Patella, S. M., Grazieschi, G., Gatta, V., Marcucci, E. & Carrese, S. (2021). The Adoption of Green Vehicles in Last-mile Logistics: A Systematic Review. *Sustainability, 13*(1), 6. https://doi.org/10.3390/su13010006

[43] Interreg OKS funding body and link. See https://interreg-oks.eu/da/larkannaoss/projektbanken/projekt/value4sea.1924.html. Last visited on May 2, 2024.

Authors' contacts:

Kenn Steger-Jensen, Associate Professor
Aalborg University, M&P,
Fibigerstrede 16, DK 9220 Aalborg, Denmark
+45 2420 5094, E-mail: kenn@mp.aau.dk
and
Professor, University of South-Eastern Norway,
Department of Maritime Operations,
Campus Vestfold, Raveien 215, 3184 Borre, Norway
E-mail: kenn.steger-jensen@usn.no

Hans-Henrik Hvolby, Senior Researcher, PhD
(Corresponding author)
UCN act2learn, Department of Business and Industry,
Sofiendalsvej 60, DK 9200 Aalborg, Denmark
+45 7269 1764, E-mail: hhv@ucn.dk
https://orcid.org/0000-0002-5574-5216

Sven Vestergaard, Senior Researcher
UCN act2learn, Department of Business and Industry,
Sofiendalsvej 60, DK 9200 Aalborg, Denmark
+45 7269 1733, E-mail: svv@ucn.dk
https://orcid.org/0000-0003-0317-173X

Mihai Neagoe, PhD
Foresion.net
Mount Gambier, Australia
mihai.neagoe@foresion.net

Carsten Svensson, Senior Project Officer
International Monetary Fund,
Washington DC, United States
csvensson@imf.org

ISSN 1846-6168 (Print), ISSN 1848-5588 (Online)
https://doi.org/10.31803/tg-20240512171214

Preliminary communication
Received: 2024-05-12, Accepted: 2024-05-20

Technology or Organization: What is More Important for Artificial Intelligence Adoption?

Aman Pathak*, Veena Bansal

Abstract: Artificial intelligence (AI) technology is different from all other technologies that organizations have adopted in the past. A systematic literature review revealed that technological, organizational, and environmental factors have been explored to assess an organization's readiness for adopting a new technology. In this work, we have focused on the first two factors. From many subfactors, we selected 13 subfactors based on the discussion with the domain experts. Experts also provided ranks of the two factors and their sub-factors. These ranks are used to calculate the global ranking of the factors and their sub-factors. The top three subfactors from the technology context are the following - capabilities of AI, compatibility, and complexity of AI systems. The top two subfactors from the organization context are technology infrastructure & skilled workforce and support from the top management. The results reveal that organization and technology are equally important.

Keywords: AI capabilities; R-SWARA; Systematic Literature Review; TOE framework; Top management support

1 INTRODUCTION

This article focuses on Artificial Intelligence (AI), an innovative technology. We asked 30 technically savvy adults to define AI. Following are the consolidated responses –

- Tools that are part of our daily life.
- Tools that increase our productivity, perform repetitive tasks, and reduce human effort.
- AI has a cognitive ability to perform complex tasks that involve reasoning and decision-making.

We further asked them if they use AI. Everyone gave an affirmative answer. Some went to the extent of saying that life now revolves around AI. The following AI applications were mentioned: google-assistant, ChatGPT, recommendation engines in eCommerce websites, email spam filters, movie recommendations, Instagram filters, help in coding, and check grammar. It is interesting to note that no critical application of AI was mentioned.

However, the scenario changes when adopting AI at the organizational level. AI systems provide various business benefits, such as enhancing efficiency by automating repetitive processes, better decision-making by predicting multiple business metrics, improving customer service, and content generation [2,6]. The adoption of AI in Indian financial services organizations is growing and is currently at a moderate stage [8]. The business intention for adopting AI is to enhance the efficiency of backend operations and improve customer service through interactive AI conversational agents [7, 11].

Adopting an innovation such as AI is a multistage process [26]. Some studies treat this as a three-stage process consisting of the following stages: Pre-adoption, Adoption, Post-adoption, or Usage [23]. These stages have been expanded to the following five stages: Knowledge, Persuasion, Decision, Implementation, and Confirmation [22]. During the pre-adoption phase, an organization checks its preparedness for the technology. Factors that affect an organization's readiness for adopting AI have been studied using the technology-organization-environment (TOE) framework [1, 11]. However, AI is an emergent technology, and its adoption has yet to be widely studied. The extant literature is scarce on studies determining the factors and their importance in assessing organizations' readiness. The objectives of the study are the following –

- Determine the relative importance of the technology and organization factors in affecting AI adoption intention.
- Determine if Technology and Organization are two independent realities.

We use a multicriteria decision-making approach, Rough Stepwise Weight Assessment Ratio Analysis (R-SWARA) Method [27], to obtain the weight scores of the factors affecting AI adoption intention. We perform unpaired t-tests to obtain difference between the subfactors. The literature review is presented in Section 2, the research methodology, data analysis & results in Section 3, and the discussion & conclusion in Section 4.

2 LITERATURE REVIEW

Our focus in this work is to study the relative importance of two factors, namely organization and technology, concerning an organization's readiness for adopting AI. The technology-organization-environment (TOE) [25] framework has been widely used in the literature to study an organization's readiness. We have not incorporated environmental factors because earlier studies indicate it to be a less critical construct [20]. We conducted a systematic literature review to extract subfactors of organization and technology factors that affect AI adoption intention.

2.1 Organization and Its Subfactors

An organization is an entity that is represented by its constituent. For instance, the top management is an essential constituent of an organization. With the support of top management, an organization may be able to adopt AI. There are multiple reasons for the importance of their support. The top management support would provide the required resources for AI adoption [9]. The lack of financial capability could inhibit the intention to adopt AI [9]. The financial

resources are needed to create the assets and competency to leverage the potential of AI [13]. In addition, the human resources required for adopting AI would be made available only if the management would release them from their regular task.

The top management's understanding of AI and potential risks also facilitates the adoption of AI in an organization [11]. The understanding and preparedness of the management are influenced by the existing IT infrastructure in the organization. If the organization has computational power and data storage [13], the management will likely have a positive intention towards adopting AI. The management may have an understanding of AI and may be interested in leveraging the existing infrastructure. A related subfactor, the data ecosystem consisting of adequate data security and privacy measures, positively affects AI adoption [20].

An organization with the IT infrastructure and the data ecosystem will likely have digitally skilled human resources that may positively impact AI adoption intention [24]. The right talent and digitally skilled human resources are needed to transform business processes through AI. However, employees may lack an understanding of AI's impact and may fear job security [4, 13]. The fear of change could inhibit an organization's intention to adopt AI.

One of the reasons for not adopting AI is uncertainty about the return on investment in AI due to a lack of understanding of AI technology, business needs, and technology vendors [2]. Uncertainty about the return on investment could inhibit the intention to adopt AI [2]. Clear AI-business case involves identifying suitable tasks for AI and performance evaluation matrices [26]. Assessing AI-process fit could positively affect AI adoption [13].

2.2 Technology and Its Subfactors

AI has been categorized based on its capability into three broad classes as follows:

Automation Intelligence - AI possesses automation intelligence if it can be used to automate processes. Examples of such systems include chatbots, customer service, and customer claim settlement [7]. Automation intelligence requires a large amount of training data. These are the least complex and expensive among the three variants of AI [2, 19]. Business teams may own and use automation intelligence for good ROI [15, 18].

Analytical Intelligence - Analytical AI systems create a cognitive representation of the world based on the patterns in data that represent prior learning and experience to make predictions [2, 12, 14]. Virtual insurance advisors [18], fraud detection in insurance claims [7], predicting consumer buying behaviour, and accurate actuarial modelling [2] are some of the examples of analytical intelligence.

Empathetic Intelligence - Empathetic Intelligence (a.k.a Human-Inspired AI) [12, 14] can recognize and incorporate human emotions. Despite numerous applications of empathetic intelligence [2, 12], it continues to be more of a theoretical concept. We told ChatGPT, one of the most talked about intelligent systems, the following, “The gift for my son arrived late; his birthday and party were yesterday.” The responses from ChatGPT were annoying in the least:

- Explain the situation to your son and let him know that the gift is on its way.
- Assure him that you still want to celebrate his birthday and make it special.
- Schedule another time to celebrate your son's birthday when the gift arrives.

ChatGPT has completely ignored the fact that his birthday was yesterday. Nevertheless, it has provided some useful advice.

In order to adopt AI, the organization must have a clear idea of its capability. If the organization finds a fit between AI and its business processes, the organization may be more inclined toward adopting AI. The compatibility of AI with the organization would affect AI adoption intention [1, 11]. AI systems are vulnerable to cyberattacks, adversarial data, and denial-of-service attacks [21]. The organization may want assurance of its data security and privacy while adopting AI [5].

A characteristic of AI systems that makes them unique is the lack of explainability. AI systems are often called black boxes due to lack of explainability and transparency. For instance, a linear multi-regression predictive model finds a linear relationship between independent variables and the dependent variable based on historical data. The coefficients of the independent variables are virtually meaningless, as the model might have used centering scaling and regularization techniques. Lack of explainability may inhibit an organization from adopting AI [3, 16].

The quality of data used for training AI systems is very important, i.e., it should be free of biases and prejudices [3]. These biases in AI systems could affect an organization's intention to adopt them. Adopting AI systems may be complex due to the effort required to learn and work with them [9, 11]. A related issue is trust in AI systems, especially in empathetic AI systems [10]. We summarize the discussion on factors and their subfactors in Tab. 1.

3 RESEARCH METHOD, DATA ANALYSIS & RESULTS

Our objective is to assess the relative importance of the factors, Technology, and Organization directly and through their subfactors. Our first method to establish was very simple and straightforward. We asked 20 experts to rank technology and organizations according to their relative importance in adopting AI. Half of the industry experts were from the financial services sector, and the other half were from the technology consultancy sector.

The selected experts from the consultancy services area were employed in multi-national companies with revenue exceeding $ 40 million. The financial services organizations that employed the selected experts were the leading major insurance firms in the country. The work experience of all the experts was around 15 years. The profile of all 20 experts is shown in Tab. 2.

Table 1 Subfactors of Organization and Technology factors

Factor	Subfactor	References
Organization	Top management support (O1)	[9, 11]
	Technology Infrastructure & skilled workforce (O2)	[9, 24]
	Lack of Funding (O3)	[9]
	Employee Fear of Change (O4)	[4, 13]
	Uncertainty of the Return on Investment (O5)	[2]
	Strategic alignment of AI systems & Unclear Business case (O6)	[26]
Technology	Compatibility of AI (T1)	[1, 11]
	Security and Privacy Concern (T2)	[5, 21]
	Capability of AI (T3)	[12]
	Lack of Explainability (T4)	[3, 16]
	Inherent Biases (T5)	[3]
	Complexity of AI (T6)	[9, 11]
	Anthropomorphism of AI system (T7)	[10]

Table 2 Profile of the selected experts

Expert	Job profile	Experience	Sector
E1	Technical support manager	14	Technology Solution provider
E2	Manager - IT delivery	14	Strategy & Consulting
E3	Senior manager - Digital transformation	18	IT consultancy
E4	Manager - Digital products	15	Cloud services
E5	Senior consultant - Digital transformation	15	IT consultancy
E6	Technical architect	16	IT consultancy
E7	Senior consultant - Digital transformation	14	IT consultancy
E8	Manager - Client relationship	15	Technology Solution
E9	Senior Analyst - R&D	15	IT consultancy
E10	Solutions architect	16	IT consultancy
E11	Regional manager - Marketing & Sales	14	Life Insurance
E12	Regional manager - Operations	15	Life Insurance
E13	Manager - Underwriting	14	Life Insurance
E14	Zonal manager	14	Life Insurance
E15	Divisional manager - Operations	15	General Insurance
E16	Senior manager - Claims	17	General Insurance
E17	Manager - Risk & Audit	13	General Insurance
E18	Manager - KYC & AML	14	Ecommerce - Insurance
E19	Manager - Business strategy	13	Health Insurance
E20	Manager - Customer service	15	Health Insurance

Three experts pointed out that we had left Individual and Environment out, which are also important factors. These three experts ranked Individual, Environment, Technology, and Organization factors 1, 2, 3, and 4, respectively. Respecting their opinion, we retained their ranks. All other experts ranked Technology and Organization between 1 and 2. The average ranks were 1.7 and 1.8, with standard deviations 0.98 and 0.83 for Technology and Organization, respectively. The difference between these two turned out to be insignificant at a 95% confidence level. Thus, the relative importance of technology and organization factors could not be established through this simple exercise for AI adoption.

We added the subfactors under technology and organization factors and consulted experts again to help us drop some subfactors. We started with 20 subfactors spread across Organization and Technology and reduced them to 13 with the help of experts. Experts were asked individually to rank these subfactors. Ranks were converted into weight scores using R-SWARA. Experts ranked the subfactors in order of importance. The ranks provided by 20 experts were used as the input for R-SWARA.

R-SWARA (Rough Stepwise Weight Assessment Ratio Analysis Method) [27] is a multi-criteria decision-making (MCDM) method that calculates the weights of the factors based on individual rankings. R-SWARA has been used earlier to rank the critical success factors for AI adoption in the food supply chain [4] and healthcare supply chain [17].

The analysis shows that Technology and Organization weigh 0.463 and 0.303, respectively. The weight scores based on combined individual ranks for the subfactors are also obtained using R-SWARA. These weight scores represent ranks of subfactors.

The weight scores and respective ranks for the subfactors obtained from R-SWARA are presented in Tab. 3. The high-ranking sub-factors from the technology factor were the capability of AI (T3), compatibility of AI (T1), and complexity of AI (T6). The high-ranking sub-factors from the organization factor were top management support (O1) and technology infrastructure & skilled workforce (O2) in the organization. We retained the top three sub-factors of Technology and two subfactors of Organization for further analysis.

We further investigated and assessed the impact of the technology and organization subfactors on readiness for adopting AI. When we look at AI technology standing in an organization, we feel that it is the technology that has to be ready for our organization. At the same time, when we spoke to a few technology experts and consultants, they suggested that an organization should be ready for the technology. There is a possibility that both are right.

Table 3 Ranking of affecting subfactors affecting AI adoption

Factor	Subfactor	Weight	Rank
Organization (Weight = 0.303)	Top management support (O1)	0.357	1
	Technology Infrastructure & skilled workforce (O2)	0.268	2
	Lack of Funding (O3)	0.029	6
	Employee Fear of Change (O4)	0.188	3
	Uncertainty of the Return on Investment (O5)	0.059	5
	Strategic alignment of AI systems & Unclear Business case (O6)	0.110	4
Technology (Weight = 0.463)	Compatibility of AI (T1)	0.263	2
	Security and Privacy Concern (T2)	0.113	4
	Capability of AI (T3)	0.328	1
	Lack of Explainability (T4)	0.064	5
	Inherent Biases (T5)	0.034	6
	Complexity of AI (T6)	0.183	3
	Anthropomorphism of AI system (T7)	0.017	7

We treated the selected subfactors as constructs and measured those using scales from the literature. Top management support (TMS), Technology infrastructure & skilled workforce (TI), Capability of AI (AICapa), Compatibility of AI (CB), and Complexity of AI (CX) are the constructs. Top management support (TMS) and Technology infrastructure & skilled workforce (TI) were under the organization context. The capability of AI (AICapa), compatibility of AI (CB), and complexity of AI (CX) were under the technology context.

A questionnaire was constructed, and a survey was conducted. The 7-point Likert scale was used. The respondents were professionals from the financial services sector. The questionnaire was sent to 858 professionals. A total of 327 responses were received.

Demographic profile of respondents: The respondents from the financial services sector worked as operations managers, IT systems managers, business development managers, area managers, and zonal area managers. There were 214 (65.44%) male respondents and 113 (34.55%) female respondents. Most respondents (59.02%) were from (25-34) years age group, and 23.85% of respondents were from (35-44) years age group. All the respondents had around 15 years of work experience.

The top management support (TMS) was measured using three variables, namely, TMS1: support to employees; TMS2: providing necessary resources; and TMS3: promoting AI as a strategic priority. Technology infrastructure & skilled workforce (TI) were measured using four variables, namely, TI1: high-quality data; TI2: IT infrastructure; TI3: multi-disciplinary teams; and TI4: data governance policy. The capability of AI (AICapa) was measured using three variables: AuCapa: automation capability, AnaCapa: analytical capability, and EmCapa: empathetic capability. Similarly, measurement scales for the compatibility of AI (CB) and complexity of AI (CX) were taken from the extant literature, consisting of three variables each.

They responded to measurement variables of each of the five constructs based on their perception using a 7-point Likert scale. They were providing responses to questions such as:

- Top management would provide necessary support to the employees for AI adoption.
- Our organization has high quality data.
- My organization thinks implementing AI is not simple.
- Our organization will use AI analytics for information driven decision making.
- AI will be compatible with our organization's work practices.
- My organization thinks implementing AI is not simple.

We tested convergent, divergent, and discriminant validity to check validity of our measurement scales.

The convergent and divergent validity of the measurement scales were found above the 0.7 threshold. The discriminant validity of the measurement scales was measured using the Heterotrait-monotrait ratio (HTMT). All values were below the 0.9 threshold.

A series of t-tests were performed between various pairs of measurement variables wherein one variable was taken from the Organization constructs, such as TMS1, TMS2, etc., and the other one was taken from the Technology constructs, such as AuCapa, AnaCapa, etc. A person from an organization considering adopting AI may feel that the organization will arrange the required resources, but implementing AI will be complex. On the other hand, a person from the IT industry may feel that implementing AI is not really complex, but the organization may lack resources. In order to avoid such a situation, we included personnel from the organization who were considering AI deployment. We hypothesized that most of the respondents would be assured of the readiness of their organizations and would be skeptical about AI technology. The results of comparing pairs using t-tests are shown in Table 4. The first variable belongs to the organization, and the second variable belongs to the Technology. There were 63 pairs, out of which 27 pairs showed a difference that was statistically significant. In other words, in more than half of pairs, there was no significant difference in perception of technology and organization variables. We are not able to draw a conclusion that the technology factors are more important than the organization factors.

4 DISCUSSION AND CONCLUSION

We started with the assumption that AI as a technology is probably challenging to adopt, and our respondents will help us confirm our hypothesis. The extant literature has utilized the full Technology-Organization-Environment (T-O-E) framework to assess factors that influence the readiness of an organization for AI. The T-O-E framework has been extended to the T-O-E-H framework by including Human factors [1, 9, 20]. The overall picture indicates that Technology and Organization factors dominate compared to environment and human factors.

We zoomed into the Technology and Organization factors and discovered that both are important. Our findings are in conformance with earlier findings. We discovered that capability of AI, compatibility of AI, and complexity of AI are important in affecting AI adoption. The top management support, technology infrastructure & skilled workforce in an organization are important in affecting AI adoption. If an organization has a strong IT ecosystem, the top management is positively influenced to adopt AI. These findings are also in synchronization with extant literature.

Top management support to the employees (TMS1) is significantly different from 4 variables, High-quality data (TI1) is significantly different from 5 variables, and Data ecosystem (TI4) is significantly different from 7 variables across technology subfactors. Collectively, 27 variable pairs between Organization-Technology are significantly different from each other out of a total of 63 pairs. This indicates that both technology and organizations are two realities that are significantly different from each other. The implication is that an organization should comprehend the various aspects of AI before deploying it. The focus should be on identifying

relevant capabilities of AI systems and assessing their perceived compatibility and complexity.

An organization should evaluate AI compatibility, complexity issues, and IT infrastructure. The organization should explore the technology and its appropriateness before committing to it. In brief, an organization needs to comprehend technology.

Adopting AI technology requires certain characteristics in the organization. For example, Top Management Support, the presence of a data ecosystem, and IT infrastructure are required to undertake an AI project. Thus, to adopt AI, an organization needs to pay attention to both the technology and the organizational factors.

Table 4 Result of t-tests of TMS and TI paired with CB, CX and AICapa

Organization	Technology	Two-tailed *p*-value	Conclusion
TMS1	CB1	0.0083	Statistically significant
	CX3	0.0117	Statistically significant
	AuCapa	0.0474	Statistically significant
	AnaCapa	0.0001	Very statistically significant
TMS2	CB1	0.0212	Statistically significant
	CX3	0.0281	Statistically significant
	AnaCapa	0.0001	Very statistically significant
TMS3	CB2	0.0132	Statistically significant
	AnaCapa	0.0001	Very statistically significant
	EmCapa	0.0402	Statistically significant
TI1	CB2	0.0001	Very statistically significant
	CB3	0.0303	Statistically significant
	CX1	0.004	Statistically significant
	CX2	0.0143	Statistically significant
	EmCapa	0.0001	Very statistically significant
TI2	CB2	0.0003	Very statistically significant
	CX1	0.039	Statistically significant
	AnaCapa	0.0099	Statistically significant
	EmCapa	0.0014	Statistically significant
TI3	AnaCapa	0.0001	Very statistically significant
TI4	CB1	0.0001	Very statistically significant
	CB3	0.0019	Statistically significant
	CX1	0.0254	Statistically significant
	CX2	0.007	Statistically significant
	CX3	0.0001	Very statistically significant
	AuCapa	0.0001	Very statistically significant
	AnaCapa	0.0001	Very statistically significant

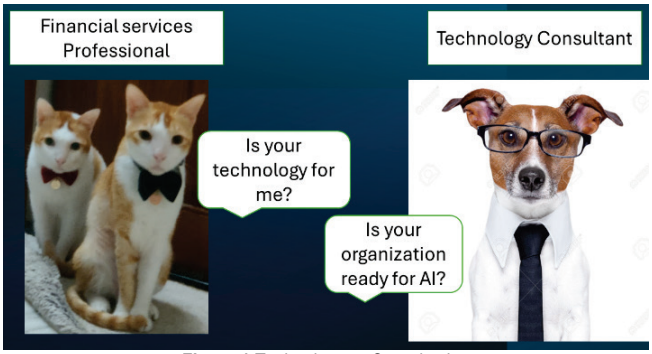


Figure 1 Technology or Organization

We discussed our findings with AI consultants informally. One of the consultants confirmed that if the top management doesn't show commitment and support, we don't undertake their AI project. In addition, we also check their existing IT infrastructure. If an organization doesn't have an IT infrastructure, we conclude that there will be a lack of understanding of AI among management and employees. This is a red flag, and we hesitate to work with such an organization. We summarize our findings in Fig. 1. Organization and IT consultants should both evaluate each other before embarking on AI deployment. When implementing new software applications in a company, it is always a problem whether to adapt the software to the company's organization or to adapt the company's organization to the software requirements. Namely, employees and managers always face covert or open resistance to change because they fear their current and future positions in the company's hierarchy.

5 REFERENCES

[1] Alsheibani, S., Messom, C. & Cheung, Y. (2020). Re-thinking the competitive landscape of artificial intelligence. *Proceedings of the 53rd Hawaii International Conference on System Sciences*. https://doi.org/10.24251/HICSS.2020.718

[2] Davenport, T. H. & Ronanki, R. (2018). Artificial intelligence for the real world. *Harvard business review, 96*(1), 108-116.

