TEHNIČKI GLASNIK

TEHNIČKI GLASNIK / TECHNICAL JOURNAL – GODIŠTE / VOLUME 20 – BROJ / ISSUE 3

RUJAN 2026 / SEPTEMBER 2026 – STRANICA / PAGES 343-510

/

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 20 Issue 3
Varaždin, September 2026 Pages 343-510

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):
June 25, 2026	July 6, 2026	Septemeber 15, 2026

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-20240708150316

Original scientific paper
Received: 2024-07-08, Accepted: 2025-04-15

Low-Cost Attitude Determination Using Two GPS Antennas and IMU Data

Habib Kanberoğlu (Habib Ghanbarpourasl)*, Mukaddes Bolat

Abstract: This study explores methods for precise attitude estimation using carrier phases within global positioning systems (GPS). Carrier phase data is preferred over pseudorange data for short baseline distances due to its lower noise levels. However, resolving ambiguities through integer ambiguity resolution methods is essential for effectively utilizing carrier phases. One widely adopted approach, the LAMBDA method, is introduced, alongside a solution tailored for low-cost GPS systems and receivers with higher error. The novel method primarily focuses on utilizing only the fractional parts of the carrier phases while integrating constraints such as pitch and baseline constraints to enhance ambiguity detection accuracy. Comparative analysis with the LAMBDA approach demonstrates that the proposed method achieves remarkable accuracy, even in environments characterized by multipath interference.

Keywords: Attitude determination; GPS carrier phase; Integer ambiguity; LAMBDA method

1 INTRODUCTION

Satellite-based global positioning systems (GPS) boast diverse applications, ranging from providing precise navigation solutions for aircraft and maritime vessels to facilitating kinematic positioning across varying scales, from short distances to several kilometers. Moreover, rotation estimation becomes attainable with the appropriate configuration of GPS antennas within a unified system. GPS antennas are used to detect yaw, pitch, and roll by using measurements representing the distance between the receiver and satellites. Pseudorange measurements, which involve measuring the time it takes for the satellite signal to travel from the satellite to the receiver, can provide distance measurements. However, they are subject to errors caused by factors such as atmospheric delays and receiver clock errors. Carrier phase measurements, on the other hand, offer much higher precision because they directly measure the phase of the carrier wave. While carrier phase measurements offer higher precision, they pose the challenge of resolving the integer ambiguity [1]. The integer ambiguity arises because the receiver cannot directly determine how many complete cycles the carrier wave has traveled [2]. Instead, it measures the total phase change since the signal was transmitted. This total phase change includes both the fractional part (which can be accurately measured) and the integer part (which cannot be directly measured). Various methods have been developed to resolve these integer ambiguities and enable accurate positioning and rotation using carrier phase measurements in GPS applications.

These methods are typically categorized into two main groups. Motion-based techniques require the accumulation of data over a specific period during which significant changes occur either in the visible GPS constellation or in the observable platform rotations. Conversely, search-based methods rely on single epoch measurements to identify the most probable ambiguity combination, acknowledging that noise levels may sometimes lead to incorrect ambiguity identification [3].

Some of the earliest applied integer ambiguity resolution methods are the LSAST [4], FASF [5] and AFM [6], methods. LSAST and FASF are search-based methods that rely on sequential least squares estimation and operate under the assumption of normally distributed errors within a properly defined ambiguity search space. As stated by [7], these methods always include the correct set of ambiguities within the search space and obtain the least squares solution. D. Kim and R. B. Langley applied the OMEGA method by modifying the LSAST method to follow a more efficient process. It adopts two search space reduction processes (a scaling and a scanning process) to transform the search space and filter ambiguity candidates at multiple search levels. To achieve maximum efficiency, an optimization procedure is applied in a closed form before the search-validation step, determining parameters that minimize candidates under specific conditions [8].

On the other hand, the AFM method is a motion-based technique that utilizes only the fractional components of instantaneous carrier phase measurements. As a result, the ambiguity function values remain unaffected by full-cycle changes or cycle slips in the carrier phase. The adaptive nature inherent in AFM is characterized by a nonlinear and highly complex structure. Consequently, the process of searching for its maximum value demands a significant amount of computational capacity, creating a bottleneck in GPS attitude determination [9]. To reduce the computational burden, some enhanced AFM methods have been developed. Ref. [10] aims to minimize the number of grid points that need to be tested and determine the optimal search step size while [11] focuses on minimizing the number of test points to reduce computational load.

The ARCE method, or Ambiguity Resolution with Constraint Equation, is a sophisticated motion-based technique utilized to establish the connection between integer ambiguities. This correlation can be established separately from the user's location. These ambiguities consist of three distinct elements that are utilized to derive the user's position. Upon determining the independent elements, the remaining dependent elements can be calculated using the constraint equation. This technique provides a direct method for determining the dependent elements without requiring knowledge of the user's location. Furthermore, the

covariance information obtained from these ambiguities can be used as an additional criterion for rejection [12].

The aforementioned techniques exhibit both merits and demerits. Among the primary challenges encountered with integer ambiguity resolution methods are the computational overhead and intricate structures. These limitations often translate into inaccuracies within real-time attitude determination applications. Presently, the widely adopted LAMBDA method effectively mitigates this computational burden. LAMBDA method [13], which stands for "Least-squares Ambiguity Decorrelation Adjustment". LAMBDA involves a two-step process to compute integer least squares (ILS) ambiguity estimates. In the first step, the LAMBDA method employs Z-transformation to decorrelate ambiguities. This decorrelation step is crucial as it facilitates the effective search for the integer candidate solving the integer minimization problem. The second step in the LAMBDA method involves the search process, which is facilitated by the elimination of correlations obtained in the first step [14]. Test results for rotation estimation using the LAMBDA method can be found in [15].To enhance the success rate of the LAMBDA method, several baseline-constrained versions have been proposed [16–18]. While the original formulation of the LAMBDA method is applicable to both unconstrained and linearly constrained GNSS models, its baseline-constrained version, MLAMBDA [19], can be applied to second-order constrained GNSS models as well [2].

In this study, as observed, the LAMBDA method yields superior results in expensive GPS devices and suitable environments, while its accuracy diminishes in cheaper GPS devices and multipath environments. It is well known that other search-based and motion-based methods struggle with computational overhead and complex structures, as mentioned above. This study aims to achieve low-error attitude estimation even in multipath environments. To accomplish this, a New method has been developed, incorporating constraints such as baseline constraint and rotation constraint to accurately determine integer ambiguities.

1.1 GNNS Observation Model

It is important to note that there are multiple sources of error in carrier phase and pseudorange measurements, including ionospheric, tropospheric, satellite clock, receiver clock, and multipath errors. The undifferenced code and carrier phase measurements can be modeled as [20]:

$$p_a^s = \rho_a^s + c(dt_a - dt^s) + I_a^s + T_a^s + M_{a,P}^s + \varepsilon_{a,P}^s \phi_a^s \quad (1)$$

$$\phi_a^s = \rho_a^s + c(dt_a - dt^s) + I_a^s + T_a^s + M_{a,\phi}^s + \varepsilon_{a,\phi}^s + \lambda N_a^s \quad (2)$$

Where superscript s stays for representation of s-th satellite, and subscript a stays for a-th receiver. p_a^s and ϕ_a^s are the pseudorange and carrier phase measurements from s-th satellite to a-th receiver, ρ_a^s is the geometric range, dt_a and dt^s are clock biases of the a-th receiver and the s-th satellite, I_a^s and T_a^s are the ionospheric and trophospheric delay between the s-th satellite and the a-th receiver, $M_{a,P}^s$ and $M_{a,\phi}^s$ are the multi-path errors for the code and carrier phase of the a-th receiver, $\varepsilon_{a,P}^s$ and $\varepsilon_{a,\phi}^s$ are the errors that cannot be modeled and are assumed to be white Gaussian noises, λ is the wavelength and N_r^s is the unknown integer ambiguity.

To elimination the clock difference of satellites, we can compare signals from two antennas. Atmosphere and satellite errors are eliminated by single difference of measurements between receivers for short baselines. After a single difference (SD) between receivers *a* and *b*, the model can be written as:

$$\Delta p_{rb}^s = p_r^s - p_{rb}^s = \Delta\rho_{rb}^s + c(\Delta dt_{rb}) + M_{rb,P}^s + \varepsilon_{rb,P}^s \quad (3)$$

$$\Delta\phi_{rb}^s = \phi_r^s - \phi_b^s = \Delta\rho_{rb}^s + c(\Delta dt_{rb}) + M_{rb,\phi}^s + \varepsilon_{rb,\phi}^s - \lambda\Delta N_{rb}^s \quad (4)$$

Where Δp_{ab}^s and $\Delta\phi_{ab}^s$ represents the SD pseudorange and carrier phase measurements between a-th and b-th receiver respectively. Moreover, to eliminate receiver's clock biases, we perform double difference between two satellites measurements, resulting in the following double differenced (DD) measurements

$$\nabla\Delta p_{rb}^{sk} = \Delta\rho_{rb}^s - \Delta\rho_{rb}^k = \Delta\rho_{rb}^{sk} + M_{rb,P}^{sk} + \varepsilon_{rb,P}^{sk} \quad (5)$$

$$\nabla\Delta\phi_{rb}^{sk} = \Delta\rho_{rb}^s - \Delta\rho_{rb}^k = \Delta\rho_{rb}^{sk} + M_{rb,\phi}^{sk} + \varepsilon_{rb,\phi}^{sk} - \lambda\Delta N_{rb}^{sk} \quad (6)$$

Where, $\nabla\Delta p_{ab}^{sk}$ and $\nabla\Delta\phi_{ab}^{sk}$ represent the DD pseudorange and carrier phase measurements between the a-th and b-th receivers, and the s-th and k-th satellites. The linearized observation equation for both Eqs. (5) and (6) can be written in the following form [14]:

$$y = Aa + Br^n + \eta \quad (7)$$

Where y is the DD pseudorange and carrier phase measurements. The vector a is the DD carrier phase ambiguity which is an integer vector. Vector r^n is the baseline vector in local navigation frame which is represented by its three component *x*, *y* and *z*. A is a constant matrix, and matrix B is function of satellite unit vectors and wavelength. The vectors and matrices in Eq. (7) can be expressed as:

$$y=\begin{bmatrix}\nabla\Delta p_{rb}^{s1}-\nabla\Delta p_{rb}^{s2}\\ \nabla\Delta p_{rb}^{s1}-\nabla\Delta p_{rb}^{s3}\\ \vdots\\ \nabla\Delta p_{rb}^{s1}-\nabla\Delta p_{rb}^{sn}\\ \nabla\Delta \phi_{rb}^{s1}-\nabla\Delta \phi_{rb}^{s2}\\ \nabla\Delta \phi_{rb}^{s1}-\nabla\Delta \phi_{rb}^{s3}\\ \vdots\\ \nabla\Delta \phi_{rb}^{s1}-\nabla\Delta \phi_{rb}^{sn}\end{bmatrix} B=\begin{bmatrix}(e^{s1}-e^{s2})^T\\ (e^{s1}-e^{s3})^T\\ \vdots\\ (e^{s1}-e^{sn})^T\\ \frac{(e^{s1}-e^{s2})^T}{\lambda}\\ \frac{(e^{s1}-e^{s3})^T}{\lambda}\\ \vdots\\ \frac{(e^{s1}-e^{sn})^T}{\lambda}\end{bmatrix} A=\begin{bmatrix}0 & \cdots & 0\\ \vdots & \ddots & \vdots\\ 0 & \cdots & 0\\ 1 & \cdots & 0\\ \vdots & \ddots & \vdots\\ 0 & \cdots & 1\end{bmatrix} r^n=\begin{bmatrix}b_x\\ b_y\\ b_z\end{bmatrix} a=\begin{bmatrix}a^{s1}-a^{s2}\\ a^{s1}-a^{s3}\\ \vdots\\ a^{s1}-a^{sn}\end{bmatrix} \eta=\begin{bmatrix}\eta^{s1}-\eta^{s2}\\ \eta^{s1}-\eta^{s3}\\ \vdots\\ \eta^{s1}-\eta^{sn}\end{bmatrix}$$

Where n is the number of visible satellites. If there are n visible satellites, there are $n-1$ DD carrier phases and pseudoranges when the satellite 1 is selected as the pivot satellite. η is supposed to be white and Gaussian with zero mean and covariance Q_y. The matrix A consists of two parts: the first part is associated with the pseudorange measurements, while the second part which is represented by a unit matrix, is associated with carrier phase measurements. This choice of a unit matrix reflects the fact that carrier phase measurements inherently contain integer ambiguity.

The aim of this study is to find the integer ambiguity vector a and estimate the relative position vector between antennas by fixing the carrier phases, and then attitude determination will be applicable.

2 ATTITUDE DETERMINATION

The positions and distances of the antennas in the body frame are shown in Fig. 1. The position of the antennas are fixed and when the antenna 1 is taken as a reference antenna, the position vectors of antenna 2 and antenna 3 with respect to antenna 1 is:

$$r_2^b=\begin{bmatrix}L_{12}\\ 0\\ 0\end{bmatrix}\ r_3^b=\begin{bmatrix}L_{13}\cos\alpha\\ L_{13}\sin\alpha\\ 0\end{bmatrix} \tag{8}$$

Where L12 is the distance between antenna 1 and antenna 2, and L13 is the distance between antenna 1 and antenna 3, α is the angle between antenna 1, antenna 2 and antenna 3, r_2^b and r_3^b are the position of antenna 2 and 3 with respect to the antenna 1 in body frame.

The rotation matrix between body and navigation frame is defined as [21]:

$$C_b^n=\begin{bmatrix}\cos\psi & -\sin\psi & 0\\ \cos\psi & \sin\psi & 0\\ 0 & 0 & 1\end{bmatrix}\begin{bmatrix}\cos\theta & 0 & \sin\theta\\ 0 & 1 & 0\\ -\sin\theta & 0 & \cos\theta\end{bmatrix}\begin{bmatrix}1 & 0 & 0\\ 0 & \cos\varphi & -\sin\varphi\\ 0 & 0 & \cos\varphi\end{bmatrix}=C_1^nC_2^1C_b^2 \tag{9}$$

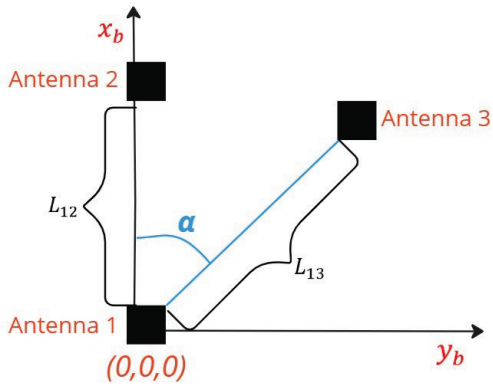


Figure 1 Position of antennas in the body reference system.

Where ψ, θ, and ϕ are the heading, pitch, and roll angles of the body frame with respect to the navigation frame, respectively. With reference to Fig. 1, relation between r^b and $r^n$is:

$$r_2^n=C_b^nb_2,\ r_2^b=\begin{bmatrix}L_{12}\\ 0\\ 0\end{bmatrix},\ r_2^n=\begin{bmatrix}x_2\\ y_2\\ z_2\end{bmatrix} \tag{10}$$

By substituting Eq. (9) into Eq. (10), one can write:

$$\begin{bmatrix}L_{12}\cos\psi\cos\theta\\ L_{12}\sin\psi\cos\theta\\ -L_{12}\sin\theta\end{bmatrix}=\begin{bmatrix}x_2\\ y_2\\ z_2\end{bmatrix} \tag{11}$$

From Eq. (11), the heading and pitch angles are found as follows:

$$\psi=\tan^{-1}\left(\frac{y_2}{x_2}\right) \tag{12}$$

$$\theta=\sin^{-1}\left(\frac{-z_2}{L_{12}}\right) \tag{13}$$

When we are using two antennas with configuration showed in Fig. 1, just heading and pitch angles are observable. If three antennas are positioned, then it becomes possible to estimate roll, pitch, and heading angles by three GPS antennas. If the rotation angles are known from Eq. (12) and Eq. (13), one can write:

$$r_3^b = \begin{bmatrix} L_{13}\cos\alpha \\ L_{13}\sin\alpha \\ 0 \end{bmatrix},\ r_2^3 = C_1^2 C_n^1 r_3^n \tag{14}$$

Then r_3^2 and r_3^b are known vectors, then:

$$r_3^b = C_2^b r_3^2,\ r_3^n = \begin{bmatrix} x_3 \\ y_3 \\ z_3 \end{bmatrix},\ r_3^b = \begin{bmatrix} L_{13}\cos\alpha \\ L_{13}\cos\varphi\sin\alpha \\ L_{13}\sin\psi\sin\alpha \end{bmatrix} \tag{15}$$

By comparing equivalent components, it resulted as:

$$\phi = \tan^{-1}\left(\frac{z_3}{y_3}\right) \tag{16}$$

Algebraic method is a classic method for attitude determination [22]. If two or more vectors are known in both frames, the rotation matrix between these frames can be found. For example, let r_1^b and r_2^b be two known vectors in the body frame. The equivalent of these vectors in the navigation frame are r_1^n and r_2^n. The transformation matrix C_b^n is calculated as follows:

$$N = C_b^N B \tag{17}$$

$$B = [r_1^b \quad r_2^b \quad r_1^b \times r_2^b],\ N = [r_1^n \quad r_2^n \quad r_1^n \times r_2^n] \tag{18}$$

Using least squares (LS) method, the rotation matrix can be estimated as:

$$C_B^N = (NB^T BB^T)^{-1} \tag{19}$$

After these calculations, the transformation matrix may not remain orthonormal. To make it orthonormal [23]:

$$\hat{C}_b^n = C_b^n (chol((C_b^n)^T C_b^n)^{-1} \tag{20}$$

chol is the 'cholesky factorization' of a matrix. After calculation of transformation matrix, the roll, heading, and pitch angles can be calculated as follows:

$$\emptyset = \tan^{-1}\left(\frac{\hat{C}_b^n(3,2)}{\hat{C}_b^n(3,3)}\right) \tag{21}$$

$$\psi = \tan^{-1}\left(\frac{\hat{C}_b^n(2,1)}{\hat{C}_b^n(1,1)}\right) \tag{22}$$

3 THE INTEGER AMBIGUITY RESOLUTION METHODS

In this section, the LAMBDA method, one of the integer ambiguity resolution methods that we mentioned in the introduction section, and our method are introduced.

3.1 LAMBDA

The LAMBDA Method is a well-known and widely used method for resolving integer ambiguities in carrier phase measurements. In the LAMBDA method, the least squares principle is often used to estimate the baseline and integer DD ambiguities [14]:

$$\min_{a,b} \| y - Aa - Bb \|_{Q_y}^2,\quad a \in Z^n, b \in R^p \tag{23}$$

Where $\|e\|_A^2$ is the second norm of a vector with weighting matrix A as:

$$\|e\|_A^2 = (e)^T A^{-1}(e) \tag{24}$$

Q_y is the covariance matrix of GNSS observations. These types of least squares problems are called integer least squares (ILS) and differ from standard least squares due to the integer behavior of (a ∈ Zn) [19]. The LAMBDA Method involves three main steps: float solution, integer ambiguity estimation, and fixed solution. In the float solution step, the integer aspect of the ambiguities a is ignored, and floating point solutions of a and the baseline vector r^n are calculated, and suppose the least squares solution of the problem defined by Eq. (7) is:

$$\begin{bmatrix} \hat{b} \\ \hat{a} \end{bmatrix}, \begin{bmatrix} Q_{\hat{b}\hat{b}} & Q_{\hat{b}\hat{a}} \\ Q_{\hat{a}\hat{b}} & Q_{\hat{a}\hat{a}} \end{bmatrix} \tag{25}$$

Where $\hat{a}$ and $\hat{b}$ are float solutions, $Q_{\hat{b}\hat{b}}$ and $Q_{\hat{a}\hat{a}}$ are covariance, and $Q_{\hat{a}\hat{b}}$ is cross-covariance between a and b.

In the integer ambiguity estimation step, the float solution $\hat{a}$ is used to calculate the corresponding integer ambiguity estimate:

$$\breve{a} = S(\hat{a}) \tag{26}$$

Where $S: R^n \rightarrow Z^n$ is an integer mapping from the n-dimensional space of real values to the n-dimensional space of integers. In this step, the mapping function S provides alternative options besides the ILS in Eq. (22), which covers various integer estimation methodologies. Prominent alternatives include integer bootstrap (IB) [24] and integer rounding (IR) [24]. Nonetheless, ILS stands out as the most efficacious among these approaches [25]. However, when ILS is used, the calculation can take a lot of time. The precision of the elements of the ambiguity vector can be increased, while simultaneously greatly reducing the correlation between ambiguities. This reparameterization is called Z-transformation, and it transforms the original DD ambiguities into a new set of ambiguities as follows [26]:

$$\hat{z} = Z^T \hat{a} \tag{27}$$

Where $\hat{a}$ is the estimation of a in the first step, Z is the transformation matrix and $\hat{z}$ is the new set of ambiguities. The corresponding covariance matrices are [27]:

$$Q_{\hat{z}\hat{z}} = Z^T Q_{\hat{a}\hat{a}} Z,\ \ Q_{\hat{b}\hat{z}} = Q_{\hat{b}\hat{a}} Z \tag{28}$$

An integer search is required to compute the ILS solution. After reparameterization, the search space is defined with the transformed ambiguities as follows [14]:

$$F(z) = (\hat{z} - z)^T Q_{\hat{z}\hat{z}}^{-1} (\hat{z} - z) \leq \chi^2, \quad z \in \mathbb{Z}^n \tag{29}$$

This search space is bounded by a hyper-ellipsoid, centered at zˆ, characterized by the covariance matrix $Q_{\hat{z}\hat{z}}$, and its dimensions determined by χ^2 . The application of the triangular decomposition of $Q_{\hat{z}\hat{z}}$, allows for the rewriting of the left-hand side of the quadratic inequality in Eq. (29) as a sum-of-weighted squares [19]:

$$\sum_{i=1}^{n} \frac{(\hat{z}_{i|I} - z_i)^2}{\sigma_{i|I}^2} \leq \chi^2 \tag{30}$$

The left side of the above equation corresponds to the conditional least-squares estimator, denoted as $\hat{z}_{i|I}$, which emerges when conditioning is applied to the integers z_i and $i = 1, 2, \ldots, n-1$. Employing the sum-of-squares format the subsequent step involves establishing n intervals for the search. These sequential intervals are defined as:

$$\begin{aligned} (\hat{z}_1 - z_1)^2 &\leq \sigma_1^2 \chi^2 \\ (\hat{z}_{2|1} - z_2)^2 &\leq \sigma_{2|1}^2 \left(\chi^2 - \frac{(\hat{z}_1 - z_1)^2}{\sigma_1^2}\right) \\ &\vdots \end{aligned} \tag{31}$$

Where $\sigma_i^2, i = 1,2,\ldots,n-1$are variance of i-th integer ambiguity. After this extensive integer estimation step, the fixed solution step will start. Here, the estimated integer ambiguities are employed to update the baseline vectors. The solution can be implemented directly as follows [14]:

$$\check{b} = \hat{b} - Q_{\hat{b}\hat{z}} Q_{\hat{z}\hat{z}}^{-1} (\hat{z} - \check{z}) \tag{32}$$

Where $\hat{b}$ is the float baseline vector, $\hat{z}$ is the transformed and float ambiguity vector, $\check{z}$ is the transformed and estimated true ambiguity vector and $\check{b}$ is the updated true baseline vector.

3.2 New Method

The method employed in this study can be categorized as search-based, as it seeks to determine the integer ambiguity within defined limits. However, selecting an appropriate limit can be a challenging task as it is crucial to balance computational burden and accuracy. Excessive computational costs may yield imprecise outcomes, while a narrow range may result in inaccuracies due to the potential exclusion of the correct solution. Therefore, careful consideration is essential when deciding on an appropriate search range to ensure that reliable and accurate results are obtained. To prevent this:

We are using only fractional parts of the carrier phases. Maximum of search area is around the L/λ [20].

Rotation constraints which are calculated by the IMU data [27] were added to pull the correct ambiguity from candidates.

The steps of the proposed method are as follows:

Step 1: Removal of the integer part of the DD carrier phase measurements

Thus, integer ambiguities will be in the range of the baseline length.

$$\nabla\Delta\widetilde{\emptyset} = \nabla\Delta\emptyset - \text{round}(\nabla\Delta\emptyset) \tag{33}$$

Here $\nabla\Delta\widetilde{\emptyset}$ is the fractional part of the DD carrier phase measurements and $\nabla\Delta\emptyset$ is the original carrier phase measurements.

Step 2: Candidate assignment.

The search area for any ambiguity is:

$$-\frac{L}{\lambda} < a_i < +\frac{L}{\lambda} \tag{34}$$

In this step candidate of integer ambiguities are assigned.

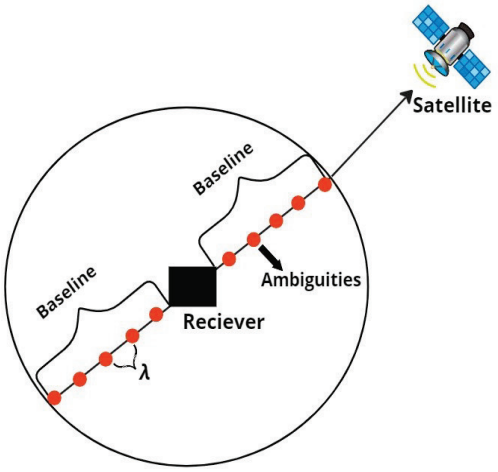


Figure 2 Diagram of possible ambiguities after fractional parts of DD carrier phases remain.

Step 3: Candidate base vectors.

For any candidate of a, the base vector is calculated using the LS method:

$$\hat{b}_j = (B^T R^{-1} B)^{-1} B^T R^{-1} (y - A\hat{a}_j), j = 1,2,\ldots,N_p \tag{35}$$

Where N_p is the number of total integer ambiguity candidates, R is the covariance matrix of DD carrier phase and pseudorange measurements. In the search step, instead of $\hat{a}$ in Eq. (34), all possible values shown in Fig. 2, are substituted one by one. For each of these candidate ambiguities, $\hat{b}_j(a)$ is calculated. When a low-cost GPS is used, the covariance of the base vector and integer ambiguity in float solution is high. Then when the search area is selected as shown in Step 2, the total number of candidates will reduce significantly. Although the number of candidates and the computational load have been reduced by taking fractional parts of the DD carrier phases, the number of candidates is still high due to the high error of low-cost GPSs. Therefore, it is reduced by adding constraint conditions to the candidate $\hat{b}(\hat{a})$.

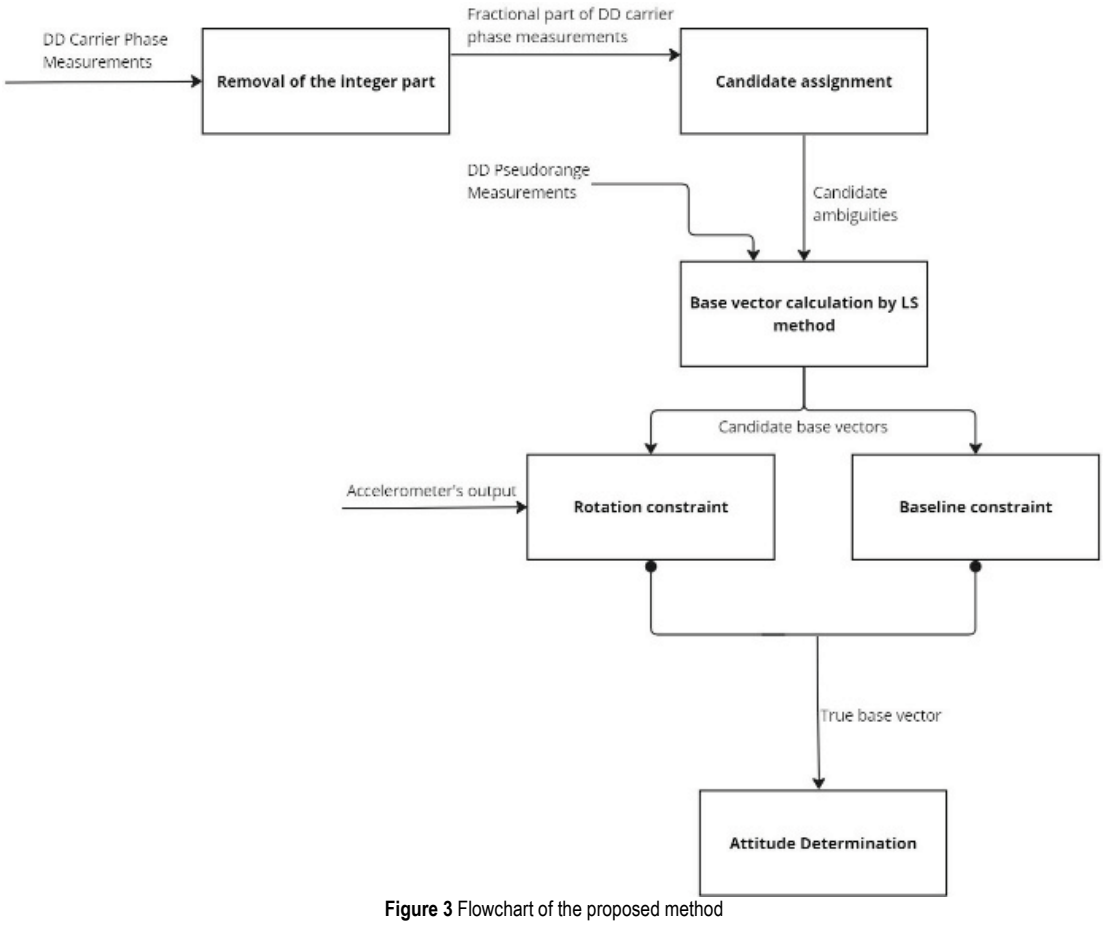


Figure 3 Flowchart of the proposed method

Baseline Constraint: In our case, the antennas remain stationary during rotation. Therefore, the baseline norm does not change. Then for sieving of candidates of the base vector, we can use the following constraint:

$$\ell - \delta\ell < \left|\hat{b}_j(a)\right| < \ell + \delta\ell \, , j = 1,2, \dots, N_p \tag{36}$$

Rotation constraint: Since the pitch angle is determined by accelerometers, it is compared by pitch angle determined by any base vector. When the error remains lower than a threshold, those candidates will be accepted.

$$\theta_a = \tan^{-1}\left(\frac{f_x}{\sqrt{(f_y)^2+(b_z)^2}}\right) \tag{37}$$

$$\theta_b = \tan^{-1}\left(\frac{b_z}{\sqrt{(b_y)^2+(b_x)^2}}\right) \tag{38}$$

Where f denotes the output of the accelerometer. Meanwhile, θ_a represents the pitch angle calculated by the accelerometer, and θ_b signifies the pitch angle calculated by the base vector. The threshold value is determined according to the difference of these two pitch angles.

$$|\theta_b - \theta_a| < \Delta\theta \tag{39}$$

The chart below depicts the flowchart of the proposed method.

4 ANALYSIS OF TESTS

This section includes the conditions for conducting the tests required to apply the methods mentioned in Section 3.2, the information on the materials used, and the analyses of the tests performed. Additionally, the results of the LAMBDA method and the new proposed method are compared.

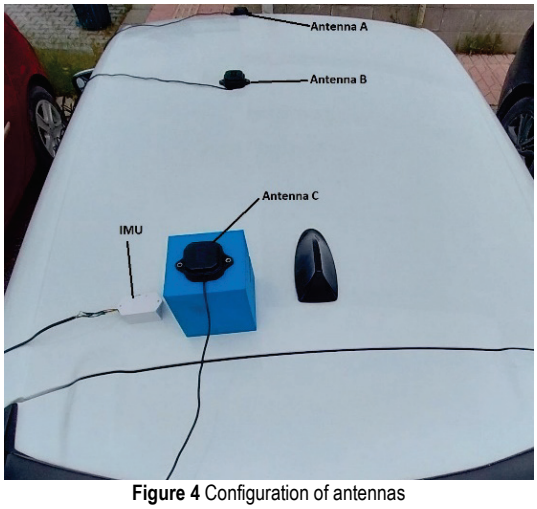


Figure 4 Configuration of antennas

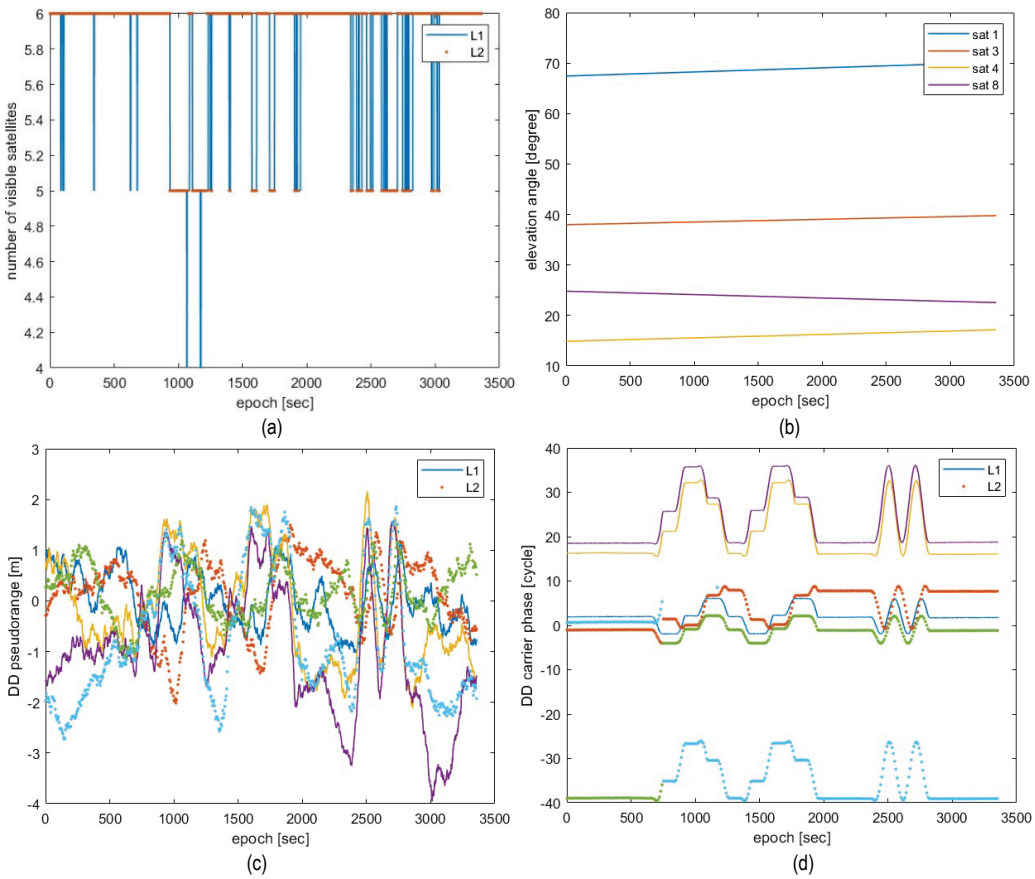


Figure 5 (a) number of visible satellites, (b) elevation angles of visible satellites, (c) DD pseudorange measurements, (d) DD carrier phase measurements.

The tests were conducted using a U-blox ZED-F9P GNSS receiver, an MPU-9255 IMU, and an STM32F407 electronic board. This electronic board was used to read data from four GNSS receivers and one IMU. The STM32CUBEIDE software was utilized to obtain the necessary data packets from the GNSS receivers and IMU mounted on the electronic board. After acquiring the data, the mentioned algorithms were implemented using MATLAB.

Moreover, this section includes graphs for visible satellites, elevation angles, pseudorange and carrier phase measurements, and rotation for both methods. Given that the test was conducted in a multipath environment, only satellites with elevation angles exceeding 10 degrees were processed to mitigate multipath errors, resulting in a low number of visible satellites. The measurements and the count of visible satellites are presented separately for the L1 band.

The test was conducted with a baseline length of 0.75 meters and a fixed pitch angle of 20 degrees throughout. The system's X direction was initially aligned with the north direction, and the vehicle underwent a complete rotation, turning 90 degrees west, 90 degrees south, 90 degrees east, and finally returning to the north direction. Subsequently, another full circle was completed, this time without any stops, moving from north to north.

As seen in Fig. 5a, the number of visible satellites varies during the test as the vehicle moves. Although the maximum visible satellite count is 6, only 4 satellites are selected from each band because 4 of these satellites have elevation angles greater than 10 degrees. Since the suitable satellites for both the L1 and L2 bands are the same, their elevation angles are identical. We previously mentioned that pseudorange measurements are noisy and misleading, as observed in Fig. 5c, while sudden changes in measurements due to integer ambiguities in carrier phases are seen in Fig. 5d. Additionally, carrier phase noise is minimal.

Due to the high error of the receivers, when estimating rotation using the LAMBDA method, as depicted in Figs. 6a and 6c, the pitch angle exhibits notably poor accuracy. Furthermore, within Figs. 6d and 6e, trajectory graphs depict the base vector corrected using both the LAMBDA and New methods. In the New method, the base vector closely approximates the norm value of 0.75 meters. However, discrepancies are evident in the LAMBDA method, manifesting as errors.

Table 1 Methods and mean error values (degree)

	0.75 meter baseline
LAMBDA	Heading: 30 deg Pitch: 40 deg
New Method	Heading: 3 deg Pitch: 5 deg

In Tab. 1, the mean error values of heading and pitch angles for two separate methods are given. As can be seen, our method is much more successful than the LAMBDA method. The LAMBDA method gives poor results, especially in multipath conditions with few satellites.

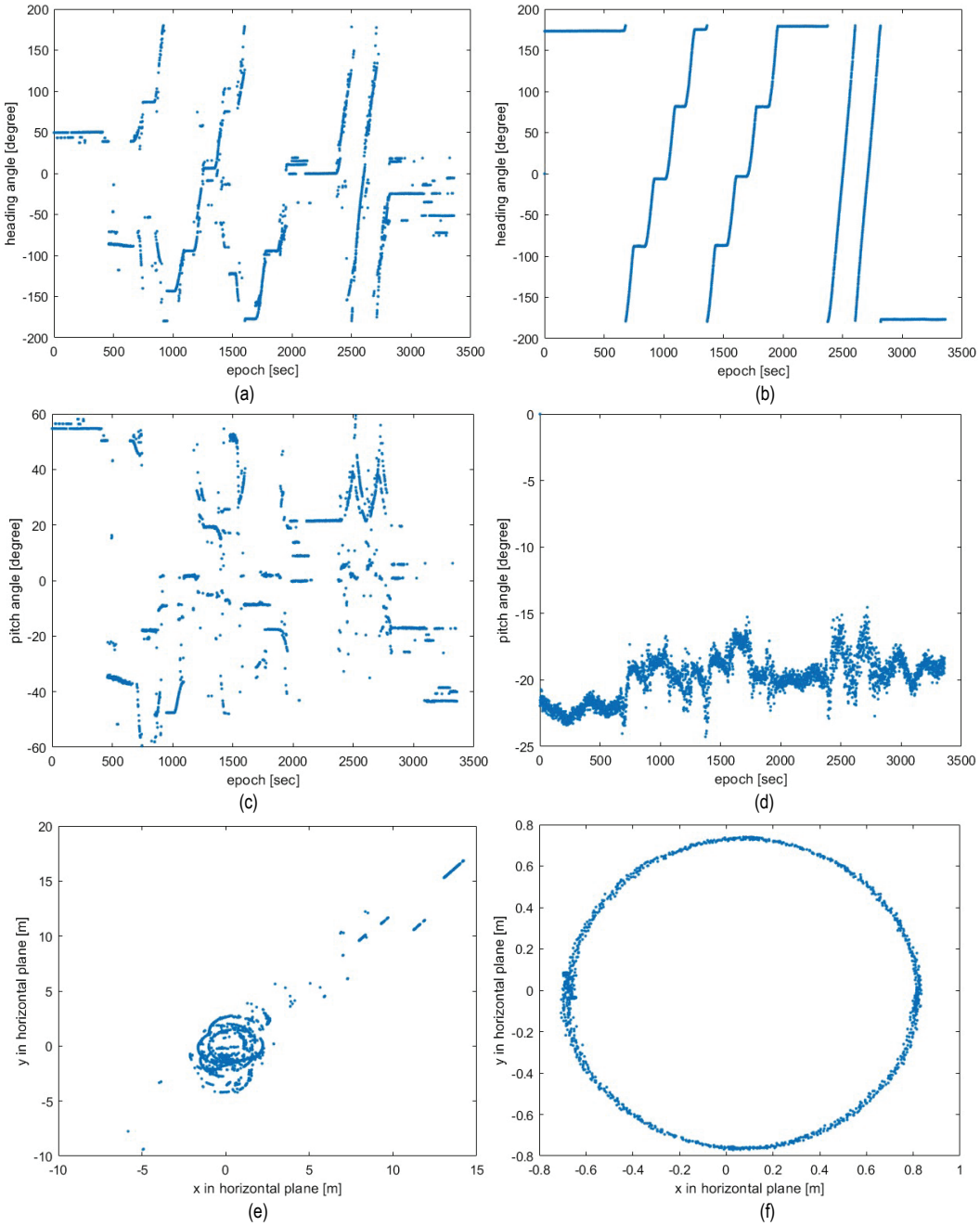


Figure 6 (a) heading angle by LAMBDA method, (b) heading angle by New method, (c) pitch angle by LAMBDA method, (d) pitch angle by New method, (e) fixed trajectory by LAMBDA method, (f) fixed trajectory by New method

5 CONCLUSIONS

In this study, a comparison between the LAMBDA method and the new method in detecting and correcting integer ambiguities in carrier phases was conducted. While the LAMBDA method is widely used in carrier phase applications, it exhibits poor performance, especially in environments prone to multipath effects. On the other hand, the new method, incorporating baseline and rotation constraints, demonstrates superior performance even in challenging multipath-rich environments.

In the conducted tests, the new method consistently outperforms the LAMBDA method, showing significantly reduced mean error values in heading and pitch angles. This improvement is particularly noteworthy considering the adverse conditions tested, which included multipath effects from buildings, trees, and vehicles. Moreover, even with antennas and receivers exhibiting elevated error rates, the new method still demonstrates considerable accuracy improvements.

The potential of the new method could be further enhanced, particularly in environments free from multipath interference, such as those achievable with drones or aircraft. Additionally, employing more precise antennas and receivers could lead to even better results.

6 REFERENCES

[1] Subirana, J. S., Zornoza, J., & Hernández-Pajares, M. (2013). *GNSS data processing: Fundamentals and algorithms*. European Space Agency.

[2] Buist, P. J. (2007). The baseline constrained LAMBDA method for single epoch, single frequency attitude determination applications. In *Proceedings of the 20th International Technical Meeting of the Satellite Division of the Institute of Navigation (ION GNSS)* (pp. 2962–2973).

[3] Fan, S., Zhang, K., & Wu, F. (2005). Ambiguity resolution in GPS-based, low-cost attitude determination. *Journal of Global Positioning Systems, 4*(1–2), 207–214. https://doi.org/10.5081/jgps.4.1.207

[4] Hatch, R. (1991). Instantaneous ambiguity resolution. In *Kinematic systems in geodesy, surveying, and remote sensing: Symposium No. 107* (pp. 299–308). Springer. https://doi.org/10.1007/978-1-4612-3102-8_27

[5] Chen, D. (1993). Fast ambiguity search filter (FASF): A novel concept for GPS ambiguity resolution. In *Proceedings of the 6th International Technical Meeting of the Satellite Division of the Institute of Navigation (ION GPS)* (pp. 781–787). Salt Lake City, UT.

[6] Counselman, C. C., & Gourevitch, S. A. (1981). Miniature interferometer terminals for Earth surveying: Ambiguity and multipath with global positioning system. *IEEE Transactions on Geoscience and Remote Sensing, GE-19*(4), 244–252. https://doi.org/10.1109/TGRS.1981.350379

[7] Lu, G. (1995). *Development of a GPS multi-antenna system for attitude determination* (Master's thesis). University of Calgary, Calgary, Canada.

[8] Kim, D., & Langley, R. B. (1999). An optimized least-squares technique for improving ambiguity resolution and computational efficiency. In *Proceedings of the 12th International Technical Meeting of the Satellite Division of the Institute of Navigation (ION GPS)* (pp. 1579–1588). Nashville.

[9] Wang, Y., Zhan, X., & Zhang, Y. (2007). Improved ambiguity function method based on analytical resolution for GPS attitude determination. *Measurement Science and Technology, 18*(9), 2985–2990. https://doi.org/10.1088/0957-0233/18/9/032

[10] Han, S., & Rizos, C. (1996). Improving the computational efficiency of the ambiguity function algorithm. *Journal of Geodesy, 70*(6), 330–341. https://doi.org/10.1007/BF00868185

[11] Erickson, C. (1992). *Investigations of C/A code and carrier measurements and techniques for rapid static GPS surveys* (Master's thesis). University of Calgary, Calgary, Canada.

[12] Park, C., Kim, I., Lee, J. G., & Jee, G. (1996). Efficient ambiguity resolution using constraint equation. In *Proceedings of Position, Location and Navigation Symposium - PLANS '96* (pp. 277–284). IEEE. https://doi.org/10.1109/PLANS.1996.509089

[13] Teunissen, P. J. G. (1994). A new method for fast carrier phase ambiguity estimation. In *Proceedings of 1994 IEEE Position, Location and Navigation Symposium - PLANS'94* (pp. 562–573). IEEE. https://doi.org/10.1109/PLANS.1994.303362

[14] Verhagen, S., & Li, B. (2012). *LAMBDA software package: MATLAB implementation* (Version 3.0). Perth, Australia.

[15] Bejeryd, J. (2007). *GPS-based attitude determination*. Institutionen för systemteknik, Linköping.

[16] Ma, L., Zhu, F., Liu, W., Lu, L., Lou, Y., & Zhang, X. (2022). VC-LAMBDA: A baseline vector constrained LAMBDA method for integer least-squares estimation. *Journal of Geodesy, 96*(9), 59. https://doi.org/10.1007/s00190-022-01644-7

[17] Liu, X., Chen, G., Zhang, Q., & Zhang, S. (2017). Improved single-epoch single-frequency PAR LAMBDA algorithm with baseline constraints for the BeiDou navigation satellite system. *IET Radar, Sonar & Navigation, 11*(10), 1549–1557. https://doi.org/10.1049/iet-rsn.2017.0048

[18] Teunissen, P. J. G. (2010). Integer least-squares theory for the GNSS compass. *Journal of Geodesy, 84*(7), 433–447. https://doi.org/10.1007/s00190-010-0380-8

[19] Park, C., & Teunissen, P. J. G. (2009). Integer least squares with quadratic equality constraints and its application to GNSS attitude determination systems. *International Journal of Control, Automation and Systems, 7*(4), 566–576. https://doi.org/10.1007/s12555-009-0408-0

[20] Li, W., Fan, P., Cui, X., Zhao, S., Ma, T., & Lu, M. (2018). A low-cost INS-integratable GNSS ultra-short baseline attitude determination system. *Sensors, 18*(7), 2114. https://doi.org/10.3390/s18072114

[21] Shoemake, K. (1994). Euler angle conversion. In *Graphics Gems* (pp. 222–229). https://doi.org/10.1016/B978-0-12-336156-1.50030-6

[22] Sidi, M. J. (1997). *Spacecraft dynamics and control: A practical engineering approach* (Vol. 7). Cambridge University Press.

[23] Ghanbarpour, A. H., Pourtakdoust, S. H., & Samani, M. (2009). A new non-linear algorithm for complete pre-flight calibration of magnetometers in the geomagnetic field domain. *Proceedings of the Institution of Mechanical Engineers, Part G: Journal of Aerospace Engineering, 223*(6), 729–739.

[24] Teunissen, P. J. G. (1998). Success probability of integer GPS ambiguity rounding and bootstrapping. *Journal of Geodesy, 72*, 606–612. https://doi.org/10.1007/s001900050199

[25] Teunissen, P. J. G. (1999). An optimality property of the integer least-squares estimator. *Journal of Geodesy, 73*(11), 587–593. https://doi.org/10.1007/s001900050269

[26] De Jonge, P., & Tiberius, C. (1998). *The LAMBDA method for integer ambiguity estimation: Implementation aspects* (Publications of the Delft Geodetic Computing Center, No. 12).

[27] Eling, C., Klingbeil, L., & Kuhlmann, H. (2015). Real-time single-frequency GPS/MEMS-IMU attitude determination of lightweight UAVs. *Sensors, 15*(10), 26212–26235.

Authors' contacts:

Habib Kanberoğlu (Habib Ghanbarpourasl)
(Corresponding author)
Department of Computer Engineering,
Turkish Aeronautical Association University,
Bahçekapı Mahallesi Okul Sk. No: 11, 06790 Etimesgut-Ankara, Turkey
hghanbarpourasl@thk.edu.tr

Mukaddes Bolat
Department of Mechanical and Aeronautical Engineering,
Turkish Aeronautical Association University,
Bahçekapı Mahallesi Okul Sk. No: 11, 06790 Etimesgut-Ankara, Turkey

ISSN 1846-6168 (Print), ISSN 1848-5588 (Online)
https://doi.org/10.31803/tg-20240808075046

Original scientific paper
Received: 2024-08-08, Accepted: 2025-04-15

Study the Differential Cross-Section of Scattering Potential (Yukawa, Coulomb and Square Well) around PEMFC

Saddam Husain Dhobi*, Kishori Yadav, Suresh Prasad Gupta, Jeevan Jyoti Nakarmi, Ajay Kumar Jha

Abstract: The objective of this work is to study the differential cross section (DCS) around-surface the electrode of proton exchange membrane (PEMFC) for stability and performance of PEMFC. For this considered a system of electrons proton and hydrogen particle around electrode formed around-surface electrode due to inlet of H_2 and O_2 in PEMFC system. After the particle formed around they go interaction and the region of interaction is defined by S-matrix theory. S-matrix formulation was formulated by Kroll and Watson (1973) for a low-frequency response. After formulation of DCS was calculated for possible three case around electrode of PEMFC. The three case is consideration of Yukawa potential, Coulomb potential and square well potential was calculated found that the DCS of Yukawa potential is less than coulomb potential, and coulomb potential DCS is less than the square well potential at the same interaction strength. The DCS for considered potential in this work ranges widely 10^{-10} a. u. to 1.4×10^{8} a. u. The DCS was studied for both electron and meson at an ideal using square well potential and found that the DCS for meson is greater than an electron, this is because of mass of meson is greater than an electron. The DCS found stable in higher energy range and higher scattering angle indicate system of PEMFC is stable and this may imply to remove the complexity and stability of PEMFC system when the system of PEMFC single cell is design with consideration of DCS and diffusion angle of inlet fuel inside PEMFC.

Keywords: Coulomb potential and square well potential; DCS; Kroll and Watson; PEMFC; S-Matrix; Yukawa potential

1 INTRODUCTION

Electron was discovered by J. J. Thompson in 1897 AD and discovery help the development of physics beyond classical physics. Elastic scattering of alpha particles with foil led to the discovery of the nucleus by Rutherford in 1911 AD. In addition, the discovery also led to the development of the planetary model of the atomic structure. Bohr proposed a model of hydrogen in 1912 AD which also led to the development of new laws and formulation in the field of physics with the assumption that electron in a hydrogen atom moves in a circular orbit due to the Coulomb force between the electron and the nucleus with quantization of angular momentum of the electron and emitter of radiation when electron jump from higher to lower quantum number. A scattering experiment in 1923 AD by Compton [1-5] confirm the existence of photons and the experiment was verified and supported by Bothe and Geiger [6].

One of the most studied phenomena of laser-atom interaction is that of photo-ionization. In photoionization, an electron is removed from a parent atom or ion living behind a single or multiply charged ion. At conventional intensities, only one photon is absorbed in an elementary event of the interaction of light with substance. The energy $\hbar\omega$ of the photon coincides with the difference between the energy levels of the relevant atom [6, 7]. At high intensities of laser, two or more photons may be absorbed in an elementary event of interaction. In this case, light not only the frequency ω but also of the frequencies $\frac{\omega}{2}$, $\frac{\omega}{3}$ etc. may be absorbed. Such absorption is called multi-photon (2 photons, 3 photons, etc.), in this case, the electric field strength is nearly equal to the microscopic field of the atom and optical property depends upon both frequency and intensity (nonlinear case).

The multi-photon ionization for a field is the transition of electrons from a bound state to a continuum state by the simultaneous absorption of more than two photons. This theory based on the scattering method was realized in 1960 AD but Kroll and Watson in 1973 AD [8] gave a new approach to S-matrix for low-frequency response for the solution of the scattering process. It can be extended not scattering but another type of process also. S-matrix formulation was invented in a search for a general non-perturbative formulation to use for strong field problems. It is a suitable approach for free-free scattering it is a suitable approach also for ionization (bound free, recombination free bound and excitation/DE excitation) [9]. The s-matrix method is more useful than another method because this formulation avoids the decoupling problem (nor loss, nor gain of energy due to a very short time).

In this method, the interaction is turned on or off is not absorbable while other methods suffer from this complication [9] S-matrix formulation can be formulated entirely in terms of quantity measurable in the laboratory. It can be applied to any process as long as the space-time domain of the interaction region is bounded. It gives unambiguous rules gauge transformation. S-matrix is related to T-matrix which helps to determine the differential cross-sectional area when electron interacts with the target atom in laser field.

The DCS in the presence of a weak laser field during inelastic scattering, finding that DCS increases when the target absorbs energy and decreases to a minimum before reaching a maximum upon energy emission at specific eV levels [10]. It also increases with the scattering angle. Additionally, the study examines thermodynamic properties of thermal electrons in laser fields, showing variations in thermodynamic energy and potential with field amplitude and temperature [11]. In the case of laser-assisted thermal electron-hydrogen atom elastic scattering, the DCS for thermal electrons was higher than for nonthermal electrons, with interference patterns observed under certain conditions and specific angles [12]. In present work authors trying to interlink DCS with PEMFC performance which help

researcher/reader for thermal management of PEMFC generated during chemical reaction and interaction between the particles formed during chemical reaction at surface of PEMFC electrode.

A PEMFC is a device that converts chemical energy into electrical energy through an electrochemical reaction between hydrogen and oxygen. It consists of an anode, where hydrogen molecules are split into protons and electrons, and a cathode, where oxygen combines with protons and electrons to form water. The protons pass through a polymer electrolyte membrane, while the electrons travel through an external circuit, generating electricity. The overall reaction produces water and heat as byproducts. The reaction around electrode form electron, proton and residue hydrogen which interacts and effects the performance of PEMFC and these electron and proton are causing production of current studied. Dhobi et al. study the DCS around electrode of PEMF and its impact on performance of PEMFC [13-15].

Despite extensive research on fuel cells and scattering phenomena, the integration of scattering principles with technological applications, particularly in PEMFCs, remains underexplored. This work aims to bridge this gap by investigating the relationship between differential cross-section (DCS) variations in weak laser fields and the performance of PEMFCs. By studying the interactions of electrons and protons around PEMFC electrodes, this research seeks to develop a comprehensive understanding of how scattering impacts the efficiency and functionality of PEMFCs, potentially extending these insights to other technological fields.

2 METHODS AND MATERIALS

2.1 Interrelation of S-Matrix and T-Matrix with Amplitude

S-matrix is a matrix element of time development operator $\tilde{T}(t, t_0)$ between unperturbed asymptotic in and out state and denoted by s or *S*. T-matrix is defined in terms of matrix element S-matrix when time tends to $-\infty$ to $+\infty$ whereas transition amplitude is the matrix element of T-matrix in a fixed time and given by [16, 17] as,

$$(S-1)_{fi} = -\frac{i}{\hbar}\int_{-\infty}^{t}\left(\Phi_f V \Psi_i\right) \mathrm{d}t \tag{1}$$

Here $\Phi_f = \phi_f(\boldsymbol{r}) \exp\left(-\frac{iE_f t}{\hbar}\right)$ and $\Psi_i = \psi_i(\boldsymbol{r}) \exp\left(-\frac{iE_i t}{\hbar}\right)$ substituting the value of Φ_i and Φ_f in Eq. (1) and using delta function, we get

$$(S-1)_{fi} = -2\pi i \delta\left(E_f - E_i\right) T_{fi} \tag{2}$$

Here $T_{fi} = \int \phi_f V \psi_i d^3r$ and defined as *T* matrix and $(S-1)_{fi}$ is and transition amplitude. For Elastic scattering there is no change in energy, therefore, the initial and final energy are equal and hence Eq. (2) become

$$(S-1)_{fi} = -2\pi i T_{fi} \tag{3}$$

This equation defines the t-matrix in terms of s-matrix elements of the S-matrix at an infinite time with scattering potential which is time-independent and equal to the Hamiltonian of a system.

2.2 Derivation of S-Matrix and Scattering Amplitude in Terms of the Transition Matrix

Consider a complete set of states $\{\Psi\}$ that satisfy the Schrodinger equation describing the atomic electron that may be in interaction with a laser beam.

$$\frac{i\partial\Psi}{\partial t} = H\Psi = (H_0 + H_I)\Psi \tag{4}$$

Since no laboratory experiment is possible to measure the outcome of an experiment in the laser field. So, as far as laboratory instruments are a concern, there is a complete set of status $\{\Phi_n\}$ that satisfy the Schrodinger equation describing an atomic electron that may be undisturbed or in interaction of the laser field.

$$\frac{i\partial\Phi}{\partial t} = H_0\Phi \tag{5}$$

Hypothetically, the laser pulse is finite, so

$$\lim_{t\to\pm\infty}[H(t) - H_0] = 0 \tag{6}$$

Then, the two complete sets of Φ states and Ψ states can be organized so that correspond at $t \to -\infty$

$$\lim_{t\to-\infty}[\Phi_n(t) - \Psi_n(t)] = 0 \tag{7}$$

After the laser interaction has been introduced, the only wave for the laboratory instrument to discover what has happened is to form an overlap of all possible final Φ_f state with the initial particular state Ψ_i. This forms *S*-matrix such as

$$S_{fi} = \lim_{t\to+\infty}(\Phi_f, \Psi_i) \tag{8}$$

Now the transition matrix, before and after scattering is obtained with the help of *S*-matrix can be obtained as

$$M_{fi} = (S-1)_{fi} = \lim_{t\to+\infty}(\Phi_f, \Psi_i) - \lim_{t\to-\infty}\left(\Phi_f, \Psi_i\right) \tag{9}$$

This is now in the form of an exact differential and represented as

$$M_{fi} = \int_{-\infty}^{\infty}\left(\frac{\partial\Phi_f}{\partial t}, \Psi_i\right)\mathrm{d}t + \int_{-\infty}^{\infty}\left(\Phi_f, \frac{\partial\Psi_i}{\partial \mathrm{t}}\right)\mathrm{d}t \tag{10}$$

Now using Eqs. (4) and (5), we get $\frac{\partial\Phi_f}{\partial t} = -iH_0\Phi_f$ and $\frac{\partial\Psi_i}{\partial \mathrm{t}} = -i(H_0 + H_I)\Psi$ and using these on Eq. (10) we get,

$$M_{fi} = -i\int_{-\infty}^{\infty}\left(\Phi_f, H_{\mathrm{I}}\Psi_i\right)\mathrm{d}t \tag{11}$$

Where H_I is a weak interaction field, since $(S-1)_{fi}$ is a transition amplitude then transition probability per unit time is given with the help of Eq. (2) as

$$W = \lim_{\tau\to\infty} \frac{|(s-1)_{fi}|^2}{\tau} = \frac{1}{\tau}\left(\lim_{\tau\to\infty} \int_{-\frac{\tau}{2}}^{\frac{\tau}{2}} \exp\left[\frac{i(E_f-E_i)t}{\hbar}\right] \frac{\mathrm{d}t}{\hbar} \times \lim_{t\to\infty} \int_{-\frac{\tau}{2}}^{\frac{\tau}{2}} \frac{\mathrm{d}t}{\hbar} \left|T_{fi}\right|^2 \right) \quad (12)$$

Now applying the delta property of delta function and solving we get,

$$W = \left(\frac{2\pi}{\hbar}\right) \delta\left(E_f - E_i\right) \left|T_{fi}\right|^2 \quad (13)$$

For transition into the continuum of final states. The total transition rate in the range $\left(E_f - \frac{\Delta E}{2}\right)$ to $\left(E_f + \frac{\Delta E}{2}\right)$is given by

$$W = \left(\frac{2\pi}{\hbar}\right) \int_{\left(E_f-\frac{\Delta E}{2}\right)}^{\left(E_f+\frac{\Delta E}{2}\right)} \mathrm{d}E \;\; \delta\left(E_f - E_i\right)\left|T_{fi}\right|^2 \rho(E) \quad (14)$$

Here $\rho(E)$ is the density of the final state and on solving Eq. (14) we get,

$$W = \left(\frac{2\pi}{\hbar}\right) \left|T_{fi}\right|^2 \rho(E) \quad (15)$$

2.3 Differential Cross-Section in Born Approximation for Yukawa Potential (YP)

Yukawa purposed potential in 1935 [18] as effective non-relativistic potential which described the strong interactions between nucleons as

$$V(r) = \frac{Ae^{-\mu r}}{r}\; A, \mu > 0 \qquad \text{In Atomic Unit} \quad (16)$$

Here A described the strength of interaction (special case equal to 1), μ is screening parameters (ranges from 0.10 to 1 in bounded case) [1] and $1/\mu$ is its range. This potential is also called Debye-Huckel potential in plasma physics, while this potential also represents as a potential of a charged particle in a weakly nonideal plasma, as well as in electrolytes and colloids, in solid-state as Thomas-Fermi potential. Pais and Jost shows 3D spherical potential such that $I = 2m\int_0^\infty dr r|V(r)|$ is finite for the bounded state is greater than unity. Since amplitude is defined by born first approximation as,

$$f_k(\theta) = -\frac{2m}{k\hbar^2} \int_0^\infty \mathrm{d}r\; r\, V(r) \sin(Kr) \quad (17)$$

Now from Eqs. (17) and (16) we get

$$f_k(\theta) = -\frac{2m_m A}{K\hbar^2} \int_0^\infty \mathrm{d}r\; e^{-\mu r} \sin(Kr) = -\frac{2m_m A}{\hbar^2(\mu^2+K^2)} \quad (18)$$

Here $K = 2k\sin\left(\frac{\theta}{2}\right)$ and m_m is meson mass. Since scattering amplitude is directly proportional to the differential cross-section. Therefore, DCS for YP in the Born approximation is equal to

$$\left(\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega}\right)_{B1}^{Y} = \left[\frac{2m_m A}{\hbar^2\left(\mu^2+4k^2\sin^2\left(\frac{\theta}{2}\right)\right)}\right]^2 \quad (19)$$

In terms of energy, the above Eq. (19) [19, 20] become

$$\left(\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega}\right)_{B1}^{Y} = \left[\frac{2m_m A}{\left(\hbar^2\mu^2+8m_m^2E^2\sin^2\left(\frac{\theta}{2}\right)\right)}\right]^2 \quad (20)$$

Here $E = \frac{\hbar^2k^2}{2m}$ and Eq. (20) gives the DCS in Born approximation with Yukawa potential. This is the case which we considered around electrode in absence of non-monochromatic wave form around electrode due to self-generated case wave due to scattering. This case is considered initial case of collision because at initial state of reaction the speed of particle formed around electrode is higher which increase the energy of electron. In addition, after time pass the number of electrons-proton formation is higher and the effect of screening is also observed around which may affect the flow of electron and decrease the current of PEMFC

2.4 Differential Cross-Section in Born Approximation for Coulomb Potential (CP)

For Coulomb/central potential is defined as $V(r) = \frac{Z_1Z_2e^2}{4\pi\epsilon_0 r} = \frac{A}{r}\; i$n atomic unit, therefore from Eq. (17) the born first amplitude for CP around electrode of PEMFC become

$$f_k(\theta) = -\frac{2mA}{k\hbar^2}\int_0^\infty \mathrm{d}r\; \sin(Kr) = -\frac{2m_eA}{k\hbar^2K} \quad (21)$$

Similarly, as above the DCS in Born approximation for CP obtained with the help of Eq. (21) [19, 20] is

$$\left(\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega}\right)_{B1}^{C} = \frac{m_e^2A^2}{16E^2\sin^4\left(\frac{\theta}{2}\right)} \quad (22)$$

Eq. (22) is similar to the Rutherford formula with the incidence kinetic energy of particles E. For larger momentum transformation that is k^2 is larger a sharp peak is obtained at a small scattering angle $(\theta \to 0)$, while at a larger scattering angle $(\theta \to \infty)$ the $4k^2\sin^2\left(\frac{\theta}{2}\right)$ is larger. The Born approximation is valid with higher incidence energy of particles in weak scattering potentials. This means the average interaction energy between the incident particle and the scattering potential is much smaller than the incidence kinetic energy of the particle. Also, this case is considered around fuel cell electrode when coulomb potential plays an important role around due to large number of electrons protons formed around electrode of fuel cell.

2.5 Differential Cross-Section in Born Approximation for Square Well Potential (SWP)

SWP in 3D for scattering is defined as

$$V(r < a) = V_0 \\ V(r > a) = 0 \tag{23}$$

Now to calculate the born first amplitude, for DCS we have from Eq. (17) and (23)

$$f_k(\theta) = -\frac{2mV_0}{k\hbar^2}\left[\frac{\sin(ka)-ka\cos(ka)}{k^2}\right] \tag{24}$$

Similarly, as above the DCS in Born Approximation for SWP in 3D with the help of Eq. (24) is

$$\left(\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega}\right)_{B1}^{SWP} = \left(\frac{2mV_0}{\hbar^2 k}\right)^2 \begin{cases} \frac{1}{9}\left(1-\frac{1}{5}k^2a^2\right) & \text{For low Eenergy } (E), ka < 1 \\ \frac{a^2}{k^2} & \text{For High Energy } (E), ka > 1 \end{cases} \tag{25}$$

Here m is the mass of considering particle in the potential well. For X-ray scattering with a neutron, the scattering cross-sections are about 10^{-24}cm^2 with the cross-sectional area per atom is about 10^{-16}cm^2 [3]. The potential define in this work is based on the atomic unit with screening parameters unit a_0^{-1} (Bohr's radius) [21]. This is another case scattering case below to condition of equation (23) around electrode inside PEMFC.

In a theory that combines both quantum and relativity theory, it is not possible to specify the position of individual particles precisely. If this is done, in field theory, one has to put off with mathematical inconsistencies which are, indeed, the main problem in quantum field theory. S-matrix theory bypasses this problem by specifying the momenta of the particles. The important new concept theory is the shift of emphasis from object to the event; its basic concern is not with the particles but with their reactions. Such a shift from object to event is required both by quantum theory and relativity. In S-matrix theory, as in field theory, the interaction forces are associated with particles but the concept of virtual particles is not used. Instead, the relation between forces and particles is based on a special property of S-matrix i.e. known as a crossing.