[3] De Bruyn, A., Viswanathan, V., Beh, Y. S., Brock, J. K. U. & Von Wangenheim, F. (2020). Artificial intelligence and marketing: Pitfalls and opportunities. *Journal of Interactive Marketing, 51*(1), 91-105. https://doi.org/10.1016/j.intmar.2020.04.007

[4] Dora, M., Kumar, A., Mangla, S. K., Pant, A. & Kamal, M. M. (2022). Critical success factors influencing artificial intelligence adoption in food supply chains. *International Journal of Production Research, 60*(14), 4621-4640. https://doi.org/10.1080/00207543.2021.1959665

[5] Dutt, R. (2020). The impact of artificial intelligence on healthcare insurances. In *Artificial intelligence in healthcare* (pp. 271-293). Academic Press. https://doi.org/10.1016/B978-0-12-818438-7.00011-3

[6] Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E. L., Jeyaraj, A., Kar, A. K., ... & Wright, R. (2023). So what if ChatGPT wrote it? Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. *International Journal of Information Management, 71*, 102642. https://doi.org/10.1016/j.ijinfomgt.2023.102642

[7] Eling, M., Nuessle, D. & Staubli, J. (2021). The impact of artificial intelligence along the insurance value chain and on the insurability of risks. *The Geneva Papers on Risk and Insurance Issues and Practice*, 1-37. https://doi.org/10.1057/s41288-020-00201-7

[8] EY India (2022). EY NASSCOM AI Adoption Index: is AI still incubating in your organization or driving innovation? https://www.ey.com/en_in/ai/ey-nasscom-ai-adoption-index-is-ai-stillincubating-in-your-organization-or-driving-innovation (accessed on January 30, 2024).

[9] Gupta, S., Ghardallou, W., Pandey, D. K. & Sahu, G. P. (2022). Artificial intelligence adoption in the insurance industry: Evidence using the technology–organization environment framework. *Research in International Business and Finance, 63*, 101757. https://doi.org/10.1016/j.ribaf.2022.101757

[10] Gursoy, D., Chi, O. H., Lu, L. & Nunkoo, R. (2019). Consumers acceptance of artificially intelligent (AI) device use in service delivery. *International Journal of Information Mgmt*. https://doi.org/10.1016/j.ijinfomgt.2019.03.008

[11] Horani, O. M., Al-Adwan, A. S., Yaseen, H., Hmoud, H., Al-Rahmi, W. M. & Alkhalifah, A. (2023). The critical determinants impacting artificial intelligence adoption at the organizational level. *Information Development*, 02666669231166889. https://doi.org/10.1177/02666669231166889

[12] Huang, M. H. & Rust, R. T. (2021). A strategic framework for artificial intelligence in marketing. *Journal of the Academy of Marketing Science, 49*, 30-50. https://doi.org/10.1007/s11747-020-00749-9

[13] Jöhnk, J., Weißert, M. & Wyrtki, K. (2021). Ready or not, AI comes—an interview study of organizational AI readiness factors. *Business & Information Systems Engineering, 63*. https://doi.org/10.1007/s12599-020-00676-7

[14] Kaplan, A. & Haenlein, M. (2019). Siri, Siri, in my hand: Who's the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence. *Business horizons, 62*(1), 15-25. https://doi.org/10.1016/j.bushor.2018.08.004

[15] Kolbjørnsrud, V., Amico, R. & Thomas, R. J. (2016). How artificial intelligence will redefine management. *Harvard Business Review, 2*(1), 3-10.

[16] Kruse, L., Wunderlich, N. & Beck, R. (2019). Artificial intelligence for the financial services industry: What challenges organizations to succeed. https://doi.org/10.24251/HICSS.2019.770

[17] Kumar, A., Mani, V., Jain, V., Gupta, H. & Venkatesh, V. G. (2023). Managing healthcare supply chain through artificial intelligence (AI): A study of critical success factors. *Computers & Industrial Engineering, 175*, 108815. https://doi.org/10.1016/j.cie.2022.108815

[18] Lamberton, C., Brigo, D. & Hoy, D. (2017). Impact of Robotics, RPA and AI on the insurance industry: challenges and opportunities. *Journal of Financial Perspectives, 4*(1).

[19] McKinsey (2017). How top tech trends will transform insurance. https://www.mckinsey.com/industries/financial-services/our-insights/how-top-tech-trends-willtransform-insurance (accessed on January 30, 2024)

[20] Pathak, A. & Bansal, V. (2024). Factors Influencing the Readiness for Artificial Intelligence Adoption in Indian Insurance Organizations. In *International Working Conference on Transfer and Diffusion of IT* (pp. 43-55). Springer, Cham. https://doi.org/10.1007/978-3-031-50192-0_5

[21] Peres, R. S., Jia, X., Lee, J., Sun, K., Colombo, A. W. & Barata, J. (2020). Industrial artificial intelligence in industry 4.0-systematic review, challenges and outlook. *IEEE Access, 8*. https://doi.org/10.1109/ACCESS.2020.3042874

[22] Rogers, E. M. (2003). *Diffusion of Innovations*, 5th ed., Free Press, New York, NY.

[23] Sepasgozar, S. M. Loosemore, M., & Davis, S. R. (2016). Conceptualising information and equipment technology adoption in construction: A critical review of existing research. *Engineering, Construction and Architectural Management, 23*(2), 158-176. https://doi.org/10.1108/ECAM-05-2015-0083

[24] Solaimani, S. & Swaak, L. (2023). Critical Success Factors in a multi-stage adoption of Artificial Intelligence: A Necessary Condition Analysis. *Journal of Engineering and Technology Management, 69*, 101760. https://doi.org/10.1016/j.jengtecman.2023.101760

[25] Tornatzky, L. G., Fleischer, M. & Chakrabarti, A. K. (1990). *Processes of technological innovation*. Lexington books.

[26] VanGiffen, B. & Ludwig, H. (2023). How Siemens Democratized Artificial Intelligence. *MIS Quarterly Executive, 22*(1), 3.

[27] Zavadskas, E. K., Stević, Ž., Tanackov, I. & Prentkovskis, O. (2018). A novel multicriteria approach–rough step-wise weight assessment ratio analysis method (R-SWARA) and its application in logistics. *Studies in Informatics and Control*. https://doi.org/10.24846/v27i1y201810

Authors' contacts:

Aman Pathak, Research Scholar
(Corresponding author)
Department of Management Sciences (DoMS),
IIT Kanpur, G66M+W5J, Kalyanpur, Kanpur, Uttar Pradesh 208016, India
amanp20@iitk.ac.in

Veena Bansal, Associate Professor
Department of Management Sciences (DoMS),
IIT Kanpur, G66M+W5J, Kalyanpur, Kanpur, Uttar Pradesh 208016, India
veena@iitk.ac.in

ISSN 1846-6168 (Print), ISSN 1848-5588 (Online)
https://doi.org/10.31803/tg-20240516182956

Preliminary communication
Received: 2024-05-16, Accepted: 2024-10-19

Experimental Determination of the Focal Distance of Cumulative Pyrotechnic Device in Launch Vehicle Separation System

Yevhen Boliubash

Abstract: Pyrotechnic devices are widely used in modern aerospace engineering. One of the most effective types of pyrotechnic devices for performing obstacle separation operations in aerospace engineering is linear shaped charges. The penetration efficiency of the cumulative jet into a metal plate varies depending on the distance to the obstacle, i.e., on the focal distance. The focal distance is a key design parameter, therefore its study is extremely important for the development of effective systems. The article presents a methodology for conducting experimental research and the results of experimental research to determine the optimal focal distance for setting a linear shaped charge with a semicylindrical shaped liner with a diameter of 5 mm on an obstacle made of aluminum alloy 2219. The method for determining the optimal focal distance, the obtained data of the optimal and recommended focal distance are described. The research results provide valuable data for optimizing the mass characteristics of the structure by applying the optimal focal distance parameter for setting the linear shaped charge. This will help in the development of more efficient and reliable separation systems in aerospace engineering. The materials of the article will be useful for scientific and technical workers working on separation processes.

Keywords: aerospace engineering; experimental research; focal distance; linear shaped charge; material strength; separation system

1 INTRODUCTION

Pyrotechnic Devices (PDs) are widely used in modern aerospace engineering. They are employed for separating launch vehicle stages, detaching spent structural elements, terminating flight by shutting down the engine, and destroying the structure in case of emergency situations [1-6].

Efforts to reduce launch vehicle (LV) mass are becoming particularly important for ensuring competitiveness in the face of ever-increasing competition in the space industry. Therefore, research aimed at developing more effective solutions plays a crucial role in this direction.

Linear shaped charges (LSCs) are one of the most effective types of PDs for performing separation operations in aerospace engineering [7]. They are characterized by high energy density, small size and weight, and simple manufacturing [3, 8]. LSCs provide efficient separation of rocket structural elements by cutting the obstacle with a cumulative jet (CJ). Currently, LSCs are used in the separation systems (SS) of the following launch vehicles: Ariane-5 (France), PSLV (India), Atlas-5/551 (USA), Minuteman III (USA), Delta IV (USA) [2-5]. As a result of their operation, the separating bodies are given relative linear and angular velocities. Therefore, in some cases, LSC-based PDs can be used directly for separating passive structural elements [3], without the use of springs, pyromechanical pushers, or impulse solid-propellant engines. This feature allows for further reduction of the mass of the launch vehicle separation system.

LSCs are elongated explosive charges (ECs) enclosed in a thin metal casing with a cumulative part (CP) in the form of an inverted "V" (wedge shape, common in the USA), or "U" (semicylindrical shape, common in Ukraine), which enhances the effect of the explosion by concentrating it along the axis of the charge in a given direction (Fig. 1) [8, 9].

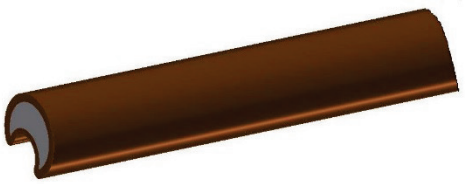


Figure 1 General view of a LSC with a semicylindrical cumulative part

The cumulative effect is achieved by using a charge with a shaped cavity directed towards the obstacle [10-12]. The schematic representation of a structural element of a launch vehicle SS with an installed LSC is shown in Fig. 2.

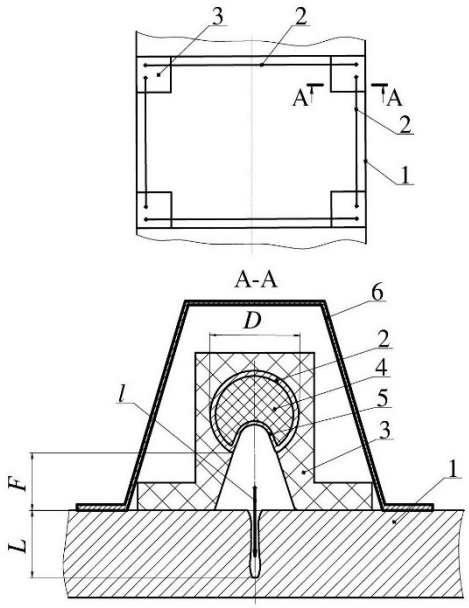


Figure 2 Schematic of a structural element of a launch vehicle SS with an installed LSC: 1 - LV body section (obstacle); 2 - LSC; 3 - LSC mounting bracket; 4 - LSC explosive, 5 - LSC cumulative part; 6 - Detonation product protection; *l* - cumulative jet; *F* - focal distance; *L* - penetration depth of the CJ

Focal distance is the distance between the obstacle and the LSC at which the cumulative jet reaches its maximum depth of penetration into the obstacle material. It is an important design parameter that affects the penetration depth of the CJ and, consequently, the separation efficiency.

Thus, determining the required focal distance is a key task in calculating the effect of cumulative charges and developing pyrotechnic separation systems based on LSCs [13-16].

Upon detonation of the explosive material (EM) in the LSC, a cumulative jet is formed. Individual elements of this jet will move normally to the surface of the cumulative part (CP) only in the immediate vicinity of the CP itself. As the jet propagates further, it will straighten in accordance with the general laws of gas dynamics. At a certain distance from the base of the CP, the cumulative jet will experience its maximum compression. This distance is defined as the focal distance (F), see Fig. 3. Beyond the focal distance, the cumulative jet rapidly degenerates due to the radial dispersion of the detonation products compressed to high pressure [7].

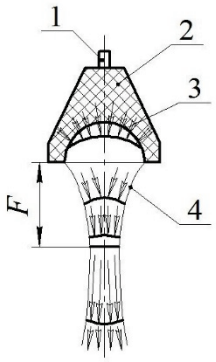


Figure 3 Schematic of the formation of a cumulative gas flow: 1 - detonator; 2 - explosive charge, 3 - semispherical cumulative part, 4 - cumulative gas flow

In general, the focal distance depends on the design of the LSC, the parameters of its cumulative part, the type and amount of explosive material (EM), the manufacturing accuracy of the LSC, and the characteristics of the obstacle, including its density. The focal distance increases with increasing EM charge power, increasing obstacle material density, and manufacturing accuracy of the cumulative charge [14]. At the optimal focal distance, maximum penetration of the CJ into the obstacle is ensured.

Optimal focal distance is the distance between the LSC and the obstacle that ensures maximum penetration of the cumulative jet into the obstacle material. The optimal focal distance can be determined experimentally or by modeling, taking into account various parameters such as the LSC design, obstacle material properties, and so on. The optimal focal distance allows for maximum separation efficiency of structural elements and, accordingly, increases the reliability and safety of rocket and space systems.

Recommended focal distance is the distance that is recommended for use in the design of separation systems. The recommended focal distance may differ from the optimal one depending on the specific conditions. Recommendations may be based on previous research, standards, empirical data, or design considerations. They may include safety factors, ease of use, design features of the SS, technological limitations in the manufacture and installation of the LSC, operating conditions of the LV, etc.

The nominal value of the distance between the LSC and the obstacle should be equal to the optimal focal distance, but this is not always achieved in practice; as a rule, it is close to it. The recommended focal distance, in contrast to the optimal one, takes into account the tolerance field, determined by the manufacturing accuracy, thermal expansion of the structure during flight, and other factors.

It is important to select the recommended focal distance of the LSC taking into account all these factors. This can be achieved through theoretical calculations, experimental research, and computer modeling.

Aluminum alloys are widely used in the manufacture of LVs due to their lightness and sufficient strength. In Ukraine, there is currently a transition to the use of aluminum alloy 2219 due to its characteristics and capabilities for improving the efficiency of rocket and space technology. Thus, the use of aluminum alloy grade 2219 contributes to a reduction in the thickness of the shell structure of the LV body compartment, which results in an increase in the volume of tanks and the weight of the payload. Therefore, there is a need to determine the focal distance of the LSC installation on obstacles made of this alloy.

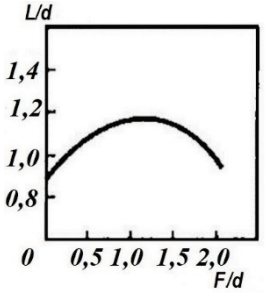


Figure 4 Dependence of the penetration depth of the CJ of a copper LSC with a semicylindrical CP, filled with hexogen, on the focal distance for a steel obstacle [7]: L – penetration depth of the CJ; *d* - diameter of the CP of the LSC; *F* – focal distance

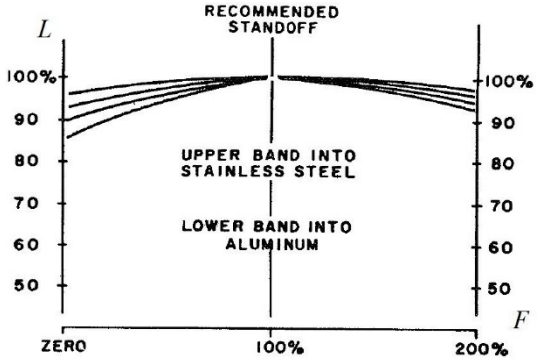


Figure 5 Penetration efficiency of the CJ of a lead LSC with a wedge-shaped CP, filled with hexogen, as a function of the focal distance [9]: *F* - Focal distance, in percent of the recommended value; *L* - Penetration depth of the CJ, in percent of the recommended focal distance; The upper graph shows the dependencies for a steel obstacle, the lower one - for an aluminum obstacle

Within the framework of this article, experimental studies are considered to determine the optimal and recommended focal distance of the LSC installation in the LV body compartment, made of aluminum alloy grade 2219, using a new method that allows you to reduce the number of tests and increase the accuracy of determining the focal distance.

The focal distance is a fundamental characteristic of the LSC [7, 10]. With a deviation of the focal distance from the optimal value, the penetration depth of the CJ decreases, in accordance with the results presented in Figs. 4 and 5.

According to [7, 12, 17, 18], , as a result of a series of studies of copper LSCs on various obstacles, the following ranges of recommended empirical ratios are known *F*/*D* (see Tab. 1).

Table 1 Recommended ranges of empirical ratios *F/D*

Ratio *F/D*	Material of the obstacle	Source
1,0...1,5	Steel	Petushkov, L. P. [7]
0,7...1,1	Steel of various grades and titanium alloys	Yefanov, V. V. [17]
0,8...1,6	Aluminum and magnesium alloys	Yefanov, V. V. [17]
1,1...1,3	Aluminum alloys grades AmG6, D16	Linnik, A. K. [18]

Analyzing the data presented in Figs. 4, 5, and Tab. 1, it can be concluded that the recommended focal distance for the installation of the LSC is within a certain range of values, while the maximum penetration depth corresponds to the most effective (optimal) focal distance.

Relevance of the research. Determining the optimal and recommended focal distance for the installation of the LSC is an important task in the design of LV separation systems using new materials.

The purpose of the research is to determine the optimal and recommended focal distance for the installation of an LSC of a certain diameter with a semicylindrical CP for the effective separation of an LV structural element, which is an obstacle made of aluminum alloy grade 2219.

Research Objectives:

1) Develop a methodology for the experimental determination of the focal distance;
2) Conduct an experimental study to determine the focal distance of the LSC for an obstacle made of aluminum alloy grade 2219;
3) Measure the depth of penetration of the CJ into the obstacle for the corresponding focal distance;
4) Analyze the research results.

2 PROBLEM SOLUTION

2.1 Methodology for the Experimental Determination of the Focal Distance

To determine the focal distance of a cumulative pyrotechnic device (LSC), a new methodology has been developed that allows:

- Reducing the number of tests by setting the LSC at an angle to the obstacle. This enables more efficient utilization of resources and time during the experimental process.
- Enhancing the accuracy of focal distance determination by obtaining CJ penetration depth results for the entire range of focal distances under investigation. This provides a more comprehensive understanding of the relationship between focal distance and CJ penetration depth.

The experimental study involved the following sequence of actions:

1) **Manufacturing of the LSC, an obstacle made of aluminum alloy grade 2219, and LSC mounting elements.** The obstacle was fabricated with a thickness exceeding the anticipated CJ penetration depth by several times. Standardized cumulative charges manufactured at the Scientific and Engineering Center "Explosive Processing of Materials" of the Paton Institute of Electric Welding (Ukraine) were employed. Brackets were also constructed to enable the LSC to be mounted on obstacle samples with high precision and control of the focal distance between the LSC and the obstacle.
2) **Installation of the LSC on the obstacle of the research object.** To identify the optimal and recommended focal distance, the LSC was installed at an angle to the obstacle. This facilitated a more comprehensive evaluation of the impact of focal distance on CJ penetration depth.
3) **Experiment.** Initiation of the LSC explosive charge. This initiated the explosive process that resulted in the formation of the CJ.
4) **Measurement of the CJ penetration depth into the obstacle.** After each experiment, the CJ penetration depth into the obstacle was measured for the corresponding focal distance values. This provided data on the relationship between focal distance and CJ penetration depth.
5) **Data analysis.** Following the experimental studies, the obtained data were systematized and analyzed using analytical methods. This enabled the identification of patterns and trends in the data, leading to the determination of the optimal and recommended focal distance for the LSC.

2.2 Research Object

The experimental setup for the study is shown in Fig. 6. The test object consists of the following components:

- Linear shaped charge (LSC) with a copper shell of 5 mm diameter filled with explosive material (hexogen 11,5 g/m);
- Metal plate made of aluminum alloy grade 2219 with a thickness of 20 mm. The alloy has the following characteristics: tensile strength σ_v = 436,4 MPa, yield strength $\sigma_{0,2}$ = 323,6 MPa, and relative elongation δ = 10%;
- Brackets for mounting the LSC and an electric detonator.

The shape of the LSC used in the experiment is standardized and corresponds to the manufacturer's

nomenclature. It is approximately cylindrical with a half-cylindrical cumulative part (see Fig. 7).

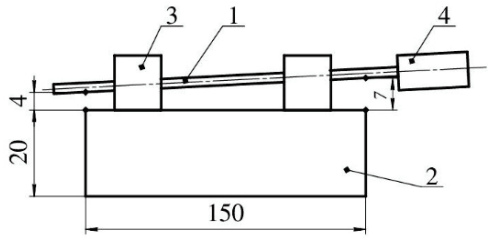


Figure 6 Scheme of the research: 1 - LSC; 2 - obstacle; 3 - LSC mounting brackets; 4 - electric detonator

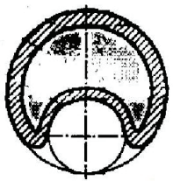


Figure 7 General view of the LSC with a semicylindrical cumulative part [7]

As shown in Fig. 6, the LSC (item 1) is mounted on the metal obstacle (item 2) using brackets (item 3) at a specific angle determined by the variation of focal distances within the investigated range. Based on the data presented in Tab. 1 and the analysis of the graphs in Figs. 4 and 5 - it was decided to conduct research in the range: F/D = 0,8 - 1,4, which corresponds to 4 to 7 mm for an LSC with a diameter of 5 mm. The LSC is detonated using an electric detonator (item 4) placed at the end of the LSC. The test specimen with the installed LSC is shown in Fig. 8.

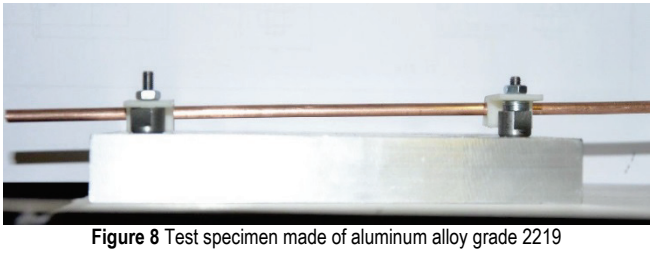


Figure 8 Test specimen made of aluminum alloy grade 2219

Figure 9 Cross-section of 2219 alloy specimens along the CS penetration line of the LSC: 1/1 – sample number in the numerator, test number in the denominator; 4 - designation of the minimum focal distance; 7 - designation of the maximum focal distance; a - trace of CJ penetration; b - trace of sample separation; c - trace of cutting with a disc cutter

To determine the penetration depth of the cumulative jet (CS), after testing, the specimens were cut from the reverse side of the CS action, along the LSC installation line, not reaching the CS penetration zone, and then one part of the specimen was separated from the other. After that, the depth of CS penetration into the obstacle was evaluated using high-resolution photography (Fig. 9).