3 RESULTS AND DISCUSSION

The DCS in Born Approximation with Coulomb potential, Yukawa potential and Square Well Potential are studies at different incidence energy of the electron, angle and screening parameters. Taking three cases of around the electrode of PEMFC and help to improve the efficiency of PEMFC knowing the interaction region of particles around electrode my minimizing the space between gas diffusion laser, electrodes and proton exchange membrane. This may help in improve of current state of efficiency [22].

3.1 DCS in Born Approximation for YP

The DCS is based on the atomic unit for meson particle mass $273m_e$ where m_e is a mass of the electron. It is the case if meson formed around the electrode due to scattering between particles around electrode of PEMFC. To study the DCS with scattering angle for YP the screening parameters $0.05a_0^{-1}$, interaction potential strength 1 a.u. and energy of incidence electron 500 a.u. From Fig. 1, it is observed that the scattering at a low angle is high and uncontrollable around the electrode of PEMFC so to control this angle also play an important role and hence to designed PEMFC with flow of inlet fuel angle also play an important role. In addition, this make systems stable which is major challenges [23] of PEMFC.

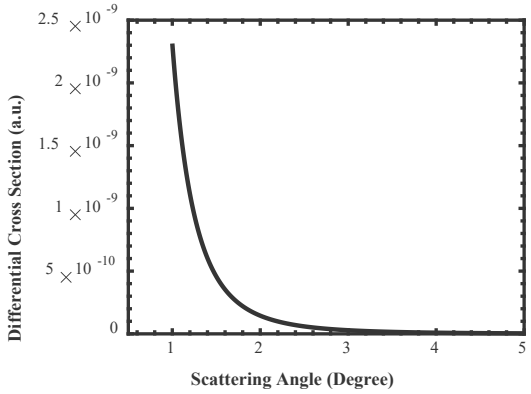


Figure 1 Scattering Angle vs DCS for YP around electrode of PEMFC

But with increasing the scattering angle the scattering goes decrease and become constant at a larger angle [24] but they study for CF_4 like molecules. Because at the large-angle the interaction between target and incidence is less therefore the DCS at large angle is small and constant so this angle may be fit for stable system for PEMFC.

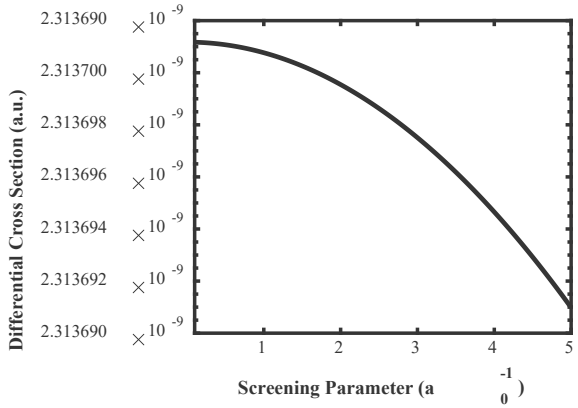


Figure 2 Screening Parameters vs DCS for YP around electrode of PEMFC

Fig. 2, show the effect of screening on DCS as we discussed the continues inlet of hydrogen form screening around the electrode and disturbance of flow of electron effect the performance of PEMFC. Therefore, this study show screening effect DCS and for a system stable lower screening region is best with higher DCS as shown in figure. The observation of DCS with scattering parameters shows the DCS decrease with the increase of screening parameters. In this case, the DCS has no such drastic change in numerical value because with increasing the screening parameters the interaction between target and electron becomes less.

Fig. 3 shows the nature of DCS in Born approximation for YP, with increasing the incidence energy of the electron to target the DCS decrease [25] but they study for He-atom. This is because at high energy of incidence the probabilities of interaction between target and incidence are low as well as target can't see the incidence at high energy. Fig. 3 is for meson incidence particles with mass $273m_e$ at one degree. This also shows that speed of electron corresponding energy effect the DCS and DCS is stable in higher energy region and hence at initial case the seep of particle is higher and performance of PEMFC is higher at initial case than after some time. So, for stable and performance as initial condition DCS management of scattering help the improvement of PEMFC.

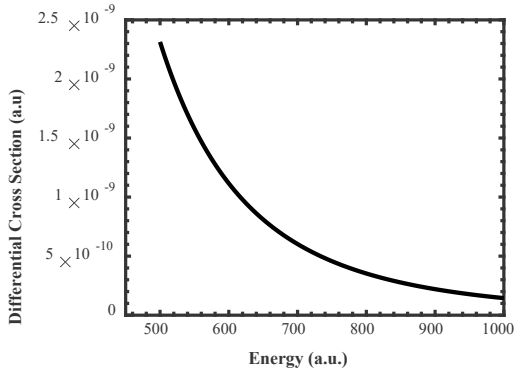


Figure 3 Incidence Energy vs DCS for YP around electrode of PEMFC

The DCS with incidence energy (500 to 1000 a. u.) of meson was ranges from 10^{-10} to 10^{-9} a. u. with 1 degree of scattering angle. The DCS with screening parameters was of the order 10^{-9} a. u. with 1 degree of scattering angle and 500a. u. incidence energy of meson. The DCS with scattering angle was ranges 10^{-10} to 10^{-9} a. u. with incidence energy of meson 500 a. u. for all DCS of YP potential strength is constant as 1 a. u.

3.2 DCS in Born Approximation for CP

Considering the case of scattering around the electrode of PEMFC the DCSC with energy of electrons show decrease as shown in Fig. 3 but for CP the DCS is higher than YP. This observation also shows that at higher energy range the system is stable hence its better knowledge how that DCS nature is for stable system of PEMFC. The DCS in Born approximation for CP was shown in Fig. 4. The observation shows that DCS is decreased with increasing incidence energy of electron with potential strength 1 a.u. at a scattering angle of 1 degree.

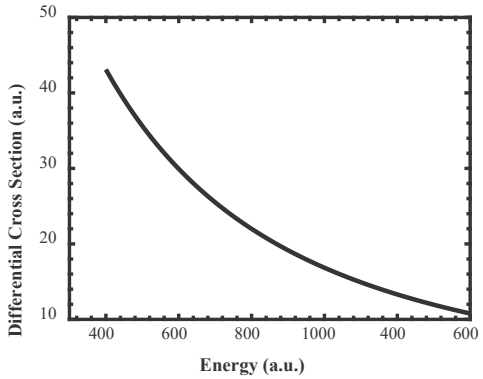


Figure 4 Energy of incidence particles vs DCS for CP around electrode of PEMFC

Similarly, the DCS with scattering angle was also observed as shown in Fig. 5. The observation shows that the DCS is high and uncontrollable/undetectable at a small angle and with increasing the scattering angle the DCS becomes constant, similar to the DCS of YP.

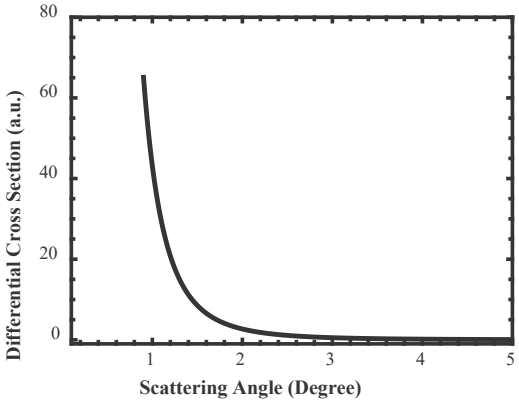


Figure 5 Scattering Angle vs DCS for CP around electrode of PEMFC

The DCS in Born approximation for CP ranged from 10 a. u. to 45 a. u. but the nature is quite different because DCS with scattering angle is sharply decreased while DCS with energy is slow. This case also shows that DCS is stable for higher scattering angle system as YP case. In compare the DCS and stable for PEMFC CP is better than YP case.

3.3 DCS in Born Approximation for SWP

The DCS in Born approximation for SWP is shown in Figs. 6 and 7 respectively from electron and meson. If we considered case around the PEMFC of particle dividing two cases, low (for time passes of reaction) and high energy (initial reaction and formation of particle). The DCS at low energy for the electron is almost zero at momentum 0.5 a. u., potential strength 1 and potential good width 0.9 Å but at high energy DCS goes decrease with increasing scattering angle at momentum 1 a. u. and potential good width 2 Å at same potential strength. The DCS at low energy for meson is almost zero at momentum 0.5 a. u. and potential strength 0.9

a.u. but at high energy, DCS goes decrease with increasing scattering angle at momentum 1 a.u. and potential strength 2 a.u. The DCS when the electron is welling in the potential region is shown in Fig. 6, while meson is in Fig. 7. The DCS for electron was observed up to 200 a.u. while for meson 1.4×10^8 a.u. at high energy of incidence of particles to the target.

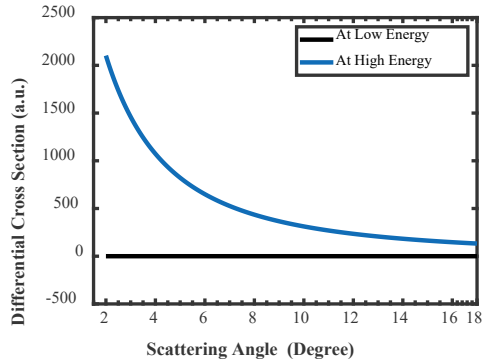


Figure 6 Scattering Angle vs DCS for SWP with the electron as welling particle in Potential Well around electrode of PEMFC

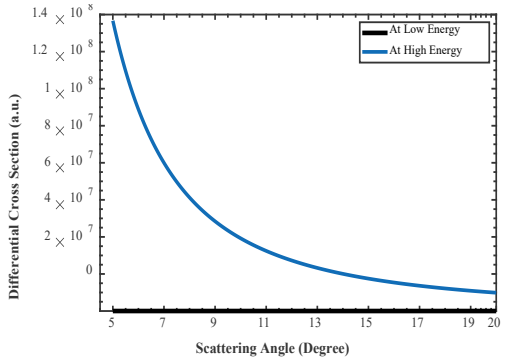


Figure 7 Scattering Angle vs DCS for SWP with meson as welling particle in Potential Well around electrode of PEMFC

On comparing the DCS Fig. 6 and Fig. 7, we can say that at lower energy a constant DCS is observed which means more stable with increasing scattering angle while at higher energy DCS decrease with increasing scattering angle and remain constant at higher scattering angle. In other hand, for stable system at higher and lower energy both are suitable but for complexity on measurement higher DCS is easy. So, the DCS of higher energy in higher scattering angle region is suggested better to design the PEMFC for thermal managements for higher performance.

4 CONCLUSIONS

The DCS in Born Approximation with Coulomb potential, Yukawa potential, and Square Well Potential are also studies with different incidence energy of electron and meson, angle, and screening parameters form around the electrode of PEMFC. The DCS for meson in YP at 1 degree was observed 10^{-10} to 10^{-9} a.u., for an electron in CP, was observed 10 a.u. to 65 a.u. for the electron in SWP was observed up to 2000 a.u. while for meson in SWP was observed 1.4×10^8 a.u. around electrode of PEMFC. The nature found the stable DCS is best for stable performance of PEMFC.

Acknowledgment

The authors would like to thanks all the faculty members of the Department of Physics, Patan Multiple Campus, Patan Dhoka, Lalitpur-4470, Nepal. Similarly, also thanks University Grant Commission Nepal (UGC Award No.: PhD-80/81-S&T-13) for providing research grant for this work.

5 REFERENCES

[1] Alhaidari, A. D. (2021). Taming the Yukawa potential singularity: improved evaluation of bound states and resonance energies. *ArXiv*. 1-2.

[2] Bhatia, A. K. (2020). Scattering and Its Applications to Various Atomic Processes: Elastic Scattering, Resonances, Photoabsorption, Rydberg States, and Opacity of the Atmosphere of the Sun and Stellar Objects. *Atoms. 8*(4), 1-30. https://doi.org/10.3390/atoms8040078

[3] Blügel, S. (2012). Scattering Theory: Born Series. *Lecture Notes of the 43rd IFF Spring School 2012*, Key Technologies, 33, 1-30.

[4] Bothe, W., & Geiger, H. (1925). Uber das Wesen des Compton effects ein experimenteller Breitrag Theorie der Strahlung. *Zeits für Phys., 32*(1), 639-663. (in German) https://doi.org/10.1007/BF01331702

[5] Compton, A. H. (1923). A Quantum Theory of the scattering of X-rays by light. Elements. *Phys. Rev., 21*, 1-6. https://doi.org/10.1103/PhysRev.21.483

[6] Compton, A. H., & Simon, A. W. (1925). Directed Quanta of Scattered X-rays. *Phys. Rev. 1925*, 26, 3-5. https://doi.org/10.1103/PhysRev.26.289

[7] Savelyeve, I. V. (1981). Physics: A General Course, Vol. III, MRI Publishers Moscow.

[8] Kroll, N. M., Watson, K. M., & Keldysh, L. (1973). Ionization in the field of a strong Electromagnetic wave, *Physical Review A, 20*(5), 1307-1314.

[9] Reiss, H. R. (2010). *High-Field Laser Physics*. ETH, Zürich, Switzerland.

[10] Dhobi, S. H., Nakarmi, J. J., Yadav, K., Gupta, S. P., Koirala, B., & Shah, A. K. (2022). Study of thermodynamics of a thermal electron in scattering. *Heliyon, 8*(12). https://doi.org/10.1016/j.heliyon.2022.e12315

[11] Dhobi, S. H., Gupta, S. P., Yadav, K., Nakarmi, J. J., & Jha, A. K. (2024). Differential Cross Section with Volkov-Thermal Wave Function in Coulomb Potential. *Atom Indonesia, 1*(1), 19-25. https://doi.org/10.55981/aij.2024.1309

[12] Dhobi, S. H., Yadav, K., Jha, A. K., Karki, B., & Nakarmi, J. J. (2022). Free Electron-Ion Interaction and Its Effect on Output Current of Permeable Exchange Membrane Hydrogen Fuel. *ECS Transactions, 107*(1), 8457. https://doi.org/10.1149/10701.8457ecst

[13] Dhobi, S. H., Gupta, S. P., Nakarmi, J. J., Koirala, B., Yadav, K., Oli, S. K., & Gurung, M. (2023). Scattering of Free Electrons with Hydrogen Atoms in Proton Exchange Membrane Fuel Cell System. *International Annals of Science, 13*(1), 22-28. https://doi.org/10.21467/ias.13.1.22-28

[14] Dhobi, S. H., Pudasaini, A., Oli, D., Khatiwada, S., Ghimire, K., Rijal, O. S., ... & Paudel, G. (2024). Scattering of Free Electrons with Hydrogen Atoms in Proton Exchange Membrane Fuel Cell. *Hadronic Journal, 47*(1).

https://doi.org/10.29083/HJ.47.01.2024/SC1
[15] ETH Lecture (2011). *Quantum Mechanics without Perturbation Theory*. Zürich, Switzerland.
[16] Frijof, C. (1976). *The Tao of physics*. 3rd Edition, Harper Collins Publishers, Fulham Palace Road, United Kingdom.
[17] Liverts, E. Z., & Mandelzweig, V. B. (2009). Analytical computation of amplification of coupling in relativistic equations with Yukawa potential. *Annals of Physics, 324*(2), 388-407. https://doi.org/10.1016/j.aop.2008.08.004
[18] Collas, P. (2021). Coulomb scattering in the Born approximation and the use of generalized functions. *ArXiv*, 5-6. https://doi.org/10.1119/10.0005453
[19] GUFAM. (2021). Lecture 5: Scattering theory, Born Approximation Born Approximation, SS2011: Introduction to Nuclear and Particle Physics, Part 2.
[20] Nayek, S., & Ghoshal, A. (2012). Dynamics of positronium formation in positron-hydrogen collisions embedded in weakly coupled plasmas. *Phys. of Plasmas, 19*, 1-14. https://doi.org/10.1063/1.4764467
[21] Harper, C. (2006). *Introduction to Mathematical physics, California State University*. Prentice Hall of India Pvt. Ltd, India.
[22] Park, S., & Lee, S. (2023). Theoretical Analysis for Improving the Efficiency of HT-PEMFC through Unreacted Hydrogen Circulation. *Applied Sciences*. https://doi.org/10.3390/app13169292
[23] Chen, L., Xu, K., Yang, Z., Yan, Z., & Dong, Z. (2022). Optimal Design and Operation of Dual-Ejector PEMFC Hydrogen Supply and Circulation System. *Energies*. https://doi.org/10.3390/en15155427
[24] Ketkar, & Bonham (1985). Small-angle elastic differential scattering cross section for 25-keV electrons scattering from helium. *Physical Review Letters, 55*(13), 1395-1397. https://doi.org/10.1103/PhysRevLett.55.1395
[25] Nagashima, Y., Hyodo, T., Fujiwara, K. & Ichimura, A. (1998). Momentum-transfer cross section for slow positronium-He scattering. *Journal of Physics B, 31*, 329-339. https://doi.org/10.1088/0953-4075/31/2/014

Authors' contacts:

Saddam Husain Dhobi
(Corresponding author)
Central Department of Physics, Tribhuvan University,
Kirtipur, Kathmandu-44600, Nepal
Department of Physics, Patan Multiple Campus, Tribhuvan University,
Lalitpur-44700, Nepal
E-mail: saddam@ran.edu.np

Kishori Yadav
Department of Physics, Patan Multiple Campus, Tribhuvan University,
Lalitpur-44700, Nepal

Suresh Prasad Gupta
Department of Physics, Patan Multiple Campus, Tribhuvan University,
Lalitpur-44700, Nepal
E-mail: guptasir@gmail.com

Jeevan Jyoti Nakarmi
Central Department of Physics, Tribhuvan University,
Kirtipur, Kathmandu-44600, Nepal
Department of Physics, Patan Multiple Campus, Tribhuvan University,
Lalitpur-44700, Nepal

Ajay Kumar Jha
Department of Mechanical and Aerospace Engineering, Institute of Engineering,
Tribhuvan University, Pulchowk Campus, Lalitpur, Nepal

ISSN 1846-6168 (Print), ISSN 1848-5588 (Online)
https://doi.org/10.31803/tg-20240819123835

Original scientific paper
Received: 2024-08-19, Accepted: 2025-03-24

Study of Gender-Specific Emotion Expressivity in Speech Using MFCC and CNN

Mayuri Bapat, Shankar M. Mali*, Chandrashekhar Patil

Abstract: Analysis of sentiment is a pivotal component of natural language processing and has recently observed noteworthy evolutions. Still, the impact of gender on expressivity of the emotion remains an undiscovered area. The proposed work utilizes a broad range of over 12000 audio data samples from four different benchmarked datasets RAVDESS, CREMA-D, TESS, and SAVEE. Convolutional Neural Network (CNN) is used to identify and detect patterns and biases according to gender. The study found mixed-gender emotion accuracy at 84.26%, female emotion accuracy at 89.40%, and male emotion accuracy at 82.70%. The proposed work aims to demonstrate that the female voice is more expressive of emotion than the male voice by examining the difference in sentiment expression between genders. This research will enhance the insights of sentiment analysis and can be useful to contrivance industries ranging from customer service to human-system interaction.

Keywords: Convolution Neural Network (CNN); feature extraction; Mel-Frequency Cepstral Coefficients (MFCC); vocal dimorphism

1 INTRODUCTION

Emotion expressivity is the ability to communicate internal emotions through various channels, including voice, facial expressions, body language, and written text. It builds connections, encourages empathy, and improves understanding of human social interaction. Emotion expressivity improves sentiment analysis accuracy by capturing nuances in spoken content. Audio files play a significant role in analyzing and quantifying emotion expressivity through voice. Measurable characteristics that are derived from audio signals are known as audio features. These features offer important information for diversely different applications of audio processing like Gender recognition, music analysis, Speech Emotion Recognition (SER), etc. Labeled training data is used to categorize features using a classifier into predefined classes based on patterns. Researchers used different classifiers such as Support Vector Machines (SVM) [3, 6, 11, 13], Convolutional Neural Networks (CNNs) [1, 12], Recurrent Neural Networks (RNNs) [7], or Long Short-Term Memory networks (LSTMs) [15], Decision Trees and Random Forests [1], Gaussian Mixture Models (GMM) [16], k-Nearest Neighbors (k-NN)[17], Ensemble Methods (e.g., AdaBoost, Gradient Boosting), Hidden Markov Models (HMM) [14], Naive Bayes Classifiers [13], Extreme Gradient Boosting (XGBoost) [19], BERT and GRU [8] which works as decision-making models that learn and infer emotional content from speech signals based on extracted features. Different techniques used for audio signal processing and can be listed as Spectral Features [1], Pitch-related Features, Zero Crossing Rate (ZCR) [1], Mel-frequency Cepstral Coefficients (MFCC) [1, 6, 7, 11, 12, 15, 25], UniSpeech-SAT [2], Harmonic to Noise Ratio (HNR) [3], Prosodic Features [7], Fundamental Frequency (F0), Spectral Features [11], Harmonic-to-Noise, Energy Features, Formant Frequencies, Statistical Features, Tonal centroid features [11], Time-Frequency Representations (e.g. Spectrograms) [18] etc.

The proposed study aims to analyze the emotion expressivity in male and female voices using Convolutional Neural Networks. This analysis is conducted using four benchmark datasets, ensuring a robust and comprehensive evaluation of vocal expressivity across genders.

The research paper's enduring sections are organized as follows: An outline of pertinent research is given, Section 2 highlights the research overview and associated research gap, and Section 3 elaborates on methodology in detail. Section 4 presents the Findings and Interpretation. Section 5 provides the concluding remark along with the possible future research.

2 LITERATURE REVIEW

Every human voice has some specific attributes like pitch, tone, pace, linguistic patterns, and emotions of voice, Emotional components present in the voice define the mental health of the human along with decisions. This section gives a brief about the current scenario of the research work done in voice or audio sentiment analysis with attention to the use of CNN.

Rezapour et al. (2023) [1] - This study aimed to classify emotions in speech using audio features and machine learning models. Researchers used a limited dataset to extract audio features using models like one-dimensional convolutional neural network (conv1D) and random forest (RF). RF with feature selection achieved higher accuracy (69%), precision (72%), and recall (84%) for fear and calm. However, similar acoustic qualities caused the misclassification of anger as happiness, disgust as sadness, and fear as sadness. Atmaja et al. (2022) [2] - This study evaluates sentiment analysis and emotion recognition from speech using self-supervised learning models. The results indicate that two types of sentiment analysis produced the greatest outcomes. On the other hand, higher-class models performed poorly on tests involving sentiment analysis and emotion perception. Performance declines could have been caused by the dataset's imbalanced nature.

Hadhami et al. (2020) [3] - The paper presents an emotion recognition system using speech signals, utilizing a two-stage approach for feature extraction and classification. It extracts a 42-dimensional vector of audio features, uses Auto-Encoder for parameter selection, and uses Support

Vector Machines as a categorizer. Trials are led at the Ryerson Multimedia Laboratory. Khan et al. (2021) [4] - The research demonstrate the effectiveness of the new method in surpassing a baseline system for sentiment analysis on audio data. It also identifies opportunities for further advancement, emphasizing the need to address ASR challenges and explore the integration of pure speech features for improved speech-based sentiment detection.

Madanian et al. (2023) [5] present a systematic review of ML-based research in SER over the past decade, specifically focusing on data processing, feature selection/extraction, and classification steps. It provides a thorough analysis of problems and their remedies, such as low classification accuracy in Orator-unconstrained experiments. Furthermore, the review furnishes guidelines for SER evaluation, emphasizing common baselines and available metrics for experimentation. Selvaraj et al. (2016) [6] - This paper presents an emotion recognition method using the MFCC approach and Radial basis function network, focusing on gender classification using a support vector machine and pitch analysis, proving more accurate than the Back Propagation Network.

Yoon et al. (2018) [7] paper introduces a groundbreaking multimodal approach that effectively combines text and audio inputs, outperforming existing methods in emotion classification. Its success in mitigating misclassification issues related to the neutral class demonstrates its promise for advancing emotion recognition in speech analysis. Yong-hun-Lee [8] used BNC64 corpus data and three types of analyses: dictionary-based, GRU-based, and BERT-based. Despite similar sentiment word usage, women used more positive words. The BERT-based analysis revealed more gender differences, supporting previous studies and highlighting the development of gender differences in sentiment analysis methods. Marianne Latinus [9] Two ERP studies revealed neural correlates of all processing of speeches. Differences in pitch between female and male voices were observed at 87 ms, with N1 and P2 showing early gender effects. P2 differentiated male from female voices regardless of pitch, suggesting that voice gender processing involves two stages: initial pitch discrimination followed by a later, more precise gender identification at P2 latency. Chiara De Amicis [10] - The study analyzed 78,000 earnings conference calls between 2004 and 2018 to compare sentiment between female and male senior managers. Results show female executives use a more positive tone and less vagueness, indicating a linguistic feature. Financial analysts also exhibit gender bias, but the stock market reaction is influenced by the call's sentiment, not the executive's gender. Ref. [11] focuses on detecting gender from voice signals using techniques to identify relevant features. It analyses voice signal features using a dataset, studies machine learning models, and uses feature selection algorithms to improve classification models. Experimental results show that sub-features are crucial for enhancing performance efficiency. Deep learning and SVM models gave 99.97% best recall value, and 100% for feature extraction techniques using SVM.

A three-layer feature extraction technique is used by Uddin et al. [12] to classify gender and region from human voices. In the first layer, it extracts the fundamental frequency, spectral entropy, spectral flatness, and mode frequency; in the second layer, it uses MFCC to map the audio data; and in the third layer, it computes LPC. The method performs better on gender and region classification using a combined dataset.

Maghilnan et al. [13] research aims to analyze sentiment in speech transcripts that distinguish between speakers to recognize the emotions of distinct speakers. It examines various methods for speaker discrimination and sentiment analysis to identify efficient algorithms for this task.

From the literature, we found that there is a need for systematic study to demonstrate gender differences using speech sentiment study and the use of advanced machine learning and deep learning approaches are necessary [8]. It is also suggested that more trials are required to find the spatial-temporal discrimination [9]. Research should consider gender differences in vocal emotion recognition using fully naturalized speech datasets to enhance conservation validity. [14]

3 METHODOLOGIES

3.1 Datasets Used in Experiment

The Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS) dataset contains 1200 files for speech and 1200 files for songs. It is a collection of multimedia types that are preferably used to study speech emotion recognition: songs. Crowd-sourced Emotional Multimodal Actors Dataset (CREMA-D) is an emotive multimodal actor dataset featuring 7,442 original clips from 91 actors. The Toronto Emotional Speech Set (TESS) analyses how age affects an individual's capacity for emotion recognition. It contains 200 neutral phrases imitating seven different emotions. Surrey Audio-Visual Expressed Emotion (SAVEE) is an emotion identification dataset consisting of 480 utterances performed by 4 male actors. To combine these datasets for analysis, preprocessing steps are essential to ensure uniformity across all sources. First, the audio files are converted to a consistent format, such as WAV, and normalized for sampling rate and bit depth. Then, emotion labels are standardized to create a unified taxonomy across datasets, aligning overlapping categories. Features such as MFCCs are extracted from the audio to maintain consistency in feature representation. Finally, the data sets are merged into a single data structure, ensuring a balanced distribution of the samples by emotions and gender to prevent bias during the project cycle, like training and evaluating the model.

Table 1 Metadata of different datasets

Sr. No.	Dataset	FA	MA	NF FV	NF MV	NE	NF	AR
1	RAVDESS	12	12	732	720	8	7356	-
2	CREMA-D	43	48	1849	2304	6	7442	20 and 74
3	TESS	2	-	2800	-	7	2800	26 and 64
4	SAVEE	-	4	-	480	7	480	27-31

3.2 Data Pre-processing

Audio pre-processing is an essential step in modifying unprocessed acoustic signals into a form appropriate for machine learning prototypes. The process begins with audio sampling, where the waveform is captured at a rate of 0.70. If required, resampling is performed to match the model's input requirements or reduce computational overhead, using a target rate. To improve signal quality, noise reduction is applied followed by silence trimming based on amplitude thresholds. These techniques help to remove irrelevant artifacts and ensure that only the relevant portions of the audio are processed, which is critical for tasks like voice expressivity analysis. The raw audio is then converted into feature representations using MFCC. MFCCs are chosen because they effectively capture the timbral qualities of speech, which is essential for recognizing subtle variations in vocal expressivity.

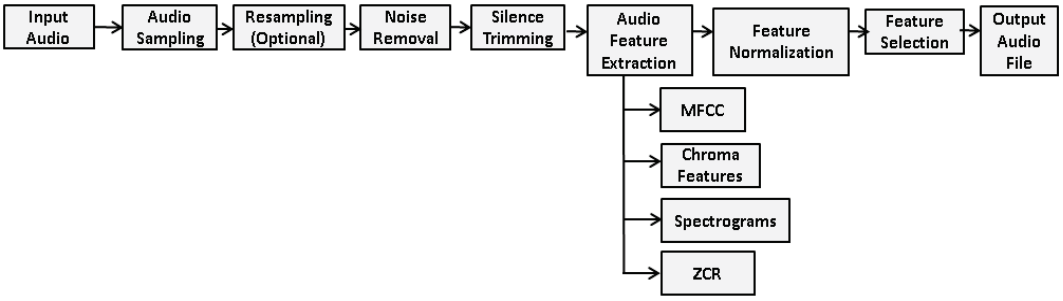


Figure 1 Audio Data Preprocessing Architecture

Table 2 Comparative Analysis of Techniques

Author	ED	MA	Dataset	FSM	AC	FSR
María et al. (2021)	7	Fully Convolutional Neural Network	EMODB, RAVDESS and TESS	Mel Spectrogram, MFCC	75.28 - RAVDESS 92.71 - EMODB, TESS -99.03	Evaluate auditory emotions broadly.
Eduyard et al. (2017)	6	CNN	Own Dataset	MFCC	71.33%	Explore cultural-language emotion identification.
Mu, Y. et al. [21]	4	CNN, BRNN Attention Model, LSTM	IEMOCAP	-	64.08 and 56.41 weighted and unweighted resp	Explore multimodal emotion identification regression.
Thomas, M. et al. [22]	-	RNN, LSTM	Malayalam datasets	-	80%	Deep learning enhances multilingual accuracy
Pavithra, P. et al. [23]	4	RNN	Own dataset	-	83%	personal assistant systems
Meenakshi, S. R. et al. [24]	-	RNN, LSTM, Word2Vec	PyAudio	-	-	Crowd Analysis
Huang, A. et al. [25]	8	SVM and HMM	RAVDESS TESS	MFCC, STFT	85%	Extend features and bidirectional LSTM training.
Jain, M. et al. [26]	4	SVM	Linguistic Data Consortium (LDC) and UGA database	MFCC, LPCC	85.085 %.	Enhance accuracy with MFCC, MEDC
Luo, Z. et al. [27]	5	Utterance-Based Parallel Neural Network	MOST	MFCC	68.72%	Integrate technologies for audio fusion
Cibau, N. E. et al. [28]	7	Autoencoder	EMO-DB	MFCC Prosodic features	70%	This can apply to digit recognition
Sahu, S. et al. [29]	4	Autoencoder SVM	IEMOCAP	Spectral, Prosody, and Energy features	58.38	Examine emotive speech in low-dimensional coding
Patel, N. et al. [30]	7	SVM, Decision tree classifier, CNN, Autoencoder	RAVDESS TESS	MFCC	96%	Replace decision trees with LSTMs/CNNs
Latif, S. et al. [31]	4	adversarial auto encoder	IEMOCAP MSP-IMPRO Librispeech	generative adversarial models	65.1	Integration of reinforcement learning expected
Hajarolasvadi, N. & Demirel, H. [32]	6	3D CNN	RML, SAVEE and eNTERFACE'05	MFCC	81.05% for SAVEE	Explore 3D architecture comparisons, data augmentation.

ED - Emotion Detected, MA - ML Algorithms, FSM - Feature Selection Method, AC - Accuracy, FSR - Further Scope of Research

These features are normalized using StandardScaler to ensure they share the same scale and distribution, which helps improve model convergence and performance. Data augmentation techniques, such as pitch shifting (with a pitch factor of 0.8), time stretching (with a factor of 0.75), and noise addition (with a sample rate of 0.70), are employed to expand the dataset. These augmentations are critical for simulating variations in voice due to factors like pitch, speed, and environmental noise. The use of data augmentation helped to address potential imbalanced classes in the dataset, ensuring

that the model learns to generalize across a broader spectrum of expressivity patterns. For example, by artificially altering the pitch and speed, the model can better recognize both male and female vocal expressivity across different conditions. Furthermore, feature determination methods are used to identify the most relevant characteristics that meaningfully affect the expressivity analysis. This ensures that the model does not overfit and focuses on the most important characteristics for classification. Finally, the pre-processed audio data are now ready for the machine learning model for further classification. In the proposed method Convolutional Neural Networks (CNN), are used which are well suited for capturing hierarchical patterns in Spectro-temporal features like MFCCs. This model will be used to analyze male and female vocal expressions, recognizing emotional and expressive cues from the processed audio data. Fig. 1 shows the audio data preprocessing architecture.

3.3 Different Approaches Used to Analyze Audio Sentiments

Convolutional Neural Networks (CNN) [32], Recurrent Neural Networks (RNN) [22-24], a combination of Convolutional and Recurrent Neural Networks (CRNN) [21, 37], Support Vector Machines (SVM) [26] and autoencoders [28-30] are mainly used and suitable for audio/voice/speech sentiment analysis. The use of one of these algorithms may depend on the characteristics of the dataset selected for the experiment and available resources. CNN master in feature extraction. The possibility of data overfitting is controlled by RNN. SVM is used to classify the emotions. Autoencoders are mainly used in unsupervised learning to focus on operations like data compression, feature learning, and dimensionality reduction. Tab. 2 gives a comparative analysis of these techniques with the scope of future research.