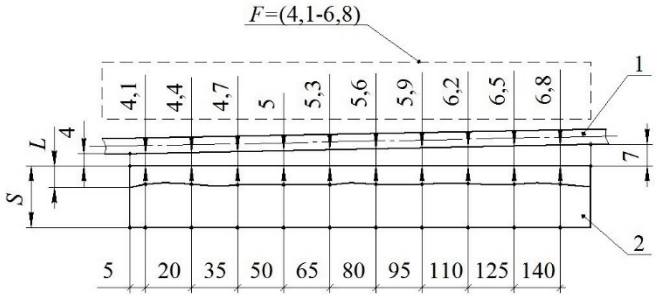


Figure 10 Measurement scheme: 1 - LSC; 2 - obstacle; S - obstacle thickness; L - measurement of CJ penetration depth; (5/20...140) - measurement points; F - range of investigated focal distances

To take measurements, photographs of the test specimen were imported into a CAD software package such as AutoCAD, where the required scale of the test specimen was obtained. After that, the CS penetration depth was measured according to the measurement scheme presented in Fig. 10. For each test, 10 measurements of the CS penetration depth (L) were made along the LSC obstacle cutting line at points from 5 to 140 mm (with a step of 15 mm).

2.3 Test Results

Three experiments were conducted. To improve the accuracy and reliability of the results, measurements were taken on both parts of each sample: Li.1 – corresponds to its first part, and Li.2 – to the second. The test results of the experiments to determine the dependence of the depth of penetration of the CJ on the focal distance of the installation of the LSC are presented in Tab. 2.

Table 2 Depth of CJ penetration of LSC into aluminum alloy grade 2219

Distance to the measurement point, mm	5	20	35	50	65	80	95	110	125	140
F	4,10	4,40	4,70	5,00	5,30	5,60	5,90	6,20	6,50	6,80
$L_{1.1}$	5,65	5,73	5,68	5,65	5,62	5,60	5,55	5,47	5,33	5,25
$L_{1.2}$	5,69	5,75	5,92	5,75	5,67	5,60	5,43	5,21	5,09	5,06
$L_{2.1}$	5,85	5,96	6,00	6,06	6,36	6,22	6,06	6,02	6,00	5,68
$L_{2.2}$	5,98	6,16	6,24	6,30	6,34	6,23	6,22	5,81	5,68	5,43
$L_{3.1}$	6,27	6,53	6,50	6,61	6,30	6,28	6,30	6,25	6,11	5,93
$L_{3.2}$	6,32	6,50	6,52	6,68	6,52	6,43	6,65	6,20	6,31	6,25
L_{min}	5,65	5,73	5,68	5,65	5,62	5,60	5,43	5,21	5,09	5,06
$L_{average}$	5,96	6,11	6,14	6,18	6,14	6,06	6,04	5,83	5,75	5,60
L_{max}	6,32	6,53	6,52	6,68	6,52	6,43	6,65	6,25	6,31	6,25

The results of the experiments in graphical form, showing the dependence of the focal distance on the depth of penetration of the CJ for the conducted experiments, are presented (see Figs. 11-12).

Fig. 11 shows the graphs of the dependence of the CJ penetration depth into the obstacle for each part of the sample ($L_{i.n}$). The penetration of the CJ demonstrates a uniform pattern in all conducted experiments. Minor deviations can be explained by the technological features of the LSC manufacturing process, including variations in the density of

the explosive substance in the casing and slight deviations in the geometry of the LSC cumulative part. This is confirmed by other studies that noted technological limitations could cause variations in the depth of CJ penetration up to 30% [14].

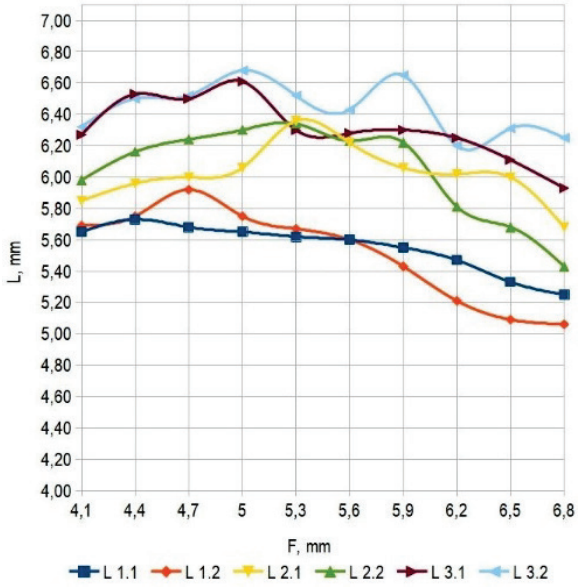


Figure 11 Dependence of focal distance on CJ penetration depth for conducted experiments: *F* – focal distance; *L* – penetration depth of CJ; $L_{i,n}$ – penetration depth of CJ, where *i* is the experiment number, *n* is the sample side number

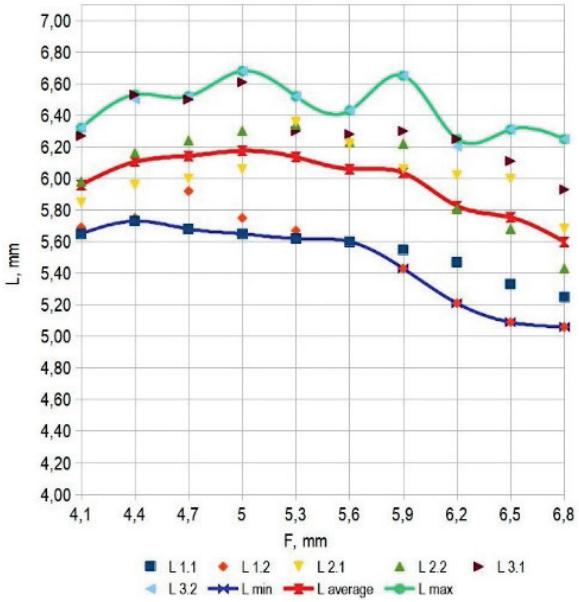


Figure 12 Dependence of focal distance on average, minimum and maximum depth of CJ penetration

For better data visualization, graphs were constructed to show the dependence of the average, minimum, and maximum CJ penetration depths on the focal distance, as shown in Fig. 12. The average line reflects the general trend, while the minimum and maximum values indicate the range of data distribution. The graphical representation of the average values is necessary to eliminate random deviations.

The graphs show that the penetration depth of the CJ initially increases with the increase in focal distance and then decreases. Thus, there is a certain dependence between the focal distance and the depth of penetration of the cumulative jet.

These graphs allow us to identify the correspondence to the empirical range of other researchers (see Tab. 1), but to accurately determine the optimal focal distance, it is necessary to establish a specific functional dependence.

2.4 Method for Establishing the Optimal Focal Distance

Graphically construct a polynomial function based on the average experimental values of CJ penetration into the obstacle. The construction process is best done in spreadsheets, such as LibreOffice, where a trend line should be constructed based on the average CJ penetration values, and the regression type should be selected as polynomial; the degree of the function should be chosen so that the approximation probability index R^2 is close to 1. An R^2 value that falls within the range (0,75 - 0,95) indicates a high level of correlation. After constructing the graph, a perpendicular should be dropped from the vertex of the function to the (x) axis, and the resulting value will be the optimal focal distance for the installation of the LSC.

The dependence of the depth of CJ penetration into the obstacle on the focal distance of the LSC installation, constructed based on the average values of CJ penetration depth, is shown in the form of a quadratic polynomial function in Fig. 13).

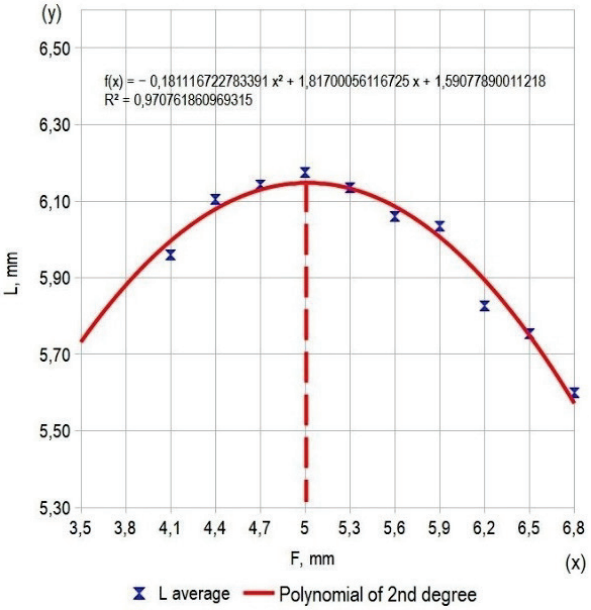


Figure 13 Dependence of the depth of CS penetration into the obstacle on the focal distance of LSC installation, represented by a 2-degree polynomial function (based on the average values of CJ penetration depth

The choice of a second-degree polynomial function for approximating the data is justified, as it provides sufficient approximation accuracy and allows for easy determination of

the function's extremum. The approximation probability index R^2 = 0,97, indicates a high level of correlation.

The conducted research has established that the dependence of CJ penetration depth on focal distance follows a quadratic polynomial law:

$$f(x) = -0,1811x^2 + 1,817x + 1,5908. \quad (1)$$

Penetration depth of a 5 mm diameter cumulative jet into an aluminum alloy 2219 obstacle at focal distances ranging from 4,1 to 6,8 mm varies from 5,06 to 6,68 mm. The average penetration depth of the cumulative jet is 5,98 mm.

As can be seen from Fig. 13, the maximum penetration depth of the cumulative jet into the obstacle is achieved at a focal distance of 5 mm.

Comparison of the obtained result with the data from other studies showed that it corresponds to the empirical coefficients for aluminum alloys [17], where the optimal focal distance varies within F/D = 0,8 - 1,6. In our experiment for aluminum alloy 2219, this value was:

$$\frac{F}{D} = \frac{5}{5} = 1. \quad (2)$$

This confirms that the proposed methodology and research results correspond to the literature data and can be used in further developments.

The optimal focal distance for a 5 mm diameter LSC is F = 5 mm and provides a cumulative jet penetration depth of L = 6,15 mm. At focal distances of F = 4,7 - 5,3 mm, the cumulative jet penetration depth is L = 6,13 mm. The efficiency of cumulative jet penetration into the obstacle for the selected range of focal distances is 99.7%.

The nominal value of the distance between the LSC and the obstacle corresponds to the optimal value and is equal to 5 mm. The recommended range of focal distances for a 5 mm diameter LSC with a semi-cylindrical cumulative jet is 5 ± 0,3 mm.

The obtained values of LSC penetration depth and optimal/recommended range of focal distances are recommended for use in the design of new products where the obstacle is aluminum alloy 2219.

Thus, a comprehensive experimental study was conducted of one of the main parameters that determine the efficiency and reliability of the separation process - the range of focal distances of the LSC installation, which made it possible to increase the efficiency, reliability, and performance of pyrotechnic devices of the separation system of rocket and space elements.

3 CONCLUSION

A new methodology for the experimental determination of the focal distance was developed, which allows reducing the number of tests by installing the LSC at an angle relative to the obstacle and increasing the accuracy of determining the focal distance by obtaining results of CJ penetration depth for the entire range of focal distances under study. This significantly reduces resource costs and the time for conducting experiments, which is an important advantage in resource-limited conditions.

An experimental study was conducted to determine the focal distance for a 5 mm diameter LSC with a semi-cylindrical cumulative part (filled with hexogen) on an obstacle made of aluminum alloy 2219.

The dependence of the CJ penetration depth into the obstacle on the focal distance of the LSC installation, which has a second-degree polynomial character, was obtained, allowing for optimization of the separation process by selecting the optimal focal distance.

As a result of the research, it was determined that with a change in focal distance from 4,1 mm to 6,8 mm, the penetration depth of the LSC cumulative jet into the obstacle material changes from 5,06 to 6,68 mm. The average penetration depth of the LSC cumulative jet is 6,11 mm.

A method for establishing the optimal focal distance is presented. The obtained value of the optimal (effective) focal distance of the LSC installation with a semi-cylindrical cumulative charge of 5 mm diameter is 5 mm; and the recommended focal distance is 5 ± 0,3 mm.

The research results provide valuable data for optimizing the mass characteristics of the structure by applying the optimal and recommended LSC installation parameter (focal distance). This will help in the development of more efficient and reliable separation systems in rocket and space technology. Optimizing the focal distance also helps to minimize mechanical loads on structural elements during the CJ penetration process into the obstacle.

The research results may be useful in the design of separation systems using new materials, such as composite materials or alloys with improved characteristics. Further experiments with new materials will expand the applicability of the proposed methodology.

Despite the successful results, this study has some limitations. It was conducted only for a specific type of LSC and obstacle material. For other types of LSCs or grades of obstacle materials, similar studies need to be conducted.

Acknowledgments

The author would like to express his sincere thanks to Yuzhnoye State Design Office for supporting this research.

4 REFERENCES

[1] Bolyubash, Ye. S. (2023). Umen'sheniye zony vozmozhnogo padeniya avariynykh RN za schet ispol'zovaniya sistem razdeleniya raketno-kosmicheskikh elementov, osnashchennykh pirotekhnicheskimi ustroystvami. *KHKHV Mezhdunarodnaya molodezhnaya nauchno-prakticheskaya konferentsiya "Chelovek i kosmos"*, p. 229. (in Ukrainian).

[2] Okhochyns′kyy, M. M. & Afanas′yev, K. A. (2013). Systemy podilu u raketniy tekhnitsi. CH. 2. Systemy viddilennya korysnykh navantazhen′ ta obtichnykiv: navchal′nyy posibnyk, p. 55. ISBN 978-5-85546-762-8 (in Russian).

[3] Kolesnikov, K. S., Kozlov, V. I. & Kokushkin, V. V. (1977). Dinamika razdeleniya stepeney letatel'nykh apparatov. M. Mashinostroyeniye, p. 224. (in Russian).
[4] Kobelev, V. N. & Milovanov, A. G. (2009). Sredstva vyvedeniya kosmicheskikh apparatov. Moskva: Izdatel'stvo "RESTART", p. 528. (in Russian).
[5] Flexible linear shaped charge (FLSC). Site: *Psemc emc*. URL https://psemc.com/products/flexible-linear-shaped-charge-flsc/ (Accessed on 05.07.2023).
[6] Kuzin, Ye. N. & Zagarskikh, V. I. (2012). Tvorcheskiy put' dlinoy pochti v polveka. *Nauchno-tekhnicheskiy zhurnal NPO im. S. A. Lavochkina. Vestnik. Vypusk, 4*(15), 157-163. (in Russian).
[7] Petushkov, V. G. (2005). Primeneniye vzryva v svarochnoy tekhnike, p. 753. ISBN 966-00-0286-6 (in Russian).
[8] Bolyubash, Ye. S. (2022). Kumulyatyvnyy efekt ta yoho vykorystannya dlya rozdilennya raketno-kosmichnykh elementiv za dopomohoyu pirotekhnichnykh prystroyiv. *Zbirnyk dopovidey Materialy XVII Naukovykh chytan' "Dniprovs'ka orbita – 2022"* (26-28 zhovtnya), 10-15. (in Ukrainian). https://doi.org/10.34079/2226-3055-2022-15-26-27-28-38
[9] James, L. S. (1984). Pyrotechnic shock: a literature survey of the linear shaped charge (LSC). *NASA TM - 82583*. Systems Dynamics Laboratory. Science and Engineering, p. 56.
[10] Baum, F. A., Stanyukovich, K. P. & Shekhter, B. I. (1959). Fizika vzryva, p. 800. (in Russian).
[11] Orlenko, L. P. (2004). Fizika vzryva. Izd. tret'ye, vol I, p. 828. (in Russian).
[12] Report of the Presidential Commission on the Space Shuttle Challenger Accident. Site: *NASA*. https://history.nasa.gov/rogersrep/v1ch9.htm (Accessed on 20.06.2023).
[13] Lim, S., Baldovi, P. & Rood, C. (2022). Jet Grouping of Linear-Shaped Charges and Penetration Performance. *Applied Sciences*, 12(24), p. 12768. https://doi.org/10.3390/app122412768
[14] Orlenko, L. P. (2004). Fizika vzryva. Izd. tret'ye, vol II, p. 644. (in Russian).
[15] Lavrent'yev, M. A. (1957). Kumulyativnyy zaryad i printsipy yego raboty. *UMN, 12*(4), 41-56. (in Russian).
[16] Bohanek V., Dobrilović M. & Škrlec V. (2014). The efficiency of linear shaped charges. *Tehnički vjesnik, 21*(3), 525-531. https://hrcak.srce.hr/123317
[17] Yefanov, V. V., Khartova, V. V. & Pichkhadze, K. M. (2014). Proyektirovaniye avtomaticheskikh kosmicheskikh apparatov dlya bazovykh nauchnykh issledovaniy. T. 3. Izd-vo MAI-PRINT, p. 464. ISBN 978-5-905646-04-1. ISBN 978-5-905646-07-2(т.3). ISBN 978-5-7035-2316-2. ISBN 978-5-7035-2319-3(т.3) (in Russian).
[18] Blyznychenko, V. V., Dzhur, Y. E. O., Krasnikova, R. D., Kuchma, L. D. & Lynnyk, A. K. (2007). Proektuvannya i konstruktsiya raket-nosiyiv, p. 504. ISBN978-966-551-218-9 (in Ukrainian).

Author's contacts:

Yevhen Boliubash, PhD student
Yuzhnoye State Design Office,
3, Kryvorizka Str., Dnipro, 49008, Ukraine
bolyubas@gmail.com

ISSN 1846-6168 (Print), ISSN 1848-5588 (Online)
https://doi.org/10.31803/tg-20220914012010

Subject review
Received: 2022-09-14, Accepted: 2024-01-02

Classification of Layers Using Artificial Neural Networks in the Province of Kurdistan (Iran)

Semko Arefpanah*, Alireza Sharafi, Alireza Gholamian

Abstract: Artificial Neural Networks (ANN) is a field that combines science, technology, and ancient and modern knowledge. It has demonstrated the ability to resolve complex engineering issues beyond the computational capacity of conventional approaches and classical mathematics. ANN has applications in various fields, including computer science, engineering science, biology and medical science, and communication science. Neural networks are particularly useful in civil engineering, particularly in geotechnical problems, where soil heterogeneity and nonlinear behavior significantly impact geotechnical phenomena. Researchers have employed ANN to address various geotechnical engineering issues, including behavioral modeling, due to their potent abilities on nonlinear and multivariate problems. In this study, information from boreholes was used to collect and classify data to describe soil strata. The outputs of the neural network showed general consistency when compared to experimental borehole data, indicating its effectiveness in estimating changes in the soil layer. This was achieved by first presenting the network with information from various boreholes.

Keywords: artificial intelligence; behavioral model; behavioral modeling; geotechnical engineering; liquefaction

1 INTRODUCTION

Over the past few decades, researchers have studied the mecanical properties of soil, but soil behavior remains poorly understood. Common methods, such as mathematical and experimental methods, use simplified to solve multi-variable geotechnical problems. However, these methods cannot describe complex soil behavior. Neural networks have been used as an alternative method to solve many geotechnical engineering problems, as they have the power to solve complex problems.

The process of drilling borefloles and performing experiments to identify soil layer structures is expensive and time-consuming. Increasing reliability in interpolating soil layer structures and properties in inter-borefloles improves the quality of geotechnical evaluation, decreases costs, and facilitates proper construction planning. Artificial neural networks (ANNs) use specific capabilities of data processing in the human brain, such as learning, data generalization, and parallel processing. Many researchers have studied ANN application in geotechnical engineering, using ANN models to predict complex soil, predict landslide risk, and model soil behavior in uniaxial strain conditions.

Analog-to-digital converters are artificial neural networks that have been successfully applied in various fields of geotechnical engineering, including pile capacity prediction, foundation settlement, soil properties and behavior, liquefaction, site characteristics, retaining structures, slope stability, and tunnel and underground opening design. Known applications include predicting drive pile capacity, liquefaction, and soil properties and behavior. However, other applications, such as structure settlement, should be carefully considered until more research is conducted.

Many areas where the feasibility of artificial neural networks has not been tested include predicting bearing capacity of shallow foundations, bored piles, and the design and drainage of sheet pile walls. Mathematical models often compensate for the lack of physical understanding by simplifying problems or incorporating assumptions into the model. Artificial neural networks, on the other hand, are data-only and can be trained on input-output data pairs to determine the model's structure and parameters.

However, artificial neural networks have limitations, including a lack of theory, the success of finding a good solution, and the ability to explain solutions. As new data becomes available, the ANN can always be updated by providing new training examples for better results.

2 BACKGROUND

Artificial neural networks (ANN) are a rapidly growing field of study, with applications in various fields such as economics, industrial engineering, automation, electronics, computer technology, medicine, intelligence, and object identification. They have also been successfully used in geographical and construction engineering due to advancements in computational science and computing power. ANNs are inspired by biological neurons, which study the brain's working system. They perform operations using the learning feature of the human brain, a unique feature compared to other computational methods. Classical administrative methods, which recognize linear relationships between dependent and independent variables, struggle with complex interplay between variables in geoscience. ANNs offer alternative solutions to these problems and provide additional tools to traditional statistical methods in geographic studies.

Artificial Neural Networks (ANNs) have been successfully applied to geotechnical engineering since the early 1990s due to their ability to model the complex behavior of geotechnical materials, which exhibit great variability [1]. Classical structural modeling based on elastic and plastic theories has limited ability to accurately model the behavior of geological materials due to formulation complexity, idealization of material behavior, and a plethora of experimental parameters [2].

ANNs have been proposed as reliable and practical alternatives for simulating the homogeneous and hysteretic behavior of geological materials [3]. Geotechnical properties

and soil behavior are controlled by factors such as mineralogy, which can be difficult to estimate with traditional statistical methods. ANNs have been used to estimate various soil properties, including pre-consolidation pressure, shear strength, stress history, expansion pressure, and lateral earth pressure [4].

ANNs have also been applied in liquefaction prediction, retaining walls, dams, explosions, mining, environmental geotechnical engineering, and rock mechanics [5]. They have been used in tunnels, underground openings, slope stability, landslides, and deep foundation pits. However, networks have limitations, such as the lack of theory to help them develop, the success of finding a good solution not always guaranteed, and the ability to explain the solution[6].

Iran, situated in the Alps-Himalayan seismic belt, faces high earthquake risk, causing significant losses. To mitigate this, measures learned from past earthquakes are implemented, including strengthening structures to resist dynamic influences and ensuring site safety to prevent geotechnical phenomena like liquefaction and landslides. Understanding soil layers below the surface is crucial for seismic response. Multiple boreholes are required to test soil layers, a time-consuming and costly process. This study uses artificial intelligence to predict soil layers and depths, and to verify network performance using common methods and available information. The output data from soil modeling is used in dynamic analysis, generating 1D, 2D, and 3D models based on soil stratification and Cartesian coordinates as input parameters. This approach reduces costs and allows for correct planning and execution of structural operations [7].