3.4 Feature Extraction Using MFCC

Feature extraction in audio files involves transforming raw data into meaningful features for easier analysis and classification by machine learning models or signal processing algorithms. Three axes-time, amplitude, and frequency represent the three dimensions of the audio stream [34]. The working of MFCC is depicted in Fig. 2.

Figure 2 Steps to perform MFCC

Audio Framing consists of Segmentation and Windowing. Segmentation divides the audio signal into short overlapping frames using techniques like the Hamming window. After that, apply a window function to each frame to minimize artifacts at frame boundaries. Fast Fourier Transform (FFT) helps to represent the frequency domain and to obtain a power spectrum. The frequency domain audio segment was converted from the temporal domain to the spectral domain using FFT. Calculation of the power spectrum from the magnitude of the FFT output is done to obtain the power spectrum. A Mel Filterbank is used to divide an audio sample into distinct Mel frequency scale frequency bands. To estimate the human perception of voice intensity Logarithmic Transformation takes the log of the filter. To Extract the high and low frequency changes from the audio signal Discrete Cosine Transform (DCT) is used. Applying DCT to log filter bank to de-correlate and coefficients helps to identify the most pertinent information. Coefficient Calculation usually discards higher-frequency coefficients, keeping a subset of the generated DCT coefficients as the final MFCCs. Feature Vector Construction is the next step, and it consists of two subsets: Vectorisation and Dynamic Features (Optional). Each frame is represented as a feature vector after vectorization aids in the selection of MFCC coefficients. Dynamic Features are an optional step support that computes delta and delta-delta coefficients to record the acceleration and rate of change of MFCCs between frames. The third stage, the final presentation, comprises the generated MFCC feature set and creates a series of feature vectors, each of which represents a distinct audio signal time window. By retaining certain spectral properties and eliminating others, these MFCC feature extraction methods seek to effectively represent the audio content.

3.5 Data Preparation

Data preparation involves preprocessing the data necessary for solving the problem of multiclass classification. Initially, all essential libraries like Scikit-learn were imported for preprocessing, evaluation, and data splitting. Secondly, labels and feature extraction are done from female voice and male voice datasets. Extracted features and labels for each gender are stored in separate variables. Using one-hot encoding categorical labels are transformed into binary matrix representations suitable for classification tasks. After concatenating the data for both genders, an 80-20 split is used to separate the merged dataset into training and testing sets. This ensures that the model is trained on a diverse dataset and evaluated on unseen data. The code further splits the datasets by gender, maintaining the same 80-20 train-test ratio. This separation allows for individual evaluation of the model's performance on female and male voices. To apply standardization, StandardScaler is used on training and testing datasets. This is achieved by scaling the features to have unit variance and zero means. This step is essential for ensuring that the structures are on a similar scale, improving the convergence of the training process. After standardization, the feature arrays are reshaped to add an extra dimension, making them compatible with the input requirements of Convolutional Neural Networks (CNNs). Lastly, the final shape of the dataset shows that the prepared data is three-dimensional, and the third dimension represents the channel of a single feature.

CNN expects input data in the form of RGB Channels,i.e. multiple channels, hence reshaping of the data is critical for CNNs. After this step, the dataset is ready to be fed into the proposed machine-learning model for training and evaluation. To illustrate probability distribution around the mean of continuous random variable Gaussian Curve or the Bell Curve is used. It is given in Eq. (1).

$$Z = \frac{(X - \mu)}{\sigma}, \tag{1}$$

where X is a normal arbitrary variable, μ is the mean of the distribution, σ is the standard deviation which measures the spread or variability of the distribution and z is the z-score over variable X. In summary, the code meticulously prepares the data for training a CNN by extracting features, one-hot encoding labels, splitting datasets, standardizing features, and reshaping arrays. These steps ensure that the data is in the optimal format for training an effective model to analyze gender-driven variations in sentiment.

3.6 Model Design

Convolutional Neural Network (CNN) is mainly used and suitable for audio or video processing [24]. It includes a sequential model, convolution layers, pooling, dropout, flattened, and dense layers. It master in feature extraction. Features may include identifying emotional changes, speakers, and their recognition, for example, speech pattern classification. In the case of voice expressiveness analysis using CNN, the process starts with input audio. The input audio is transformed into a spectrogram, which is a 2D representation of time versus frequency. A CNN applies a convolution operation wherein filters slide over the spectrogram to extract critical features such as pitch, tone, and amplitude variations. This operation generates feature maps, where each output captures meaningful characteristics related to voice expressiveness. In the proposed work, we have used multiple layers of convolution. The output layer is denoted by l and is calculated as given in Eq. (2).

$$y^{(l)}[m] = f\sum_{k=0}^{k-1} x^{(l-1)}[m-k]\cdot h^{(l)}[k] + b^{(l)}. \tag{2}$$

Where, $x^{(l-1)}$ denotes input to the lth layer, $h^{(l)}[k]$ are Learnable kernel weights for layer l, $b^{(l)}$ represents Learnable bias term for the layer l and $f(.)$ is a Non-linear activation function.

Pooling operations such as average pooling are applied to decrease the dimensionality of these feature maps while retaining the greatly significant characteristics and are continued by subsequent convolutional layers of 128 filters by wrapping up additional average pooling layers. The main task of the average pooling layer is to decrease the latitude dimensions, smooth the feature maps, and help embrace global context, which is crucial for accurate emotion detection. In the first layer of the model kernel size of 6 with 256 filters is used, then to dwindle feature map down sampling while maintaining important data average pooling is used. Subsequently, a convolutional layer with 64 filters is applied, followed by an average pooling layer that enhances key features by choosing the highest values from each pooling region.

To prevent overfitting by randomly omitting units during training, a dropout layer with a 20% dropout rate is incorporated, thus improving the model's robustness. After this, 3D feature maps are flattened into 1D vectors as a fact for fully connected condensed layers. A dense layer with 32 units combines the features nonlinearly, followed by another dropout layer with a 30% dropout frequency to decrease the overfitting of the model. The output layer, consisting of 8 units with a SoftMax activation function, provides a distribution of probabilities across the eight emotion groups, allowing multi-class organization. The Adam optimizer, well-known for its efficiency in the training of machine learning models, is used to build the model, as well as the category cross-entropy loss function. It is appropriate for solving problems involving multiple class categorizations. The accuracy metric is used to examine the performance of the model. Lastly, the features extracted are processed through wholly connected layers, which analyze and classify the features to determine specific voice expressiveness attributes. This structured approach makes CNN able to effectively model and interpret the nuances of voice expressiveness from audio data. Fig. 3 shows the CNN model architecture implemented. SVG-NN is used to create the proposed architecture.

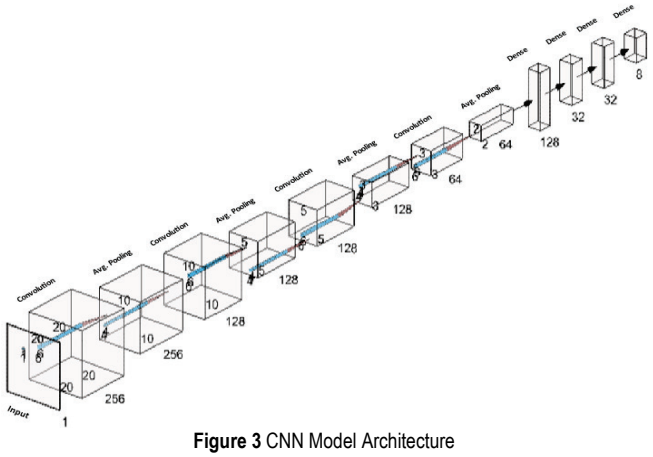


Figure 3 CNN Model Architecture

4 PERFORMANCE EVALUATION

The training features and matching labels are the inputs of the training and testing evaluation method, which is used to determine a proposed model's performance on two distinct datasets, i.e. training and testing. The results of this evaluation are stored in the score variable, which contains various evaluation metrics, such as accuracy. The print function is used to print the training and testing accuracy of the model up to two decimal places. To assess the model performance testing data is used in the second part of the model. Accuracy values give insights into the effectiveness of the final model in the prediction of male, female, and mixed-gender accuracy on training and testing data. Better

performance of the model is indicated with higher accuracy values. The generalization ability of the model is evaluated and provides valuable information about its effectiveness in real-world scenarios.

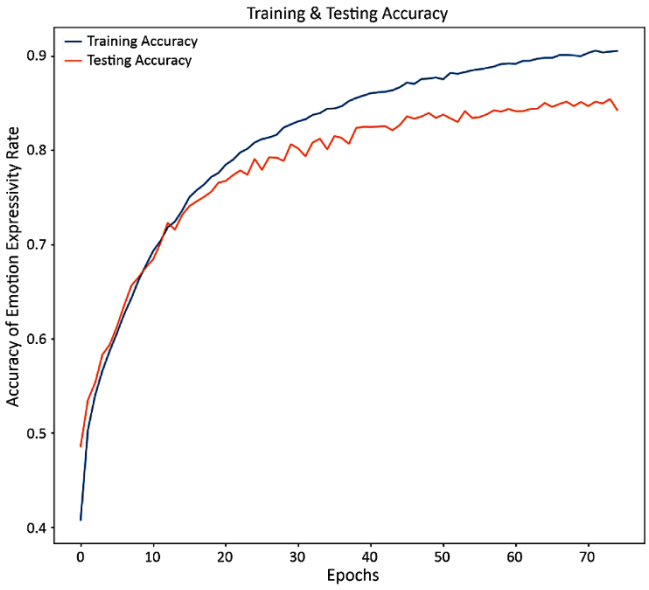


Figure 4 Accuracy across Gender

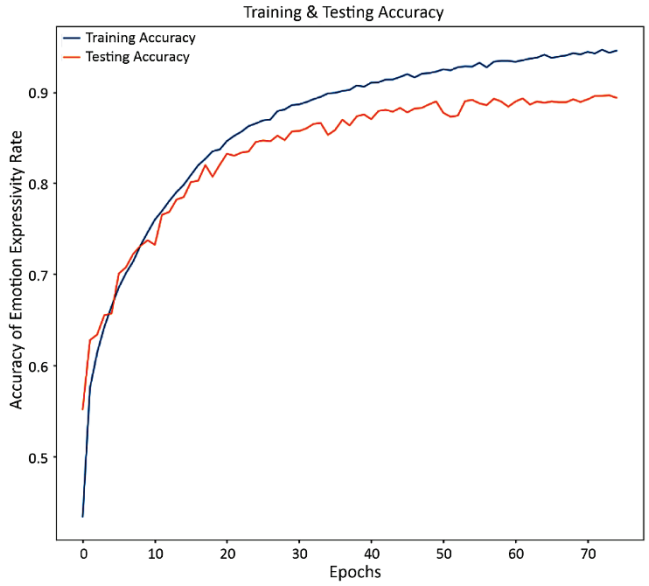


Figure 5 Female Voice Precision

The accurate emotion expressivity rate of mixed-gender, female, and male is presented in Figs. 4, 5, and 6 respectively. Distinct variations in accuracy across these categories are revealed by our analysis. It also highlights the impact of gender-specific characteristics on performance evaluation. Fig. 4 depicts the overall accuracy of emotion expressivity rate when analyzing a dataset that includes both male and female voices. The accuracy achieved in this mixed-gender scenario is 84.26%. This result provides a baseline understanding of the system's performance when gender-specific variations are not explicitly accounted for. Fig. 5 focuses on the accuracy of emotion expressivity rate specifically for female voices. The system gets an accuracy rate of 89.40% when analyzing female speech. This higher accuracy suggests that the emotional expressions in female voices are more consistently captured by the sentiment analysis model, indicating a better alignment between the attributes mined by the Mel-Frequency Cepstral Coefficients (MFCC) and the emotional content in female speech. Fig. 6 shows the accuracy of the emotion expressivity rate for male voices, which stands at 82.70%.

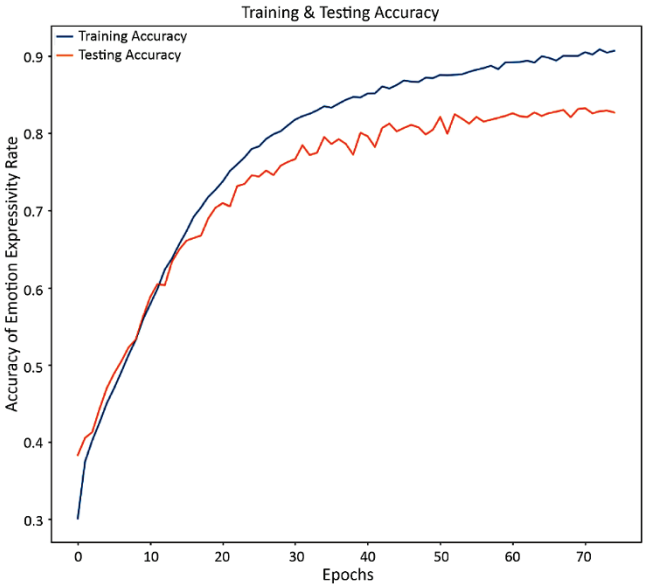


Figure 6 Male Voice Precision

From the results depicted in Figs. 4, 5, and 6, it is evident that female voices are more accurately analyzed for emotions, with a notable accuracy rate of 89.40%. These findings suggest that female voices may exhibit more pronounced or detectable emotional cues that align well with the features captured by MFCC. The disparity in accuracy rates underscores the importance of considering gender-specific characteristics in sentiment analysis.

A detailed comparative study of other research work focussed on emotion recognition of male, female, and mixed-gender is given in Tab. 3. Study shows that researchers have used different benchmark datasets like RAVDESS, EMODB, SAVEE, CASIA, etc. or some have developed their own datasets for research work. Different machine learning techniques are employed on varying sizes of voice samples. In the proposed work 12460 voice samples are used. The proposed work has a favorable outcome as CNN with average pooling is used on the integration of four datasets. It recognizes a broad spectrum of emotions and demonstrates its versatility and comprehensive approach. The proposed methodology results in the identification of expressiveness in a female voice in comparison with a male voice and from the obtained accuracy it is proved. As compared to other work proposed model provides competitive and balanced performance, making it a reliable solution for emotion recognition tasks across different gender-specific and mixed-gender datasets.

Table 3 Comparative analysis of various research studies on emotion recognition rates across male, female, and mixed-gender

Related Work	Dataset	Model	Voice Samples Used	Male Sentiment Credit Rate	Female Sentiment Credit Rate	Mix Gendered Credit Rate	Work Done
Nasaruddin, N. et al. [20]	Dataset from Kaggle	CNN with ResNet50 and ResNet101	3000	-	-	ResNet50: 99.67 ResNet101:99.82	Gender classification and detection
Kanwal, S. et al. [33]	RAVDESS EMO-DB SAVEE	SVM	1440	75.49 86.2 -	91.12 88.3 -	82.59 89.6 77.7	Speaker-dependent and speaker-independent speech recognition
Sun, T. W. [35]	CASIA	CNN	1200	-	-	84.60	Speech recognition with and without gender information.
Madhu, M. et al. [36]	RAVDESS	SVM	1440	-	-	72.02	Gender Recognition Using Speech
Singh, V. et al. [37]	RAVDESS	CNN	1440	-	-	72.07	Gender Dependent Speech Emotion Recognition System
Bhukya, S. [38]	Own Dataset	K-means algorithm	400	78	84	58	Gender Difference Identification and Genderize Speech Recognition
Dewan Arpita, H. et al. [39]	corpus of Bengali conversations	CNN	3185			99.37	Gender Identification Using Bengali Speech
Proposed Work	RAVDESS, TESS, SAVEE and CREMA-D	CNN with Average Pool	12460	82.70	89.40	84.26	Gender-Based expressivity of speech

5 CONCLUSION AND FUTURE ENHANCEMENT

We have achieved the key objective of the venture which is to identify the difference between the expressiveness of emotions of males and females using convolutional neural networks with MFCC. To fulfill this requirement, we have used four benchmark datasets that include 8 different emotions. From the accuracy mentioned in the above section, it is clear that the female voice is more expressive than the male voice. Following are the valuable contributions of the proposed research work: 1) We have developed a method to separate male and female voice files. 2) Synthetic training samples are created by adding perturbations to the initial training set. 3) The generalized CNN model was developed to capture the expressivity of the male, female, and mixed-gender voices using pitch data of the voice. 4) From the accuracy of emotion expressivity rate we have proved that female voice is more expressive than male voice. Additionally, we would like to use the proposed method in a real-time system. One can use the transfer learning technique along with the exploration of pre-trained models on voice data for sentiment analysis to improve accuracy. The proposed research can be extended by integrating speech data with different strategies like facial expressions and text. Exploring multimodal deep learning models for sentiment analysis possibly will lead to more precise and robust results. The future research directions mentioned above would contribute to further advancements in this field and enable the development of more sophisticated sentiment analysis systems with practical applications in various domains.

6 REFERENCES

[1] Mashhadi, M. M. R., & Osei-Bonsu, K. (2023). Speech emotion recognition using machine learning techniques: Feature extraction and comparison of convolutional neural network and random forest. *PLoS ONE, 18*(11), e0291500. https://doi.org/10.1371/journal.pone.0291500

[2] Atmaja, B. T., & Sasou, A. (2022). Sentiment Analysis and Emotion Recognition from Speech Using Universal Speech Representations. *Sensors, 22*(17), 6369. https://doi.org/10.3390/s22176369

[3] Aouani, H., & Ayed, Y. B. (2020). Speech Emotion Recognition with deep learning. *Procedia Computer Science, 176*, 251-260. https://doi.org/10.1016/j.procs.2020.08.027

[4] Khan, P. A., Sumanth, T., & Vardhan, K. V. (2021). Audio sentiment analysis. *International Journal of Creative Research Thoughts (IJCRT), 9*(5), e835-e837. https://ijcrt.org/papers/IJCRT2105531.pdf

[5] Madanian, S., Chen, T., Adeleye, O., Templeton, J. M., Poellabauer, C., Parry, D., & Schneider, S. L. (2023). Speech emotion recognition using machine learning — A systematic review. *Intelligent Systems with Applications, 20*, 200266. https://doi.org/10.1016/j.iswa.2023.200266

[6] Selvaraj, M., & Karthik, S. P. (2016). Human speech emotion recognition. *International Journal of Engineering and Technology (IJET), 8(1),* 311-323. https://www.researchgate.net/publication/299185942_Human_speech_emotion_recognition

[7] Yoon, S., Byun, S., & Jung, K. (2018). Multimodal Speech Emotion Recognition Using Audio and Text. *arXiv* [Cs.CL]. http://arxiv.org/abs/1810.04635 https://doi.org/10.1109/SLT.2018.8639583

[8] Lee, Y.-H., & Kim, J.-H. (2021). A sentiment analysis of men's and women's speech in the BNC64. In K. Hu, J.-B. Kim, C. Zong, & E. Chersoni (Eds.), *Proceedings of the 35th Pacific Asia Conference on Language, Information and Computation* (pp. 603-610). https://aclanthology.org/2021.paclic-1.63

[9] Latinus, M., & Taylor, M. J. (2011). Discriminating Male and Female Voices: Differentiating Pitch and Gender. *Brain Topography*, 25(2), 194-204. https://doi.org/10.1007/s10548-011-0207-9

[10] De Amicis, C., Falconieri, S., & Tastan, M. (2021). Sentiment analysis and gender differences in earnings conference calls. *Journal of Corporate Finance, 71*, 101809. https://doi.org/10.1016/j.jcorpfin.2020.101809

[11] Alkhawaldeh, R. S. (2019). DGR: Gender Recognition of Human Speech Using One-Dimensional Conventional Neural Network. *Scientific Programming*, *2019*, 1-12. https://doi.org/10.1155/2019/7213717

[12] Uddin, M. A., Pathan, R. K., Hossain, M. S., & Biswas, M. (2021). Gender and region detection from human voice using the three-layer feature extraction method with 1D CNN. *Journal of Information and Telecommunication, 6*(1), 27-42. https://doi.org/10.1080/24751839.2021.1983318

[13] Maghilnan S., & Rajesh Kumar, M. (2018). Sentiment analysis on speaker-specific speech data. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.1802.06209

[14] Lausen, A., & Schacht, A. (2018). Gender Differences in the Recognition of Vocal Emotions. *Frontiers in Psychology, 9*. https://doi.org/10.3389/fpsyg.2018.00882

[15] Schuller, B., Rigoll, G., & Lang, M. (2003). Hidden Markov Model-based Speech Emotion Recognition. ICASSP, IEEE *International Conference on Acoustics, Speech, and Signal Processing – Proceedings (ICASSP)*. 2. 401-404. https://doi.org/10.1109/ICME.2003.1220939

[16] Patel, P., Chaudhari, A. A., Pund, M. A., & Deshmukh, D. H. (2017). Speech Emotion Recognition System Using Gaussian Mixture Model and Improvement Proposed via Boosted GMM. *IRA-International Journal of Technology & Engineering, 7*(2 (S)), 56. https://doi.org/10.21013/jte.ICSESD201706

[17] Lanjewar, R. B., Mathurkar, S., & Patel, N. (2015). Implementation and Comparison of Speech Emotion Recognition System Using Gaussian Mixture Model (GMM) and k-Nearest Neighbor (k-NN) Techniques. *Procedia Computer Science, 49*, 50-57. https://doi.org/10.1016/j.procs.2015.04.226

[18] Luitel, S., & Anwar, M. (2022). Audio Sentiment Analysis using Spectrogram and Bag-of-Visual-Words. *The IEEE 23rd International Conference on Information Reuse and Integration for Data Science (IRI)*, San Diego, CA, USA, 200-205. https://doi.org/10.1109/IRI54793.2022.00052

[19] García-Ordás, M. T., Alaiz-Moretón, H., Benítez-Andrades, J. A., García-Rodríguez, I., García-Olalla, O., & Benavides, C. (2021). Sentiment analysis in non-fixed length audios using a Fully Convolutional Neural Network. *Biomedical Signal Processing and Control, 69*, 102946. https://doi.org/10.1016/j.bspc.2021.102946

[20] Nasaruddin, N., Pratama Tresma, M. A. P., Muchamad, M. K., & Fuadi, Z. (2024). Voice frequency-based gender classification using convolutional neural network for smart home. *IEEE Access, 12*, 104190-104203. https://doi.org/10.1109/ACCESS.2024.3434547

[21] Mu, Y., Gómez, L. a. H., Montes, A. C., Martínez, C. A., Wang, X., & Gao, H. (2017). Speech Emotion Recognition Using Convolutional-Recurrent Neural Networks with Attention Model. *DEStech Transactions on Computer Science and Engineering*. https://doi.org/10.12783/dtcse/cii2017/17273

[22] Thomas, M., & Latha, C. A. (2018). Sentimental analysis using recurrent neural network. *International Journal of Engineering & Technology, 7*(2.27), 88. https://doi.org/10.14419/ijet.v7i2.27.12635

[23] Pavithra, P., Priya, N., & Naveenkumar, E. (2022). Recurrent Neural Network Based Speech Emotion Detection using Deep Learning. *Journal of Science Technology and Research, 3*(1).

[24] Meenakshi, S. R., Kumar, S. D., Rajasekar, D., & Prasad, S. S. (2020, August). Sentiment analysis using recurrent neural networks. In *National Conference on Recent Advancements in Communication, 7*(08).

[25] Huang, A., & Bao, P. (2019). Human vocal sentiment analysis. *arXiv*. https://arxiv.org/abs/1905.08632

[26] Jain, M., Narayan, S., Balaji, P., Bhowmick, A., & Muthu, R. K. (2020). Speech emotion recognition using a support vector machine. *arXiv*. https://arxiv.org/abs/2002.07590.

[27] Luo, Z., Xu, H., & Chen, F. (2019, January). Audio sentiment analysis by heterogeneous signal features learned from utterance-based parallel neural networks. In *AffCon@AAAI* (pp. 80-87). https://doi.org/10.29007/7mhj

[28] Cibau, N. E., Albornoz, E. M., & Rufiner, H. L. (2013). Speech emotion recognition using a deep autoencoder. *Anales de la XV Reunion de Procesamiento de la Informacion y Control*, *16*, 934-939.

[29] Sahu, S., Gupta, R., Sivaraman, G., AbdAlmageed, W., & Espy-Wilson, C. (2018). Adversarial auto-encoders for speech-based emotion recognition. *arXiv preprint arXiv:1806.02146*. https://arxiv.org/abs/1806.02146 https://doi.org/10.21437/Interspeech.2017-1421

[30] Patel, N., Patel, S., & Mankad, S. H. (2022). Impact of autoencoder-based compact representation on emotion detection from audio. *Journal of Ambient Intelligence and Humanized Computing*, *13*(2), 867-885. https://doi.org/10.1007/s12652-021-02979-3

[31] Latif, S., Rana, R., Khalifa, S., Jurdak, R., Epps, J., & Schuller, B. W. (2020). Multi-task semi-supervised adversarial autoencoding for speech emotion recognition. *IEEE Transactions on Affective Computing*, *13*(2), 992-1004. https://doi.org/10.1109/TAFFC.2020.2983669

[32] Hajarolasvadi, N., & Demirel, H. (2019). 3D CNN-based speech emotion recognition using k-means clustering and spectrograms. *Entropy*, *21*(5), 479. https://doi.org/10.3390/e21050479

[33] Kanwal, S., & Asghar, S. (2021). Speech emotion recognition using clustering-based GA-optimized feature set. *IEEE Access*, *9*, 125830-125842. https://doi.org/10.1109/ACCESS.2021.3111659

[34] Zhang, L. M., Li, Y., Zhang, Y. T., Ng, G. W., Leau, Y. B., & Yan, H. (2023). A deep learning method using gender-specific features for emotion recognition. *Sensors*, *23*(3), 1355. https://doi.org/10.3390/s23031355

[35] Sun, T. W. (2020). End-to-end speech emotion recognition with gender information. *IEEE Access*, *8*, 152423-152438. https://doi.org/10.1109/ACCESS.2020.3017462

[36] Nashipudimath, M. M., Pillai, P., Subramanian, A., Nair, V., & Khalife, S. (2021). Voice feature extraction for gender and emotion recognition. In *ITM Web of Conferences* (Vol. 40, p. 03008). EDP Sciences. https://doi.org/10.1051/itmconf/20214003008

[37] Singh, V., & Prasad, S. (2023). Speech emotion recognition system using gender-dependent convolution neural network. *Procedia Computer Science*, *218*, 2533-2540. https://doi.org/10.1016/j.procs.2023.01.227

[38] Bhukya, S. (2018). Effect of gender on improving speech recognition system. *International Journal of Computer Applications*, *179*(14), 22-30. https://doi.org/10.5120/ijca2018916200

[39] Dewan Arpita, H., Al Ryan, A., Fahad Hossain, M., Sadekur Rahman, M., Sajjad, M., & Noor Islam Prova, N. (2025). Exploring Bengali speech for gender classification: machine learning and deep learning approaches. *Bulletin of Electrical Engineering and Informatics, 14*(1), 328-337. https://doi.org/10.11591/eei.v14i1.8146

Authors' contacts:

Mayuri Bapat
Department of Computer Science and Application,
Dr. Vishwanath Karad MIT World Peace University,
Pune, Maharashtra, India 411038
mmbapat@mitacsc.ac.in

Shankar M. Mali
(Corresponding author)
Department of Computer Science and Application,
Dr. Vishwanath Karad MIT World Peace University,
Pune, Maharashtra, India 411038
shankar.mali@mitwpu.edu.in

Chandrashekhar Patil
Department of Computer Science and Application,
Dr. Vishwanath Karad MIT World Peace University,
Pune, Maharashtra, India 411038
chpatil.mca@gmail.com

ISSN 1846-6168 (Print), ISSN 1848-5588 (Online)
https://doi.org/10.31803/tg-20240906044453

Original scientific paper
Received: 2024-09-06, Accepted: 2025-03-13

Power Flow Optimization in Cyber-Physical Systems Using Jellyfish Optimization Technique

Vijoy Kumar Peddiny*, Brajagopal Datta, Abhik Banerjee

Abstract: Cyber-physical systems (CPS) are under increasing strain, which has created a number of problems for the power system, including voltage instability, line overloads, and rising power losses. In order to solve these problems, it is essential to strategically deploy energy resources at the right nodes to maximise real and reactive powers, hence lowering losses and enhancing the voltage profile, especially in crowded networks. Grid operators can more effectively meet the demand for electricity while preserving the stability of the power system thanks in large part to power flow optimisation. However, because they are unpredictable, integrating intermittent renewable energy sources (RESs) presents particular difficulties. The power flow model put out in this study includes four different kinds of energy sources: thermal power plants (which stand in for traditional energy sources), wind power plants, battery charging stations, and solar photovoltaic power plants (which represent renewable energy sources). The goal is to assess the efficiency of the upgraded IEEE 30-bus test system through optimisation utilising an integrated strategy integrating fuzzy logic and jellyfish search optimizer (JS). Practical factors like the thermal generators' ramp rate restrictions and the erratic generation patterns of renewable energy sources are taken into account by the suggested algorithm. The algorithm effectively solves the optimum power flow (OPF) problem by minimising overall generating costs and reaching solution convergence, as shown by simulation results. The suggested method demonstrates its effectiveness in solving the difficulties of power flow optimisation by taking into account real-world scenarios and adding the complexities associated with RESs. The algorithm's effectiveness in assuring an ideal and affordable power flow solution is demonstrated by its ability to handle the intermittent nature of renewable energy sources and the limitations of thermal generators.

Keywords: Cyber-Physical System's (CPS); Fuzzy; Jellyfish; Optimum Power Flow (OPF); Smart Grid

1 INTRODUCTION

The main tool for designing and managing electric networks is now OPF [1]. The goal of the Optimal Power Flow (OPF) problem is to find the optimum solutions for a number of objectives while abiding by a number of equality and inequality conditions, system security restrictions, and the operational limitations of electrical equipment [2]. The active power outputs of generators, bus voltages at power plants, transformer tap settings, and reactive power contributions from VAR compensators are just a few of the control variables that are adjusted to achieve this. The OPF problem's main goals are to minimize total emissions (TE), minimize network losses, and minimize economic fuel costs (EFC) [3]. These goals combine technological, environmental, and economic factors to meet the expanding needs of the energy industry. Recently, academics have become increasingly eager to investigate OPF, a problem with two to three objectives. OPF remedy the purpose is to identify the best options for distinct objective functions. While also upholding a set of equity and limitations of inequality constraints, a solitary best option [4]. There must be trade-offs among two or more opposing aims. As in actuality, there is a collection of the best options, not a single one, unique best OPF solution. Consequently, the choice only the maker's preferences can determine the best course of action, and nobody else can be better to disregard alternative solutions. That is context, one of the well-known Pareto-optimal solutions dependable instruments for multi-objective optimization issues, can gather a group of ideal options.

Numerous traditional optimization techniques have been put forth since the OPF problem was first presented [5]. These techniques include of quadratic programming, mixed-integer mathematical optimization, non-linear programming, and interior-point methods [6-8].The composition of power systems is growing more complicated due to the ongoing integration of computing, communication, and control innovations in smart grids, transforming them into cyber-physical power systems (CPPS) with a close combination of data and physics. It consists of both the conventional physical grid and the power network connectivity created by data acquisition, control, and decision-making units, among other things.

Due to their quick convergence and ability to produce an ideal solution, some of these techniques have been successfully applied by the industry sector. But to use these optimization techniques, the optimization function must first be linearized. For this problem, various properties, such as the quasi, non-differentiable, and non-smooth ones, are frequently estimated for the optimization function.

Heuristic optimization approaches have also been proposed [9-12] as a remedy to this issue. In this way, the OPF is solved using a variety of heuristic methods. In [13], a sequential GA solution strategy was used to solve the OPF issue using a simple genetic algorithm (SGA), resulting in an appropriate control variable resolution that did not violate the system's restrictions. A hybrid genetic algorithm (GA) is introduced in Reference [14], integrating the capabilities of MATPOWER software's linear programming (LP) and sequential quadratic programming (SQP) techniques. This approach enhances the optimization process by leveraging both heuristic and mathematical programming methods.

Additionally, Reference [15] presents a refined genetic algorithm (RGA) designed to efficiently encode multiple control variables within a power system while maintaining a practical chromosomal length, ensuring computational feasibility and effectiveness. However, based on the analogy from nature, various global algorithms have been developed. A number of population-based heuristic techniques, including as genetic algorithms (GA), particle swarm optimization (PSO), teaching-learning-based optimization, evolutionary algorithms, and differential evolution, are frequently used to solve optimization problems. Various variations of these techniques have been investigated for the

Optimal Power Flow (OPF) problem. An electromagnetism-like mechanism optimizer, for example, has been updated and adapted for OPF. Specifically, modifications were made to the local search and force estimating phases, adding two randomly selected particles while maintaining the initialization phase. Furthermore, convergence properties for the OPF issue were improved by combining a Gaussian mutation operator with an adaptive biogeography-based optimization (BBO). To handle the same problem, a modified salp swarm algorithm (SSA) was used in another method, illustrating the variety of approaches used to optimize OPF solutions [21] that exhibited chaotic behaviour. In those reference [22-23], the management of various objective functionalities has been made possible by combining them into a single optimization problem using weighting factors, making the achieved operating point particularly sensitive to the choices of weight factors. For the best placement of fault current limiters and distributed generators in distribution networks, the coyote optimization technique (COA) has been dedicated in Ref. [24]. In this study, a fuzzy based model was used to activate the single objective optimization for the minimization of energy losses, short circuit currents, and installed devices.

The goal of this study is to complete the efforts made to identify the best solution to the OPF problem. This article's contribution can be succinctly summed up as follows:

- The jellyfish search (JS) metaheuristic optimization algorithm, which is brand-new and recently created, is integrated with fuzzy in this study to provide an efficient OPF in cyber-physical systems.
- To test the proposed algorithm's viability in finding the best solution for the OPF problem using renewable energy sources under ideal and realistic settings, it is applied to an IEEE 30 parallel coupled with communication bus system that includes two wind turbines and one solar Photovoltaic generator.

The remainder of this essay is organised in the manner below. The mathematical formalism and related appropriate restrictions used for the Objective functions are shown in Section 1.1. The new idea for applying JS and fuzzy to OPF that incorporates the indeterminate RES is established in section 2. For the four methods under consideration, Section 3 presents the simulation results of various practical case studies. Section 4 and Section 5 offers this paper's conclusions and references at the end.

1.1 Objective Function and Problem Formulation

The working parameters of the IEEE 30-bus system are listed in Tab. 1. Three separate power production resources, namely thermal power generators (TGs) with fixed outputs, solar PV generators (SPGs) with variable outputs, and wind generators (WPGs) with variable outputs, are included in the modified network. The mix of all sources and reserve power must be used to balance this variance in PV and wind outputs, therefore the total generation cost consists of operating expenses for all units, reserve expenses, and penalty expenses.

Table 1 IEEE 30 bus system details

Component	Quantity	Description
Total Buses	30	30 buses are interconnected connected
Communication Points (Wireless)	10	At Bus 1, 2, 8, 5, 11, 15, 17, 21, 22 & 28
Branches	41	These branches are connecting load, sources and buses
Steam based Generator	3	SG1, SG2, SG3, in which Bus 1 is swing bus, other at 2 & 8
Wind Turbine Source	2	WTS1, WTS2 at Bus 5 & 11
Solar Source	2	SS1, SS2 both at Bus 15
Physical Control Variables	05	Voltage of Bus 2, 5, 11, 17, 21
Cyber Control Variables	10	Weight of each communication point

A. **Cost Model for different sources used**. Fossil fuel is used to power thermal power plants. The relationship between steam generator power output in MW and the cost of fossil fuels in $/Hr is provided by Eq. (1).

$$C_{\mathrm{T}}(SG) = \sum_{i=1}^{NTG} x_i + y_i P_{SG} + z_i P_{SG}^2. \tag{1}$$

Where x, y, z shows the cost coefficient of the steam generators while i indicate the i^{th} element of the steam turbine. If the generator units are owned by the Independent System Operator (ISO), the cost function might not apply in the same way when an Independent Grid Operator (IGO) chooses to include compensation costs in the ongoing maintenance and renewal expenses or assigns these costs to the initial investment in solar PV or wind generators. In contrast, the ISO is required to pay costs based on the scheduled power contracts with private owners of solar PV or wind generators if they are privately owned. Depending on how the renewable energy assets are owned, this distinction has an impact on how costs are handled and recorded. The following is how planned power affects the actual costs of the j^{th} wind based renewable power generator:

$$C_{\mathrm{WTS}} = f_j P_{\mathrm{WTS},j}. \tag{2}$$

In a similar manner, the k^{th} solar Photovoltaic PV generator's direct cost is:

$$C_{\mathrm{SS}} = f_k P_{\mathrm{SS},k}. \tag{3}$$

Where $P_{\mathrm{WTS}}/P_{\mathrm{SS}}$ and f, respectively, stand for the planned wind/solar power and the direct cost coefficient associated with the j^{th} wind power/solar power plant.

B. **Uncertainty in the model.** Due to the sporadic nature of wind energy, two situations are possible. The first situation occurs when the wind farm's production power is lower than what was anticipated. Overestimating output power is the word used to describe this circumstance. In this

instance, the grid operators uses spinning reserve to give its customers a dependable power supply [26]. Reserve cost is the price necessary to commit the reserve producing units in order to correct the overestimation problem. The j^{th} wind energy plant's reserve cost is calculated using:

$$C_{\mathrm{WTS,\,RC}} = \\ = P_{\mathrm{WTS},j}\int_0^{P_{\mathrm{WTS}}} (\textit{relative power} \times \textit{power distribution function}). \quad (4)$$

Solar energy also has erratic and sporadic output, just as wind energy. In general, the method used to address under- and overstatement of solar power output should be the same as that used to address wind output power. For solar Photovoltaic PV plant *k*, the reserve cost is provided by:

$$C_{SS,\,\mathrm{RC}} = \\ = P_{\mathrm{SS},\,i} \times \textit{relative power} \times \textit{power distribution function}. \quad (5)$$

C. **Cyber layer interaction with grid**. In, the above cases all consist of physical layer model and uncertainty. The cyber layer model indicates the interaction with the communication network of the system. In this research will use the weight of each Communication Points and it will impact the system directly. If the weight of the communication point is more than 0.8 then will consider as trusted healthy network and if the value is less than this means unhealthy network.

$$W_{\mathrm{cp}} = \begin{cases} > 0.8 & \text{Healthy} \\ < 0.8 & \text{Unhealthy} \end{cases}. \quad (6)$$

D. **Problem statement objective**. As shown in Eqs. (1)-(5), the goal optimum for OPF is generated by incorporating all cost function models. Emission cost is disregarded in the objective function ($F1$). The objective function ($F2$) is constructed with the effect of communication points to better understand the variation in generation scheduling when uncertainty is taken into account.

Consequently, reducing overall cost is the objective function:

$$F1 = C_{\mathrm{SG,\,Total}} + C_{\mathrm{SS}} + C_{\mathrm{WTS}} + C_{\mathrm{SS,\,RC}} + C_{\mathrm{WTS,\,RC}}, \quad (7)$$

$$F2 = W_{\mathrm{cp}} \times \sum_{b=0}^{b=10} \textit{Bus power}. \quad (8)$$

The OPF objective functions are formulated while incorporating both equality and inequality constraints within the system. Notably, achieving power convergence to an optimal solution ensures that the equality constraints are inherently satisfied through power balance equations. Among the inequality constraints, generator bus voltages and active power outputs—excluding those from the swing or slack generator, which is typically assigned to bus 1—serve as self-regulating parameters.

The optimization method employed determines an optimal value for each control variable within its allowable range. Consequently, careful attention is required for inequality constraints related to the slack generator's active and reactive power limits, the reactive power output of other generators, voltage constraints on PQ buses, and transmission line capacity limitations.

In the context of a cyber-physical system (CPS), interactions between the cyber and physical layers must be considered. While cyber-layer disruptions can directly impact the physical system's performance, the reverse influence is typically constrained and does not exhibit the same level of impact.

2 PROPOSED METHEDOLOGY

The three pillars of the JFS optimizer are based on the movements of jellyfish. The first pillar holds that the jellyfishes' movements might be in their swarm or on their journey to the ocean current, and that the temporal control (TC) function can regulate these two forms by shifting between them. The second tenet is that jellyfish are attracted to certain areas once there is enough food available. The numerical objective function is used to characterise food quantity as the third pillar [27, 28]. The jellyfish colony is randomly initialised in chaotic logistic mapping, and can be expressed as follows:

$$X_i(t+1) = 4P_0(1 - X_i) \quad (9)$$

Where P_0 denotes the beginning jellyfish population, which has the potential to produce a value of P_0, and X_i conveys the i^{th} jellyfish based on chaotic value. Their motions in the ocean complete the quest for food.

In areas where there is an abundance of food, they are extremely prone to mobility. This approach is used in this situation because it can explore more effectively. This characteristic is used to determine the global optimal point. According to the flowchart for this modification in Fig. 1, the jellyfish can move in a jellyfish swarm using either passive motion or active motion. In a swarm, jellyfish move both actively (type B) and passively (type A) [29]. Jellyfish start moving passively right after the swarm forms, then after a short while, they start moving actively.

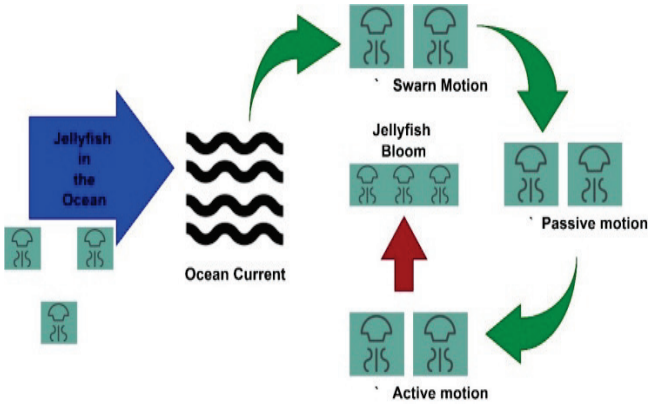


Figure 1 Jellyfish search optimization flow

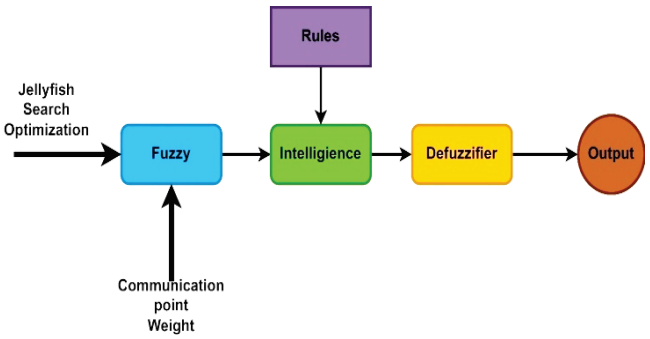


Figure 2 Integration of fuzzy with JS

Table 2 Fuzzy rule for the proposed methodology

Rule	Identification	Description
JS is H and CPW is H	Their optimal values	Healthy system with OPF
JS is H and CPW is L	Individual values	Healthy OPF with cyber disturbance
JS is L and CPW is H	With respect to Rule 1	Unhealthy
JS is L and CPW is L	With respect to Rule 1	Unhealthy with possibility of cyber disturbance
H = High; L = Low		

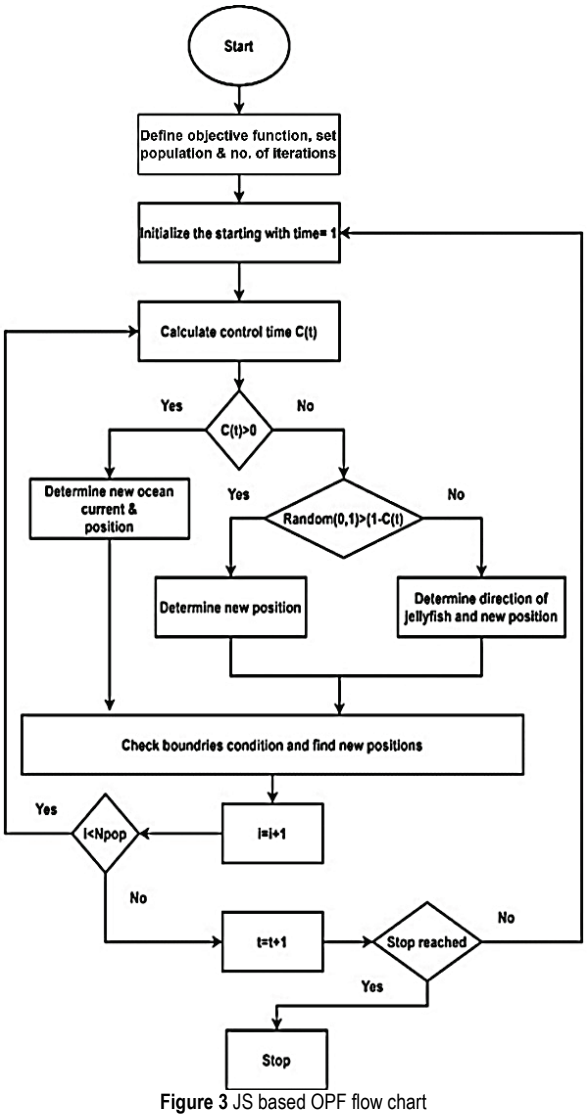


Figure 3 JS based OPF flow chart

The TC mechanism determines whether to choose type A or type B. When compared to a random number in the range [0-1], the phrase $(1 - CF(t))$ is significant in this context. The jellyfish exhibits type A motion if this value is higher than the $(1 - CF(t))$ calculation's result. When the pseudo random is less than the computed value stated, jellyfish, on the other hand, follows type Bmotion If. This concept is further explained by the fact that type A motion is favoured at start point, when the TC function rapidly decreases from 1 to 0, and type B motion is chosen as time advances.

Now the JS OPF output will serve as input for the fuzzy. The final status of the line or bus will be obtained by the interaction with fuzzy and cyber layer of the system.

3 RESULT AND DISCUSSION

When dealing with the OPF issue as previously explained, the equality and inequality requirements are taken into account. The equality conditions, which serve as load flow balancing equations, are effectively upheld by applying the Newton-Raphson algorithm (NRA). The NRA method in this article maintains the uncertainty of (4) and (5) since it establishes the program's steady state for power system operators. Since MATPOWER is employed, the NRA algorithm serves as an effective tool for visualising three-phase networks [30].

Any other limits, such as decision-making constraints and dependent variable constraints, are expressed by two operating limitations. The decision variables of the first type continue to reach their limits, and if either of them is beyond evaluation, they are arbitrarily recreated within their suitable ranges. Additionally, the objective function prolongs and penalises the second type's restrictions (dependent variables). As a result, the jellyfish solution could not be chosen in the following iteration if there were any violations of these criteria.

Table 3 System parameters used

Parameters	Values
Nominal bus voltage	100 V (L-L)
Rated frequency	60 Hz
Source Impedance	1+j1.5 Ω
Non-linear load R-L	20+j79 Ω
dc-link voltage	800 V
Coupling Inductor, L_f	3.5 mH
Overloading factor, a	1.2
DC bus voltage recovery time, t	90 ms
Variation of energy during dynamics, k	20%
Current ripple, ΔI	25% of I
Overshoot voltage, V_{os}	15% of V_{dc}
Switching frequency, f_s	5 kHz
C_{dc}	3.3 mF
L_f	4.2 mH
r_f	0.5 Ω
L_c	0.6 mH
r_c	0.425 Ω
C_f	15 μF
R_d	2.025 Ω
ω_c	50.26 rad/s
ω_n	377 rad/s
$\omega_{c,\,PLL}$	7853.98 rad/s

A well-known test case used in power systems research and analysis is the IEEE 30-bus system. It is frequently used to examine how electrical power systems behave and to evaluate different algorithms and approaches to power system analysis. This system is portrayed as a network of transmission lines and buses that act as reduced versions of an electrical grid. Please be aware that compared to actual power grids, which can be much larger and more complex, the IEEE 30-bus system is a comparatively modest and simplified model. However, it provides a useful benchmark for evaluating and creating methods and methodologies for power system analysis. Fig. 4, shows the actual model used for the simulation in the MATLAB environment.

Parameters used in the model is stated in the Tab. 3.

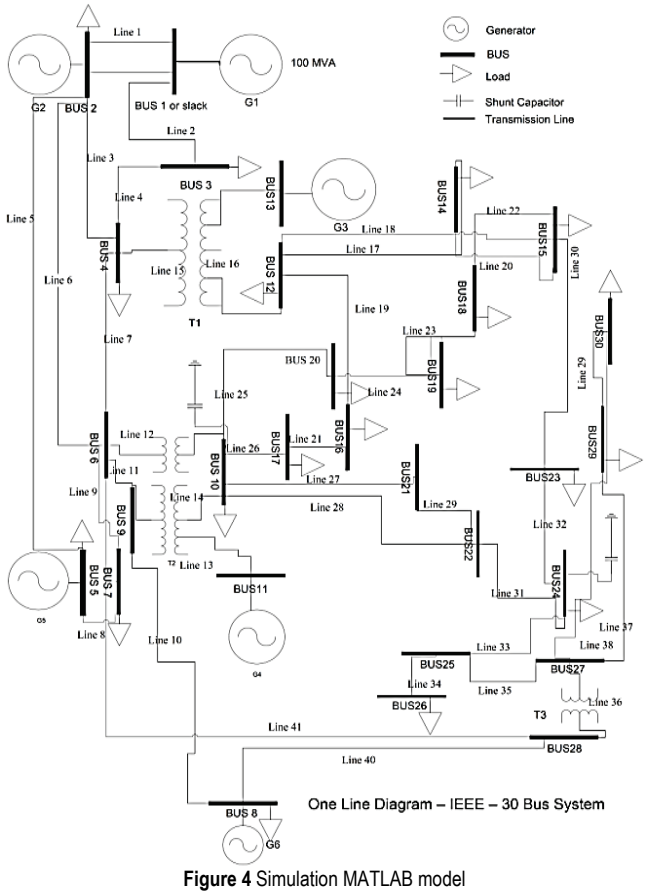


Figure 4 Simulation MATLAB model

3.1 Simulation Results

Various case studies are carried out for the modified IEEE-30 bus system in this section. The outcomes of the various case studies utilising the suggested Jellyfish search (JS) optimization technique is shown and described. ABC, CGO, FPA, and GPC are used as four distinct optimization techniques to test the efficacy of the JS optimization process. Apart from these two cases will be consider:

- Case A: Only Physical layer is taken into consider putting cyber layer at 1.
- Case B: Considering cyber and physical layer both.

The effects of changing the schedule for wind and solar power as well as PDF parameters on generation costs are investigated using case studies A and B. The goal of the subsequent case studies is to maximise the schedule generation across all sources under ideal and realistic circumstances. A minimum of 1000 iterations are used as the finishing condition after the algorithm has been run vs times, the optimal value of the optimization problem of each case study has been identified, and control variable settings have been listed.

3.1.1 Case A

In Case A, we consider only the OPF with a normal cyber layer. Figs. 5 and 6 show the simulation results for the JS. At iteration 100, we obtain the best results with a value of 15,116 for the five inputs. Ideally, when the cyber layer weight is set to 1, it implies no faults in the communication system. However, in reality, this is not the case, leading to a higher global best value. During the simulation, some parameters were affected due to misinterpretations in the communication channel.

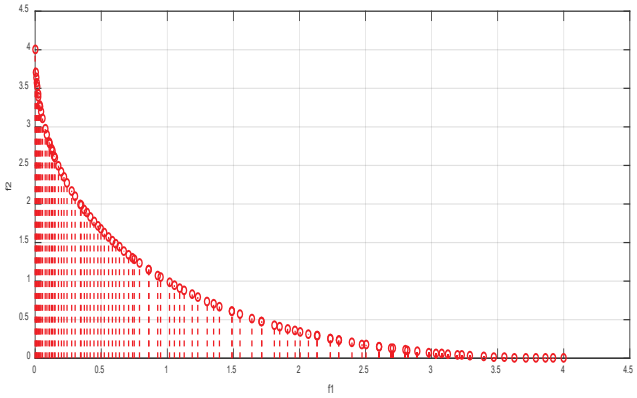


Figure 5 Plot of the objective functions

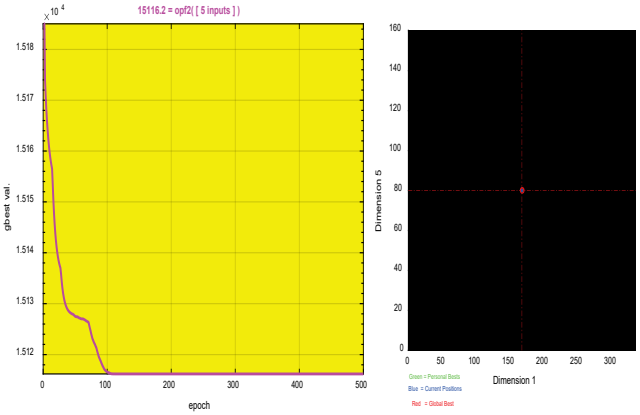


Figure 6 OPF for the case A using JS

3.1.2 Case B

In this case CPS system is consider and for every simulation the weight of the communication point is taken into consideration. So, the effect and OPF will be correctly monitored. So, Fig. 6, shows the results of the simulation by

taking into CPS system. The tuning parameters for the other methods with the proposed method is stated in the Tab. 4. Control variables with the results for the case A and B is shown in the Tab. 4. The overall cost for the simulation is shown in the Fig. 8.

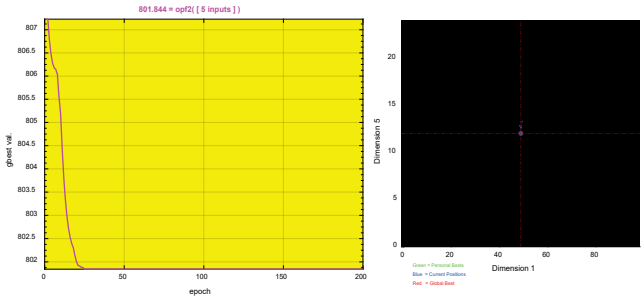


Figure 7 OPF of CPS using JS+Fuzzy

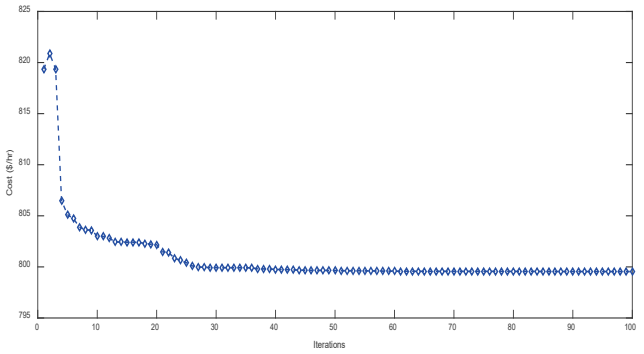


Figure 8 Cost iteration for CPS system

Table 4 Simulation parameters

Parameters	PSO	GA	ABC	JS+Fuzzy
Population	15	15	15	15
Iteration	200	200	200	200
Other parameters	Constriction factor 0.75	Mutation rate 0.1	Limit 100	Not required

Table 5 Case A simulation result

Control Variable	PSO	GA	ABC	JS+Fuzzy
CV1	1.042	1.0332	1.062	1.048
CV2	1.064	1.0414	1.034	1.069
CV3	1.053	1.0510	1.053	1.071
CV4	1.066	1.0676	1.036	1.069
CV5	1.032	1.0332	1.072	1.037

Table 6 Case B simulation result

Control Variable	PSO	GA	ABC	JS+Fuzzy
CV1	1.021	1.054	1.089	1.068
CV2	1.052	1.067	0.99	1.073
CV3	1.067	1.032	1.012	1.035
CV4	1.042	1.048	1.068	1.022
CV5	1.031	1.073	1.011	1.057

Table 7 Computational time for case A and Case B

Technique	Case A (Sec)	Case B (Sec)
PSO	2.24	2.33
GA	2.11	2.21
ABC	1.98	1.99
JS+Fuzzy	1.52	1.47

The control variable for the voltage in P.U is shown in the Tabs. 5 and 6 for both cases. The Computational time for case A and case B is shown in the Tab. 7. Fig. 9, shows the effect of CPS weight of communication point for case A and case B.

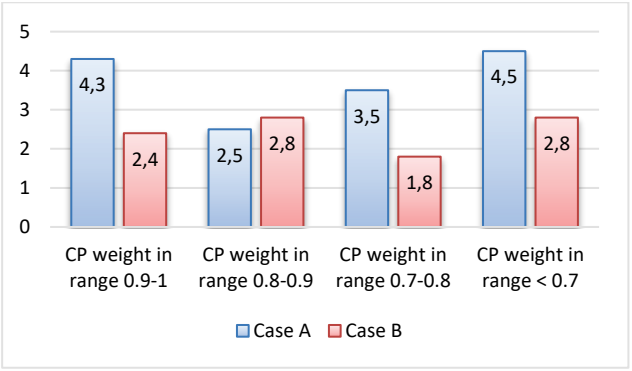


Figure 9 Case A and Case B response as per the weight change in CPS

3.2 Discussion

A comparison on different parameters is stated in the Tab. 8. The comparison is done on 5 major parameters which are stated below:

- Parameter 1: Fast convergence
- Parameter 2: Time taken
- Parameter 3: Is able to deal with CPS system
- Parameter 4: Is possible to extend to other system
- Parameter 5: Accuracy.

Table 8 Comparison [30-32]

Parameter	PSO	GA	ABC	JS+Fuzzy
P1	••	••	•••	••••
P2	•••	••	•••	••••
P3	•	•	•	••••
P4	••	••	••	••••
P5	•••	•••	•••	••••
Four dot means highest value				

4 CONCLUSION

To sum up, this research presents a new metaheuristic optimization approach that addresses the Optimal Power Flow (OPF) problem in a complex power system scenario with stochastic wind and solar energy sources by combining jellyfish optimization with a fuzzy search optimizer. The suggested algorithm provides a reliable method to minimize generation costs while taking into account variable cost coefficients and carbon emission fees. It does this by modeling the unpredictable nature of these renewable sources using Weibull and log-normal probability density functions. The thorough comparison of the suggested JS integrated Fuzzy algorithm with established techniques like ABC, PSO, and GA shows how much better it performs in terms of cost savings and convergence of the solution. These findings highlight how well the JS integrated Fuzzy approach optimizes the OPF problem, indicating that it is a useful tool for managing and planning energy systems in the future.

5 REFERENCES

[1] El-sehiemy, R. A., El-hosseini, M. A., & Hassanien, A. E. (2013). Multi-objective real-coded genetic algorithm for economic/environmental dispatch problem. *Studies in Informatics and Control, 22*(2), 113-122.

[2] Farhat, M., Kamel, S., Atallah, A. M. & Khan, B. (2021). Optimal Power Flow Solution Based on Jellyfish Search Optimization Considering Uncertainty of Renewable Energy Sources. *IEEE Access, 9*, 100911-100933. https://doi.org/10.1109/ACCESS.2021.3097006

[3] Bouchekara, H. R. E. H., Chaib, A. E., Abido, M. A., & El-Sehiemy, R. A. (2016). Optimal power flow using an improved colliding bodies optimization algorithm. *Applied Soft Computing Journal, 42*, 119-131. https://doi.org/10.1016/j.asoc.2016.01.041

[4] Chou, J. S., & Truong, D. N. (2020). Multi-objective optimization inspired by behavior of jellyfish for solving structural design problems. *Chaos, Solitons & Fractals, 135*. https://doi.org/10.1016/j.chaos.2020.109738

[5] Bai, X., & Wei, H. (2011). A semidefinite programming method with graph partitioning technique for optimal power flow problems. *International Journal of Electrical Power & Energy Systems, 33*(7), 1309-1314. https://doi.org/10.1016/j.ijepes.2011.06.003

[6] Yan, W., Yu, J., Yu, D. C., & Bhattarai, K. (2006). A new optimal reactive power flow model in rectangular form and its solution by predictor-corrector primal-dual interior point method. *IEEE Transactions on Power Systems, 21*(1), 61-67. https://doi.org/10.1109/TPWRS.2005.861978

[7] Vargas, L. S., Quintana, V. H., & Vannelli, A. (1993). A tutorial description of an interior point method and its applications to security-constrained economic dispatch. *IEEE Transactions on Power Systems, 8*(3), 1315-1324. https://doi.org/10.1109/59.260862

[8] Momoh, J. A., & Zhu, J. Z. (1999). Improved interior point method for OPF problems. *IEEE Transactions on Power Systems, 14*(3), 1114-1120. https://doi.org/10.1109/59.780938

[9] Granelli, G. P., & Montagna, M. (2000). Security-constrained economic dispatch using dual quadratic programming. *Electric Power Systems Research, 56*(1), 71-80. https://doi.org/10.1016/S0378-7796(00)00097-3

[10] Crisan, O., & Mohtadi, M. A. (1992). Efficient identification of binding inequality constraints in the optimal power flow Newton approach. *IEE Proceedings C - Generation, Transmission and Distribution, 139*(5), 365-370. https://doi.org/10.1049/ip-c.1992.0053

[11] Zehar, K., & Sayah, S. (2008). Optimal power flow with environmental constraint using a fast successive linear programming algorithm: Application to the Algerian power system. *Energy Conversion and Management, 49*(11), 3362-3366. https://doi.org/10.1016/j.enconman.2007.10.033

[12] Shafik, M. B., Chen, H., Rashed, G. I., & El-Sehiemy, R. A. (2019). Adaptive multi-objective parallel seeker optimization algorithm for incorporating TCSC devices into optimal power flow framework. *IEEE Access, 7*, 36934-36947. https://doi.org/10.1109/ACCESS.2019.2905266

[13] Reddy, S. S., & Bijwe, P. R. (2016). Efficiency improvements in meta-heuristic algorithms to solve the optimal power flow problem. *International Journal of Emerging Electric Power Systems, 17*(6), 631-647. https://doi.org/10.1515/ijeeps-2015-0216

[14] Younes, M., Rahli, M., & Abdelhakem-Koridak, L. (2007). Optimal power flow based on hybrid genetic algorithm. *Journal of Information Science and Engineering, 23*(6), 1801-1816.

[15] Paranjothi, S. R., & Anburaja, K. (2002). Optimal power flow using refined genetic algorithm. *Electric Power Components and Systems, 30*(10), 1055-1063. https://doi.org/10.1080/15325000290085343

[16] Bouchekara, H. R. E. H., Chaib, A. E., & Abido, M. A. (2018). Optimal power flow using GA with a new multi-parent crossover considering prohibited zones, valve-point effect, multi-fuels, and emission. *Electrical Engineering, 100*(1), 151-165. https://doi.org/10.1007/s00202-016-0488-9

[17] Bakirtzis, A. G., Biskas, P. N., Zoumas, C. E., & Petridis, V. (2002). Optimal power flow by enhanced genetic algorithm. *IEEE Transactions on Power Systems, 17*(2), 229-236. https://doi.org/10.1109/TPWRS.2002.1007886

[18] El Ela, A. A., Abido, M., & Spea, S. R. (2010). Optimal power flow using differential evolution algorithm. *Electrical Power and Energy Systems, 80*(7), 878-885. https://doi.org/10.1016/j.epsr.2009.12.018

[19] Sivasubramani, S., & Swarup, K. S. (2011). Sequential quadratic programming-based differential evolution algorithm for optimal power flow problem. *IET Generation, Transmission & Distribution, 5*(11), 1149. https://doi.org/10.1049/iet-gtd.2011.0046

[20] Slimani, L., & Bouktir, T. (2012). Optimal power flow solution of the Algerian electrical network using differential evolution algorithm. *TELKOMNIKA Indonesian Journal of Electrical Engineering, 10*(2), 199-210.

[21] Sayah, S., & Zehar, K. (2008). Modified differential evolution algorithm for optimal power flow with non-smooth cost functions. *Energy Conversion and Management, 49(11)*, 3036-3042. https://doi.org/10.1016/j.enconman.2008.06.014

[22] Amjady, N., & Sharifzadeh, H. (2011). Security constrained optimal power flow considering detailed generator model by a new robust differential evolution algorithm. *Electric Power Systems Research, 81(2)*, 740-749. https://doi.org/10.1016/j.epsr.2010.11.005

[23] Abido, M. A. (2002). Optimal power flow using particle swarm optimization. *International Journal of Electrical Power & Energy Systems, 24(7)*, 563-571. https://doi.org/10.1016/S0142-0615(01)00067-9

[24] Kang, Q., Zhou, M., & Xu, C. (2012). Solving optimal power flow problems subject to distributed generator failures via particle swarm intelligence. In *Proceedings of the International Conference on Advanced Mechatronics Systems (ICAMechS)* (pp. 418-423).