3 ARTIFICIAL NEURAL NETWORK

Smart systems, which have been around for 40 years, use models like behavior patterns, decision-making processes, brain structure, neural networks, and probabilistic reasoning to solve problems. They are intelligent within the scope of problem management rules, not performing all human brain functions. In geotechnical engineering, artificial intelligence is more efficient and economical than classical methods due to the availability of statistics. AI systems can predict phenomena by remembering their history and frequency, and if comprehensive information is used, they can predict behavior in other domains. This makes intelligent systems more efficient and economical than classical methods in handling statistical phenomena [8].

4 INTRODUCTION TO ARTIFICIAL NEURAL NETWORK PARAMETERS

Artificial neural networks (ANN) are models similar to human neural networks, consisting of nerve cell units connected by axons. The goal is to create a structure similar to the biological structure of the human brain, capable of learning and decision-making. The goal is to train the model by introducing the performance history and frequency of the dynamic system, store the system's performance in the model's memory, and use it for situations the model has not encountered before. Artificial intelligence is a vast field with roots in philosophy of linguistics, mathematics, psychology, and physiology [9].

This paper aims to link artificial intelligence with soil mechanics, specifically soil modeling through a powerful tool called artificial intelligence. The mathematical basis is considered and translated into an understandable algorithm by software, and soil modeling is performed in the software environment. The brain is well-prepared for parallel data processing, as it can hold knowledge from different sources in the brain and process it simultaneously.

Neural network computational methods estimate the strategic principles underlying brain processes to answer these questions and use them in computer systems. The brain is made up of a large number of interconnected units, and the inherently parallel structure of neural network systems makes them suitable for parallel machines, with further advantages in terms of speed and reliability.

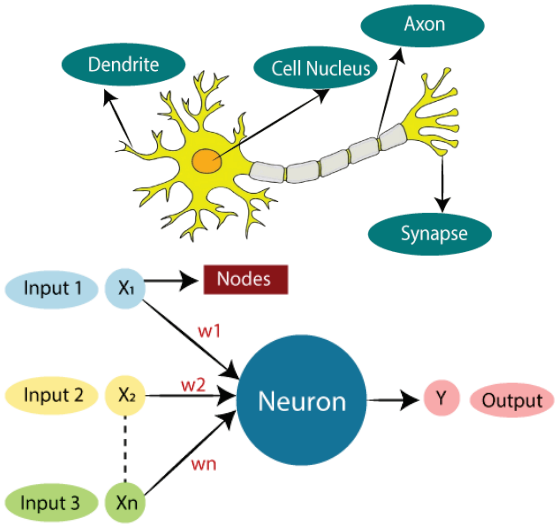


Figure 1 An illustration of the biological and artificial brain networks

Table 1 Artificial neural networks and biological neural networks' relationship

Biological Neural Network	Artificial Neural Network
Dendrites	Inputs
Cell nucleus	Nodes
Synapse	Weights
Axon	Output

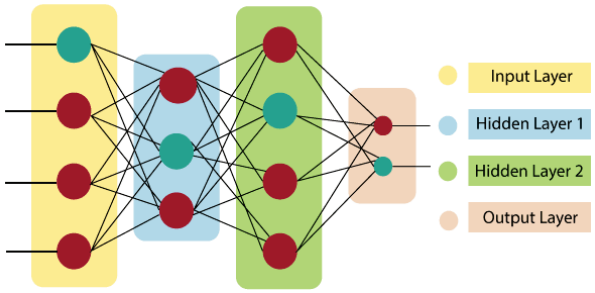


Figure 2 The three main layers of an artificial neural network

An artificial neural network (ANN) is a network of neurons, similar to biological ones in the human brain. These neurons are arranged in layers, with the strength of connection between adjacent layers represented by a

'connection strength' or 'weight'. ANNs typically consist of three layers: input, hidden, and output. In a feed-forward network, weighted connections feed activations only in the forward direction, while in a recurrent network, additional weighted connections feed previous activations back into the network. The training of the weight matrix is an essential step in developing an ANN model, where the weights are randomly initialized and updated using a specific training mechanism. Evaluation standards: Using R^2 and *RMSE* between the measured and predicted values, the accuracy of the regression equations for the development of PTFs was assessed and expressed as:

$$RMSE = \sqrt{\frac{1}{N}\sum_{k=1}^{n}\left(Z_s - Z_0\right)^2}. \quad (1)$$

The values are Z_s (observed), Z_0 (predicted), and n (number of samples).

The input layer accepts various input formats from the programmer, while the hidden layer performs calculations to identify hidden features and patterns. The output layer transforms the input through a series of transformations, resulting in the final output. The artificial neural network calculates the weighted sum of inputs, including bias, using a transfer function. This weighted total is passed to an activation function for output. Activation functions determine whether nodes should fire, with only those fired reaching the output layer. Different activation functions are available for different tasks.

$$\sum_{i=1}^{n} W_i \times X_i + b. \quad (2)$$

Artificial neural networks function as weighted directed graphs, with artificial neurons forming nodes and their outputs and inputs being represented as directed edges with weights. They receive input signals in the form of patterns and vectors, which are mathematically assigned for every n inputs. Artificial neural networks use weights to solve specific problems, representing the strength of interconnection between neurons. These weights are summarized in a computing unit. If the weighted sum is zero, bias is added to scale the output. Bias has the same input and weight as 1, with a total of weighted inputs ranging from 0 to positive infinity. To maintain the desired response, a maximum value is benchmarked, and the total weighted inputs are passed through an activation function. Activation functions can be linear or non-linear, with commonly used sets being Binary, linear, and Tan hyperbolic sigmoidal activation functions.

5 THE MAIN PARTS OF THE NEURAL NETWORK

The main parts of the neural network are:

- Structure or Architecture
- Training mechanism or learning algorithm
- Test mechanism or recall mechanism.

5.1 Structure or Architecture

When working with any neural network, its architecture or structure is what you see initially, and it depicts how the neural network is physically expressed. The appearance and internal communication of this intelligent system are determined by the neural network's topology. This part of a neural network serves as a practical way for it to store its data. More specifically, this implicitly intelligent system stores information about its mapping in its structure. Moreover, the neural network pulls the necessary data from the structure to respond to the user's query [10].

Previous studies have used the Artificial Neural Network (ANN) method to assess soil properties and classification. Ref. [11] classified soil using LL, PI, and dust content, resulting in heavy, light, hard, medium, small, and light dust soils. Ref. [12] predicted high moisture content and maximum dry thickness values using various methods, including bread, dust, golf, LL, and PL, with 126 samples. The study compared Simple-Multiple Analysis and ANN methods, revealing R^2 values ranging from 0.77-0.78 for multiple linear regression analysis, 0.74-0.82 for simple linear regression analysis, and 0.67-0.89 for ANN analysis. Refs. [13] and [14] conducted soil classification studies using Artificial Neural Networks (ANN) techniques. Hassannejad et al. found the Levenberg-Marquardt algorithm as the best for soil classification, while Tenpe and Kaur calculated OMC and MDD from LL, PL, and sieve analysis values using ANN techniques. The results showed R^2 values for OMC values were 0.85 in training, 0.76 in testing, and 0.95 in simulation. Ref. [15] used 280 data from previous studies to estimate the PI, MDD, and OMC values using ANN. The study used LL, PL, and Lime content as inputs. The R^2 value for PI was 0.91, MDD and OMC were 0.83. Reale et al. (2018) utilized CPT values to estimate soil classification using Artificial Neural Networks (ANN) on a 216-dataset dataset. They developed two ANN networks for estimating fines content (FC) and predicted for both LL and PI. The correlations showed R^2 values of 0.79 for FC, 0.85 for LL, and 0.78 for PI.

5.2 Training Mechanism or Learning Algorithm

By looking at instances from the training phase, which is an estimate of the actual map of the system under consideration, the training mechanism creates a map of the neural network. In neural network nomenclature, these samples are referred to as training samples or learning patterns, and the training mechanism stores the information pertaining to this mapping that was acquired from the samples during the training phase in the neural network's structure [16].

5.3 Test Mechanism or Neural Network Call Mechanism

The testing mechanism or neural network recall mechanism is a crucial component of an intelligent system that uses information extracted from the training mechanism to answer user questions. The neural network structure acts

as a memory in this system, and the testing mechanism retrieves this information. The efficiency of the neural network is determined during the testing phase or recall phase. The neural network receives expected samples representing the user's problem and outputs these, without any output to estimate. The recall mechanism uses the mapping obtained during the training phase to estimate the output of the test samples. The closer the network's answer to the test sample is to the real answer, the closer the graph formed by the neural network is to the problem system graph. Therefore, the efficiency of a neural network to solve a problem is deterministic [17].

6 RESEARCH ASSUMPTIONS (RESEARCH HYPOTHESES)

Research assumptions (research hypotheses) are:

- Soil stratification can be determined with acceptable accuracy using an artificial neural network.
- The soil of the plan is considered in layers.

7 SOIL MODELING

Artificial intelligence (AI) has become increasingly popular in various fields, with software like MATLAB and STATISTICA being particularly useful. The selection and use of boreholes is crucial for determining the exact location of the project area, allowing for accurate coordinates and utilization of borehole information. The choice of input parameters for a neural network to predict soil stratification, such as soil type and depth, is directly related to the coordinates and information of the selected borehole. Drill holes are selected based on the information in the identification table, ensuring no limit to the required information.

One of the most important steps in developing an artificial neural network is choosing the data used to train the network. After selecting the input and output variables and ensuring the accuracy and validity of the soil stratification data, the next step is selecting data for training and choosing from the available data. Control and testing the network during each step is essential to ensure the network's performance in predicting situations outside the training data.

In drilling modeling, a point is selected as the coordinate origin in the Cartesian coordinate system as the input parameter, and the classification information of the drilling is used as the input data. The chosen software chooses the most suitable network based on the input parameters, resulting in more network errors than the software raises.

8 HISTORY OF TECHNICAL DESCRIPTION IN SOIL DEFINITIONS

Many soil properties such as color, texture, density, moisture, or structure can be observed directly in the field. Most soil profiles show a series of rather well-separated layers. When this is due to soil formation or sexual reproduction, and not to changes in the parent material, it is said to be the horizon. These terms may be considered roughly synonymous herein. Often used to describe the individual soils on which soil classification and mapping are based. They are usually taken from soil profiles (Tab. 1), by digging sampling pits or coring exposed vertical transects using a soil auger.

Table 2 Common terms used in soil descriptions

Term	Description
PED	A single individual naturally occurring soil aggregate such as a granule or prism; also, a soil unit beneath the horizon level.
HORIZON	Relatively uniform material that extends laterally, continuously, or discontinuously throughout the pedo-unit; runs approximately parallel to the surface of the ground and differs from the related horizons in many chemicals, physical and biological properties.
PROFILE	A vertical section through the soil from the surface to the relatively unaltered parent material.
PEDON	The smallest classifiable soil unit, intended to be large enough to contain the root system of an average plant
PEDOUNIT	A selected column of soil containing sufficient material in each horizon for adequate field and laboratory characterization.
SOIL BODY	A representative specimen of a taxon[1] in place on the natural terrain.
POLYPEDON	A soiled area composed of many similar pedons that are all given the same soil name
PEDOTOPE	The space is occupied by a polyhedron.
SOIL SCAPE	The assemblage of soil bodies on a land surface in a particular landscape.

9 SOIL HORIZON DESIGNATIONS

With the work of Dokuchaev and later Sibirtsev on Chernozem soils in Ukraine towards the end of the nineteenth century, the scientific study of soil horizons was initiated (Buol et al., 1973; p. 174). Dokuchaev utilized the initial letters of the Latin alphabet to denote specific regions within soil and geological profiles. The letters A, B, and C were initially just designations, but over the ensuing decades, they came to have distinct meanings [18]:

1) Compared to the layers below, the A horizon (topsoil) often has a higher organic matter content and a darker hue. Eluviation, or the leaching of clay to the layers below, is another characteristic of it.
2) The subsoil, or B horizon, is often lighter in color than the topsoil and frequently exhibits clay illuviation or deposit, typically from the horizon or horizons above.
3) The parent material, or substrate, known as the "raw" soil (horizon C) is where the A and B horizons have developed.
4) A large portion of the later work on soil categorization was built around these concepts. The system changed as profile descriptions became more detailed. For example, new "master horizons" were developed (O for organic horizons, R for bedrock, and G for gley).

The International Society of Soil Science (ISSS) published a "Proposal for a uniform system of soil horizon designations" that forms the basis for most of today's soil classification systems. Subdivisions of these major horizons were introduced, first by the use of numeral subscripts, and later by subscript letters (e.g. b for buried or h for hummus). The emphasis of the description shifted from the (assumed)

processes which had formed the soil to the actual morphology of the profile [19].

10 SOIL CLASSIFICATION SYSTEMS

As an alternative to the soil hierarchy classification, the first soil classification system was created in Russia (Buol et al., 1973; Duchaufour, 1982). The parent material of the central high plains of Russia is rather consistent, loess-like, and exhibits a steady increase in temperature from north to south as well as a gradual increase in precipitation from east to west. Three types of soil classification systems have been developed: non-zonal, intrazonal, and zonal. Interrazonal soils are influenced locally by things like parent materials; zonal soils are immature soils like those found on recently eroded slopes; and zonal soils are soils generated mostly by vegetation and climate. Two issues need to be addressed by soil classification:

1) It should categorize higher units by assembling the primary soil types found worldwide based on the prevalence of their basic characteristics, so offering a structure that serves as the foundation for soil research.
2) The tool ought to facilitate the creation of expansive maps for pragmatic uses, like farming, where precise details of marginal significance are frequently needed for unit designation and definition.

10.1 German Bodensystematik

Developed the German soil categorization system known as the "Bodensystematik", which is a genetic system based on the idea of soil kinds (German: Typen), or what he refers to as the "characteristic stages of soil development" (op. cit.).The distinct features of such vistas as well as their distinctive arrangement characterize types. Higher categories 5 are divided into two suborders (G.: Klassen) and an order (G.: Abteilungen). The direction and extent of infiltration—the movement of materials through the soil by water—determine the top layer division primarily. Soils classified as terrestrial, semi-terrestrial (groundwater and flood), and submerged (underwater) are the basic categories. Soils are categorized at the sub-level based on comparable horizon configurations. Based on qualitative characteristics like the presence of calcium carbonate, types can be further classified into subtypes (G.: Subtype). Specifying transitions or changes between types is a crucial application of subtypes. The following primary horizons are identified by BS: F, H, L, P, T, S, M, E, and Y [12]. Only three appear in the following areas: M (colluvium, from L.: migrate), S (stagnant water-affected deposit, G.: Stauwasser), and H (peat, from humus). Subscript letters in lowercase are used to indicate horizontal scale subdivisions; genetic trait-related letters come after the primary symbol, while all other letters come before it.The German method allows for some flexibility with regard to slope; for example, strata that demonstrate the effect of many soil formation processes can be described by combining up to three primary symbols. Likewise, combinations of up to three different kinds of symbols can be used to define variations. Moreover, soil surveyors have some leeway to interpret soil profiles genetically because the BS classification system considers the complete profile. The most significant distinction is seen in the USDA's "soil taxonomy", whose "diagnostic" scope is mostly determined by physical or chemical measures [20].

10.2 Soil Classification

Fine-grained soils are defined as those with a grain size of less than 0.075 mm, and more than 50% of their dry weight should be finer than 0.075 mm. These soils are a mixture of clay and silt grains, with the clay fraction being the size limit between the two. The plasticity properties of silt and clay are better separators than particle size [21] initially defined six 'Limits of consistency' to classify fine-grained soils, but in present engineering applications, only three of these limits are used: liquid (LL), plastic (PL), and shrinkage (SL). These limits are used to understand soil mechanics and physical properties, such as swelling and shrinkage potentials, shear strength, and compressibility.

Researchers have used these limits to investigate the fundamental properties of soils, with studies showing significant relationships between clay rate, LL, PL, and PI values. Studies in Canada and Nigeria have reported significant relationships with clay rate, LL, PL, and PI values. Additionally, clay content, montmorillonite ratio, and organic matter ratio have a weighty effect on LL and PI. In conclusion, fine-grained soils are classified using Atterberg limits, which can be used to understand soil mechanics and physical properties. Since early 1990, Artificial Neural Networks (ANN) have been widely used in civil engineering and geotechnical engineering studies. ANN is commonly used to estimate the compaction and uplift of pile foundations, axial and lateral load capacities, drilled poles, foundation settlements, and anchors embedment. One study by Goh et al. (1995) used 93 data for training and 74 data for testing, obtaining a nonlinear relationship with a correlation coefficient of 0.97 for training and 0.91 for testing. The prediction of settlements in foundations is affected by uncertainties, similar to other complex issues in geotechnics. Researchers tested settlement prediction using ANN, using 79 data sets for training and 10 datasets for testing. Five parameters were used as input values: net pressure, average standard penetration test (SPT) values, foundation width, foundation form, and foundation depth [22].

ANN is applied to various earth sciences applications such as retaining walls, dams, earthquakes, geographical information systems, mining, geoenvironmental engineering, petroleum engineering, and rock mechanics. Traditional statistical methods may be insufficient due to interactions between variables, and predicting physical properties of soil like mineralogy, porosity, water content, grain size, etc. with statistical methods is difficult. ANN algorithms can estimate or determine various soil characteristics, including soil classification.

10.3 USDA Soil Classification

Before the release of the (relatively) comprehensive "Soil Taxonomy" (ST) (Soil Surveyors, 1975; most recent revision 1995), the United States Department of Agriculture (USDA) soil classification system 6 underwent a number of

"approximations". Rather than particular characteristics of any pathogenic process, families—the taxonomic layers beneath subgroups—differ substantially according to useful standards. Take soil temperature, parent material, or texture, for instance. Because it is primarily functional in nature, the definition of soil series is comparable to that of house. In order to provide more specific information, the series level—which is the lowest level of "soil taxonomy"—can nevertheless be further broken down into stages according to factors like slope or stoniness. The fact that ST prohibits soil classes at the series level from crossing any hierarchical boundaries with classes above them is crucial to understand. Among all soil classification systems, the USDA Soil Taxonomy is distinct in that it specifies a whole new set of terms for the upper classes. Despite seeming odd at first, these phrases encode the large group, suborder, and order they represent directly. They also avoid the confusion-causing issue of giving terms that are already in use new meanings. From ancient tropical soils disturbed by millennia of earthworm activity to Arctic permafrost, the USDA's Soil Taxonomy system is made to function with soils from all over the world. By dividing categories using diagnostic ranges, it also aims for objectivity. Still, there are many who disagree with the system. Thus, comparatively uniform natural soil bodies that are dispersed throughout several soil series cannot be shown on a soil map as a single plot unit. Consequently, Nettleton et al. It is advised to lift this limitation. He proposes looking into non-hierarchical classification systems because, in contrast to Nettleton et al., he does not see a way out of this predicament. Fitzpatrick, who created his classification of non-hierarchical soils because he was so against any kind of soil hierarchy, made a similar argument with great vigor. Most likely the most intricate and sophisticated soil classification system is the USDA "Soil Classification" utilized in the United States. It makes the claim to be broadly applicable and acts as a benchmark for contrasting with other systems. In addition to containing water, air, and living things, soil is the end result of the transformation of minerals and organic materials on the surface of the Earth. It exhibits its morphological organization and has grown and is still growing as a result of environmental influences. Higher plants develop in soil, which also provides the foundation for animal and human survival. Given that soil expands both in space and time, it is a four-dimensional system. Everyone has a general notion of what soil is, yet soil experts' opinions can frequently differ greatly from ours.Experts view soil as a complicated system with a distinct structure, while laypeople typically think of the "earth," the black surface layer that aids in plant growth in fields and gardens [23].

10.4 French Référentiel Pédologique

"Référentiel Pédologique" (RP) is a type system as opposed to a rank system, like the USDA's "Soil Classification" and the German "Bodensystematik" [24]. The system's goal is to achieve a distinct division between the conceptual domain and the actual reality. In RP language, the solum, or vertical section of soil visible in a trench or pit, represents the reality soil mantle. The reference horizon and the conceptual solum or reference are the two key ideas connected to this reality. In contrast to USDA diagnostic ranges, reference ranges are conceptual ranges that are defined in morphogenetic typology, allowing for educational interpretations. In stark contrast to the USDA system, the RP places more weight on pedagogical judgments and less on exact definitions. The USDA's "soil taxonomy" is a prime example of an unstable system since minor variations can have a significant effect. For example, the diagnostic ranges used to classify soils as hiatus require that a specific percentage of topsoil organic carbon be present. One comparable soil beyond that limit and one below it will wind up in entirely different regions of the USDA system. One more way that RP differs from other systems is that it may be made to adjust to the specific needs of local classification by simply adding (local) references.

Table 3 Classification of the desired test soil

Clay	C	CL	CH	-
Silt	M	ML	MH	-
Sand	SM	SC	SW	SP

Table 4 Details of a typical borehole in the Saqqez zone

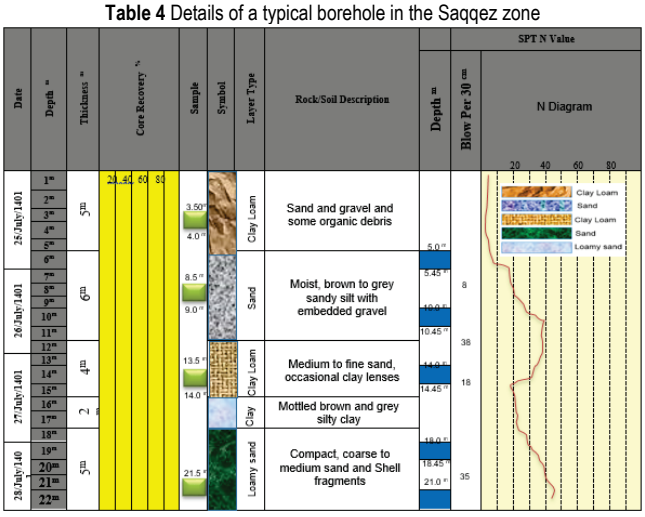
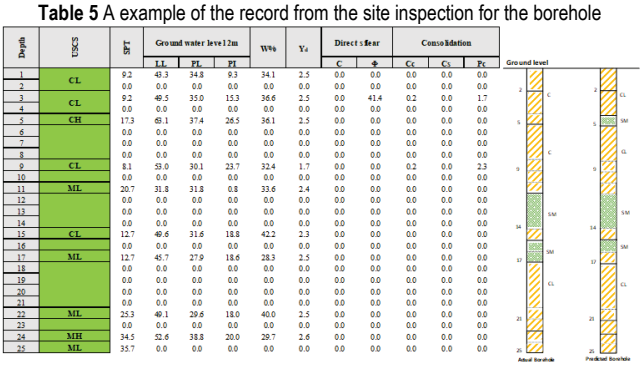


Table 5 A example of the record from the site inspection for the borehole



A few soil parameters, including bulk density, water saturation percentage, sand, silt, and clay, were included in the input data for the prediction of FC and PWP. However, the cation exchange capacity projection used organic carbon

and clay as inputs. found a substantial relationship between CEC and these inputs. shows the RMSE plot for every one of the many neurons in the buried layer. The best model for every soil attribute was produced using two neurons, as these data showed. For cation exchange capacity, correlation coefficients of 0.86 and 6.43 were found; for water percentage at permanent wilting point, 0.82 and 4.58; and for water % at field capacity, 0.86 and 6.1. Using MATLAB software, multi regression was calculated for three soil train data sets.