[25] El-Sehiemy, R. A., Selim, F., Bentouati, B., & Abido, M. A. (2020). A novel multi-objective hybrid particle swarm and salp optimization algorithm for technical-economical-environmental operation in power systems. *Energy, 193*, 116817. https://doi.org/10.1016/j.energy.2019.116817

[26] Meng, A., et al. (2021). A high-performance crisscross search-based grey wolf optimizer for solving optimal power flow problem. *Energy, 225*, 120211. https://doi.org/10.1016/j.energy.2021.120211

[27] Abbasi, M., Abbasi, E., & Mohammadi-Ivatloo, B. (2020). Single and multi-objective optimal power flow using a new differential-based harmony search algorithm. *Journal of Ambient Intelligence and Humanized Computing, 1*, 3. https://doi.org/10.1007/s12652-020-02089-6

[28] Rao, N. T., Sankar, M. M., Rao, S. P., & Rao, B. S. (2020). Comparative study of Pareto optimal multi-objective cuckoo search algorithm and multi-objective particle swarm optimization for power loss minimization incorporating UPFC. *Journal of Ambient Intelligence and Humanized Computing, 12*, 1069-1080. https://doi.org/10.1007/s12652-020-02142-4

[29] El-Sattar, S. A., Kamel, S., El-Sehiemy, R. A., Jurado, F., & Yu, J. (2019). Single- and multi-objective optimal power flow frameworks using Jaya optimization technique. *Neural Computing and Applications, 31(12)*, 8787-8806. https://doi.org/10.1007/s00521-019-04194-w

[30] Warid, W. (2020). Optimal power flow using the AMTPG-Jaya algorithm. *Applied Soft Computing Journal, 91*, 106252. https://doi.org/10.1016/j.asoc.2020.106252

[31] Bentouati, B., Khelifi, A., Shaheen, A. M., & El-Sehiemy, R. A. (2020). An enhanced moth swarm algorithm for efficient energy management-based multi-dimensions OPF problem. *Journal of Ambient Intelligence and Humanized Computing.* https://doi.org/10.1007/s12652-020-02692-7

[32] Niknam, T., Narimani, M. R., Jabbari, M., & Malekpour, A. R. (2011). A modified shuffle frog leaping algorithm for multi-objective optimal power flow. *Energy, 36*(11), 6420-6432. https://doi.org/10.1016/j.energy.2011.09.027

Authors' contacts:

Vijoy Kumar Peddiny
(Corresponding Author)
National Institute of Technology,
Jote-791113, Arunachal Pradesh, India
vijoy49@gmail.com

Brajagopal Datta
National Institute of Technology,
Jote-791113, Arunachal Pradesh, India
Brajagopal@nitap.ac.in

Abhik Banerjee
National Institute of Technology,
Jote-791113, Arunachal Pradesh, India

ISSN 1846-6168 (Print), ISSN 1848-5588 (Online)
https://doi.org/10.31803/tg-20240910214215

Original scientific paper
Received: 2024-09-10, Accepted: 2025-02-21

Detection and Classification of Growth Stages in Rice Using Artificial Neural Networks

Durodola Folasade*, Owoeye Samuel, Kamil-Bello Furqan, Daodu Sakira, Makinde Kayode, Folaranmi Olaniyi

Abstract: This project focused on the development of a machine learning model to classify rice plants based on their growth stages, specifically identifying whether the plants are in the "Raw" or "Ripe" stage. The research was conducted using a dataset obtained from Roboflow, which provided annotated images of rice plants. The dataset was divided into training, testing, and validation subsets to ensure the model's robustness and generalization capability. The project involved a comprehensive data preparation process, which included consolidating images into class-based folders, handling file conflicts, and stratifying the dataset into appropriate splits for training, testing, and validation. Several convolutional neural networks (CNN) architectures were explored, including ResNet50, InceptionV3, and MobileNetV2, each leveraging transfer learning from pre-trained models on the ImageNet dataset. ResNet50 achieved an accuracy of 87.3% with a log loss of 0.33, demonstrating good performance but with some misclassifications between similar classes. InceptionV3 outperformed the other models, achieving an accuracy of 95.1% and a log loss of 0.13, indicating superior classification capability and better calibration of predicted probabilities. MobileNetV2 also performed well with an accuracy of 93.5% and a log loss of 0.22, offering a balance between accuracy and computational efficiency. The results highlight InceptionV3 as the most effective model for this task, with a strong ability to differentiate between the rice growth stages. The findings underscore the importance of model selection and data preparation in developing accurate and reliable machine-learning models for agricultural applications. The project demonstrates the potential of CNNs in improving agricultural practices through precise crop monitoring and classification.

Keywords: convolutional neural networks (CNNs); image classification; InceptionV3; machine learning; MobileNetV2; ResNet50; rice plant

1 INTRODUCTION

Rice is a crucial food source worldwide, particularly in regions such as Asia, Africa, and Latin America, serving as a primary dietary staple for billions of people [1]. With the growing global population and changing climatic conditions, there is increasing pressure on rice production systems to enhance production [2]. Traditional methods of monitoring rice growth stages are labour-intensive, time-consuming, and prone to human error, making them inadequate for meeting the rising demands and ensuring food security [3].

Artificial Intelligence (AI), specifically Artificial Neural Networks (ANNs), has shown significant potential in transforming agriculture. ANNs can process large datasets and detect complex patterns, making them ideal for precision agriculture applications [4]. By employing ANNs, we can develop automated systems capable of accurately detecting and monitoring the growth stages of rice, leading to improved crop management, optimized resource use, and increased yields [5, 6]. This innovation contributes to global food security and promotes sustainable agricultural practices.

In addition to boosting productivity, applying ANNs in rice growth stage detection can significantly reduce the environmental impact of rice farming. Precise identification of growth stages allows for the optimized use of water, fertilizers, and pesticides, reducing wastage and environmental degradation [7]. This precision also minimizes the runoff of harmful chemicals into water bodies, protecting aquatic ecosystems and promoting eco-friendly agricultural practices.

Furthermore, ANNs offer valuable tools for adapting to climate variability and change. As climate change presents unpredictable challenges to agriculture, real-time monitoring of rice growth stages becomes essential. ANNs can help predict and mitigate the effects of adverse weather conditions by providing timely insights, enabling farmers to take proactive measures [8]. This adaptability is crucial for maintaining rice productivity and ensuring food security in the face of changing climatic patterns.

The economic benefits of deploying ANN-based systems in rice agriculture are also noteworthy. Automation of growth stage monitoring can lead to significant cost savings by reducing the need for extensive manual labour and enabling more precise agricultural interventions. For smallholder farmers, who play a significant role in global rice production, these savings can be crucial, allowing them to reinvest in their operations and improve their livelihoods [9].

Despite these advantages, traditional methods of monitoring rice growth stages remain laborious, time-consuming, and prone to errors. These methods do not provide real-time data, which is essential for making timely agricultural interventions. Inaccurate and inconsistent monitoring can lead to suboptimal resource use, affecting crop health and yield [5]. There is a pressing need for an automated, reliable, and efficient system that can accurately detect and monitor rice growth stages, handle variability in growth patterns, and provide farmers with accurate, real-time information to enhance decision-making and improve overall crop management practices.

The integration of ANNs into rice farming offers a comprehensive solution to agronomic, environmental, and economic challenges in rice production. This study aims to utilize artificial neural networks (ANNs) for the prediction of rice growth stages through the examination of phenological data and image-based inputs, specifically targeting the vegetative, reproductive, and ripening phases. The ultimate goal is to equip farmers with practical insights that facilitate improved crop management, optimized resource use, and enhanced productivity.

2 LITERATURE REVIEW

The use of technological advancements and methodologies in utilizing artificial neural networks (ANNs) for detecting rice growth stages can not be over-emphasised.

It highlights the shift from traditional monitoring methods to modern precision agriculture techniques, emphasizing the role of ANNs in enhancing agricultural practices [10].

Precision agriculture represents a significant advancement in farming, using technologies like GPS, remote sensing, and data analytics to monitor and manage crop growth with high precision. These tools enable targeted interventions, such as precise irrigation and fertilization, reducing waste, and promoting sustainable farming practices. Machine learning and AI further enhance these capabilities, offering predictive models and decision support systems for efficient farming [11-13].

ANNs have become a powerful tool in agriculture, particularly for analyzing large datasets and identifying complex patterns. In rice farming, ANNs can accurately detect and classify growth stages by processing vast amounts of data from various sources. This capability is crucial for developing automated systems that provide real-time information, reducing reliance on manual labour and increasing productivity [14, 15].

Convolutional Neural Networks (CNNs), a subset of ANNs, are particularly effective for image recognition tasks, making them ideal for analyzing images of rice at different growth stages. CNNs automatically detect and learn features from images, enabling precise classification and monitoring of rice growth. Training CNNs with diverse image datasets ensures the model's robustness and reliability in real-world applications [16, 17].

The success of an ANN-based system depends on the quality and diversity of the training data. Effective data collection involves gathering images and sensor data from multiple sources, ensuring a comprehensive view of the rice fields. Preprocessing techniques, such as data cleaning and augmentation, prepare the data for training, improving the model's robustness and accuracy [18, 19].

Training an ANN model involves feeding preprocessed data into the network and adjusting model parameters to minimize prediction errors. Selecting an appropriate neural network architecture is crucial for capturing the complex features in rice crop images. Advanced machine learning techniques are used to enhance performance, and the model's accuracy is validated using separate data to ensure reliability in real-world applications [20, 21].

3 METHODOLOGY

Accurately identifying rice growth stages is essential for effective agricultural management, as it directly impacts crop yield, resource utilization, and farming efficiency [6]. Traditionally, determining these stages has relied on manual observation and expert judgment, which can lead to variability and inefficiencies. However, advancements in artificial intelligence and machine learning present an opportunity for a more precise, automated approach to growth stage detection [21]. This chapter outlines the methodology for designing, developing, and implementing an artificial neural network (ANN)-based system to detect rice growth stages. The system aims to provide farmers with a reliable tool to automate the identification process, thereby enabling timely and informed decision-making. By integrating this system into a user-friendly platform, farmers can upload images of their rice crops and receive instant feedback on the current growth stage, optimizing their farming practices and improving productivity.

The proposed rice growth stage detection system uses artificial neural networks to analyze and classify images of rice crops, determining the specific growth stage based on visual features [14]. A convolutional neural network (CNN), tailored for image recognition tasks, serves as the system's core component [22]. The CNN is trained on a diverse dataset of rice crop images at various growth stages, captured under different environmental conditions [23]. The system processes these images to extract crucial features like texture, colour, and shape, which are essential for distinguishing between different growth [24]. The system is integrated into a web-based platform that allows farmers to upload images of their crops. Once uploaded, the CNN processes the image and provides immediate feedback on the rice's current growth stage, along with suggestions for appropriate farming practices based on the identified stage. This real-time, automated analysis significantly reduces the need for manual monitoring and expert intervention [25]. The machine learning workflow is shown in Fig. 1.

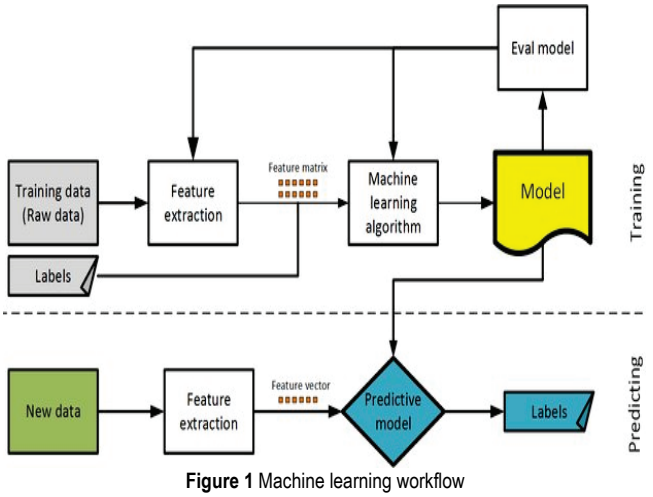


Figure 1 Machine learning workflow

The system's effectiveness hinges on the quality of the dataset used to train the machine-learning models [16]. A comprehensive dataset of rice crop images at various growth stages, captured under diverse environmental conditions, was compiled from multiple sources. To enhance the model's robustness, extensive data augmentation techniques, including random rotation, scaling, and colour adjustments, were employed [26]. Three prominent CNN architectures—ResNet, MobileNet, and InceptionNet—were selected for comparison, each offering distinct advantages in accuracy, computational efficiency, and scalability [27-29]. The training process involved hyperparameter tuning, transfer learning, and regularization techniques to ensure optimal performance. The models were continuously evaluated using metrics such as accuracy, precision, recall, and F1-score to select the best-performing architecture for deployment.

Data collection involved gathering a diverse set of rice images at different growth stages and under varying environmental conditions. The images were sourced from

publicly available agricultural datasets, custom-captured images from rice fields, and contributions from agricultural research institutions. This diverse dataset enhances the model's ability to generalize and perform well in real-world applications [30]. Preprocessing steps included image resizing, noise reduction, and data augmentation to ensure compatibility with the neural network architectures [31] [32]. Exploratory Data Analysis (EDA) was conducted to understand the dataset's underlying patterns, relationships, and potential issues [33]. EDA helped identify any imbalances in the dataset and informed decisions on model selection and training strategies.

The dataset was split into training, validation, and test sets, with 70% of the data used for training, 20% for validation, and 10% reserved for testing. Normalization was applied to standardize the input values, ensuring that all input features contributed equally to the model's learning process [34]. The test site for system deployment was carefully selected to represent typical rice-growing environments, considering factors such as rice variety, environmental conditions, and field layout [35-37]. The system, including cameras, sensors, and processing units, was installed and configured according to the experimental requirements, with the selected machine-learning models evaluated on a validation set to ensure their accuracy and reliability in classifying rice growth stages.

The system's performance was evaluated through controlled experiments focusing on key metrics such as classification accuracy, model comparison, environmental adaptability, and operational efficiency. The performance of ResNet, MobileNet, and InceptionNet models was compared to determine which provided the best balance of accuracy and computational efficiency. The system's adaptability to varying environmental conditions, including different lighting and weather patterns, was assessed, along with its operational efficiency, including power consumption and maintenance requirements [23]. User feedback was collected from farmers and agricultural experts to evaluate the system's usability, reliability, and effectiveness in supporting rice growth stage monitoring. The experimental trials were conducted over an entire growing season, with continuous monitoring and data logging to capture the system's performance across different growth stages and environmental conditions.

The data collected during the experimental trials was analyzed to identify areas for improvement and optimize the system's performance. The analysis included evaluating the effectiveness of each machine learning model, the accuracy of growth stage classification, and the overall system's reliability and adaptability. The data preprocessing involved cleaning, organizing, and formatting the raw data to facilitate thorough analysis. Trends and correlations between growth stages, environmental conditions, and system performance metrics were identified, leading to model refinement and operational modifications to enhance system performance. The continuous refinement and optimization of the rice growth stage detection system aimed to ensure its effectiveness and reliability in real-world agricultural settings, ultimately providing farmers with a valuable tool to improve crop management practices.

4 IMPLEMENTATION AND RESULT

The implementation of this research involves developing a machine-learning model by following a systematic workflow, beginning with data collection and preparation. A dataset from Roboflow, containing images of rice plants labeled as "Raw" or "Ripe," was used. This data was split into three subsets—training, testing, and validation—to ensure the model's generalization. The dataset preparation, shown in Fig. 2, included consolidating all images into a unified directory structure, separating them into "Raw" and "Ripe" categories for easier processing during the subsequent stages of model training.

```
import os
import shutil


def combine_images_to_class_folders(dataset_path, output_folder):
    """
    Combine all images from 'train', 'test', and 'valid' folders into 'Raw' and 'Ripe' subfolders
    inside a single output folder.

    :param dataset_path: The root path of the dataset containing 'train', 'test', and 'valid' folders.
    :param output_folder: The path to the folder where all images will be combined into 'Raw' and 'Ripe' subfolders.
    """
    # Ensure the output folder exists
    if not os.path.exists(output_folder):
        os.makedirs(output_folder)
```

Figure 2 Data preparation

```
import os
import shutil


def combine_images_to_class_folders(dataset_path, output_folder):
    """
    Combine all images from 'train', 'test', and 'valid' folders into 'Raw' and 'Ripe' subfolders
    inside a single output folder.

    :param dataset_path: The root path of the dataset containing 'train', 'test', and 'valid' folders.
    :param output_folder: The path to the folder where all images will be combined into 'Raw' and 'Ripe' subfolders.
    """
    # Ensure the output folder exists
    if not os.path.exists(output_folder):
        os.makedirs(output_folder)
```

Figure 3 Random sampling

Following data consolidation, the workflow handled potential file conflicts, such as duplicate file names, by appending numerical suffixes to ensure unique filenames. This process helped maintain data integrity, preventing data loss due to overwriting. Additionally, random file sampling, shown in Fig. 3 was implemented to create smaller, balanced subsets for testing or validation. Fig. 4 shows the stratified data splitting, which ensured that each class had equal representation across the training, testing, and validation sets, which is critical for building a robust model capable of generalizing well on unseen data.

The research implemented multiple deep learning models, including ResNet50, InceptionNet, and MobileNetV2, leveraging pre-trained architectures.

ResNet50 utilized transfer learning, retaining some layers for feature extraction while fine-tuning others to adapt to the new dataset. A Global Average Pooling (GAP) layer was appended, followed by Dense layers with Dropout to avoid overfitting. Similarly, InceptionNet, based on InceptionV3, captured multi-scale features through its inception modules. The model retained valuable feature extraction capabilities while adapting to the new classification task, also incorporating GAP and Dropout layers to prevent overfitting.

```
mport os
mport shutil
rom sklearn.model_selection import train_test_split

ef stratified_split_dataset(dataset_path, output_path, train_ratio=0.6, test_ratio=0.3, val_ratio=0.1):
    """
    Split the dataset into train, test, and validation sets with stratified sampling.

    :param dataset_path: Path to the dataset containing class subfolders.
    :param output_path: Path to the output folder where the split dataset will be stored.
    :param train_ratio: Ratio of the dataset to be used for training (default is 0.6).
    :param test_ratio: Ratio of the dataset to be used for testing (default is 0.3).
    :param val_ratio: Ratio of the dataset to be used for validation (default is 0.1).
    """
    # Ensure the output folders exist
    for split in ['train', 'test', 'val']:
        split_path = os.path.join(output_path, split)
        if not os.path.exists(split_path):
            os.makedirs(split_path)

    # Get class subfolders
    classes = [d for d in os.listdir(dataset_path) if os.path.isdir(os.path.join(dataset_path, d))]

    for cls in classes:
        class_path = os.path.join(dataset_path, cls)
        images = [f for f in os.listdir(class_path) if os.path.isfile(os.path.join(class_path, f))]
```

Figure 4 Data splitting

```
[ ] def resnet_model(no_layers = 5):
        resnet_model = ResNet50(include_top=False, weights='imagenet')
        for layer in resnet_model.layers[:-no_layers]:
            layer.trainable = False
        resnet_x = resnet_model.output
        x = GlobalAveragePooling2D()(resnet_x)
        #  add a fully-connected dense llayer as head
        x=Dropout(0.2)(x)
        x = Dense(512, activation='relu')(x)
        x=Dropout(0.2)(x)
        x = Dense(64, activation='relu')(x)
        x=Dropout(0.2)(x)
        # and a logistic layer
        predictions = Dense(3, activation='softmax')(x)
        model = Model(inputs=resnet_model.input, outputs=predictions, name="resnet")
        return model
```

```
[ ] resnet_model = resnet_model()


    resnet_model.compile(optimizer='adam', loss='categorical_crossentropy',metrics=['acc'])
    callback = EarlyStopping(monitor='val_acc', patience=5, restore_best_weights=True)
    history = resnet_model.fit(X_train, y_train, validation_data=(X_val,y_val), batch_size=64, epochs=20,callbacks=[callback])
```

Figure 5 RestNet implementation

A comparative analysis of the performance of three models—ResNet50, InceptionV3, and MobileNetV2—on a rice growth dataset, evaluating their strengths, weaknesses, and overall effectiveness. As shown in Figs. 8 and 9 the ResNet50 model achieved an accuracy of 87.3% on the test set, with an F1 score of 86.9%, indicating a good balance between precision and recall. However, the model struggled with differentiating between certain classes, such as misclassifying instances of 'Vegetative' as 'Reproductive' and confusing 'Ripening' with 'Vegetative.' The training curves showed a smooth increase in accuracy, plateauing at 90–92%, which suggests some level of overfitting despite effective learning and generalization.

```
[ ] def inception_model(no_layers = 5):
        inception_model=InceptionV3(include_top=False, weights='imagenet') # importation of model with pretrained weights and dropping the head
        for layer in inception_model.layers[:-no_layers]:
            layer.trainable = False # freezing the layers
        inception_x = inception_model.output
        x = GlobalAveragePooling2D()(inception_x)
        #  add a fully-connected dense llayer as head
        x=Dropout(0.2)(x)
        x = Dense(512, activation='relu')(x)
        x=Dropout(0.2)(x)
        x = Dense(64, activation='relu')(x)
        x=Dropout(0.2)(x)
        # and a logistic layer
        predictions = Dense(3, activation='softmax')(x)
        model = Model(inputs=inception_model.input, outputs=predictions, name="inception")
        return model
```

```
[ ] inception_model = inception_model()
```

Figure 6 Inception net implementation

```
[ ] def mobilenet_model(no_layers=5):
        # Load the MobileNetV2 model with pretrained weights, without the top layers
        mobilenet_model = MobileNetV2(include_top=False, weights='imagenet')

        # Freeze all layers except the last `no_layers`
        for layer in mobilenet_model.layers[:-no_layers]:
            layer.trainable = False  # Freezing the layers

        # Adding custom layers on top of the MobileNetV2 model
        mobilenet_x = mobilenet_model.output
        x = GlobalAveragePooling2D()(mobilenet_x)

        # Add fully-connected dense layers as head
        x = Dropout(0.2)(x)
        x = Dense(512, activation='relu')(x)
        x = Dropout(0.2)(x)
        x = Dense(64, activation='relu')(x)
        x = Dropout(0.2)(x)

        # Add a final softmax layer for classification (3 classes)
        predictions = Dense(3, activation='softmax')(x)

        # Define the complete model
        model = Model(inputs=mobilenet_model.input, outputs=predictions, name="mobilenet")

        return model
```

Figure 7 Mobile net implementation

```
Accuracy score: 0.8727544910179641
F1 score: 0.8697522597704368
Log loss score: 0.33019455157768657
Precision score: 0.8794540663153969
Recall score: 0.8727544910179641
```

Figure 8 ResNet results breakdown

Fig. 10 shows the confusion matrix revealed minimal misclassifications, demonstrating InceptionV3's superiority in-class differentiation.

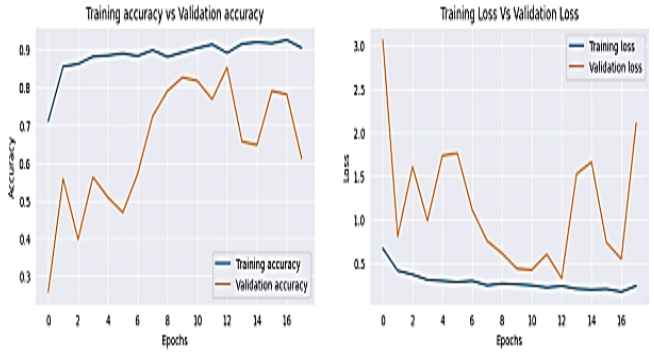


Figure 9 Accuracy and loss plot for the ResNet50

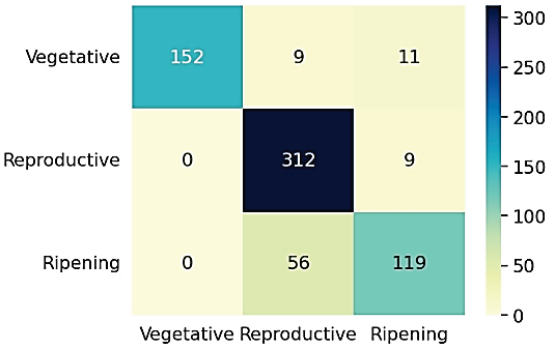


Figure 10 Confusion matrix

Figs. 11 and 12 show that the InceptionV3 model outperformed ResNet50 significantly, achieving a test accuracy of 95.1% and an F1 score of 94.8%. Its lower log loss of 0.13 indicated well-calibrated predicted probabilities, making it a highly reliable model for this classification task. Fig. 10 showed the confusion matrix, which revealed minimal misclassifications, demonstrating InceptionV3's superiority in-class differentiation. The training curves for InceptionV3 showed rapid improvement in accuracy, which stabilized at a high level, with the loss curve reflecting excellent fitting and generalization.

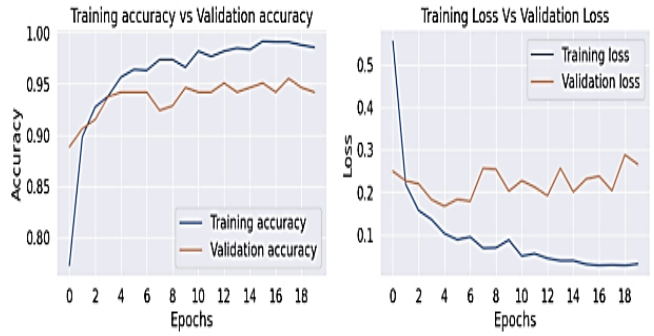


Figure 11 Accuracy and loss plot for the InceptionV3

```
Accuracy score: 0.9670658682634731
F1 score: 0.9669153790378413
Log loss score: 0.10637995986050977
Precision score: 0.9669677661612094
Recall score: 0.9670658682634731
```

Figure 12 Results break down for inception net

The MobileNetV2 model (Figs. 13 and 14) also performed well, with an accuracy of 93.5% and an F1 score of 93.1%, positioning it between InceptionV3 and ResNet50 in terms of performance. The log loss for MobileNetV2 was 0.22, better than ResNet50 but not as good as InceptionV3, indicating some room for improvement in probability calibration. While the confusion matrix highlighted some misclassifications, particularly in the 'Ripening' class, MobileNetV2 showed better stability in model performance compared to ResNet50, with steady decreases in loss and strong generalization capabilities.

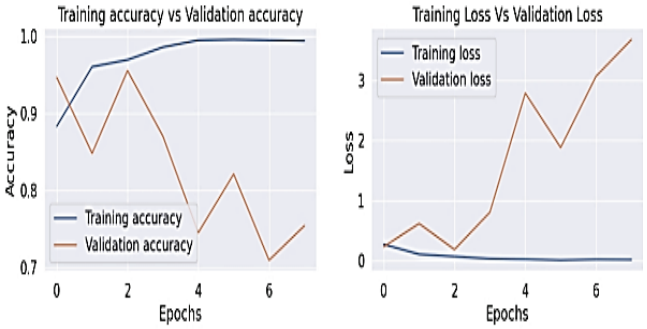


Figure 13 MobileNet accuracy and loss

```
Accuracy score: 0.9341317365269461
F1 score: 0.9322727281344163
Log loss score: 0.24570493742749838
Precision score: 0.9364948262052012
Recall score: 0.9341317365269461
```

Figure 14 MobileNet results

In comparing ResNet50, InceptionV3, and MobileNetV2, InceptionV3 emerges as the most effective model, achieving the highest accuracy (95.1%), F1 score (94.8%), and the lowest log loss (0.13), indicating superior probability calibration and reliable classification with minimal misclassifications. ResNet50 struggles with class differentiation and shows signs of overfitting, making it less reliable overall. Thus, InceptionV3 is the best model, with MobileNetV2 as a strong alternative.

MobileNetV2 was chosen for its efficiency on mobile and embedded devices, employing depthwise separable convolutions to reduce computational demands. The model was pre-trained on ImageNet, and adjustments were made to the final layers to suit the specific task of classifying rice growth stages. The model was compiled with a loss function, optimizer, and evaluation metrics tailored for multi-class classification. The training involved refining the model's predictions over multiple epochs, with validation against a separate dataset to ensure generalization and performance.

In summary, the workflow—from data preparation to model implementation—was designed to maintain data integrity, ensure balanced datasets, and build efficient models for classifying rice growth stages. Each model—ResNet50, InceptionNet, and MobileNetV2—was tailored for the task, utilizing pre-trained architectures and advanced techniques like transfer learning, GAP, and Dropout to achieve accurate and generalizable results.

5 CONCLUSION

This study analyzed and evaluated the performance of three advanced deep learning models—ResNet50, InceptionV3, and MobileNetV2—on rice plant classification. Among these, InceptionV3 proved to be the most effective, achieving the highest accuracy (95.1%) and F1 score (94.8%), along with excellent probability calibration. Its low misclassification rates make it the best model for differentiating between rice plant classes.

MobileNetV2, while slightly less powerful than InceptionV3, still performed well with an accuracy of 93.5% and an F1 score of 93.1%. It strikes a good balance between performance and efficiency, making it suitable for deployment in resource-constrained environments. Despite its higher log loss of 0.22 compared to InceptionV3, MobileNetV2 remains a reliable model.

ResNet50, although effective, lagged behind the other models with an accuracy of 87.3% and an F1 score of 86.9%. Its higher log loss of 0.33 and difficulties in class differentiation indicate that it requires significant refinement to compete with InceptionV3 and MobileNetV2.

Based on these findings, InceptionV3 is recommended as the preferred model for rice plant classification due to its superior accuracy, high F1 score, and well-calibrated probabilities, making it ideal for reliable and precise classification tasks.

To further enhance model performance, increasing the size and diversity of the dataset used for training and testing is essential. A more comprehensive dataset will help improve the models' ability to generalize, reduce misclassifications, and provide more accurate and reliable classification results.

Additionally, techniques such as transfer learning, fine-tuning of pre-trained models, and the development of hybrid models that combine the strengths of different architectures are recommended. Continuous improvement efforts, including advanced data augmentation and robust preprocessing techniques, will further advance the field of rice plant classification, enabling more effective solutions in precision agriculture. In conclusion, the research effectively highlights the transformative potential of AI in agriculture, InceptionV3 is hereby recommended as the most effective model for rice growth stage classification

6 REFERENCES

[1] Muthayya, S., Sugimoto, J. D., Montgomery, S. & Maberly, G. F. (2014). An overview of global rice production, supply, trade, and consumption. *Ann. N. Y. Acad. Sci.*, 1324, 7-14. https://doi.org/10.1111/nyas.12540

[2] Wassermann, R., Jagadish, S. V. K., Heuer, S., Ismail, A., Redona, E., Serraj, R., ... & Sumfleth, K. (2009). Climate change affecting rice production: the physiological and agronomic basis for possible adaptation strategies. *Advances in agronomy, 101*, 59-122. https://doi.org/10.1016/S0065-2113(08)00802-X

[3] Wilfred, O. R. & Consultant, L. (2006). Final survey report on the status of rice production, processing and marketing in Uganda. A report submitted to the Embassy of Japan in Uganda through JICA and Sasakawa Africa Association-Uganda, Kampala.

[4] Sharma, A., Jain, A., Gupta, P. & Chowdary, V. (2020). Machine learning applications for precision agriculture: A comprehensive review. *IEEE Access, 9*, 4843-4873. https://doi.org/10.1109/ACCESS.2020.3048415

[5] Liu, H., Wang, H., Zhou, X. & Li, Z. (2019). Crop growth stage detection using deep learning techniques for smart agriculture. *Sensors, 19*(7), 2008.

[6] Zhang, K., Ge, X., Shen, P., Li, W., Liu, X., Cao, Q., ... & Tian, Y. (2019). Predicting rice grain yield based on dynamic changes in vegetation indexes during early to mid-growth stages. *Remote sensing, 11*(4), 387. https://doi.org/10.3390/rs11040387

[7] Nabavi-Pelesaraei, A., Rafiee, S., Mohtasebi, S. S., Hosseinzadeh-Bandbafha, H. & Chau, K. W. (2018). Integration of artificial intelligence methods and life cycle assessment to predict energy output and environmental impacts of paddy production. *Science of the total environment*, 631, 1279-1294. https://doi.org/10.1016/j.scitotenv.2018.03.088

[8] Kourgialas, N. N., Dokou, Z. & Karatzas, G. P. (2015). Statistical analysis and ANN modelling for predicting hydrological extremes under climate change scenarios: The example of a small Mediterranean agro-watershed. *Journal of Environmental Management, 154*, 86-101. https://doi.org/10.1016/j.jenvman.2015.02.034

[9] Fan, S. & Rue, C. (2020). The role of smallholder farms in a changing world. *The role of smallholder farms in food and nutrition security*, 13-28. https://doi.org/10.1007/978-3-030-42148-9_2

[10] Adewusi, A. O., Asuzu, O. F., Olorunsogo, T., Iwuanyanwu, C., Adaga, E. & Daraojimba, D. O. (2024). AI in precision agriculture: A review of technologies for sustainable farming practices. *World Journal of Advanced Research and Reviews, 21*(1), 2276-2285. https://doi.org/10.30574/wjarr.2024.21.1.0314

[11] Paul, K., Chatterjee, S. S., Pai, P., Varshney, A., Juikar, S., Prasad, V., ... & Dasgupta, S. (2022). Viable smart sensors and their application in data-driven agriculture. *Computers and Electronics in Agriculture, 198*, 107096. https://doi.org/10.1016/j.compag.2022.107096

[12] Schimmelpfennig, D. (2016). Farm profits and adoption of precision agriculture. *USDA Economic Research Report*.

[13] Mulla, D. J. (2013). Twenty five years of remote sensing in precision agriculture: Key advances and remaining knowledge gaps. Biosystems Engineering, 114(4), 358-371. https://doi.org/10.1016/j.biosystemseng.2012.08.009

[14] Alfred, R., Obit, J. H., Chin, C. P. Y., Haviluddin, H. & Lim, Y. (2021). Towards paddy rice smart farming: a review on big data, machine learning, and rice production tasks. *IEEE Access, 9*, 50358-50380. https://doi.org/10.1109/ACCESS.2021.3069449

[15] Javaid, M., Haleem, A., Singh, R. P. & Suman, R. (2022). Enhancing smart farming through the applications of Agriculture 4.0 technologies. *International Journal of Intelligent Networks, 3*, 150-164. https://doi.org/10.1016/j.ijin.2022.09.004

[16] Tan, S., Liu, J., Lu, H., Lan, M., Yu, J., Liao, G., ... & Ma, X. (2022). Machine learning approaches for rice seedling growth stages detection. *Frontiers in Plant Science, 13*, 914771. https://doi.org/10.3389/fpls.2022.914771

[17] Saha, H. N., Roy, R., Chakraborty, M. & Sarkar, C. (2021). IoT-enabled agricultural system application, challenges and security issues. *Agricultural informatics: automation using the IoT and machine learning*, 223-247. https://doi.org/10.1002/9781119769231.ch11

[18] Albahri, A. S., Khaleel, Y. L., Habeeb, M. A., Ismael, R. D., Hameed, Q. A., Deveci, M., ... & Alzubaidi, L. (2024). A systematic review of trustworthy artificial intelligence applications in natural disasters. *Computers and Electrical Engineering, 118*, 109409. https://doi.org/10.1016/j.compeleceng.2024.109409

[19] Zameer, A., Arshad, J., Khan, A. & Raja, M. A. Z. (2017). Intelligent and robust prediction of short-term wind power using a genetic programming-based ensemble of neural networks. *Energy conversion and management, 134*, 361-372. https://doi.org/10.1016/j.enconman.2016.12.032

[20] Patrício, D. I. & Rieder, R. (2018). Computer vision and artificial intelligence in precision agriculture for grain crops: A systematic review. *Computers and electronics in agriculture, 153*, 69-81. https://doi.org/10.1016/j.compag.2018.08.001

[21] Shaikh, T. A., Rasool, T. & Lone, F. R. (2022). Towards leveraging the role of machine learning and artificial intelligence in precision agriculture and smart farming. *Computers and Electronics in Agriculture, 198*, 107119. https://doi.org/10.1016/j.compag.2022.107119

[22] Li, Y. & Chao, X. (2020). ANN-based continual classification in agriculture. *Agriculture, 10*(5), 178. https://doi.org/10.3390/agriculture10050178

[23] Anami, B. S., Malvade, N. N. & Palaiah, S. (2020). Deep learning approach for recognition and classification of yield affecting paddy crop stresses using field images. *Artificial intelligence in agriculture, 4*, 12-20. https://doi.org/10.1016/j.aiia.2020.03.001

[24] Iqbal, Z., Khan, M. A., Sharif, M., Shah, J. H., Ur Rehman, M. H. & Javed, K. (2018). Automated detection and classification of citrus plant diseases using image processing techniques: A review. *Computers and electronics in agriculture, 153*, 12-32. https://doi.org/10.1016/j.compag.2018.07.032

[25] Subeesh, A. & Mehta, C. R. (2021). Automation and digitization of agriculture using artificial intelligence and the Internet of Things. *Artificial Intelligence in Agriculture, 5*, 278-291. https://doi.org/10.1016/j.aiia.2021.11.004

[26] Mumuni, A. & Mumuni, F. (2022). Data augmentation: A comprehensive survey of modern approaches. *Array, 16*, 100258. https://doi.org/10.1016/j.array.2022.100258

[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*, 770-778. https://doi.org/10.1109/CVPR.2016.90

[28] Howard, A. G. (2017). MobileNets: Efficient convolutional neural networks for mobile vision applications. arXiv preprint *arXiv*: 1704.04861.

[29] Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J. & Wojna, Z. (2016). Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2818-2826. https://doi.org/10.1109/CVPR.2016.308

[30] Oza, N. C. & Tumer, K. (2008). Classifier ensembles: Select real-world applications. *Information fusion, 9*(1), 4-20. https://doi.org/10.1016/j.inffus.2007.07.002

[31] Tabik, S., Peralta, D., Herrera-Poyatos, A. & Herrera, F. (2017). A snapshot of image pre-processing for convolutional neural networks: a case study of MNIST. *International Journal of Computational Intelligence Systems, 10*(1), 555-568. https://doi.org/10.2991/ijcis.2017.10.1.38

[32] Mumtaz, W., Rasheed, S. & Irfan, A. (2021). Review of challenges associated with the EEG artefact removal methods. *Biomedical Signal Processing and Control, 68*, 102741. https://doi.org/10.1016/j.bspc.2021.102741

[33] Milo, T. & Somech, A. (2020, June). Automating exploratory data analysis via machine learning: An overview. In *Proceedings of the 2020 ACM SIGMOD international conference on management of data*, 2617-2622. https://doi.org/10.1145/3318464.3383126

[34] Burger, W. & Burge, M. J. (2013). *Principles of digital image processing: advanced methods*. Springer Science & Business Media. https://doi.org/10.1007/978-1-84882-919-0

[35] Nelson, A., Setiyono, T., Rala, A. B., Quicho, E. D., Raviz, J. V., Abonete, P. J., ... & Ninh, N. H. (2014). Towards an operational SAR-based rice monitoring system in Asia: Examples from 13 demonstration sites across Asia in the RIICE project. *Remote Sensing, 6*(11), 10773-10812. https://doi.org/10.3390/rs61110773

[36] Ferentinos, K. P., Katsoulas, N., Tzounis, A., Bartzanas, T. & Kittas, C. (2017). Wireless sensor networks for greenhouse climate and plant condition assessment. *Biosystems Engineering, 153*, 70-81. https://doi.org/10.1016/j.biosystemseng.2016.11.005

[37] Wang, R., Cheng, T. & Hu, L. (2015). Effect of wide–narrow row arrangement and plant density on yield and radiation use efficiency of mechanized direct-seeded canola in Central China. *Field Crops Research, 172*, 42-52. https://doi.org/10.1016/j.fcr.2014.12.005

Author's contacts:

Durodola Folasade
(Corresponding Author)
Department of Mechatronics Engineering,
Federal University of Agriculture,
Alabata Road, Abeokuta, Ogun State, 110111, Nigeria
durodolafo@funaab.edu.ng

Owoeye Samuel
Department of Mechatronics Engineering,
Federal University of Agriculture,
Alabata Road, Abeokuta, Ogun State, 110111, Nigeria

Kamil- Bello Furqan
Department of Mechatronics Engineering,
Federal University of Agriculture,
Alabata Road, Abeokuta, Ogun State, 110111, Nigeria

Daodu Sakira
Department of Mechatronics Engineering,
Federal University of Agriculture,
Alabata Road, Abeokuta, Ogun State, 110111, Nigeria

Makinde Kayode
Department of Mechatronics Engineering,
Federal University of Agriculture,
Alabata Road, Abeokuta, Ogun State, 110111, Nigeria

Folarami Olaniyi
Department of Mechatronics Engineering,
Federal University of Agriculture,
Alabata Road, Abeokuta, Ogun State, 110111, Nigeria

ISSN 1846-6168 (Print), ISSN 1848-5588 (Online)
https://doi.org/10.31803/tg-20250108214321

Original scientific paper
Received: 2025-01-08, Accepted: 2025-03-12

Georectification of the Historical Geological Map of Bosnia and Herzegovina: Effectiveness Evaluation of Different Approaches

Nedim Tuno

Abstract: The Geological Overview Map of Bosnia and Herzegovina, at a scale of 1:200,000, is the first comprehensive and detailed geological map meticulously created for the entire country. This paper examines the mathematical and geodetic basis of the map and evaluates its positional accuracy. A method was developed to compute coordinate corrections for arbitrary points based on the differences between the Austro-Hungarian and Yugoslav triangulations, in order to determine the position of any point on the old geological map with respect to the current Bosnian-Herzegovinian state coordinate system. After resolving the issues related to map projection, mathematical and geodetic foundations of the map and coordinate systems, the georectification procedure was applied to the map scan. The results show that the TPS transformation effectively minimizes geometric distortions caused by scanning errors, deformations of the map's base material, and systematic errors from the original map elaboration. After applying the georectification, most remaining positional distortions were attributed to random errors.

Keywords: accuracy assessment; geological map; georectification; transformation; triangulation

1 INTRODUCTION

Maps are reliable records of the space and time in which they were created, making them a valuable source of data for various spatial analyses. In this context, historical cartographic materials—such as old military maps, cadastral maps, topographic maps, thematic maps, and other specialized maps—are essential for retrospective analysis of past landscapes [1]. For this reason, maps can be seen as a reflection of past eras and serve as an important part of global cultural heritage [2]. A geological map is a graphical representation of the terrain, including the age of rocks, their composition, interrelationships, and other significant geological phenomena. Old geological maps are invaluable sources of information about the state of geological knowledge at a specific time. Integrating historical geological maps into a GIS database enables the analysis and comparison of past geological data and knowledge with contemporary findings [3]. The study of old geological maps is extremely important, not only for their historical significance in the development of geological mapping but also from a practical perspective [4].

The first map to visually represent actual geological field observations for the territory of Bosnia and Herzegovina (B&H) was created by Ami Boué. His geological travels across the Balkan Peninsula between 1836 and 1838 brought him to Bosnia and Herzegovina on two occasions, and in 1840 he published his comprehensive work, *La Turquie d'Europe*. Attracted by the rich mineral resources of Bosnia and Herzegovina, the new Austro-Hungarian administration began preparations for geological exploration immediately after occupying the country in 1878. Their goal was to create a detailed geological map of the territory, based on a precise topographic foundation. This mapping was carried out in the summer of 1879 by a team of distinguished geologists from the Geological Institute (German: *k.k. Geologische Reichsanstalt*) in Vienna, including Edmund von Mojsisovics, Emil Tietze, and Alexander Bittner. Based on the data they collected with the assistance of Croatian geologist Đuro Pilar, the Geological Overview Map of Bosnia and Herzegovina (German: *Geologische Übersichtkarte von Bosnien-Hercegovina*) at a scale of 1:576,000 was created. This map was a supplement to the renowned General Geological Map of the Austro-Hungarian Monarchy by Franz Ritter von Hauer, published in 1880. The map was created using 20 colors to represent the geological formations of Bosnia and Herzegovina. To outline the boundaries and distribution of each formation, simple sketching with a compass was employed, carefully following the topographic base of the General map of Bosnia, Herzegovina, Serbia and Montenegro at a scale of 1:300,000 published in 1876. In 1887, the map *Geologische Erzlagerstätten — Karte von Bosnien* was created at a scale of 1:300,000 by the Austrian geologist Bruno Walter. Heinrich Foullon von Norbeeck published a map of the Gornji Vakuf area at a scale of 1:75,000, which is regarded as the first detailed geological survey of a region in Bosnia and Herzegovina. A new era in the study of Bosnia and Herzegovina's geology began with the arrival of Friedrich Katzer in Sarajevo. He published the first exceptionally high-quality geological maps at a scale of 1:75,000 (covering Doboj, Zenica, Jajce and Jezero, Prozor, Mostar, and Bugojno), as well as one map at a scale of 1:200,000 (Donja Tuzla), in his Geological Guide to Bosnia and Herzegovina, published in 1903 [5, 6, 7]. His most important cartographic achievement was *the Geological Overview Map of Bosnia and Herzegovina* (GOMB&H), created at a scale of 1:200,000. This significant piece of Bosnian-Herzegovinian geological cartography remains insufficiently studied, much like other maps produced during the Austro-Hungarian rule in Bosnia and Herzegovina [8-10]. Brief overviews of the production and main features of the GOMB&H were provided by [6] and [7]. A bibliographic description of a single map sheet, following the International Standard Bibliographic Description of Cartographic Materials (ISBD-CM), was provided by [11].

To effectively analyze and utilize geospatial data from historical cartographic documents, it is essential to georectify them. The goal of georectification is to establish a relationship between the coordinate system of the digital

image (scanned map) and the modern reference system through georeferencing, while correcting image deformations using geometric transformation. Recent studies [12-14] have applied various georectification methods to historical maps based on 19th-century Habsburg military surveys. These topographic cartographic materials were used as the base maps for the geological maps, ensuring that the projections of the geological maps were essentially the same as those of the topographic maps [15].

The purpose of this paper is to identify an efficient method for georectifying the *Geological Overview Map of Bosnia and Herzegovina* at a scale of 1:200,000, ensuring reliable results. To achieve this, a detailed analysis of the map's mathematical and geodetic foundation is essential. The application of an appropriate geometric correction method should result in satisfactory positional accuracy for the transformed map data. This will enable the efficient integration of various existing data into a unified geospatial database.

2 MATERIAL AND METHODS

2.1 Geological Overview Map of Bosnia and Herzegovina at a scale of 1: 200,000

The Geological Overview Map of Bosnia and Herzegovina (German: *Geologische Übersichtskarte von Bosnien-Hercegovina*), at a scale of 1:200,000, was created under the initiative of Austrian geologist Friedrich Katzer, who is regarded as the founder of the geology in Bosnia and Herzegovina [16].

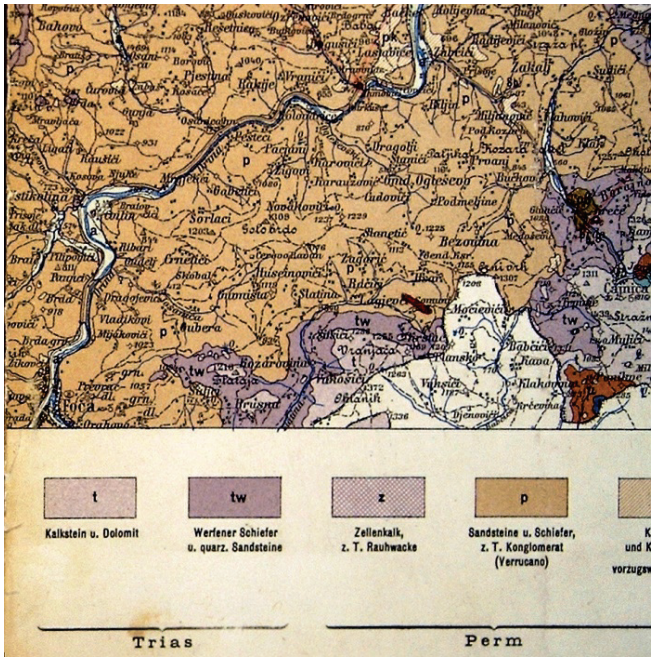


Figure 1 Clip of the Geological Overview Map of Bosnia and Herzegovina, Sheet I - Sarajevo (map is kept in Bosniak Institute Sarajevo, under the title statement "Geologische Ubersichtskarte von Bosnien - Hercegovina. 1. Sechstelblatt : Sarajevo. - Masstab 1:1.200.000", Maps B&H B, ID: 89082634).

The first sheet of this map (Sarajevo) was published in 1907 by the Geological Sector of the State Mining Directorate in Sarajevo (Fig. 1). The second and third sheets (Tuzla and Banjaluka) were published in 1910 and 1921, respectively. The fourth and fifth sheets - Travnik and Ljubuški were completed after Katzer's death by T. Jakšić and M. Milojković and published in 1929. Finally, the sixth sheet (Mostar) was produced by a group of geologists in 1953. Thus, Bosnia and Herzegovina became the first republic of the former Yugoslavia to have a comprehensive geological map covering its entire territory. The creation of this map, with interruptions, took 55 years, from Katzer's arrival in Sarajevo in 1898 until the publication of the last sheet. For many years, it served as the primary reference for a wide range of regional projects, including those in construction, mining, land reclamation, petroleum geology, hydrogeology, engineering geology, and more [7].

2.2 Georectification of Scanned Maps

The research conducted in this study involved the georectification of the fifth sheet of GOMB&H (Ljubuški), produced by Geological Institute Sarajevo in 1929 (Fig. 2), which is kept in the Archive of the Faculty of Civil Engineering, University of Sarajevo. To accomplish this, the map sheet was converted into a digital raster format using a large-format rotary scanner, resulting in a TIFF (Tagged Image File Format) image file.

Figure 2 Clip of the scanned Geological Overview Map of Bosnia and Herzegovina, Sheet V – Ljubuški (source: Faculty of Civil Engineering, Department of Geodesy and Geoinformatics, University of Sarajevo)

The scanned map contains both the errors present in the original paper map and additional scanning errors. Therefore, it is necessary to geometrically correct the map image and align it with the real world, i.e., to georeference the raster dataset. When selecting an appropriate georectification procedure for a scanned map, it is important to consider how the original analogue map was created and the types of errors that may have occurred during that process. Distortions in analogue maps can result from errors in the primary geodetic survey data, mapping inaccuracies, and deformations of the map's supporting material (such as the paper) due to factors like temperature, humidity, aging, and folding [17]. To correct the raster map deformation, it is necessary to reposition the pixels from their original locations to the

corresponding ground coordinates. This requires using an appropriate mathematical model to eliminate distortions and applying a coordinate transformation. The transformation parameters are determined based on tie points (common or control points), and in the next step, all map points are transformed accordingly [3]. The investigation of the georectification of GOMB&H involved the use of various global transformation functions, including similarity, affine, projective, and polynomial transformations. Additionally, a locally sensitive transformation based on the Thin-Plate Spline (TPS) radial basis function was also explored. Detailed descriptions of these transformation methods can be found in the literature on geometric correction of raster digital images, e.g. [18].

2.3 Definition of Tie Points

The raster coordinate system of the scanned map and the map target coordinate system are connected with the tie points. Geometric deformations of the scanned map are determined and eliminated on the basis of these points. Tie points are usually defined by the intersections of the grid formed by lines of latitude and longitude (graticule). Grid intersections are the best control points as these are easily seen on the map, and their coordinates are usually precisely known [19].

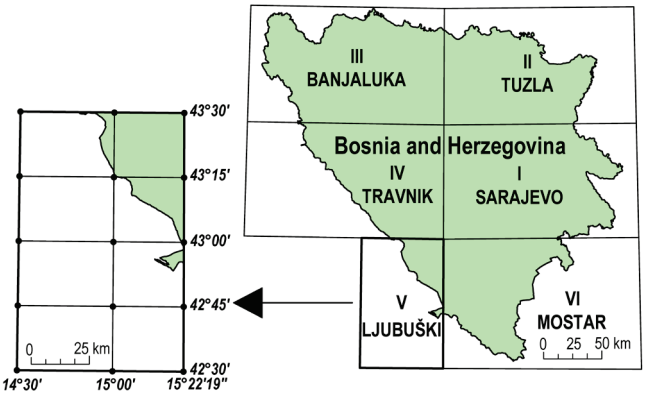


Figure 3 Map graticule of the GOMB&H Sheet V (left) and the index sheet of the 1:200,000 scale Geological Overview Map of Bosnia and Herzegovina (right)

The fifth sheet of GOMB&H contains a graticule which shows parallels every 15'. Meridians are spaced at the varying intervals. In total, 15 points representing the graticule intersections (tick marks) are available on this map sheet (Fig. 3). Since the graticule represents the mathematical framework upon which the whole map's content is based [20], it follows that graticule intersections should be clearly defined with respect to the coordinate system assigned to the geological map, as well as to the system in which the map needs to be converted. Geodetic and mathematical basis of the GOMB&H are identical to the geodetic and mathematical basis of the 1:200,000 scale General map of Central Europe (germ. *Generalkarte von Mitteleuropa*), produced by the Austro-Hungarian Military Geographic Institute (MGI). This map was created using the polyhedric projection [21].

2.4 Evaluating the Geometric Accuracy of Map Transformations

The quality of spatial information in the geometrically corrected raster image of GOMB&H is directly linked to the georectification method applied. Therefore, it is essential to evaluate the 2D (horizontal or planimetric) component of positional accuracy for all map transformation variants [22]. Absolute or external positional accuracy is defined as the accuracy of the position of features within a spatial reference system. Through this accuracy assessment, it becomes possible to determine the extent to which geometric distortions—such as those caused by differential paper shrinkage, map defects like tears, folds, and creases, or non-material-related issues—have been corrected [23].

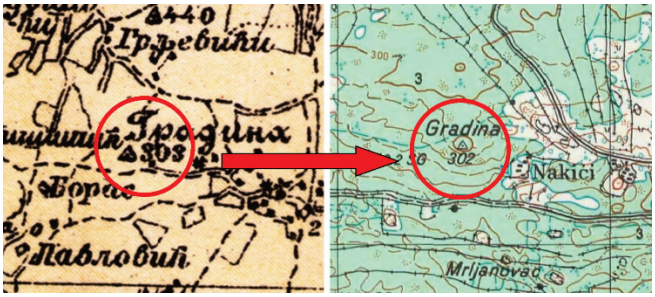


Figure 4 Trigonometrical point at Gradina Hill, 8 km north of Ljubuški – plotted on the GOMB&H (left) and on a more recent topographic map at a scale of 1:25,000 (sheet 574-1-3, Buhovo) (right) (source: Faculty of Civil Engineering, Department of Geodesy and Geoinformatics, University of Sarajevo)

This paper uses a direct 2D accuracy evaluation method, as outlined in ISO 19157 (Geographic Information – Data Quality). The process involves comparing the dataset being assessed with an appropriate set of true or reference values (the reference dataset). To carry out this comparison, check points were defined within the GOMB&H dataset, where the coordinates in the target coordinate system are known to a higher degree of accuracy than those in the dataset being evaluated. This approach allows for the determination of the closeness between the coordinate values obtained from the georectified map and those accepted as the true corresponding positions [24]. In this context, triangulation stations (used as the basis for topographic surveying) can be employed. These stations are depicted on the map with a small triangle symbol. For the purpose of evaluating the accuracy of the transformation, 29 trigonometrical points have been selected, located within the territory of Bosnia and Herzegovina and the neighboring region of Dalmatia. These triangulation stations are positioned on summits to ensure visibility and prominence, with their locations coinciding with points of a more modern trigonometric network (Fig. 4). Specifically, the marks of these triangulation stations were later used in a geodetic survey aimed at establishing the trigonometric network of the former Yugoslavia. As a result, these points became part of a new triangulation network that served as the foundation for the state survey. Consequently, the coordinates of these points are known with significantly greater accuracy than those that could be obtained through map-based measurements. Therefore, the required accuracy

criterion for the reference data was met, allowing the triangulation stations to be used as check points.

Since the ISO 19157 defines only general principles and procedures, the US National Standard for Spatial Data Accuracy (NSSDA) was applied in this research. The NSSDA uses root-mean-square error (*RMSE*) to estimate positional accuracy at 68.3 % confidence level [25]. The *RMSE* is calculated with the use of remained positional distortions d_{EN}:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n} d_{ENi}^2}, \quad (1)$$

where: $d_{EN} = \sqrt{d_E^2 + d_N^2}$, $d_E = E^{\text{ref}} - E^{\text{tr}}$, $d_N = N^{\text{ref}} - N^{\text{tr}}$ - differences between reference and transformed coordinates along *E* and *N* axis.

3 RESULTS AND DISCUSSION

As mentioned earlier, the GOMB&H sheet contains 15 points representing the intersections of the graticule. Using these tie points, various transformation models were applied, where the Stereographic projection and the Hermannskogel datum were used. Before applying the transformation, longitudes were corrected because the GOMB&H was created using the 'Parisian' mapping system, which is based on the Prime Meridian of Paris, with an offset from Greenwich of $\Delta\lambda = 2°20'13.98"$E.

Since the application of different transformation functions results in varying positions for each pixel in a corrected raster image, it is essential to assess the planimetric accuracy of the geometric transformations used to georeference the scanned map. To accomplish this, 29 predefined check points, based on triangulation stations, were used to evaluate the actual effects of different transformation models. Table 1 presents the key accuracy indicators for the transformed check points, including distortion residuals along the *E* and *N* axes, as well as the *RMSE*.

The worst results were obtained using the similarity (Helmert) transformation. The accuracy of this method is 1.26 mm at the map scale (on the map sheet), which corresponds to 253 m at full scale (on the ground). More than one-third of the check points exhibit significant positional distortions of 250 m or more, with a maximum error of 444 m (or 2.22 mm on the map). The analysis revealed that the other transformation methods outperformed Helmert's method, achieving roughly twice the accuracy of the similarity transformation. However, it is clear that all of these methods had limitations in eliminating systematic distortions from the geological map. None of the transformations used were able to remove significant systematic errors from the georeferenced map. This conclusion is reinforced by the normal distribution curve of the remaining distortions along the *N*-axis following the TPS transformation (Fig. 5), which was generated using the 'fitmethis' MATLAB function [26]. The noticeable shift in the distribution curve, indicative of the remaining systematic errors, is readily observable. The achieved accuracy of 0.65 mm on the map using the affine transformation, and 0.75 mm with the TPS and second-order polynomial transformations, is notably lower than the accuracy reported in [27]. In that study, the accuracy of the Habsburg Special Map (scale 1:75,000; SM75) was evaluated. After correcting for distortions caused by paper drying and shrinkage, the planimetric accuracy of the SM75 at the map scale was 0.4 mm, which corresponds to the plotting accuracy on paper. It is important to note that each sheet of the 1:200,000-scale General Map of Central Europe was created by combining eight 1:75,000 sheets into a single mosaic [13].

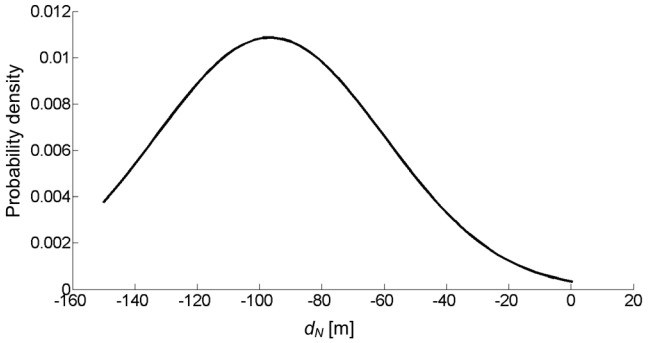


Figure 5 Distribution curve of remaining distortions d_N after applying the TPS transformation

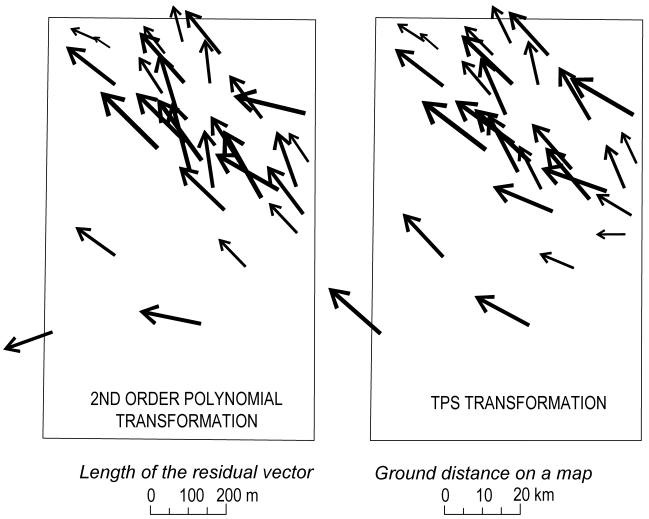


Figure 6 Remaining distortions of check points after applying 2nd order polynomial transformation (left) and TPS transformation (right)

It is therefore clear that the remaining distortions in the GOMB&H are not caused by uncorrected scanning errors or map sheet shrinkage. This is supported by the graphical representation of the positional errors at the check points. To illustrate the local geometric characteristics of the transformed map, planimetric accuracy was visualized using residual vectors (Fig. 6). Each residual vector represents the difference between the reference coordinates and the transformed coordinates of the triangulation station. Fig. 6 clearly demonstrates that, after the geometric transformation, certain positional distortions remain at the check points, indicating the overall homogeneity of the transformed data.

It also shows that the majority of deviations following the TPS and second-order polynomial transformations exhibit consistent propagation, with a noticeable uniform pattern in the vector orientation. These residual vectors reveal significant systematic errors that persist in the transformed map, i.e. the content of the GOMB&H is shifted approximately 150 meters to the northwest relative to the graticule tick marks.

3.1 Method for Enhancing the Georectification of GOMB&H

Since it was shown that transformations based on the GOMB&H graticule defined directly in the MGI 1901 system yield unsatisfactory planimetric accuracy, an alternative approach to improve the georectification quality was investigated. It is important to note that the topographic content of the GOMB&H was created using data collected during the Third Military Mapping Survey of Austria-Hungary. This survey relied on a network of trigonometric points, with coordinates referenced to the Vienna University Datum and the Bessel 1841 ellipsoid.

On the other hand, the current coordinate system (MGI 1901) used in Bosnia and Herzegovina is based on the Bessel ellipsoid and the Hermannskogel Datum. It is essential to point out that the coordinates of triangulation stations in the Vienna University System were computed using preliminary adjusted regional networks with different starting points, while the Yugoslav triangulation represents a homogeneous, final adjusted network oriented to the Hermannskogel [28]. Therefore, nonuniform disparities in point coordinates arise from differences between the geodetic basis of the GOMB&H and the more modern Yugoslav triangulation. The disparities at identical stations are significantly larger than what would be expected as normal errors in coordinate determination. For this research, a method was developed to compute the graticule intersections of GOMB&H with respect to the Bosnian-Herzegovinian state coordinate system (Fig. 7). The procedure is based on the coordinate differences between the Vienna University system and MGI 1901 [29]. The coordinate corrections at an arbitrary point are determined by linear interpolation between neighboring isolines of the differences in latitude and longitude, $\Delta\varphi$ and $\Delta\lambda$. This approach defines a set of 15 tie points (map grid intersections) with corrected coordinates in the target coordinate system, which were then used for the map's georectification.

Tab. 2 presents the main 2D accuracy indicators for the transformed check points, which correspond to the 29 previously defined triangulation stations. The analysis of the values presented in Tab. 2 highlights the limitations of the 4-parameter similarity, 6-parameter affine, and 8-parameter projective transformations in eliminating geometric distortions from the geological map. Statistical indicators reveal that most positional errors could not be corrected by these transformation functions, demonstrating the persistence of significant distortions after applying the similarity, affine, and projective transformations. Moreover, there is no significant improvement in the correction of local systematic errors compared to the values in Tab. 1. This is further confirmed by the residual vectors resulting from these transformations (Fig. 9 and Fig. 10). The vectors are grouped and generally point in the same direction, indicating systematic influences or errors. The vectors exhibit high magnitudes (up to 333 m), and it is clear that the variations in their values are quite pronounced.

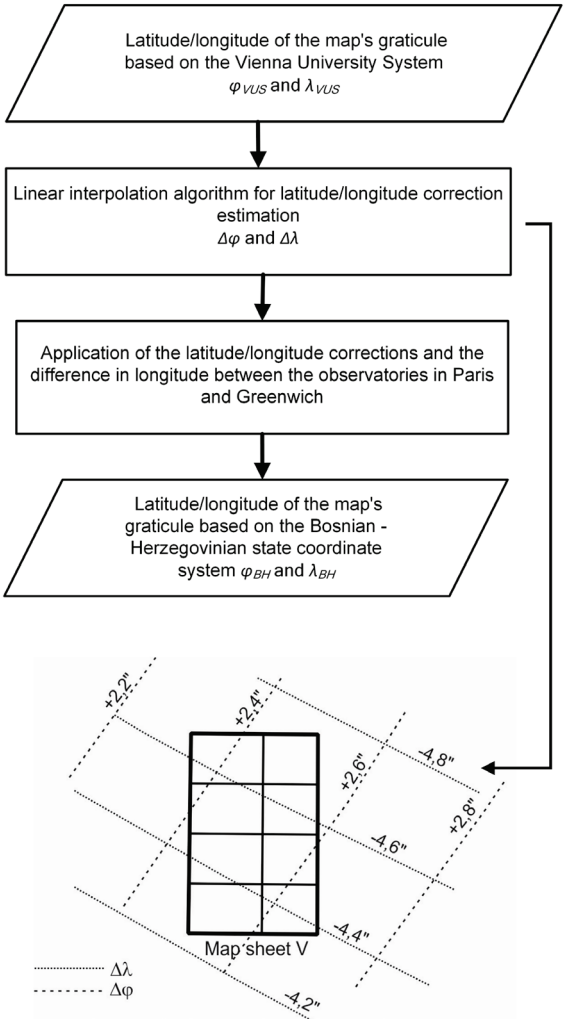


Figure 7 Flowchart for determining the corrected GOMB&H graticule coordinates

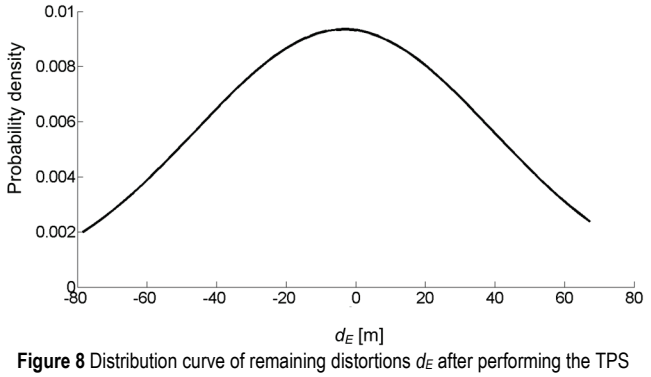


Figure 8 Distribution curve of remaining distortions d_E after performing the TPS transformation

The geometric correction of the GOMB&H using a 2nd order polynomial transformation increases the number of transformation parameters to 12, which positively impacts the elimination of systematic distortions. This transformation

significantly reduces deformations at the check points, resulting in more homogeneous outcomes. This suggests that the uniform layout and spacing of the graticule lines in the geological map help effectively eliminate most of the systematic errors after applying the polynomial transformation. Importantly, no large, undesired positional distortions remain, with the *RMSE* in this case around 70 m, or 0.35 mm at the map scale.