11 PREDICTION OF A BOREHOLE AND COMPARISON WITH ACTUAL BOREHOLE INFORMATION

Ultimately, the following graphically illustrated result was generated by running the application at a point where the layering information is available. When it comes to accurately predicting the type of soil, the shape is perfect, and its association with depth accuracy is also rather good. Naturally, modeling cannot adequately convey the presence of local lenses, which are the shape of the SM layer seen at a depth of six meters. It is not very significant in light of the objectives of this study.

Table 6 Statistical information for the FC and PWP test data sets

Neurons in hidden layers		Clay	Silt	Sand	SP	OC	Bd	FC	PWP
Training set	Min	13.8	19.7	4.7	29.9	0.0	1.4	18.9	6.0
	Max	89.4	86.6	55.0	102.6	39.1	2.0	55.4	32.1
	Mean	47.4	51.6	19.5	60.6	1.3	1.7	36.4	18.7
	Std	18.3	13.9	10.9	15.1	4.1	0.1	7.5	6.2
RMSE (%)		17.8	19.9	18.1	19.7	17.4	17.6	17.8	17.5

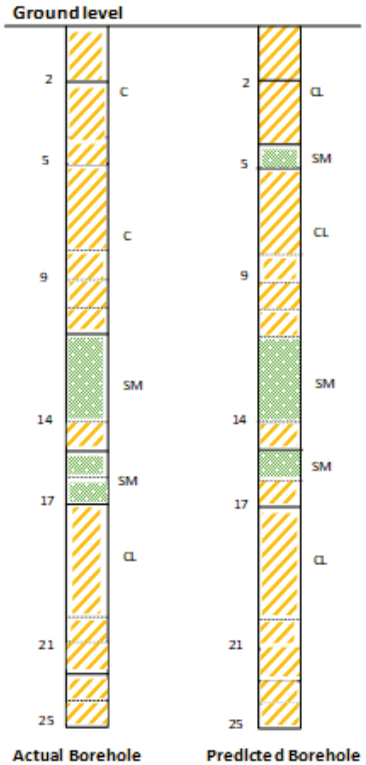


Figure 3 Borehole and comparison with actual information

The proposed neural network is designed to predict soil layers in boreholes at a site. The network uses coordination (x, y, z) to recognize the class of layer. The results show a high level of accuracy in predicting soil layers. Three performance measurements are used to evaluate the classification strategy: correctly classified instances for total accuracy, true positive rate (TPR) and false positive rate (FPR). The confusion matrix is used to assess the classification strategy. The precision of the network is indicated by the main diagonal, where higher values correspond to more accuracy. Soil layer depth and class are predicted using a multi-layer perceptron (MLP). The training approach uses supervised learning and is based on back propagation. The accuracy of the ANN model Soil layer depth and class are predicted using a multi-layer perceptron (MLP). The training approach uses supervised learning and is based on back propagation. The accuracy of the ANN model in predicting soil layers was up to 86%. The actual data and the ANN's forecast are compared to determine the accuracy values.

According to the shape in the forecast in the soil type is perfectly correct and in the case of depth is accompanied by good accuracy. The SM layer, which is at a depth of 5.5 meters, is in the form of a vertical lens.

12 CONCLUSION

The researchers created a model using Artificial Neural Networks (ANNs) to forecast the soil composition of Saqez City, Iran. Saqez City's 100 square kilometers of data were gathered, and the data was then trained to classify various soil strata at various depths. At a given depth, the model could calculate each layer's thickness. In the initial stage of geotechnical projects, site research was made possible by the model's 85% prediction accuracy. When input examples fall inside the trained range, the accuracy of the model is enhanced. Subsequent studies will concentrate on the accuracy of ANN models in data set extrapolation. Because of the precision of the model, fewer boreholes need to be drilled, potentially saving time and money.

1) According to the results obtained, the information obtained from the artificial neural network in the investigation of geotechnical phenomena such as estimating the thickness of the earth's layers, compiling the field operation plan. Interpretation of land conditions and soil classification using its characteristics has acceptable accuracy.
2) Using a closed environment instead of using data dispersion is associated with better results.
3) With the increase of boreholes during network training, the forecasting accuracy also increases.
4) Placing local lenses in the training data category increases the overall error and is recommended to be removed in the modeling.
5) When we have the concentration of boreholes in one area and the point where we want to determine the layering of the soil is also located in the same area, it is recommended to do modeling around this area in a closed environment and from boreholes far away from this area. Do not use it in modeling.
6) In cases where the boreholes are almost uniformly

distributed if we have a concentration of boreholes at a certain point and the point where we want to determine the soil stratification is far from the location of the local concentration, it is recommended to have one borehole from the concentration point in the modeling. Should be used, otherwise, the changes made in the weight function will increase the error.

7) Local lenses are not seen in the software forecast.

13 REFERENCES

[1] Gunaydin, O. (2009) Estimation of soil compaction parameters by using statistical analyses and artificial neural networks. *Environ Geol, 57*, 203-215. https://doi.org/10.1007/s00254-008-1300-6

[2] Alavi, A. H., Gandomi, A. H., Mollahassani, A., Heshmati, A. A. & Rashed, A. (2010). Modeling of maximum dry density and optimum moisture content of stabilized soil using artificial neural networks. *J Plant Nutr Soil Sci, 173*, 368-379. https://doi.org/10.1002/jpln.200800233

[3] Sivakugan, N., Eckersley, J. D. & Li, H. (1998). Settlement predictions using neural networks. *Australian Civil Engineering Transactions, CE40*, 49-52.

[4] Sterling, R. L. & Lee, C. A. (1992). A neural network - expert system hybrid approach for tunnel design. Proc., *33rd United-States Symp. Rock Mechanics*, Tillerson, J. R. & Wawerisk, W. R. eds., Brookfield VT, Balkema, 501-510.

[5] Teh, C. I., Wong, K. S., Goh, A. T. C. & Jaritngam, S. (1997). Prediction of pile capacity using neural networks. *J. Computing in Civil Engineering, ASCE, 11*(2), 129-138. https://doi.org/10.1061/(ASCE)0887-3801(1997)11:2(129)

[6] Banimahd, M. (2002). Prediction of Undrained behavior of fine sands using Artificial Neural Networks. *Master's thesis*, Tarbiat Modarres University, Tehran, Iran.

[7] Tutumluer, E. & Seyhan, U. (1998). Neural network modeling of anisotropic aggregate behavior from repeated load triaxial tests. *Transportation Research Record 1615*, National Research Council, Washington, D.C. https://doi.org/10.3141/1615-12

[8] Zhu, J. H., Zaman, M. M. & Anderson, S. A. (1998). Modelling of shearing behavior of a residual soil with a recurrent neural network. *Int. J. Numerical and Analytical Methods in Geomechanics, 22*(8), 671-687.

[9] Sarmadian, F. & Taghizadeh Mehrjardi, R. (2008). Modeling of some soil properties using artificial neural network and multivariate regression in Gorgan Province, North of Iran. *Global Journal of Environmental Research, 2*(1), 30-35.

[10] Cai, J. & Zhao, J. (1998). Computerization of rock engineering systems using neural networks with an expert system. *Rock Mechanics and Rock Engineering, 31*, 135-152. https://doi.org/10.1007/s006030050015

[11] Ambrozic, T. & Turk, G. (2003). Prediction of subsidence due to underground mining by artificial neural networks. *Computers & Geosciences, 29*, 627-637. https://doi.org/10.1016/S0098-3004(03)00044-X

[12] Santos, O. & Celestino, T. (2008). Artificial neural networks analysis of Sao Paulo subway tunnel settlement data. *Tunneling and underground space technology, 23*, 481-491. https://doi.org/10.1016/j.tust.2007.07.002

[13] Ylmaz, I. & Yuksek, A. (2008). An example of artificial neural network (ANN) application for indirect estimation of rock parameters. *Rock Mechanics and Rock Engineering, 41*, 781-795. https://doi.org/10.1007/s00603-007-0138-7

[14] Penumadu, D. & Jean-Lou, C. (1997). Geometrical modeling using artificial neural networks. *Artificial Neural Networks for Civil Engineers: Fundamentals and Applications, ASCE*, 160-184.

[15] Choobbasti, A., Farrokhzad, F., Rahim Mashaie, S. & Azar, P. (2015). Mapping of soil layers using artificial neural network (case study of Babol, northern Iran). *Journal of the South African Institution of Civil Engineering, 57*(1), 59-66. https://doi.org/10.17159/2309-8775/2015/v57n1a6

[16] McBratney, A. B., Mendonca Santos, M. L., & Minasny, B. (2003). On digital soil mapping. *Geoderma, 117*(1-2), 3-52. https://doi.org/10.1016/S0016-7061(03)00223-4

[17] Cavazza, L., Patruno, A. & Cirillo, E. (2007). Field capacity in soils with a yearly oscillating water table. *Biosystems Engineering, 98*, 364-370. https://doi.org/10.1016/j.biosystemseng.2007.07.001

[18] Givi, J., Prasherb, S. O. & Patel, R. M. (2004). Evaluation of pedotransfer functions in predicting the soil water contents at field capacity and wilting point. *Agricultural Water Management, 70*, 83-96. https://doi.org/10.1016/j.agwat.2004.06.009

[19] Sys, C., van Ranst, E. & Debaveye, J. (1991). Land evaluation Part I. Principal Land evaluation and Crop production calculation. General Administration for Development, Cooperation, *Agric Pub., 1*(7), 247.

[20] Mermoud, A. & Xu, D. (2006). Comparative analysis of three methods to generate soil hydraulic functions. *Soil and Tillage Research, 87*, 89-100. https://doi.org/10.1016/j.still.2005.02.034

[21] Schaap, M. G. & Leij, F. J. (1998). Using neural networks to predict soil water retention and soil hydraulic conductivity. *Soil and Tillage Research, 47*, 37-42. https://doi.org/10.1016/S0167-1987(98)00070-1

[22] Schaap, M. G., Leij, F. J. & van Genuchten, M. Th. (1998). Neural network analysis for hierarchical prediction of soil hydraulic properties. *Soil Science Society of America Journal, 62*, 847-855. https://doi.org/10.2136/sssaj1998.03615995006200040001x

[23] Blake, G. R. & Hartge, K. H. (1986). *Bulk density*. In: Klute, A. (Ed.). Methods of Soil Analysis. Part 1, second ed. Agron. Monogr. 9. ASA and SSSA, Madison, WI, 363-375. https://doi.org/10.2136/sssabookser5.1.2ed.c13

[24] Cassel, D. K. & Nielsen, D. R. (1986). *Field capacity and available water capacity*. In: Klute, A. (Ed.). Methods of Soil Analysis. Part 1, 2nd Edn. Agron. Monogr. 9. ASA and SSSA, Madison, WI, 901-926. https://doi.org/10.2136/sssabookser5.1.2ed.c36

Authors' contacts:

Semko Arefpanah
(Corresponding author)
Department of Civil Engineering,
Science and Research Branch, Islamic Azad University,
Tehran, Iran
semko.arefpanah@srbiau.ac.ir
s.arefpanah@gmail.com

Alireza Sharafi
Department of Civil Engineering,
Science and Research Branch, Islamic Azad University,
Tehran, Iran
alireza.sharafi@srbiau.ac.ir

Alireza Gholamian
Department of Civil Engineering,
Islamic Azad University, Hamadan Branch, Iran
Alireza_gholamian87@yahoo.com

ISSN 1846-6168 (Print), ISSN 1848-5588 (Online)
https://doi.org/10.31803/tg-20240510170107

Subject review
Received: 2024-05-10, Accepted: 2024-05-15

Comparative Analysis of Employee Training Using Conventional Methods and Virtual Reality

David Krákora, Petr Hořejší*, Andrea Šimerová

Abstract: This paper deals with an innovative method of employee training. Its concept is based on an experiment comparing a conventional training method with virtual reality (VR) training. For the evaluation, a questionnaire was designed to assess how much probands remembered immediately after the training. The area of training was the handling of sharp objects in a healthcare environment. The length of both types of training was 30 minutes). The conventional training was in the form of a presentation, the innovative one was virtual reality with guided interaction with objects. The probands were college students. A total of 42 samples were measured, where each proband was trained on only one variant. The main objective was to compare the effectiveness of each training method as a function of time and its effectiveness in terms of memorability. It showed that probands who acquired knowledge using the VR variant remembered more information. It was also found that prior experience with VR did not yield better results.

Keywords: Industry 4.0; new training method; safety training; user acceptance; virtual reality; virtual training

1 INTRODUCTION

Industry 4.0 has brought innovations and opportunities for companies to apply new methods that offer a different perspective on their current strategies. Many organisations have started working on modernising their standards, processes and practices. [1] One of the key input pillars for the application of the new concept is the digitisation of corporate data. The right digitalisation can create a competitive advantage if companies can choose the right methods and processes. This enables businesses to streamline their processes, find new solutions and use their resources efficiently, as well as avoiding waste and trying to reduce costs. [2, 3] With the right digitalisation[4], it is possible to react flexibly to change, create competitive advantages and streamline human resources. Another important feature of Industry 4.0 is integration, which can involve processes throughout the entire organisation (that is, from the production process to the core management). [3, 5]

The concept of virtual reality, which in the sense of Industry 4.0 is closely related to the above concepts, can be explained in several ways. In general, it is a technology that creates an environment using a computer screen where the user experiences the sensation of being surrounded by another space with which they can interact. [6] It is thus a medium that allows people to interact effectively with a digital 3D environment in real time using their natural sensory abilities and skills. [7, 8] Immersion and presence are key features of VR [9]. Immersion is the objective ability of VR to provide the user with faithful, complex and sensory stimuli. Presence is the subjective experience of being present in the virtual world, even when the user knows that the stimuli do not exist in a physical sense in the virtual space [8, 10]. Various devices are able to create this illusion to such a degree that the user can ignore their real surroundings [11]. From the point of view of the engineering industry, an interesting application is the possibility of simulating real processes without any risk exposure [12].

One of the latest concepts that may be of interest to businesses is virtual training (VT). This is the digitisation of a real process into VR for teaching and/or training purposes. The main objective is to impart the necessary skills to the employees and create quality experience and knowledge for the trainees [13, 14]. This method creates the conditions and experience needed to minimise human error in a safe environment. Risky activities can be explained and tasks that require extra attention can be demonstrated. This makes it an interesting tool for training people [13, 15]. Such training is easily repeatable, in an identical setting, with all output data (e.g., evaluation documents) easily recorded.

One area in which VT could be used is in the training of employees. In-house training is a key element in the development and growth of employees, as well as in improving the efficiency and competitiveness of the enterprise as a whole [16]. The disadvantage of VT is that its actual creation as well as implementation can take a few months. Overall, the development of VT can be divided into three phases:

1) Pre-project - mapping of the selected process and scenario preparation
2) Project - environment modelling and application development (setting up logic and writing scripts).
3) Final phase - the implementation period, when the details are fine-tuned with the customer and introduced into the company. [17, 18]

This is also linked to the level of acquisition costs. As this is a new technology, a high initial investment in both hardware and software is required. Further investment is needed in professional staff for equipment maintenance and technical support. Any technical system can experience malfunctions and failures which can cause interruptions in training and loss of staff performance [19, 20]. Another disadvantage is that VR can cause health problems for users. It is important to consider the possibility of Motion Sickness during development and try to eliminate it as much as possible [23].

Well-designed and implemented training can bring many benefits, such as increased productivity, higher employee motivation, reduced staff turnover and improved market positioning. From an Occupational Health and Safety (OHS) perspective, the elimination of workplace accidents can be a key criterion. It is this type of training that is essential for

companies to provide their employees with safe conditions in the workplace and to be able to eliminate all risks [22–24]. Using VR for occupational safety training is therefore effective for several reasons. One of the main factors is the possibility of realistic simulations. Due to the interactivity, participants are actively involved in the training, this results in more effective retention of information. Also, the immediate feedback allows them to hone their skills and reactions to different situations. It is therefore an interesting tool for any type of training [25].

1.1 Related Work

A study by Radianti and colleagues (2020) explores the use of VR in the context of education. They investigated the prerequisites for its successful implementation in educational processes for different subjects. The result of their findings was that VR has had positive feedback in the form of practical applications [26]. Also relevant is a study [27] that focused on the use of VR during primary school teaching. This study involved 24 students who participated in a virtual field trip. From this it can be concluded that the current young generation is open to using these digital tools. A strong sense of presence and effective VR immersion provide training in a safe and controlled environment [28]. The effectiveness of digital tools itself is captured in [29], where the authors focused on comparing traditional and computer-based tools by analysing 18 studies. The results showed a positive effect and outcomes when a new (digital tool) was used. Half of the cases had a 20% higher effectiveness and positive attitude recorded by the probands. This result is supported in [15], where immersion and presence are considered as the main advantages of VR for training health professionals. For example, through VR, fear of heights can be overcome. A total of 49 volunteers participated in the experiment [30], with those who underwent VR immersion showing a 75% reduction in their fear of heights. From an industry perspective, even very dangerous jobs such as mining can be simulated, where more than 80% of the participants believed that the depicted hazards were real. [31] In the construction industry, a comparison was made with conventional hazardous work training (e.g. working at heights). The result was a total training time of more than 20 hours over a month for new entrants [32]. In the questionnaire, more than 42% of people thought that lack of knowledge and poor-quality training was the main source of accidents. Compared to the conventional method, participants showed a higher level of knowledge in the field. Thanks to VR, they were able to remember more products and their assembly procedures. The limitations of this study were that these were very simple assemblies. In contrast, Kalkan [33] focused his study on the performance of complex assembly tasks and product quality. His goal was to use VR to reduce the training and knowledge transfer time. The results showed that they were able to reduce training time by 25% and achieved 28% better performance through VR.

In [34] , the standard learning method is compared with VR. The result of this study is that users who trained using VR committed fewer errors than in real buddy training. Another finding was the advantage that VT is shorter, more efficient and less dependent on the trainer.

HSE (explain this abbreviation, is it OHS?) managers who drew on the experience of 540 workers praised the variety of the system, the interactive elements, and the overall attractiveness. They also appreciated the shared learning functionality [35]. Zawadzki focused his research on demonstrating the advantages of VT over traditional training methods. This research was supported by a total of 20 respondents that he tested for each form of training and then compared their results. The outcome was that using VT, respondents had better times in terms of product identification and lower assembly times. He also mentioned some disadvantages. The main disadvantage cited was the high initial investment, which has the potential to pay for itself within a few years. [36] Another comparison was made using fire protection as a test area. Through research and data collection using questionnaires, it was found that VR application is more effective than the conventional method. A total of 80% of the respondents in the questionnaire identified it as an innovative and more effective method [37]. OHS elements appear in most training processes. Electrocution is another hazardous area that is appropriate to virtualise. Based on preliminary validation, it was found that VR application has a strong influence on human behaviour and understanding of OHS knowledge. People were more cautious when the hazard level was higher in VT, the same as if they were in reality [38].

The article [39] focused on a comparative approach between the conventional method and the use of VR in VT. In this analysis, the researchers found that probands who received training in VR showed better knowledge transfer and production. The results were very positive for hands-on learning. On the other hand, the results were not significant for theoretical learning.

The paper [40] reports on training opportunities in OHS in the field of electrical work. Here, the authors suggest VT as the optimal training method. Other important findings that were found are that training in VR has higher efficiency compared to the conventional method, reducing the time for knowledge and experience generation. Another advantage is better applicability in practice and some sources also report a reduction in the cost required for training.[41–43] In general, virtual training makes it easier to create new experiences, build skills and transfer the know-how of the process correctly, resulting in more efficient work with human resources. [44]. Overall, the claim that VR training allows the acquisition of tangible knowledge through the form of experiencing certain situations and then solving them is considered a great advantage. It is the immediate feedback that provides error correction that leads to a better understanding of the issue at hand [45].

1.2 Hypothesis and Research Questions

Based on the above, it can be concluded that VT in the industrial sector is an effective tool and can be more cost-effective in terms of capacity, time, and therefore money spent, than conventional methods. In this paper, we focus on VT in the field of OSH, specifically on the level of memorisation compared to conventional training. Testing of

the memorability of a selected process will be performed. Verification of memorability will be done by asking control questions on the selected process.

Based on the standardised training of accident prevention when handling sharps, a questionnaire was developed, and control questions were asked to the respondents. Each of the responses was then converted into scores to allow for better comparison between the two groups (standard and VR training).

Thus, three research questions were set, with a null hypothesis (H0) and an alternative hypothesis (H1) for each question - see below:

Q1: Which group of probands received a higher score?
Q2: Do probands with higher VR experience show better results?
Q3: Do computer game players show better results in VR?

Question Q1 is the main question and Q2 and Q3 are supplementary questions.

2 METHODOLOGY

The aim of this study is to carry out an experimental investigation to compare the effectiveness of VR training versus conventional training in terms of memorisation of the training content in the immediate post-training period. To meet the requirement of the possibility of universality of using the training in different enterprises, a training course that focuses on occupational safety and health was chosen. In the context of this study, this training was specified as training on working with sharp objects and the correct procedure for sharp injuries. We compare two types of training:

- 2D – using a presentation.
- VR – using virtual reality.

In the case of VR training, the training is digitised, and an interactive VT course is created based on interactive user involvement. One of the main advantages of this method is the hands-on execution of tasks, which allows users to acquire skills through realistic simulations and exercises. The primary objective is to assess the memorisation rate of VR compared to the traditional method. The aim is to demonstrate the benefits of a VR tool for employee training.

2.1 Conventional Training (2D)

Probands were given a standard presentation for job training, handling of sharp objects and subsequent injury procedures using frontal teaching. The presentation was prepared as the main aid which was designed to provide probands with comprehensive and understandable information about working with and handling sharps. The risks and correct procedures for sharps injuries were also listed. The presentation included pictures that served as visual support. Their main function was to simply explain the issue, describe a specific case, capture the probands' attention and try to make it more memorable. The training was carried out by a person experienced in this type of work and in training other staff. In addition to the presentation itself, it was necessary to have a room with projection technology. Finally, interactive elements in the form of a discussion were used where the probands could ask questions or have things explained in greater detail.

The total length of the conventional training was set at 30 minutes. This timeframe allowed ample space for in-depth analysis of the topic and interactive discussion. During this time, participants had the opportunity to ask questions about uncertainties and discuss practical applications of the skills learned. For authenticity of the experience, the traditional training was organised so that all the participants from this group attended at the same time. This group dynamic was intended to encourage active participation of the participants and allow them to share their thoughts and experiences in real time. At the end, each proband completed a questionnaire that was the same for both groups.

2.2 Virtual Reality (VR)

For this variant, an application was developed to represent the training in virtual reality. The main function of this application was the simulation of selected processes. The input data for the creation of the VR training was a standardised process of handling, working with and accident prevention with sharp objects. This process was divided into activities that build on each other and represent interdependent steps. Within the simulation, this means that if the first step is not completed, the next steps cannot be started. Another important point is the division of the elements into static and dynamic elements (i.e. those that are interacted with). This type of division has a big impact on performance, as static elements are in one place all the time and can be rendered to increase the fluidity of VR. The application was developed in Unity 3D software. Unity is a very popular development engine that allows the creation of interactive 2D and 3D applications (including VR). A lot of emphasis was placed on the look and quality of the environment to represent the real situation. This element was intended to use as much immersion as possible and to approximate a real situation that may occur in practice. In Fig. 1 – A sample environment from the VR app, you can see that this is a surgery where a syringe is being handled. For better orientation, the user observes everything from a 3rd person perspective and solves the problem that arises for their VR colleague. At the end, each proband completed a questionnaire that was the same for both groups.

Figure 1 A sample environment from the VR app

A virtual guide was used throughout the application to explain each step and also to tell them what to do in which step. Visual management elements were also used to improve navigation in the environment. These elements can be seen in Fig. 2 – Elements of visual management in VR, where the active objects (those that the user can manipulate) are highlighted in blue. The places where they are to be inserted are then highlighted in green.

Figure 2 Elements of visual management in VR

Oculus Quest 2 goggles were used to transmit the virtual environment. This is an All-in-One device, meaning that it does not require any additional communication devices or tracking stations. The main point of these goggles is to make working with them as easy as possible and to be able to provide high quality rendering with maximum comfort. The controls or just your hands can be used to operate them.The integrated motion tracking technology means that the goggles can record the movements of the hands and transmit them to the virtual environment. A time schedule was set with the group that participated in the VR training, with 30 minutes planned for each proband. In the first 10 minutes, the proband had the basics of using the technology explained to them, they were shown the functionality of the hardware and how to set up its interface. In the second part, the proband adjusted the optimal size, improved the sharpness, and increased or decreased the volume of the sound if needed. This was followed by the actual training. 15 minutes per proband was reserved for this, with the length of the "walk-through" application being half that. This provided a time buffer in case any complications arose. At the end, the proband completed a questionnaire. As part of hygiene, the goggles had a rubber cover and were wiped with disinfectant wipes after each use. If any of the probands wore dioptric glasses, a special attachment was used. This hardware has its limits in use, so it was important to have a well-lit room. It was also important to keep the room free of mirrors and other materials that create reflections and could impair hand tracking functionality. In the event of excessive sunlight, it was essential to be able to shutter the windows.

2.3 Research Sample

The study included a total of 42 probands. They were selected from a population of college students and their age range was between 19 and 23 years. This age range ensures a relatively homogeneous sample in terms of learning ability. The probands are students of different majors, which means that their educational backgrounds and experience with the subject matter may vary. This could influence their perception and recall of the information presented during the training. However, most of them had university education in common, which may help to minimise the influence of these variations on the results of the study. We see the above as a limiting factor of the study. The probands were proportionally and randomly divided into two treatment groups to ensure that each group received only one variation of training. This step was essential to compare the effectiveness of the two methods, as it allows us to eliminate possible bias in the results due to awareness of the issue

2.4 Questionnaire

The questionnaire consisted of three sections in total. The first section of the questionnaire focused on basic information about the proband, such as age, gender, and experience with virtual reality. It had a total of 4 questions. This section of the questionnaire aimed to provide demographic data that may influence the participants' reactions and behaviours during the training. For the question focusing on virtual reality experience, a scale was established based on the number of uses, with the lowest value being 0 and the highest being 50 or more. The PC gamer question then focused on the number of hours per week the proband spends playing games. The lowest value was 0 hours and the highest was 20 or more.

The second section contained follow-up questions that related to the actual issues that the training focused on. There were three types of these questions and there were 6 in total. The first section had only one permissible answer. There were three of these questions. For example, a question was "What part of the sharp object should be grasped?" The second section had one or more answers, here the probands could tick any number of answers and there were two questions of this type. The last type was a question where the user assigned a ranking to each activity, according to how they would deal with a sharps injury. This section was used to test whether the participants were able to correctly understand and apply the information given to them during the training.

The third section of the questionnaire dealt with the participants' subjective evaluation of the training they had just received. There were 3 questions in total. Overall, the use of the questionnaire and its structured layout allowed for a systematic evaluation of the results and a comprehensive picture of the effectiveness and benefits of virtual training compared to the traditional training method.

2.5 Monitored Data

From the data collected, several key evaluations were obtained that focused on different aspects of training effectiveness. These evaluations include memorability, effectiveness, and familiarity with the use of VR, which was needed to complete the questionnaire. Memorability is an

important factor that measures the ability of participants to remember and reproduce the information and skills that were presented during the training. In this context, the ability to remember was compared between the two groups of participants who received different versions of the training. The results of these tests provided information on how well participants remembered the information learned during the training. Participants' interest and satisfaction are subjective aspects that reflect their perception and evaluation of the training. These evaluations included collecting feedback from participants regarding their personal interest in VT. Interest and satisfaction can be key indicators of training effectiveness, as a positive experience can lead to better engagement and retention of information. These aspects of the evaluation provide a comprehensive view of the effectiveness of the training and allow for the identification of strengths and weaknesses of each method. The combination of objective and subjective measures provides a comprehensive picture of how effective and efficient different approaches to training are on a given issue.

3 RESULTS

In the evaluation we worked with two variants of knowledge testing:

- Variant 1: VR testing with 19 respondents.
- Variant 2: Conventional testing, in paper form (2D) with 21 respondents.

Table 1 Number of correct and incorrect answers

Row labels	Paper	VR	Total sum
Correct	19	18	37
Wrong	2	1	3
Total sum	**21**	**19**	**40**

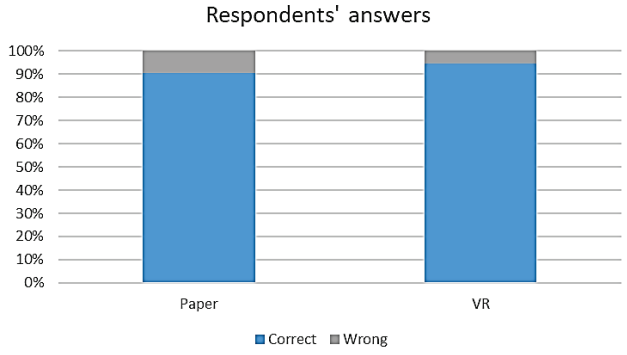


Figure 3 Number of correct and incorrect answers

To display the values, an association table was created, supplemented by a graph. As part of the evaluation process, we formulated three questions (Q1, Q2 and Q3) which are further used to generate three hypotheses. In conducting statistical tests, we set the alpha level of significance at 0.05. This level determines whether we are able to reject the null hypothesis. We obtain the results of this statistic by calculating the p-value, which is the result of the statistical tests performed [46, 47].

First, the answers of each proband were scored. The scoring was done as follows:

Question - only one correct answer. (Total 3 questions)

- Correct 1
- Wrong 0

Question - multiple correct answers. (Total 2 questions)

- Correct 1
- 1 wrong 0.5
- 2 or more wrong 0

Question - ordering of events. (Total 1 question)

- Correct 1
- 1-2 mistakes 0.5
- 3 and more mistakes 0

A new variable was created to represent the total number of points scored by the proband for all scoring questions. The maximum number of points obtained in the testing was 6. In the 2D group an average of 3.73 points was obtained and in the VR group an average of 4.01 points was obtained. Here, descriptive statistics were first calculated, and because the aim was to test the agreement of the mean scores between the forms of testing (2D vs. VR), the normality of the data (i.e., the assumption for the use of parametric tests) was examined using the Shapiro-Wilk (S-W) test [46, 47]. After performing the S-W test, the p-value for the VR variant was calculated as 0.122859473. The scores using VR showed a normal distribution (p-value > 0.05). For the 2D variant, the p-value is calculated as 0.015485487. The score using 2D does not have a normal distribution (p-value < 0.05). Because the normal distribution does not emerge for both sets, we used the nonparametric Mann-Whitney (M-W) test for agreement of medians to test for differences between the sets. Other assumptions for the use of this test include independence of the two samples ($n_1 > n_2$), equal shape, and identity of variances [46].

Q1: Which group of probands received a higher score?

H0: The VR group did not score higher on the knowledge test immediately after exposure compared to the 2D group.

H1: The VR group received a higher score on the knowledge test immediately after exposure compared to the 2D group.

The median score for the VR group is 4.5 and the median score for the 2D group is 4. Due to the small sample size of respondents (n_1 = 19, n_2 = 21), the test criterion value according to the classical procedure for this test was chosen to be U = 90, which was compared with the critical value from the table for the Mann-Whitney test (U0.05(19;21) = 126). [48] The p-value for the one-sided hypothesis was calculated as 0.00123962. The test criterion is less than the critical value found. H0 is rejected. Statistical evaluation showed a significant difference in the median scores of comparable groups of probands. The probands who used VR for testing have significantly higher scores than the probands who used the 2D form. This result suggests that switching to virtual reality may have an impact on the test results, which

may be important when deciding on the preferred form of testing in each context.

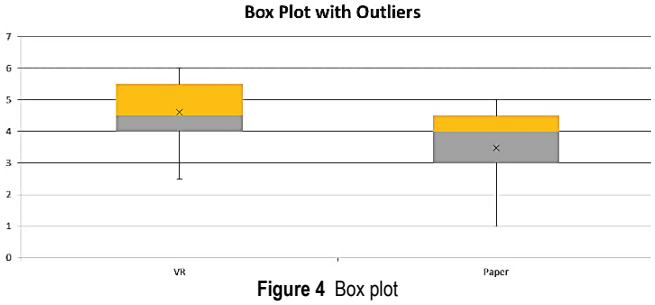


Figure 4 Box plot

A box plot was created as part of the evaluation. It is a graphical tool for displaying data [49]. In this case, the mean, median and quartiles are displayed. If there were outliers, they would also be displayed here.

Q2: Do probands with more VR experience show the best results?

H0: Subjects with higher VR experience do not perform better on the VR knowledge test.

H1: Subjects with higher VR experience show better results in the VR group knowledge test.

The higher VR experience of each proband is specified using a categorical (ordinal) variable (i.e., ordering is possible). Since these are not specific numerical values (i.e., the data do not meet the assumptions of normality), it is not appropriate to use Pearson's correlation coefficient for evaluation purposes, so we used Spearman's correlation coefficient to determine the existence of a relationship. This coefficient is not based on the measured values but on their order. In the calculation, the observed empirical values, which are ranked from lowest to highest, are replaced by their order [46, 50].

The coefficient value was calculated as –0.27377, indicating a possible weak inverse correlation. Due to the low number of respondents (n = 19), this value is not yet statistically significant. The p-value was calculated as 0.25674. H0 is not rejected. It can be concluded that more experience with VR does not yield better results in testing.

Q3: Do PC gamers perform better in VR?

H0: Computer game players do not outperform the VR group on the knowledge test. H1: Computer game players perform better on the VR group knowledge test.

First, the previously created new statistical variable (i.e., gaming scores) was divided into 2 samples. The first sample consists of the total score counts of probands who do not play PC games (i.e., non-gamers) and the second sample consists of probands who play PC games (i.e., gamers). The normality of the data was examined using the S-W test. In this case, the data can be considered as a selection from a normal distribution. The p-value was set as 0.06469 for probands who do not play PC games and 0.1565 for probands who play PC games. In both cases, the p-value is higher than the alpha significance level. For this reason, the F-test for agreement of variances was used, which confirmed a higher p-value (0.86507108) than the alpha level. Thus, the variances are comparable. Therefore, a two-sample t test for agreement of means was used, mainly because of its potential use for sampling smaller ranges [46].

The average was 4.9 for probands who do not play PC games and 4.3 for probands who play PC games. The p-value of the test was set at 0.91625, which is more than the significance level of the alpha. We cannot confirm H0 that gamers have better results. Probands who play PC games do not show better scores than probands who do not play PC games.

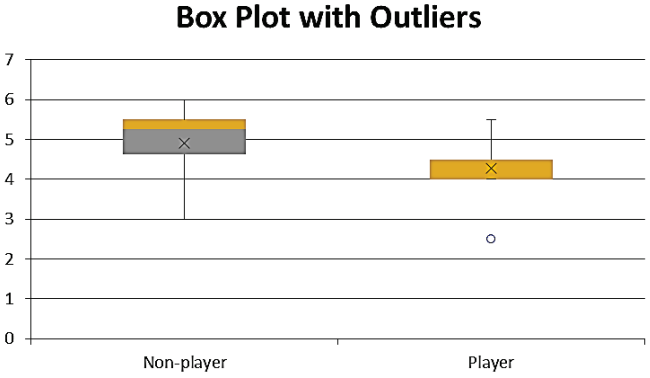


Figure 5 Box plot with Outliers

4 DISCUSSION

The innovation and digitalisation of training has been studied in the past. So has the use of virtual reality in different sectors and phases of training. With our experimental variations, we achieved identical designs - conventional). The research confirmed that the group that underwent VR showed a higher success rate in the final test compared to the 2D group. Similar results were found in the study [39] that looked at the implementation of VR in practical parts of training. Their analysis focuses mostly on the benefits that VR has over standard types of training. Among the main benefits, the authors included hands-on involvement in processes, the possibility of handling expensive or less available tools, and another important perspective is on the user experience and user engagement. VR offers the opportunity to enter an environment that can be more realistic and interactive than traditional training. Users can immerse themselves in situations and environments that would otherwise be inaccessible. This interactivity can lead to higher engagement and better retention of the experience. [28] In terms of learning efficiency and information retention, studies suggest that virtual reality may be more effective than 2D methods. More realistic environments and opportunities for practice and repetition can lead to deeper learning and better retention of information. This can be particularly useful in areas such as medical training, technical practice or safety training where hands-on experience is key [27, 36].

However, there are financial costs to consider, which can be substantial for businesses. Implementing and developing VR can be financially challenging.[27] On the other hand, although the initial costs may be higher, VR can bring long-term savings in training and education. Improved training efficiency and reduced training costs for new employees can provide significant benefits in the long run [35].

The results of our study can be compared with those of [33], where the knowledge gained from a selected process was compared. The probands who were trained by VR performed 28% better in terms of memorisation and number of errors on the selected segment. This research can also be compared with [37], where *a* questionnaire was also found to be a more effective method than traditional training. At the same time, probands received it very positively and considered it to be an innovative training method [40].

Another important element that VR brings is presence and zooming in on dangerous situations. This allows participants to experience dangerous situations without putting their health at risk [31].

Overall, VR is more effective than conventional methods. However, further work could be done to compare memorisation across different time intervals, when probands for a given process would be tested after certain periods of time. This could compare the memorisation of the process with the time elapsed since training [41, 42].

In the context of our study, we use a small number of respondents. Further research would be to expand the number to 100+ respondents for each training area. It would also be useful to explore correlations in age groups and occupational distribution.

5 CONCLUSION

The education system has evolved over the centuries, and it has always adapted to available technologies and the needs of students. We are now on the threshold of further exciting developments, and it is the duty of researchers, educators and teachers to embrace and prepare for them. The generation that are just beginning their education have been online all their lives. The digital world is as important and immersive as the real one. Educating Generation Z is challenging and requires a very different approach to maximise its effectiveness and engagement.

There are many benefits to using VR technology in education. First, VR provides superior visualisation that cannot be obtained in a traditional classroom. It reflects a world in which the younger generation feels comfortable. It is inclusive and allows everyone to participate in the learning process. It provides virtually unlimited access to information. Modern technology used for training increases participation and stimulates collaboration and engagement. It is used for highly effective blended learning, encouraging self-study and individual pursuit of knowledge. Another advantage is the active involvement of users in the training itself. Through interactions, the users' knowledge of the process can be built more effectively. An important aspect is the simulation of risky and dangerous processes, where users can try out the correct safety procedures without putting their lives in danger.

Although the use of modern technology in the educational environment is clearly beneficial, it is not without risks and dangers. One of the main problems is the lack of flexibility. During traditional training, students can ask questions, receive answers, and participate in this discussion. Using a virtual reality headset with specific software, students must follow the rules and cannot do anything other than what is pre-set. In addition, too much focus on digital learning solutions could upset the balance between teaching hard and soft skills.

Therefore, it is essential to find the right balance between traditional training and the use of VR to avoid eliminating the benefits of both methods. Although at first glance VR seems like the right method for future training, the human contact factor must not be forgotten. There are just some parts of training that are better for people if they include the human factor, personal contact and the opportunity to ask for more information. As mentioned in several studies it is effective to supplement standard training methods with different VR components.

The conclusion of this study clearly points to the superiority of VR training as a more effective training tool. However, for such a conclusion, a more detailed study on a larger sample of probands through different psycho-graphical segmentations would be needed. Moreover, VT is not perfect; it is most suitable for teaching procedures and activities that can be algorithmized. Research suggests that through VR it is possible to achieve not only higher levels of participant engagement but also improved memorisation of the material presented. This important factor, which highlights VR's ability to transfer information and skills with greater efficiency, has major implications for the effectiveness of training and education. At the same time, the results of this study motivate further research. One possibility is to extend this type of training to more areas. Where the applications would be focused on processes other than just handling sharp objects. There is currently research being done at our university on the long-term memorization of these training methods. This research is working with a larger sample of probands. Possible future studies could further explore the specific mechanisms behind VR's success in training and identify strategies to maximise its potential. Beyond this, further research could focus on the long-term impacts of VR training, including the lasting memorisation of the knowledge and skills acquired compared to traditional methods.

Acknowledgement

This work was supported by the Internal Science Foundation of the University of West Bohemia under Grant SGS-2024-032 "Environmentally sustainable production".

Informed Consent Statement

Informed consent was obtained from all the subjects involved in the study.

6 REFERENCES

[1] Leong, W. D., Teng, S. Y., How, B. S., Ngan, S. L., Rahman, A. A., Tan, C. P., Ponnambalam, S. G. & Lam, H. L. (2020). Enhancing the adaptability: Lean and green strategy towards the Industry Revolution 4.0. *Journal of Cleaner Production, 273*, 122870. https://doi.org/10.1016/j.jclepro.2020.122870

[2] Reski, N. & Alissandrakis, A. (2020). Open data exploration in virtual reality: a comparative study of input technology. *Virtual Reality, 24*(1), 1-22. https://doi.org/10.1007/s10055-019-00378-w

[3] Gillani, F., Chatha, K. A., Sadiq Jajja, M. S. & Farooq, S. (2020). Implementation of digital manufacturing technologies: Antecedents and consequences. *International Journal of Production Economics, 229*, 107748. https://doi.org/10.1016/j.ijpe.2020.107748

[4] Choi, S., Jung, K. & Noh, S. D. (2015). Virtual reality applications in manufacturing industries: Past research, present findings, and future directions. *Concurrent Engineering Research and Applications, 23*(1), 40-63. (Accessed 2024-03-19) https://doi.org/10.1177/1063293X14568814

[5] Armstrong, M. & Taylor, S. (2017). *Armstrong's Handbook of Human Resource Management Practice* (Fourteenth).

[6] Steuer, J. (1992). Defining Virtual Reality: Dimensions Determining Telepresence. *Journal of Communication, 42*(4), 73-93. https://doi.org/10.1111/j.1460-2466.1992.tb00812.x

[7] Riva, G. (2002). Virtual reality for health care: The status of research. *CyberPsychology and Behavior, 5*(3), 219-225. (Accessed 2024-03-20) https://doi.org/10.1089/109493102760147213

[8] Wilson, C. J. & Soranzo, A. (2015).The Use of Virtual Reality in Psychology: A Case Study in Visual Perception. *Computational and Mathematical Methods in Medicine, 2015*, Article ID 151702. https://doi.org/10.1155/2015/151702

[9] Mcmahan, R. P., Bowman, D. A., Zielinski, D. J. & Brady, R. B. (2012). Evaluating display fidelity and interaction fidelity in a virtual reality game. *IEEE Transactions on Visualization and Computer Graphics, 18*(4), 626-633. https://doi.org/10.1109/TVCG.2012.43

[10] Sanchez-Vives, M. V. & Slater, M. (2005). From presence to consciousness through virtual reality. *Nat Rev Neurosci, 6*, 332-339. https://doi.org/10.1038/nrn1651

[11] Freeman, D., Reeve, S., Robinson, A., et al. (2017). Virtual reality in the assessment, understanding, and treatment of mental health disorders. *Psychological Medicine, 47*(14), 2393-2400. https://doi.org/10.1017/S003329171700040X

[12] Hacmun, I., Regev, D. & Salomon, R. (2018). The Principles of Art Therapy in Virtual Reality. *Frontiers in Psychology, 9*, 398925 (Accessed 2024-03-20). https://doi.org/10.3389/fpsyg.2018.02082

[13] Di Pietro, R., Campanile, D. G. & Distefano, S. (2019). Virtual Study Partner: A Cognitive Training Tool in Education. *2019 IEEE International Conference on Smart Computing (SMARTCOMP)*, Washington, DC, USA, 192-197. https://doi.org/10.1109/SMARTCOMP.2019.00052

[14] Hugget, C. (2018). *Virtual Training Basics*. American Society for Training & Development.

[15] Mantovani, F., Castelnuovo, G., Gaggioli, A. & Riva, G. (2003). Virtual reality training for health-care professionals. *CyberPsychology and Behavior, 6*(4), 389-395. (Accessed 2024-03-20). https://doi.org/10.1089/109493103322278772

[16] Slavković, A. & Slavković, V. (2019). The importance of training in contemporary organizations. *Menadzment u hotelijerstvu i turizmu, 7*(2), 115-125. (in Serbian) https://doi.org/10.5937/menhottur1902115S

[17] Oberhauser, R., Baehre, M. & Sousa, P. (2023). VR-EvoEA+BP: Using Virtual Reality to Visualize Enterprise Context Dynamics Related to Enterprise Evolution and Business Processes. In: Shishkov, B. (eds) *Business Modeling and Software Design, BMSD 2023. Lecture Notes in Business Information Processing, vol 483*. Springer, Cham. (Accessed 2024-03-25) https://doi.org/10.1007/978-3-031-36757-1_7

[18] Heinonen, M., Ojanen, V. & Hannola, L. (2017). Adoption of VR and AR technologies in the enterprise. *ISPIM Innovation Conference, 95*. https://lutpub.lut.fi/handle/10024/ 144092 (Accessed 2024-03-25)

[19] Checa, D. & Bustillo, A. (2020). Advantages and limits of virtual reality in learning processes: Briviesca in the fifteenth century. *Virtual Reality, 24*(1), 151-161. (Accessed 2024-03-25). https://doi.org/10.1007/s10055-019-00389-7

[20] Kumar, N., Vidyapeeth, N. B., Deshmukh, J., Gavade, B., Tandale, P. G. & Nrip, N. K. (2023). Virtual Reality in Education. *International Journal of Multidisciplinary Research Transactions, 5*(3), 41-47. https://doi.org/10.5281/zenodo.7748468

[21] Chattha, U. A., Janjua, U. I., Anwar, Madni, F., T. M., Cheema, M. F. & Janjua, S. I. (2020). Motion Sickness in Virtual Reality: An Empirical Evaluation. *IEEE Access, 8*, 130486-130499. https://doi.org/10.1109/ACCESS.2020.3007076

[22] Daniel, E. (2016). The Usefulness of Qualitative and Quantitative Approaches and Methods in Researching Problem-Solving Ability in Science Education Curriculum. *Journal of Education and Practice, 7*(15), 91-100. (Accessed 2024-03-20) https://files.eric.ed.gov/fulltext/EJ1103224.pdf

[23] Colligan, M. J. & Cohen, A. (2009). The role of training in promoting workplace safety and health. In Barling, J. & Frone, M. R. (Eds.). *The psychology of workplace safety* (pp. 223-248). American Psychological Association. https://doi.org/10.1037/10662-011

[24] Ishwarya, R. (2023). A Study on Workplace Safety and Worker Protection in Organization. *International Journal of Creative Research Thoughts (IJCRT), 11*, c572-c578. https://ijcrt.org/papers/IJCRT2305338.pdf (Accessed 2024-03-20)

[25] Scorgie, D., Feng, Z., Paes, D., Parisi, F., Yiu, T. W. & Lovreglio, R. (2024). Virtual reality for safety training: A systematic literature review and meta-analysis. *Safety Science, 171*, 106372. https://doi.org/10.1016/j.ssci.2023.106372

[26] Radianti, J., Majchrzak, T. A., Fromm, J. & Wohlgenannt, I. (2020). A systematic review of immersive virtual reality applications for higher education: Design elements, lessons learned, and research agenda. *Computers and Education, 147*, 103778. https://doi.org/10.1016/j.compedu.2019.103778

[27] Cheng, K. H. & Tsai, C. C. (2019). A case study of immersive virtual field trips in an elementary classroom: Students' learning experience and teacher-student interaction behaviors. *Computers and Education, 140*, 103600. https://doi.org/10.1016/j.compedu.2019.103600

[28] Carruth, D. W. (2017). Virtual reality for education and workforce training. In: *ICETA 2017 - 15th IEEE International Conference on Emerging eLearning Technologies and Applications, Proceedings*. Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/ICETA.2017.8102472

[29] Gao, Y., Gonzalez, V. A. & Yiu, T. W. (2019). The effectiveness of traditional tools and computer-aided technologies for health and safety training in the construction sector: A systematic review. *Computers and Education, 138*, 101-115. https://doi.org/10.1016/j.compedu.2019.05.003

[30] Freeman, D., Haselton, P., Freeman, J., Spanlang, B., Kishore, S., Albery, E., Denne, M., Brown, P., Slater, M. & Nickless, A.

(2018). Automated psychological therapy using immersive virtual reality for treatment of fear of heights: a single-blind, parallel-group, randomised controlled trial. *The Lancet Psychiatry, 5*(8), 625-632. https://doi.org/10.1016/S2215-0366(18)30226-8

[31] Van Wyk, E. & De Villiers, R. (2009). Virtual reality training applications for the mining industry. In: *Proceedings of AFRIGRAPH 2009: 6th International Conference on Computer Graphics, Virtual Reality, Visualisation and Interaction in Africa*, 53-64. https://doi.org/10.1145/1503454.1503465

[32] Rokooei, S., Shojaei, A., Alvanchi, A., Azad, R. & Didehvar, N. (2023). Virtual reality application for construction safety training. *Safety Science, 157*, 105925. https://doi.org/10.1016/j.ssci.2022.105925

[33] Kalkan, Ö. K., Karabulut, Ş. & Höke, G. (2021). Effect of Virtual Reality-Based Training on Complex Industrial Assembly Task Performance. *Arabian Journal for Science and Engineering, 46*(12), 12697-12708. (Accessed 2024-03-21) https://doi.org/10.1007/s13369-021-06138-w

[34] Abidi, M. H., Al-Ahmari, A., Ahmad, A., Ameen, W. & Alkhalefah, H. (2019). Assessment of virtual reality-based manufacturing assembly training system. *International Journal of Advanced Manufacturing Technology, 105*(9), 3743-3759. (Accessed 2024-03-21) https://doi.org/10.1007/s00170-019-03801-3

[35] Rauh, S. F., Koller, M., Schäfer, P., Bogdan, C. & Viberg, O. (2021). MR On-SeT: A Mixed Reality Occupational Health and Safety Training for World-Wide Distribution. *International Journal of Emerging Technologies in Learning, 16*(5), 163-185. (Accessed 2024-03-21) https://doi.org/10.3991/ijet.v16i05.19661

[36] Zawadzki, P., Zywicki, K., Bun, P. & Gorski, F. (2020). Employee Training in an Intelligent Factory Using Virtual Reality. *IEEE Access, 8*, 135110-135117. https://doi.org/10.1109/ACCESS.2020.3010439

[37] Rahmalan, H. (2020). Development of Virtual Reality Training for Fire Safety Education. *International Journal of Advanced Trends in Computer Science and Engineering, 9*(4), 5906-5912. https://doi.org/10.30534/ijatcse/2020/253942020

[38] Zhao, D., Mccoy, A., Kleiner, B. & Feng, Y. (2016). Integrating safety culture into OSH risk mitigation: a pilot study on the electrical safety. *Journal of Civil Engineering and Management, 22*(6), 800-807. https://doi.org/10.3846/13923730.2014.914099

[39] Avveduto, G., Tanca, C., Lorenzini, C., Tecchia, F., Carrozzino, M. & Bergamasco, M. (2017). Safety training using virtual reality: A comparative approach. In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Springer Verlag, 148-163. (Accessed 2024-03-24) https://doi.org/10.1007/978-3-319-60922-5_11

[40] Zhao, D. & Lucas, J. (2015). Virtual reality simulation for construction safety promotion. *International Journal of Injury Control and Safety Promotion, 22*(1), 57-67. (Accessed 2024-03-21) https://doi.org/10.1080/17457300.2013.861853

[41] Joshi, S., Hamilton, M., Warren, R., Faucett, D., Tian, W., Wang, Y. & Ma, J. (2021). Implementing Virtual Reality technology for safety training in the precast/prestressed concrete industry. *Applied Ergonomics, 90*, 103286. https://doi.org/10.1016/j.apergo.2020.103286

[42] Górski, F., Zawadzki, P., Buń, P. & Starzyńska, B. (2018). Virtual reality training of hard and soft skills in production. In: *Proceedings - Web3D 2018: 23rd International ACM Conference on 3D Web Technology*. Association for Computing Machinery, Inc. (Accessed 2024-03-21) https://doi.org/10.1145/3208806.3219787

[43] Zhao, J., Wallgrün, J. O., Lafemina, P. C., Normandeau, J. & Klippel, A. (2019). Harnessing the power of immersive virtual reality - visualization and analysis of 3D earth science data sets. *Geo-Spatial Information Science, 22*(4), 237-250. https://doi.org/10.1080/10095020.2019.1621544

[44] Akdere, M., Jiang, Y. & Acheson, K. (2023). To simulate or not to simulate? Comparing the effectiveness of video-based training versus virtual reality-based simulations on interpersonal skills development. *Human Resource Development Quarterly, 34*(4), 437-462. (Accessed 2024-03-21) https://doi.org/10.1002/hrdq.21470

[45] Ahir, K., Govani, K., Gajera, R. & Shah, M. (2020). Application on Virtual Reality for Enhanced Education Learning, Military Training and Sports. *Augmented Human Research, 5*(1). (Accessed 2024-03-20) https://doi.org/10.1007/s41133-019-0025-2

[46] Svoboda, M., Gangur, M. & Mičudová. K. (2019). *Statistické zpracování dat*. Západočeská univerzita v Plzni. (in Czech)

[47] Svatošová, L. & Kába, B. (2020). *Statistické metody II*. Česká zemědělská univerzita v Praze, Katedra statistiky. (in Czech)

[48] Svatošová, L. & Prášilová, M. 82016). *Statistické metody v příkladech*. Česká zemědělská univerzita v Praze, Katedra statistiky. (in Czech)

[49] Mičudová, K., Gangur, M., Svoboda, M. & Říhová, P. (2017). *Základy statistiky a pravděpodobnosti*. Západočeská univerzita, Ekonomická fakulta, 225. (in Czech)

[50] Svatošová, L. & Kába, B. (2014). *Statistické metody I*. Česká zemědělská univerzita v Praze, Katedra statistiky. (in Czech)

Authors' contacts:

David Krákora, Ing.
University of West Bohemia,
Univerzitní 2732,
301 00 Plzeň 3, Czech Republic
krakorad@fst.zcu.cz

Petr Hořejší, Doc. Ing. PhD
(Corresponding author)
University of West Bohemia,
Univerzitní 2732,
301 00 Plzeň 3, Czech Republic
tucnak@fst.zcu.cz

Andrea Šimerová, Ing.
University of West Bohemia,
Univerzitní 2732,
301 00 Plzeň 3, Czech Republic
simerova@fst.zcu.cz

ISSN 1846-6168 (Print), ISSN 1848-5588 (Online)
https://doi.org/10.31803/tg-20230823142320

Professional paper
Received: 2023-08-23, Accepted: 2024-04-26

Designing and Prototyping of a Hybrid Robotic Lawnmower

Martin De Bona, Denis Kotarski*

Abstract: In this paper, the designing and prototyping of a hybrid robotic lawnmower is presented. The main system components required for functionality are briefly described. All parts of the system are selected in order to enable coordinated operation of electrical and mechanical components. Robotic lawnmower prototyping is preceded by a design process that consists of a phase of selecting the main components and CAD modeling of parts and components. The goal is to position components and parts of the chassis with maximum compactness. The main subsystems of the prototype, the differential drive configuration, and the hybrid subsystem with a belt transmission for starting the alternator and the lawnmower blade are shown. The parts of the subsystem were tested and the final testing of the robotic lawnmower in the case of remote control was carried out. This prototype was made with total costs many times lower than similar commercial robotic lawnmowers.

Keywords: CAD; differential drive configuration; hybrid robotic lawnmower; prototyping; transmission

1 INTRODUCTION

Lawn mowing is considered a boring and tedious routine household task, therefore it is one of the more promising robotic fields. One of the classifications of lawnmowers is according to the axis of rotation of the blade, where two main types are lawnmowers with a horizontal reel and rotary vertical lawnmowers which will be further discussed in this paper. The blades on the rotor achieve a clean cut and are adapted to different cutting heights. Furthermore, with regard to the drive and the energy source, there are manual, electric, or internal combustion engine (ICE) mowers. Rotary mowers are mostly powered by an ICE or an electric motor, where the main power source is engaged for rotating the cutting blades. The opening on the mower body is used to throw out the mowed grass, and some versions have an attached collection basket. There are already many commercial robots for mowing grass, and many prototypes [1, 2]. Some predictions indicate that robotic lawnmowers will further push the boundaries in robotics, as a kind of continuation after the great expansion of robotic vacuum cleaners. The significance of such research lies in the potentially large market in the field of personal robotics [3].

These types of robots usually come with similar features such as the ability to work within a defined space and the ability to avoid static obstacles. In the initial stages of robotic lawnmower research, modeling and motion mechanisms were considered, as shown in the paper [4]. A robotic lawnmower can be remotely controlled, semi-autonomous [5], or, in the most sophisticated version, an autonomous robot used to cut grass. Over the years, there have been numerous advances in lawnmower technology. In the early stages of using robotic lawnmowers, a sensor wire was used to define the boundaries of the mowed area, just as Lawrence Bellinger did in his patent, back in 1969 [6]. In this case, the wire needs to be placed around the edges of the lawn area as well as around any static obstacles. This requires additional work by the operator, which causes a higher cost of mowing. The industry also got involved in the development of robotic lawnmowers, and Husqvarna presented the Solar Mower lawnmower model in 1995, and a step further was taken in 1998 when the same company presented the Automower with a rechargeable battery and the ability to work in various weather conditions.

Nevertheless, this area of robotics is still suitable for further development [7], especially from the aspect of autonomy, given the methods of positioning, detection, and avoidance of obstacles. Due to specific circumstances, fully autonomous lawnmowers are often replaced with human control and remote control. Obstacle avoidance algorithms work with static obstacles, but there are more challenges with dynamic obstacles that are common in mowing grass. Although there are many problems, there are more and more solutions. Today's models of robotic lawnmowers can be controlled in a variety of ways, such as mobile application control. Unlike the models used in the past, modern lawnmowers do not need physical borders along the edge of the plot but use sensor systems that include GPS, cameras, distance sensors, and other types of sensors [8, 9]. With the help of a sensor package, they determine the work area intended for mowing and avoid obstacles alongside other features. In paper [10], a solar photovoltaic (PV) panel-powered smart lawn mower, employing an Internet of Things (IoT)-based control system is presented. New features increase the price and robotic lawnmower systems are expensive, which is probably the main barrier for a larger number of users. From an economic point of view, the paper [11] presents an evaluation of life cycle influence and costs associated with autonomous solutions in the agricultural sector.

As control systems and sensors advance, hardware also progresses in parallel. Hardware solutions primarily rely on factors such as the terrain configuration, plot size, and the dimensions of the grass and plants. For instance, robotic lawnmowers find application in orchards [12, 13], artichoke fields [14], and various other crop fields. Although there are numerous works related to the design of robotic lawnmowers, they are mostly conceptual designs or scaled prototypes [15]. As an illustration, the paper [16] introduced the conceptual synthesis of a multifunctional lawnmower and demonstrated a prototype. Additionally, the same group of authors presented in the paper [17] the conceptual design of a smart lawnmower for uneven grassland using the Theory of Inventive Problem Solving (TRIZ). A review of the literature

revealed a lack of papers in which the process of designing a robotic lawnmower is described in more detail.

The goal of this research is to create a prototype of a hybrid robotic lawnmower based on an ICE that is used to turn the lawnmower's blade using a right-angle gear reducer and to propel an alternator through which the batteries are charged with electricity. A differential drive configuration, made with two electric motor drives and four wheels, was chosen for the drive system that realizes the movement of the lawnmower. After selecting the components, the construction process was carried out and the key assemblies of the system were presented. The created prototype was extensively tested and successfully performed the tasks. The considered hybrid system has great autonomy in terms of the length of mowing time and is suitable for use in larger areas such as orchards. The main advantage of the prototype compared to commercial lawnmowers is the much lower price.

2 THE MAIN COMPONENTS OF A HYBRID ROBOTIC LAWNMOWER PROTOTYPE

The presented lawnmower prototype works according to the operating principle of hybrid systems known in the automotive industry, where different types of energy and drive components are used. The gasoline four-stroke ICE is the main driver of all systems and at the same time the only source of kinetic energy necessary for the operation of the lawnmower blade, which converts from the chemical energy of 95-octane gasoline. A clone of the Honda GX200 engine was selected, which has a volume of 196 cc, a power of 5.5 kW, and a torque of 13.2 Nm with a horizontal output rotating shaft. The engine operation is controlled by a servomotor that moves the engine's throttle lever. Fig. 1 shows a 3D subassembly of the motor with a double pulley, servomotor, and rubber bumpers that mitigate unwanted vibrations that are transmitted to other parts of the chassis.

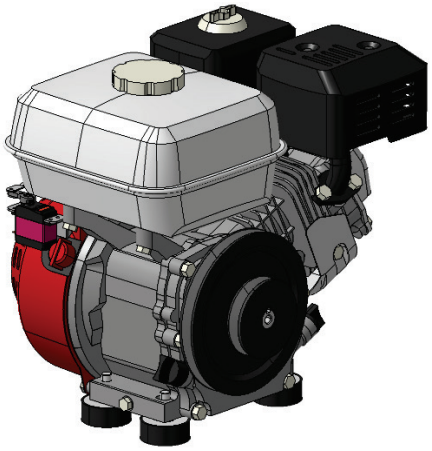


Figure 1 Internal combustion engine subassembly with servo motor, rubber bumpers, and a double pulley

The key part of the lawnmower, intended for the transfer of kinetic energy, necessary for turning the blade during mowing, is the angle gear reducer (manufacturer Coberg, model L-155J). For this prototype, a reducer with a transmission angle of 90°, and a gear ratio of 1:1.14, is used. The rotation axis of the gasoline engine rotates around the horizontal central axis, therefore, the rotation is transmitted to the vertical axis of the blade using the internal gears of the reducer. The rotation of the blade in a counterclockwise (CCW) direction is enabled. The input shaft of the angle gear reducer is a cardan shaft with 6 teeth, used in agricultural machines, while the output shaft is 25 mm in diameter. Fig. 2 shows a 3D model of the reducer in its original state on the upper part of the picture, and with modifications where the shafts are shortened and adapted, on the lower part of the picture.

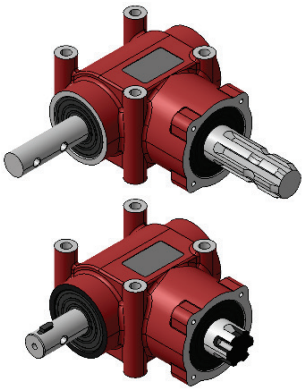


Figure 2 Angle gear reducer original version and revision

The main source of electricity for the movement of the lawnmower is the electricity generator, the so-called alternator. It is driven via a belt drive by a gasoline engine and converts the received kinetic energy into electricity. An alternator with an output DC voltage of 24 V was selected, and the charging current is 35 A, which is enough to start the electric motors of the drive configuration, as well as to recharge the lawnmower's battery at the same time. Two batteries are placed within the housing of the prototype. The voltage of these batteries is 12 V, and they have a capacity of 3 Ah. The batteries are connected in series, considering that the required voltage of the lawnmower electric motors is 24 V.

The prototype drive configuration consists of two electric motor actuators (manufacturer Vevor, model BY1016Z) which convert electrical energy into kinetic energy. Electric motors are collector designs, with graphite brushes. Considered actuators have a built-in gearbox to reduce the angle velocity on the output shaft, and to simultaneously increase output torque. According to the specifications, the maximum number of revolutions is 309 RPM, the nominal power of the motor is 350 W, and the operating voltage is 24 V. The motors are independent of each other, and they are mounted separately on the lawnmower as a left and right motor. The motors are controlled by a driver (H-bridge) which receives a PWM signal based on which it controls the electric motors. This type of driver (control unit) is based on the principle of changing the polarity of DC collector motors, which can turn in one direction or the other. Drive configuration consists of two single-channel drivers.

In the testing phase of the hybrid robotic lawnmower prototype, a remote control system will be used, where information is transmitted via a wireless communication link,

using electromagnetic waves. Components were selected that are compatible with the six-channel Flysky FS-i6 transmitter. For the complete control of this prototype, five communication channels are used, two for electric motors, and one channel each for the linear actuator of the belt tensioner, the servomotor that controls the operation of the gasoline engine, and the relay for shutting down the gasoline engine.

3 DESIGNING A HYBRID ROBOTIC LAWNMOWER SYSTEM

3.1 Designing the Drive Configuration of the Robotic Lawnmower

The most common drive configuration of unmanned ground vehicles (UGV) or ground mobile robots is a differential drive where two actuators are used for the robot's movements. The differential drive enables translational motion but also rotation in place around the vertical axis, which is convenient considering the purpose of the presented robot. Most of these configurations consist of two wheels that are directly or via a transmission connected to the actuator. In this paper, the prototype is made as a differential drive with four wheels, as shown in Fig. 3. Different wheel velocities achieve rotation, i.e., turning the mower, while the same velocities cause straight-line motion. During straight-line motion, all four axles are equally loaded at the same time. Turning the lawnmower in operation is very simple, thus faster and more efficient mowing is possible, given that no uncut marks are left in place when turning. All-wheel drive contributes to mowing on uneven and hilly terrain, and in case of minimal slippage, all wheels exert equal forces and moments. The front and rear wheels on one side are therefore synchronized and are maximally close to each other for the compactness of the lawnmower. It is necessary to design a system for transferring energy from the electric motor to the wheels.

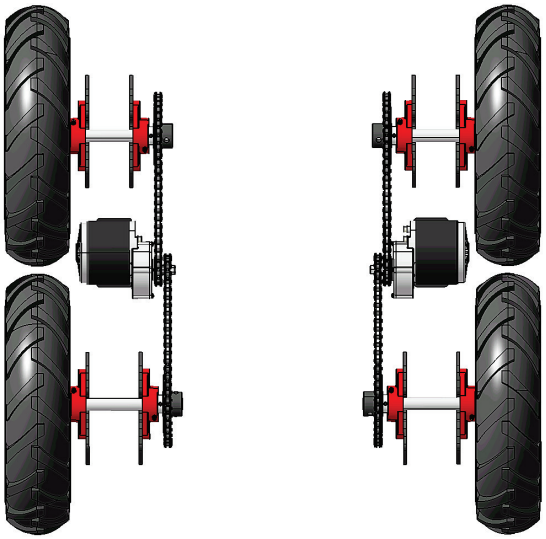


Figure 3 Prototype differential drive configuration [18]

Chain transmission is selected for the transmission of moments from the shaft of the electric motor to the shafts of the prototype driving wheels, which forms the basis of the differential drive of the robotic lawnmower. Three key components of chain transmission are the sprocket as the torque transmitter, the sprocket that receives the torque, and between them a flexible chain link. The chain transmission is relatively easy to install, has high efficiency, and compared to gears, works better under shock load conditions. In chain transmission, the working load is distributed among the teeth, and proper lubrication is required. A total of four chains of equal length are required for the proposed differential drive configuration, as each wheel uses a separate chain link. The front and rear wheels of one electric motor on each side of the lawnmower are driven synchronously. Since two wheels are synchronized on one side, it is necessary to design and manufacture a double sprocket of the electric motor shaft consisting of 410H sprockets with 10 teeth. In order to obtain a slower speed of movement of the mower and higher torque, the sprocket of each wheel has 32 teeth, therefore the ratio of teeth between the sprockets is 10:32. Larger sprockets are bolted to the wheel shafts via flanges located at the ends of the axles and attached with a pin. The sprockets of the wheel shaft and the electric motor are connected by 415H chains. Fig. 4 shows the subassembly of one side of the differential drive configuration of the proposed robotic lawnmower prototype.

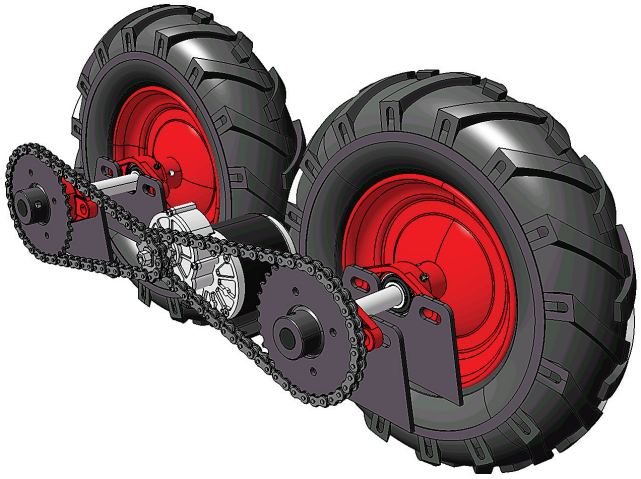


Figure 4 Subassembly of one side of the differential drive configuration [18]

As shown, one side of the mower's differential drive consists of two wheels, drive shafts, pairs of bearings, flanges, and a chain transmission connected to the electric motor actuator. The wheels consist of a metal rim on which a pneumatic tire under pressure is mounted. The width of the tire is 100 mm, while the outer diameter is 400 mm. The wheel drive shaft is attached, using a split pin, to the axle hole of the wheel. It is made by turning from a structural steel round bar, and the shaft diameter is 22 mm on the wheel side and 20 mm on the bearing side. The shaft is connected to the main housing of the robotic lawnmower utilizing bearings that serve as support for the moving parts of the robot and for the proper direction of kinetic energy. The model used for the prototype is a ball bearing in a housing, model UCFL204, which is mounted with M12 screws on the bearing supports. Vertical bearing support was constructed, taking into account

the dimensions and shape of the bearing housing, and enabling adjustment of the chain transmission tension. Simplified expression for the translational speed of the wheel

$$v_V = r_V \cdot i_{CT} \cdot \omega_M, \quad (1)$$

where r_V is the wheel radius, i_{CT} is the gear ratio of the chain transmission, and ω_M is the angular velocity of the DC motor with gearbox. In the case of the declared motor speed of 309 RPM, at the declared torque of 13 Nm, it turns out that the translational speed is $v_V = 2$ m/s.

3.2 Designing the Lawnmower Blade Mechanism

The angle gear reducer is a mechanical assembly for the transmission of kinetic energy, where a blade adapter is mounted on the output shaft of the reducer. Before mounting on the lower vertical output shaft, the blade adapter must be redesigned and manufactured as shown in Fig. 5. The lawnmower blade is a mechanical metal piece with sharp edges and is mounted with screws on the blade adapter. The blade model of this prototype has a symmetrical shape, length of 580 mm, width of 55 mm, and thickness of 5 mm. The blade rotates around the vertical axis and, passing through the grass and vegetation cuts the plants hit by the sharpened edge. The proposed prototype does not use a grass collector, therefore the blade is made without rear ejectors. Due to a rotary motion, centrifugal forces throw the cut plants out of the blade area.

Figure 5 Lawnmower blade mounting

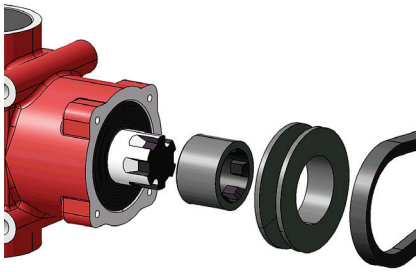


Figure 6 Angle gear reducer pulley mounting

The transmission of kinetic energy is made possible by the internal parts of the angle reducer, which transfer the rotation of the input shaft to the rotation of the blade around the vertical axis of the lawnmower. To transfer the kinetic energy from the gasoline engine to the reducer, a belt transmission was chosen. The transmission ratio of the belt transmission between the gasoline engine and the reducer is 1:1. Subassembly exploded view of the angle reducer input shaft, pulley, and section of the belt is shown in Fig. 6.

3.3 Designing the Belt Transmission of a Hybrid Lawnmower

From the aspect of designing the hybrid system of the robotic lawnmower, it is necessary to design the belt transmission [19] that connects the gasoline engine with the angle reducer and the alternator, therefore two belts are needed. The belt transmission system consists of conical pulleys and V-belts. Two 13 mm wide V-belts were chosen, models AVX 13x1100 and AVX 13×670, where the shorter belt for the angle reducer is 670 mm long and the longer belt for the alternator is 1100 mm long. The belts are reinforced with additional outer plating for strength and toughness. The basic pulley is a specially made double pulley and is mounted on the output shaft of the gasoline engine, from which it transmits moments to the other belt transmission parts. As stated, the transmission ratio between the gasoline engine pulleys and the angle reducer is 1:1, and the outer diameters of the pulleys are 80 mm. A ratio of 2:1 was chosen for torque transmission to the alternator. The pulley on the alternator is 80 mm in diameter, while that on the gasoline engine is 160 mm. The reason for such a ratio is the need for a higher nominal rotational speed of the alternator. The electrical energy of the prototype must be sufficient for the electric drive of all its components, and simultaneously for charging two batteries.

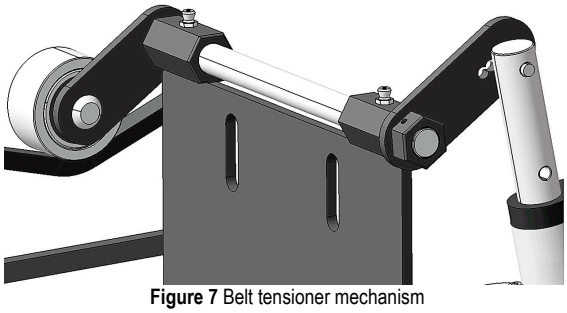


Figure 7 Belt tensioner mechanism

Belt tension is important when designing a belt transmission. The alternator belt is permanently tensioned, and it is tightened by moving the alternator away from the gasoline engine, via a lever that is an integral part of the alternator support. The belt tensioner of the angle reducer is specially designed keeping in mind the engaging and disengaging of the lawnmower blade. This reducer belt tensioner model consists of a lever that presses the belt and is mounted on the reducer bracket as shown in Fig. 7. It is mounted on top of the angle reducer bracket using two drilled M14 extended metric nuts. One end of the tensioner moves the linear actuator via a lever, while on the other side, on another lever, there are ball bearings inside a metal cover. The cover freely rotates on the outside of the angle reducer belt return connection, and in this way performs belt tensioning and therefore blade rotation. A used linear actuator is a device that performs work via an output shaft

that pulls or pushes the tensioner lever. It consists of an electric motor and gears, and the maximum pushing force of this prototype actuator is 800 N. The linear actuator is controlled by the electronic speed controller (ESC) based on the received signal via the remote control receiver. Fig. 8 shows the main components and parts that are mounted on the robot chassis. If we assume that there are no losses, expression for the moment of the blade is

$$M_{\mathrm{B}} = i_{\mathrm{AR}} \cdot i_{\mathrm{BT}} \cdot M_{\mathrm{ICE}}, \tag{2}$$

where i_{AR} is the ratio of the angle gear reducer, i_{BT} is the ratio of the belt transmission, and M_{ICE} is the ICE torque.

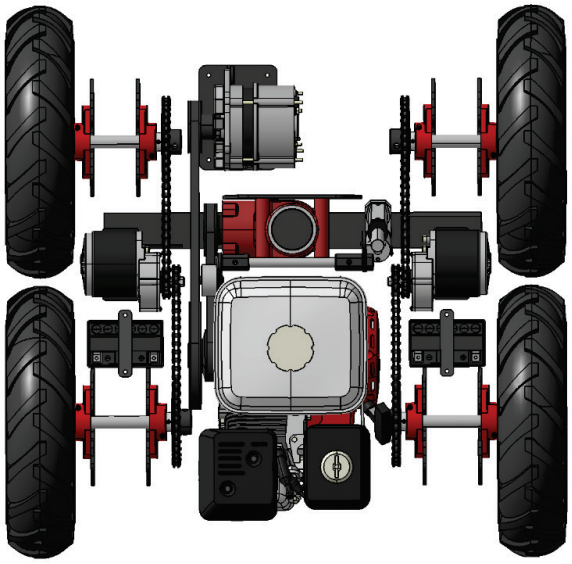


Figure 8 Arrangement of system components

3.4 Chassis of the Hybrid Robotic Lawnmower Prototype

The chassis of the hybrid robotic lawnmower is designed, which has the function of carrying all the components of the system. The back and front of the housing are completely open, which allows the grass and plants to be ejected smoothly, and prevents the blade from being blocked. A hole with a diameter of 62 mm is cut in the middle of the housing, through which the lower vertical shaft of the angle gear reducer passes. The chassis dimensions are 630 mm wide and 700 mm long. The vertical shaft bearing supports, as well as the angle reducer supports, are welded to the housing. Each wheel axle contains two supports. The brackets of the angle reducer also have the role of moving the reducer along the vertical axis, and this consequently determines the cutting height of the blade. For the production, a template was made which is shown in Fig. 9.

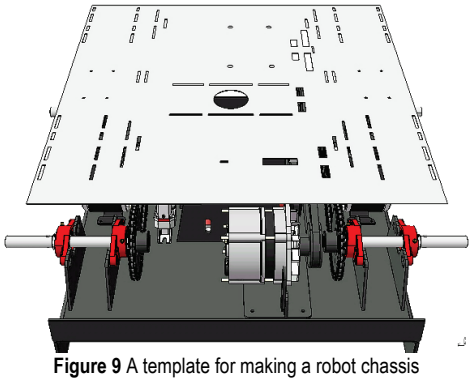


Figure 9 A template for making a robot chassis

The system components are mounted to the chassis of the lawnmower. The compactness of the entire system is achieved, and Fig. 10 shows the assembly of the hybrid robotic lawnmower. The height of the lower part of the chassis, measured from the ground, is 125 mm. Therefore, the height of the blade is set at approximately half the height of the chassis. Since the height of the blade can be moved, the lowest position of the blade is 58 mm, while the highest cutting height of the blade is 83 mm.

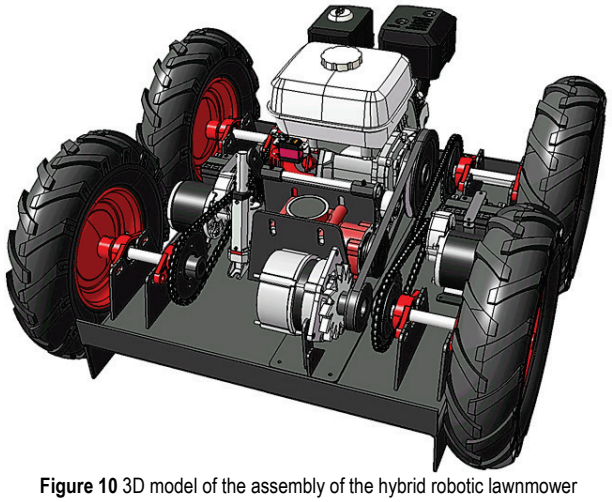


Figure 10 3D model of the assembly of the hybrid robotic lawnmower

4 PROTOTYPING AND TESTING A ROBOTIC LAWNMOWER SYSTEM

In the prototyping phase, modification and production of the necessary parts were carried out according to the 3D models and related documentation. First, chain and belt transmission components modifications and parts needed to mount the components on the chassis of the robotic lawnmower were made. Turning and laser cutting were used in the production phase. Then, laser cutting, bending, and welding of the supports were used in the stage of making the chassis of the robotic lawnmower. The final external dimensions of the lawnmower chassis after bending are 630 mm wide and 700 mm long. After the parts of the system were made, the assembly of the components and parts was done, and Fig. 11 shows the final assembly of the hybrid robotic lawnmower.

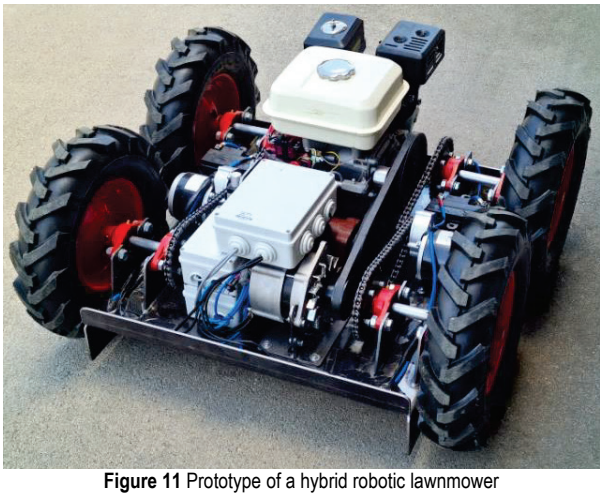


Figure 11 Prototype of a hybrid robotic lawnmower

After the assembly of the prototype, for the purpose of final implementation, testing of individual mechanical and electrical subsystems of the lawnmower prototype was carried out via remote control. The drive configuration of the robot was controlled using two channels of the remote control, one channel was used for the left wheels and the other channel for the right wheels. Gasoline engine control is tested for throttle regulation using a servo motor, and the switch for turning off the gasoline engine using a relay. Next, the engaging of the lawnmower blade was tested using a linear actuator and belt tensioner mechanism. After all the systems responded positively to the tests, the lawnmower was tested in real conditions. Fig. 12 shows the use of the presented prototype of the hybrid robotic lawnmower. Extensive testing of the prototype was done in conditions of different terrains and grass heights, and the prototype successfully performed the task of mowing. Through testing, minor defects were eliminated and segments of the prototype that were exposed during mowing were observed, for which protective housings will be made in future upgrades.

5 CONCLUSION

This paper presents the design and prototyping of a hybrid robotic lawnmower. The main components of the system are listed and explained. Parts of the electrical and mechanical systems of the prototype are connected and are mutually compatible when performing tasks. The process of designing the system through the phase of CAD modeling in the CATIA software package is presented. With regard to the selected components, chain, and belt transmission systems were designed that connect the electrical and mechanical systems. The complete prototype was made with total costs of less than 2,000 euros, which is many times less than commercial solutions, and further system upgrades are possible. After the manufacturing and assembly phase, all subsystems of this prototype were tested, and they all responded positively and successfully to the test. In further work, the increase in the degree of autonomy of the lawnmower prototype will be investigated in the form of the use of different sensors and control systems.

Figure 12 Prototype during testing

6 REFERENCES

[1] Daniyan, I., Balogun, V., Adeodu, A., Oladapo, B., Peter, J. K. & Mpofu, K. (2020). Development and Performance Evaluation of a Robot for Lawn Mowing. *Procedia Manufacturing*, 49, 42-48. https://doi.org/10.1016/j.promfg.2020.06.009.

[2] Balakrishna, K. & Rajesh, N. (2022). Design of remote monitored solar powered grasscutter robot with obstacle avoidance using IoT. *Global Transitions Proceedings*, 3(1), 109-113. https://doi.org/10.1016/j.gltp.2022.04.023.

[3] Hicks, R. & Hall, E. (2000). A Survey of Robot Lawn Mowers. *Proceedings of the SPIE*, 4197, 262-269. https://doi.org/10.1117/12.403770.

[4] Baloch, T. M. & Kae, T. T. C. (2008). Design and modelling a prototype of a robotic lawn mower. *International Symposium on Information Technology*, Kuala Lumpur, Malaysia, 1-5. https://doi.org/10.1109/ITSIM.2008.4631924.

[5] Patterson, A. E., Yuan, Y. & Norris, W. R. (2019). Development of User-Integrated Semi-Autonomous Lawn Mowing Systems: A Systems Engineering Perspective and Proposed Architecture. *AgriEngineering*, 1, 453-474. https://doi.org/10.3390/agriengineering1030033.

[6] United States Patents. United States Patent Inventor Spencer Lawrence Bellinger Sunnyvale, Calif. (https://patents.google.com/patent/US3550714A/en)

[7] Oberti, R. & Shapiro, A. (2016). Advances in Robotic Agriculture for Crops. *Biosystems Engineering*, 146, 1-2. https://doi.org/10.1016/j.biosystemseng.2016.05.010.

[8] Duan, S., Ma, H. Zhang, J. & Liu, G. (2022). Development of an Automatic Lawnmower with Real-Time Computer Vision for Obstacle Avoidance. *International Journal of Computational Methods, 19*(8). https://doi.org/10.1142/S0219876221420019.

[9] Franzius, M., Dunn, M., Einecke, N. & Dirnberger, R. (2017). Embedded Robust Visual Obstacle Detection on Autonomous Lawn Mowers. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 361-369. https://doi.org/10.1109/CVPRW.2017.50.

[10] Khaled Maaiuf, R., Tayyab, T., Adnan, K., Jehangir, A., Aun, H., Iftikhar, R., Ateeq Ur, R. & Seada, H. (2022). Implementation of an IoT-Based Solar-Powered Smart Lawn Mower. Wireless Communications and Mobile Computing, 2022, 1-12. https://doi.org/10.1155/2022/1971902.

[11] Saidani, M., Pan, E., Kim, H., Greenlee, A., Wattonville, J., Yannou, B., Leroy, Y. & Cluzel, F. (2020). Assessing the environmental and economic sustainability of autonomous systems: A case study in the agricultural industry. *Procedia CIRP*, 90, 209-214. https://doi.org/10.1016/j.procir.2020.02.123.

[12] Hossain, M. Z. & Komatsuzaki, M. (2021). Weed Management and Economic Analysis of a Robotic Lawnmower: A Case Study in a Japanese Pear Orchard. *Agriculture*, 11, 113. https://doi.org/10.3390/agriculture11020113.

[13] Li, J., Wang, S., Zhang, W., Li, H., Zeng, Y., Wang, T., Fei, K., Qiu, X., Jiang, R., Mai, C., et al. (2023). Research on Path Tracking for an Orchard Mowing Robot Based on Cascaded Model Predictive Control and Anti-Slip Drive Control. *Agronomy*, 13, 1395. https://doi.org/10.3390/agronomy13051395.

[14] Gagliardi, L., Sportelli, M., Frasconi, C., Pirchio, M., Peruzzi, A., Raffaelli, M. & Fontanelli, M. (2021). Evaluation of Autonomous Mowers Weed Control Effect in Globe Artichoke Field. *Applied Sciences, 11*(24). https://doi.org/10.3390/app112411658.

[15] Wijewickrama, D. R. D., Karunanayaka, K. M. H., Senadheera, H. W. P. & Godamulla, T. M. (2017). Fabrication of an Autonomous Lawn Mower Prototype with Path Planning and Obstacle Avoiding Capabilities. *CINEC Academic Journal*, 2, 30-33. https://doi.org/10.4038/caj.v2i0.51.

[16] Kang, C. Q., Ng, P. K. & Liew, K. W. (2021). The Conceptual Synthesis and Development of a Multifunctional Lawnmower. *Inventions*, 6, 38. https://doi.org/10.3390/inventions6020038.

[17] Kang, C. Q., Ng, P. K. & Liew, K. W. (2022). A TRIZ-Integrated Conceptual Design Process of a Smart Lawnmower for Uneven Grassland. *Agronomy*, 12, 2728. https://doi.org/10.3390/agronomy12112728.

[18] De Bona, M. & Kotarski, D. (2022). Koncept hibridne robotske kosilice. *Politehnika i Dizajn*, 10(02). (in Croatian) https://doi.org/10.19279/TVZ.PD.2022-10-2-03.

[19] Kong, L. & Parker, R. G. (2005). Steady Mechanics of Belt-Pulley Systems. *Journal of Applied Mechanics, 72*(1), 25-34. https://doi.org/10.1115/1.1827251.

Authors' contacts:

Martin De Bona
Karlovac University of Applied Sciences,
Josipa Jurja Strossmayera 9, 47000, Karlovac, Croatia
debonamartin@gmail.com

Denis Kotarski, PhD
(Corresponding author)
Karlovac University of Applied Sciences,
Josipa Jurja Strossmayera 9, 47000, Karlovac, Croatia
denis.kotarski@vuka.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

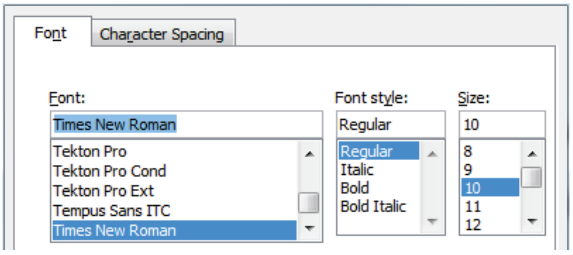


Figure 1 Text under the figure [1]
(Style: Arial Narraw, 8pt, Align Centre)

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

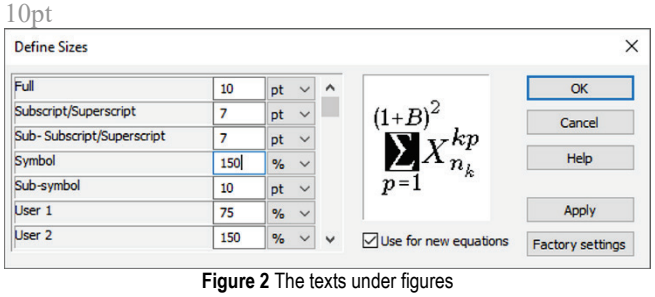


Figure 2 The texts under figures
(Style: Arial Narraw, 8pt, Align Centre)

10pt

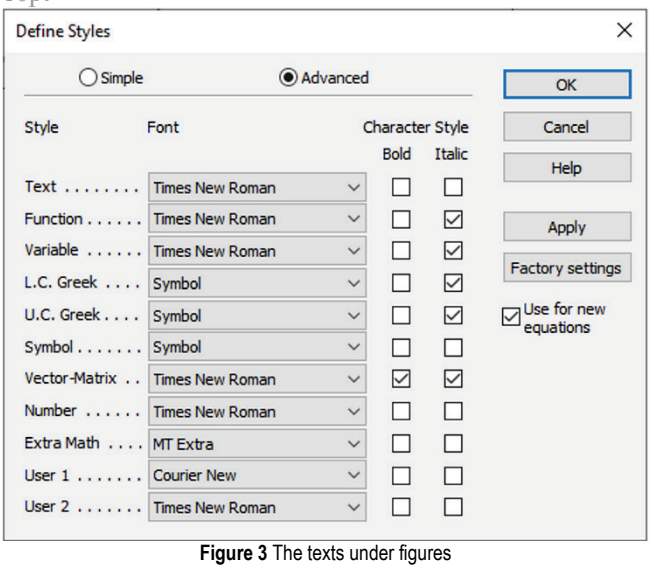


Figure 3 The texts under figures
(Style: Arial Narraw, 8pt, Align Centre)

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.


WIT
CONFERENCES
Call for Papers
HPSM/OPTI
2025
12th International Conference on High Performance and Optimum Design of Structures and Materials
10 - 12 June 2025 | Edinburgh, UK
Organised by
Wessex Institute, UK
University of A Coruña, Spain
Sponsored by
WIT Transactions on the Built Environment
www.witconferences.com/hpsm2025

TEHNIČKI GLASNIK / TECHNICAL JOURNAL – GODIŠTE / VOLUME 19 – BROJ / ISSUE 1

OŽUJAK 2025 / MARCH 2025 – STRANICA / PAGES 1-172

Sveučilište Sjever

SVEUČILIŠTE SJEVER / UNIVERSITY NORTH – CROATIA – EUROPE

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

TEHNICKIGLASNIK@UNIN.HR – HTTP://TEHNICKIGLASNIK.UNIN.HR