Table 1 Summary of residual statistics for 29 transformed checkpoints across different transformation methods

Indicator		Similarity			Affine			Projective			Polynomial 2nd order			TPS		
		d_E	d_N	d_{EN}	d_E	d_N	d_{EN}	d_E	d_N	d_{EN}	d_E	d_N	d_{EN}	d_E	d_N	d_{EN}
Minimum [m]		–114	–406	103	–71	–244	15	–112	–201	23	13	–179	50	26	–150	62
Average [m]		79	–144	236	38	–76	111	12	–97	128	89	–101	145	100	–97	144
Maximum [m]		420	115	444	202	61	263	250	71	251	188	46	211	171	0	211
Range [m]		535	521	341	273	306	248	362	272	227	175	224	160	145	151	148
RMSE [m]		150	203	253	66	108	127	75	118	140	100	113	151	108	103	150
Remaining distortion distribution %	0-50 m	31.0	17.2	0.0	62.1	37.9	24.1	59	28	7	20.7	17.2	0.0	17.2	13.8	0.0
	50-100 m	27.6	13.8	0.0	27.6	27.6	17.2	28	17	28	41.4	31.0	17.2	34.5	34.5	17.2
	100-150 m	10.3	17.2	17.2	6.9	13.8	34.5	7	31	31	27.6	34.5	27.6	31.0	48.3	24.1
	150-200 m	10.3	13.8	24.1	0.0	13.8	13.8	3	21	28	10.3	17.2	48.3	17.2	3.4	51.7
	200-250 m	10.3	17.2	20.7	3.4	6.9	6.9	3	3	3	0.0	0.0	6.9	0.0	0.0	6.9
	> 250 m	10.3	20.7	37.9	0.0	0.0	3.4	0	0	3	0.0	0.0	0.0	0.0	0.0	0.0

Table 2 Summary of residual statistics for 29 transformed checkpoints across different transformation methods (improved solution)

Indicator		Similarity			Affine			Projective			Polynomial 2nd order			TPS		
		d_E	d_N	d_{EN}	d_E	d_N	d_{EN}	d_E	d_N	d_{EN}	d_E	d_N	d_{EN}	d_E	d_N	d_{EN}
Minimum [m]		–215	–325	34	–174	–163	30	–215	–120	43	–91	–99	8	–78	–71	28
Average [m]		–23	–64	188	–65	4	107	–91	–16	129	–14	–21	66	–3	–16	56
Maximum [m]		317	195	333	97	141	174	145	154	221	83	129	133	67	80	86
Range [m]		532	520	299	271	304	144	360	274	178	174	228	125	146	151	58
RMSE [m]		130	157	204	84	76	114	118	69	136	47	54	72	43	40	58
Remaining distortion distribution %	0-25 m	17.2	3.4	0.0	10.3	24.1	0.0	3.4	13.8	0.0	34.5	31.0	3.4	31.0	34.5	0.0
	25-50 m	6.9	13.8	3.4	20.7	24.1	6.9	13.8	34.5	3.4	27.6	31.0	31.0	34.5	44.8	37.9
	50-75 m	13.8	10.3	6.9	17.2	10.3	13.8	3.4	24.1	6.9	24.1	20.7	20.7	31.0	17.2	44.8
	75-100 m	10.3	10.3	0.0	34.5	24.1	27.6	24.1	13.8	13.8	13.8	13.8	34.5	3.4	3.4	17.2
	100-125 m	13.8	10.3	10.3	10.3	6.9	20.7	20.7	10.3	27.6	0.0	0.0	6.9	0.0	0.0	0.0
	> 125 m	37.9	51.7	79.3	6.9	10.3	31.0	34.5	3.4	48.3	0.0	3.4	3.4	0.0	0.0	0.0

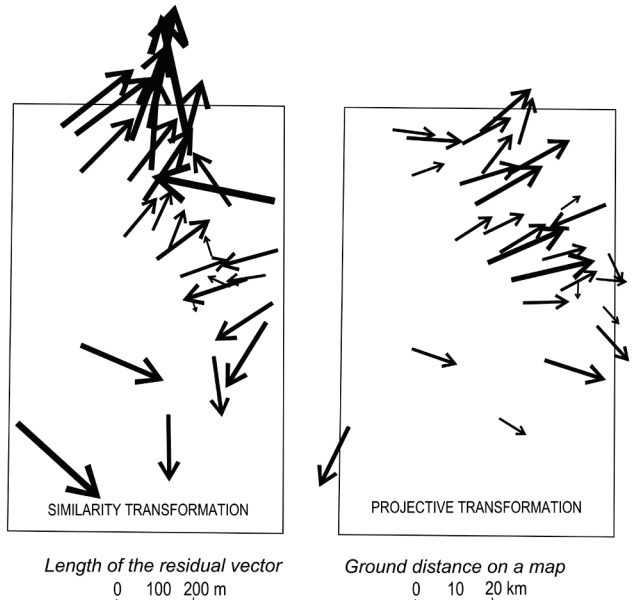


Figure 9 Remaining distortions of check points after applying similarity transformation (left) and projective transformation (right)

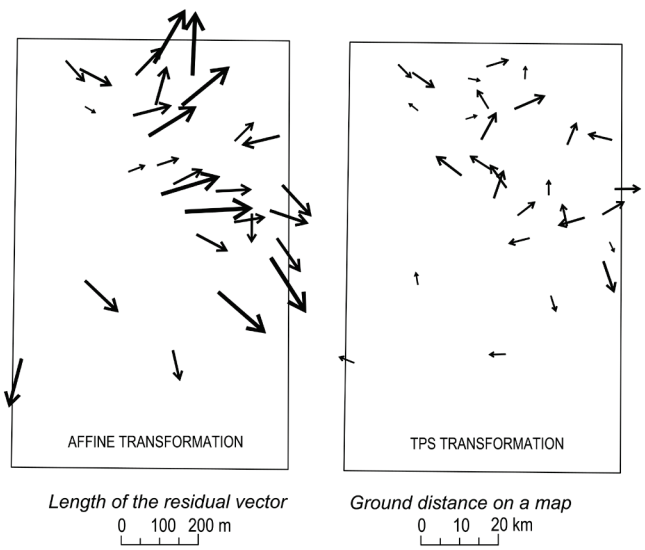


Figure 10 Remaining distortions of check points after applying affine transformation (left) and TPS transformation (right)

The best results in terms of transformation quality were achieved using the TPS transformation (defined by 64 parameters), with an *RMSE* of 0.29 mm at the map scale. The distribution curve shows that the differences between the theoretical and transformed coordinates of the check points have an arithmetic mean close to zero (Fig. 8). After eliminating significant systematic errors using the TPS

transformation, 100 % of the check points showed a d_{EN} of less than 0.5 mm at the map scale. In comparison, the second-order polynomial transformation achieved this for 86 % of the check points, affine transformation for 48 %, projective transformation for 24 %, and similarity transformation for only 10 %. The maximum positional deviation resulting from the TPS transformation was reduced by a factor of four compared to the value obtained using the similarity transformation. This suggests that the uniform arrangement and spacing of the graticule lines in the GOMB&H enable locally sensitive transformations to effectively eliminate systematic errors from the map. The residual vectors on the map georectified using the TPS method were significantly reduced (Fig. 10) compared to those in Fig. 6. Fig. 9 shows that the changes in the vectors' orientation and intensity are moderate, confirming that choosing an appropriate transformation model can effectively eliminate a large portion of the systematic distortions caused by map drafting errors, aging, scanning, and other factors

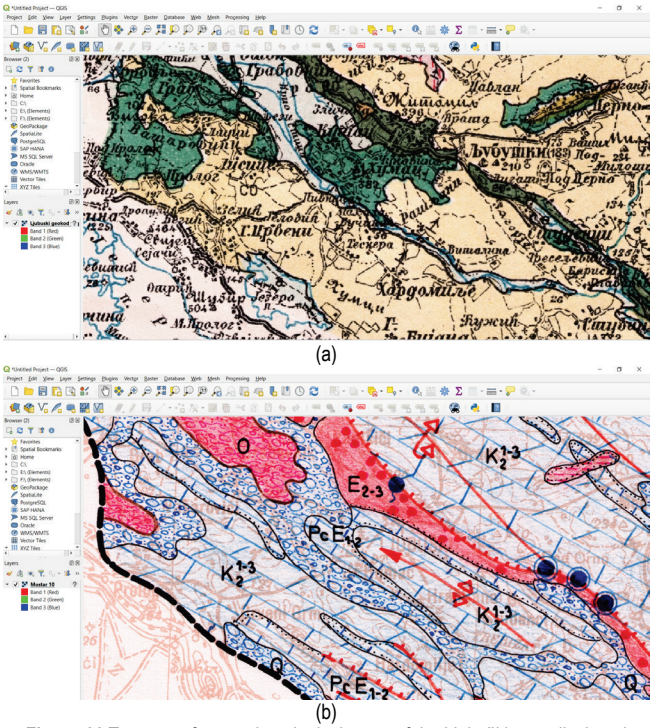


Figure 11 Two georeferenced geological maps of the Ljubuški area displayed within the QGIS environment: (a) GOMB&H; (b) Geological map of Bosnia and Herzegovina at a 1:200,000 scale, produced by the Geological Institute in Belgrade, 1967 (source: University of Sarajevo - Faculty of Civil Engineering, Department of Geodesy and Geoinformatics)

The TPS transformation of the scanned raster map produced a geometrically corrected map, now aligned with the official Bosnian-Herzegovinian state coordinate system. The accuracy of well-defined points of detail plotted on the map is expected to be within ±0.5 mm. This enables the creation of a cartographic database that allows data from the GOMB&H to be compared, overlaid, and integrated with geological data from its more modern counterparts (Fig. 11). Of particular interest is the comparison of maps created at different times that depict the same geographic area.

4 CONCLUSION

This research emphasizes the critical importance of selecting the correct transformation base (control points) and the appropriate functional transformation model in effectively eliminating systematic errors caused by factors such as paper deformation, scanning irregularities, drafting and reproduction errors, and misaligned coordinate systems. The study confirmed that the quality of map georeferencing depends largely on the precise definition of control points in both the original map's coordinate system and the target coordinate system. Nonuniform discrepancies in point coordinates, resulting from differences between the geodetic basis of the Geological Overview Map of Bosnia and Herzegovina (GOMB&H) and the more modern triangulation, were successfully identified. It was demonstrated that transformations based on the GOMB&H graticule, which is directly defined within the MGI 1901 system, resulted in unsatisfactory planimetric accuracy, with the *RMSE* reaching 0.65 mm at the map scale. This was because the topographic content of the map, including the triangulation stations, was referenced to the Vienna University Datum rather than the Hermannskogel datum. The resulting discrepancies at the same stations were irregular and considerably larger (up to 1.3 mm on the map) than expected normal errors in map measurement. As a result, the coordinates of the intersections of the meridians and parallels on the map were adjusted to align with the Bosnian-Herzegovinian state coordinate system. This correction yielded satisfactory transformation results, with an *RMSE* of 0.29 mm at the map scale. This level of accuracy can be achieved if all available graticule intersections are incorporated into the transformation model.

The study also demonstrated that selecting the appropriate functional model in the transformation process is crucial for achieving reliable georectification results. Statistical and graphical indicators clearly show how suitable transformation functions can reduce geometric distortions in the map content. Systematic errors in the scanned map cannot be effectively corrected by global transformations relying on a small number of parameters. The assessment of geometric accuracy revealed poor results when georectification was performed using common procedures, such as similarity, projective, or affine transformations. In contrast, more complex polynomial transformations, and especially the locally sensitive TPS transformation—proved successful in reducing significant geometric distortions in the geological map.

By using the correct approach for the geometric processing of the GOMB&H sheet—printed 95 years ago on standard cartographic paper and preserved under poor conditions—it was determined that the remaining distortions in the transformed points were random. This resulted in satisfactory positional accuracy for the georeferenced map, with a representative sample of triangulation stations plotted in their true positions at map scale to an accuracy better than ±0.5 mm.

5 REFERENCES

[1] Yang, Y., Zhang, S., Yang, J., Chang, L., Bu, K., & Xing, X. (2014). A review of historical reconstruction methods of land use/land cover. *Journal of Geographical Sciences, 24*(4), 746–766. https://doi.org/10.1007/s11442-014-1117-z

[2] Seemann, P. (2011). Cartometric Analysis of Planimetric Accuracy of Topographic Maps 1:100 000 Created in S-46 Coordinate System. *Acta Montanistica Slovaca, 16*(4), 253-261. https://actamont.tuke.sk/pdf/2011/n4/4seemann.pdf

[3] Korodi, E., Bartos-Elekes, Z., Rus. I., & Haidu, I. (2016). Cartographic analysis of some old Transylvanian geological maps from the second half of the nineteenth century. *Acta Geodaetica et Geophysica Hungarica, 51*(3), 405-419. https://doi.org/10.1007/s40328-015-0133-5

[4] Korodi, E., Bartos-Elekes, Z., & Haidu, I. (2017). The Anina (Steierdorf) coal mining district in Banat (Romania) on some old geological maps (1850-1884). *Austrian Journal of Earth Sciences*, *110*(2), 16 p. https://doi.org/10.17738/ajes.2017.0016

[5] Gašparović, R. (1970). *Bosna i Hercegovina na geografskim kartama od prvih početaka do kraja XIX vijeka*. Sarajevo: Akademija nauka i umjetnosti Bosne i Hercegovine. (in Croatian).

[6] Ćorić, S. (1999). Die geologische Erforschung von Bosnien und der Herzegowina und der grundle-gende Beitrag der österreichischen Geologen. *Abhandlungen der Geologische Bundesanstalt, 56*(1), 117-152. (in German)

[7] Čičić, S., & Redžepović, R. (2006). O naučnom naslijeđu u geologiji Bosne i Hercegovine. *Geološki glasnik, 36*, 1-28. (in Bosnian)

[8] Arnautović, K. (2013). Topografski znaci na katastarskim planovima austro-ugarskog premjera Bosne i Hercegovine, *Geodetski glasnik, 44*, 72-85. (in Bosnian) https://doi.org/10.58817/2233-1786.2013.47.44.72

[9] Topoljak, J., Tuno, N., Mulahusić, A., Husić, A., & Fekeža-Martinović, L. (2015). Ostaci srednjevjekovne utvrde Kaštel (Fenarlik) u kartografskim izvorima XIX i XX vijeka. *Geodetski glasnik, 46*, 55-74. (in Bosnian) https://doi.org/10.58817/2233-1786.2015.49.46.55

[10] Tuno, N., & Mulahusić, A. (2015). Generalna karta Bosne i Hercegovine razmjere 1:150 000. *Geodetski glasnik, 46*, 28-47. (in Bosnian) https://doi.org/10.58817/2233-1786.2015.49.46.28

[11] Mulahusić, A. (2007). Kartografija Sarajeva. Sarajevo: Univerzitet u Sarajevu – Građevinski fakultet (in Bosnian)

[12] Molnár, G., & Timár, G. (2009). Mosaicking of the 1:75 000 sheets of the Third Military Survey of the Habsburg Empire. *Acta Geodaetica et Geophysica Hungarica 44*(1), 115-120. https://doi.org/10.1556/AGeod.44.2009.1.11

[13] Timár, G., Molnár, G., & Craciunescu, V. (2007). Georeference of the 1:200,000 'degree maps' of Central Europe (about 1910). *Geophysical Research Abstracts, Vol. 11, EGU2009-2574, 2009*. https://doi.org/10.13140/2.1.4108.6404

[14] Affek, A. (2013). Georeferencing of historical maps using GIS, as exemplified by the Austrian Military Surveys of Galicia. *Geographia Polonica, 86*(4), 375-390. https://doi.org/110.7163/GPol.2013.30

[15] Galambos, C. (2009). Development of color signs and projections of the Hungarian archive geological maps. *Acta Geodaetica et Geophysica Hungarica, 44*, 131-140. https://doi.org/10.1556/AGeod.44.2009.1.13

[16] Hrvatović, H. (2006). *Geological Guidebook through Bosnia and Herzegovina*. Sarajevo: Geological Survey of Federation Bosnia and Herzegovina.

[17] Tuno, N., Mulahusić, A., & Kogoj, D. (2017). Improving the Positional Accuracy of Digital Cadastral Maps through Optimal Geometric Transformation. *Journal of Surveying Engineering*, *143*(3), 05017002. https://doi.org/10.1061/(asce)su.1943-5428.0000217

[18] Suri, J. S., Wilson, D. L., & Laxminarayan, S. (Eds) (2005). *The Handbook of Medical Image Analysis – Volume III: Registration Models*. New York: Kluwer Academic/Plenum Publishers. https://doi.org/10.1007/b104807

[19] Mccloy, K. R. (1995). *Resource Management Information System - Process and Practice*. London: Taylor & Francis.

[20] Maling, D. H. (2013). *Measurements from Maps: Principles and Methods of Cartometry*. Burlington: Elsevier.

[21] Kretschmer, I. (1991). The mapping of Austria in the twentieth century. *Imago Mundi: The International Journal for the History of Cartography*, *43*(1), 9–20. https://doi.org/10.1080/03085699108592715

[22] Tuno, N., Mulahusić, A., & Topoljak, J. (2018). Komparativna analiza različitih pristupa korigiranju geometrijskih distorzija stare šumarske karte. *Šumarski list, 142*(1-2), 287-295. (in Croatian) https://doi.org/10.31298/sl.142.5-6.3

[23] Jongepier, I., Soens, T., Temmerman, S., & Missiaen, T. (2016). Assessing the Planimetric Accuracy of Historical Maps (Sixteenth to Nineteenth Centuries). New Methods and Potential for Coastal Landscape Reconstruction. *The Cartographic Journal, 53*(2), 114-132. https://doi.org/10.1179/1743277414y.0000000095

[24] Tuno, N., Mulahusić, A., Topoljak, J., & Kogoj, D. (2017). Vrednotenje postopkov za ocenjevanje točnosti georeferenciranja geodetskih načrtov. *Geodetski vestnik, 61*(3), 373-386. (in Slovenian) https://doi.org/10.15292//geodetski-vestnik.2017.03.373-386

[25] FGDC (1998). Geospatial Positioning Accuracy Standards, Part 3: National Standard for Spatial Data Accuracy (NSSDA). Washington, DC: Federal Geographic Data Committee. Retrieved from http://www.fgdc.gov/standards/projects/FGDC-standards-projects/accuracy/part3/chapter3

[26] de Castro, F. (2018). Fitmethis, MATLAB Central File Exchange. Retrieved from https://www.mathworks.com/matlabcentral/fileexchange/40167-fitmethis

[27] Čechurová, M., & Veverka, B. (2009). Cartometric analysis of the Czechoslovak version of 1:75 000 scale sheets of the Third Military Survey (1918-1956). *Acta Geodaetica et Geophysica Hungarica, 44*, 121-130. https://doi.org/10.1556/AGeod.44.2009.1.12

[28] Mugnier, C. (2013). Grids & Datums - Boznia and Herzegovina. *Photogrammetric Engineering & Remote Sensing, 79*(3), 229-230.

[29] Tuno, N. (2014). Optimalne metode geometrijske obrade digitalnih geodetskih planova. *PhD thesis*. Sarajevo: Univerzitet u Sarajevu – Građevinski fakultet (in Bosnian)

Author's contacts:

Nedim Tuno, Full Professor
University of Sarajevo, Faculty of Civil Engineering,
Patriotske lige 30, 71000 Sarajevo, Bosnia and Herzegovina
E-mail: nedim_tuno@gf.unsa.ba

ISSN 1846-6168 (Print), ISSN 1848-5588 (Online)
https://doi.org/10.31803/tg-20250127200641

Original scientific paper
Received: 2025-01-27, Accepted: 2025-03-04

The Influence of Copper Plate Dimensions on Temperature Distribution during TIG Welding

Mihael Zelić*, Matija Bušić, Tomislav Veliki

Abstract: This research investigates the influence of copper plate dimensions on temperature distribution during TIG (Tungsten Inert Gas) welding of Cu-ETP (Electrolytic Tough Pitch Copper). When welding copper and lightly alloyed copper alloys with high thermal conductivities, the appropriate type of current and shielding gas must be selected to ensure sufficient heat input. This high heat input compensates for the rapid heat dissipation and cooling in the localized weld zone. The required heat input is determined by the dimensions of the copper sheets to be welded. Experimental measurements using thermocouples were performed on plates with identical thicknesses but varying dimensions, while numerical simulations via ANSYS complemented the analysis. The results highlight the significant impact of plate dimensions on heat dissipation and temperature gradients, providing valuable insights for improving welding settings.

Keywords: Cu-ETP; heat conduction; numerical simulations; thermal conductivity; TIG welding

1 INTRODUCTION

Copper and copper alloys are widely used in many industrial applications for their high electrical and thermal conductivities, good combinations of strength and ductility, and excellent corrosion resistance [1]. The welding of copper and copper alloys is a critical process in producing various components for the chemical and beverages industry, such as high-voltage power transformers, nuclear waste processing, etc. Copper's exceptional thermal and electrical conductivity makes it widely used, but these properties present significant challenges during welding [2]. Copper's high thermal conductivity rapidly dissipates heat from the welding zone, making it difficult to maintain optimal welding temperatures [2]. Additionally, copper oxidizes quickly at high temperatures, potentially compromising weld quality and causing defects [2].

Copper is not subject to phase transformations and cannot be hardened through heat treatment, a characteristic shared by many of its alloys. This characteristic poses additional challenges in maintaining mechanical strength during welding processes [3]. This challenge is further exacerbated by the fact that thermal exposure induces softening in cold-worked copper alloys, resulting in a pronounced reduction in mechanical strength within the heat-affected zone [3].

TIG welding, which employs a non-consumable tungsten electrode in an inert gas atmosphere, has proven to be an effective method for welding copper [2]. This technique ensures precise heat control and minimal weld contamination, which is critical for achieving high-quality joints with high purity and minimal inclusions. This is particularly important in industries where reliability and durability are paramount [3]. Welding generates substantial heat, significantly influencing the base material's microstructure and mechanical properties [3]. The direction and extent of heat distribution in the welded metal are closely linked to the residual stresses and deformations that develop during the cooling phase. Therefore, predicting heat distribution within the material is essential, especially for optimizing weld quality and mitigating defects.

From a practical standpoint, analytical solutions offer the benefit of exceptional computational speed [4]. However, their limited accuracy often becomes a constraint when applied to complex engineering problems such as welding simulations [4], prompting researchers to explore more refined numerical methods. Consequently, numerous researchers have sought to improve analytical approaches, but these efforts have achieved only marginal gains in precision [5-8]. Numerical solutions, on the other hand, provide a higher degree of accuracy but require substantial computational resources [9].

The finite element method (FEM) was initially applied in welding simulations for static and dynamic analyses and later evolved to tackle more intricate issues involving geometric and material nonlinearities. Early studies, such as those by Ueda and Yamakawa [9] and Hibbit and Marcal [10] in the 1970s, pioneered the integration of thermal and mechanical analyses, addressing large deformations and nonlinear stress conditions. Marcal [11] underscored the exceptional complexity of welding as a nonlinear problem in structural mechanics. Goldak et al. [12] noted that the challenges faced by these pioneers discouraged many experts from entering the field. Despite these obstacles, tools like MARC, developed by Marcal and Hibbit, and ABAQUS, created by Hibbit, emerged as foundational programs for advanced welding simulations, becoming essential for nonlinear analysis.

Over the last 30 years, welding simulations have drawn significant research interest, particularly in addressing complex challenges like copper welding. This focus has led to numerous publications and reviews, such as [11–16], which provide valuable insights into progress and best practices. Notably, Lindgren's extensive reviews [14–16] offer a thorough summary of past studies, the current state of the field, and specific guidelines for modeling welding processes to achieve targeted simulation objectives. These efforts have contributed significantly to advancing the understanding and practical implementation of welding simulations. The aim of this research is to analyze heat input, heat conduction, and temperature distribution during TIG welding of copper plates. The focus will be on examining

temperature distribution across varying dimensions of copper plates, analyzing welding parameters' influence on weld quality, and employing numerical methods to simulate thermal processes. The study will include theoretical analysis, experimental measurements, and simulations to comprehensively understand the process. The experimental component of the study will involve detailed measurements of temperature distribution on copper plates during TIG welding using thermocouples at selected points. Simulations will be conducted using ANSYS to accurately model and compare results, with the collected data being used to optimize welding parameters to achieve the highest weld quality.

2 EXPERIMENT

In this experiment, the temperature as a function of time was measured during heating to simulate the TIG welding process for Copper ETP plates with various dimensions. Measurements were taken at 10 positions on each plate, as illustrated in Fig. 1 below.

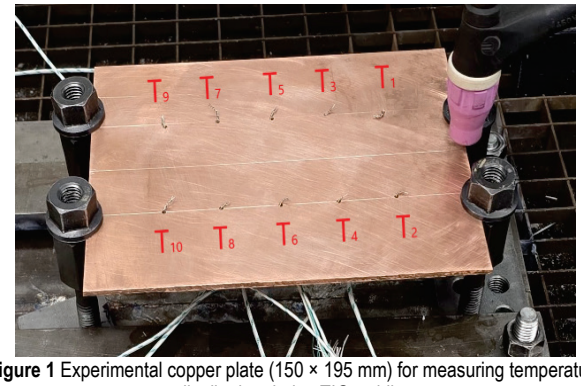


Figure 1 Experimental copper plate (150 × 195 mm) for measuring temperature distribution during TIG welding

The temperature was recorded using K-type thermocouples connected to a precision temperature measurement device (Fluke 1586A Super-DAQ). The test plates featured designated positions (T1, T2, T3, T4, T5, T6, T7, T8, T9, T10) indicating the locations where the thermocouples were fixed, allowing for indirect temperature determination via voltage differences. The experiment used two plates of equal length but different widths (Figs. 2 and 3).

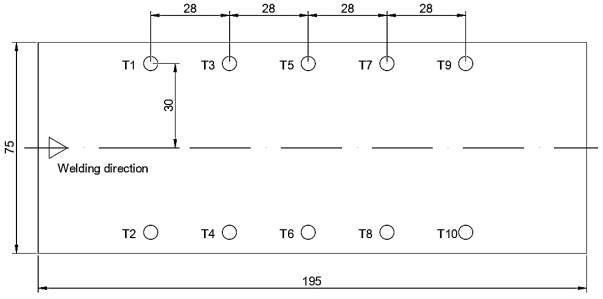


Figure 2 Experimental copper plate (75× 195 mm) for measuring temperature distribution during TIG welding

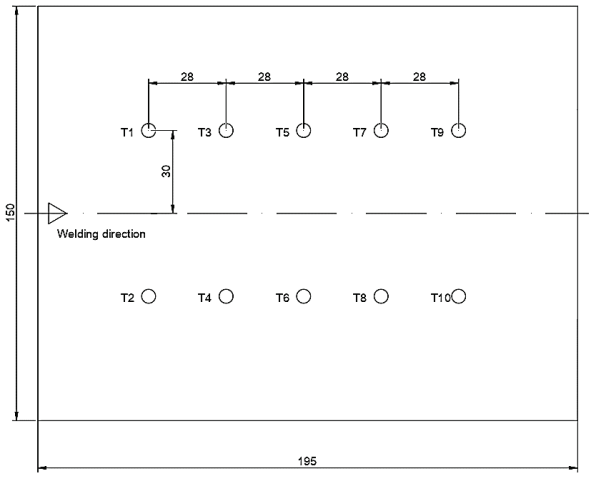


Figure 3 Experimental copper plate (150 × 195 mm) for testing temperature distribution during TIG welding

In both cases, the thermocouples were positioned at an equal distance of 30 mm from the heat input line. The thickness of both plates was identical, measuring 3 mm. The TIG torch was guided using an automated PROMOTECH Rail Bull system.

2.1 Welding Base Material

The plates used in this experiment were made of Cu-ETP copper, in the form of sheets with a thickness of 3 mm and dimensions of 195 × 150 mm and 195 × 75 mm. The material certificate confirms the chemical composition of the Cu-ETP copper used, as listed in Tab. 1 below [17].

Table 1 The material chemical composition [17]

Composition			
Cu	O	Bi	Pb
min 99.9 %	max 0.04 %	max 0.0005 %	max 0.005 %

Tab. 2 provides thermophysical data for the Cu ETP material used in the experiment.

Table 2 Thermophysical data for the Cu ETP [17]

Property of Cu - ETP	Value	Condition
Density, ρ	8.9 g/cm^3	At 20 °C
Thermal Conductivity, k	394 W/m·K	At 20 °C
Specific Heat Capacity, c	0.386 J/g·K	At 20 °C
Emissivity, ε	Up to 0.78	Oxidized Copper
Convective Heat Transfer Coefficient, a_c	5-25 W/m^2·K	Free Convection (Air)

2.2 Experimental Methods and Equipment

The following equipment was used in the experiment:

- Precision Temperature Scanner: Fluke 1586A Super-DAQ, used for accurate temperature measurements and data logging.
- K-Type Thermocouples: Employed for real-time monitoring of temperature distribution on the copper plates.
- TIG Welding Machine: VARTIG 005 AC/DC, providing a stable and controllable heat source for the TIG process.

- Automated Electrode Guiding System: PROMOTECH Rail Bull ensures consistent and precise movement of the welding electrode along the specified welding path.

2.2.1 The Measuring Equipment and Uncertainties

The 1586A Super-DAQ, Fig. 4, is a low voltage 6 ½ digits multimeter that enables precise measurement of temperature and electrical values and, when combined with the internal modules, up to 40 channels can be scanned [18].

Figure 4 Fluke 1586A Super-DAQ, [18]

Measurement was performed with 10 Thermocouple Type K class 1 inserted in their respective holes on the plate and connected to the internal 1586-2586 High-Capacity Modul. Cold Junction Compensation was internal, and the delay time between channels was 0.5 s, enabling scanning at a more accurate medium speed mode. Reading was stored in the internal memory of the multimeter. The uncertainty of the temperature acquisition was in the mentioned setup 0.6 K over the whole range according to the manufacturer specification [18]. Thermocouples were new, unused 0.5 mm Type K from Omega Engineering, mineral fiber insulated, and suitable for measurement up to 1200 °C. Uncertainty of such thermocouples is 1.5 K, according to the [19]. To minimize the uncertainty of the measurement, equipment was calibrated as a temperature indicator (both the thermocouples and the multimeter) according to the [20] in the same configuration as it was to be used for the measurement. Calibration was performed in the range from 30 °C to 350 °C, and the resolution of the reading on the multimeter was 0.01 K. The Thermal source in the range of 30 °C to 150 °C was Druck Premium Calibrator PTC165 [21] with a calibration uncertainty of 0.250 K (K = 2, P = 98 %). All thermocouples were within the range after the stabilization of the thermal source and internal CJC. The range from 200 °C to 350 °C, the Additel 875 Dry Well Calibrator with calibration block was used [22]. Uncertainty of the temperature source was 0.6 K (K = 2, P = 98 %), and again, readings of all thermocouples were within the range. Uncertainty of the thermal zone as determined by the calibration includes stability, radial and axial gradients, and loading effect. The inhomogeneity effect of the thermocouple wire was not considered, as the thermocouples were new and used only for the experiment. Based on the abovementioned uncertainty, the measured temperature in the experiment is estimated to be 0.6 K (K = 2, P = 98 %), and no correction on measured values has been applied.

2.2.2 VARTIG Welding Sources

Vartig welding sources are designed for the TIG process, with the additional capability of welding with coated electrodes. They are suitable for welding structural and stainless steels, aluminum, and copper. The devices feature a digital display of welding parameters and LED indicators for various operating modes [23]. Tab. 3 presents the welding parameters, which were identical for both plates [23].

Table 3 Welding parameters applied in experiments.

Welding parameter	Type/ value
Current type	Direct current (DC)
Electrode polarity	Electrode negative (EN, SP)
Current intensity	120 A
W electrode diameter	3.2 mm
W electrode type	Wla 15
Vertex angle of the W electrode	45 °
Electrode to workpiece angle	90 °
Shielding gas	Argon 4.8
Shielding gas flow rate	12 l/min
Welding position	PA (down hand, flat)
Arc voltage	14.8 V
Welding speed	5.3 mm/s

The gas flow was 12 l/min. Samples were welded using pure argon. This shielding gas, which has the trademark name Argon 4.8 (with a purity of 99,998%), is classified as I1 according to standard EN ISO 14175:2008 [23].

2.2.3 Automated Electrode Guiding Device

The PROMOTECH Rail Bull device, Fig. 5, automates the linear movement of the TIG torch and is designed for TIG-MIG/MAG welding or plasma cutting.

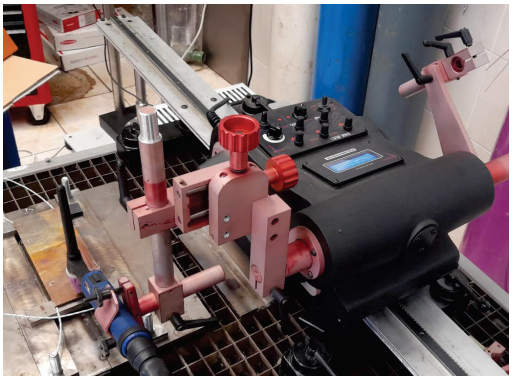


Figure 5 Automated TIG Torch Guiding in the Experiment

The device moves over ferromagnetic and non-ferromagnetic materials using a rack-and-rail drive system. The rails can be attached to flat or curved surfaces using magnetic or vacuum units, with the minimum curvature radius for the semi-flexible track being 5 m [23].

2.3 Results of the Experiment

In Fig. 6, the temperature distributions over time can be seen for a plate with dimensions of 75 × 195 mm, while in Fig. 7, the temperature distribution is shown for a plate with dimensions of 150 × 195 mm.

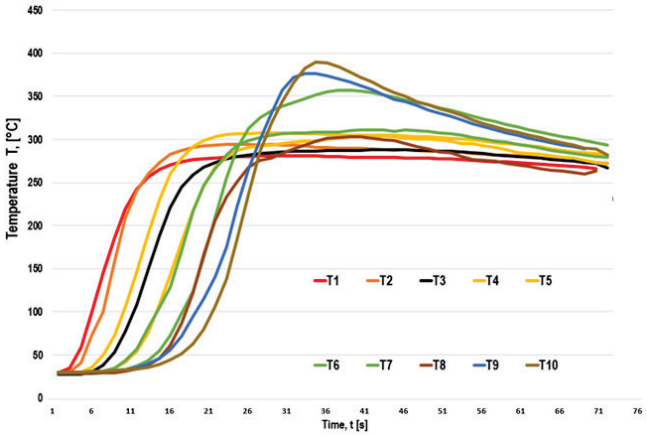


Figure 6 Graph of the plate (75 × 195 mm) temperature dependence on time for the thermocouples

The maximum temperature achieved by the smaller plate is T_{max} = 390 °C, recorded at thermocouple 10. An initial increase in temperature is observed, following an exponential heating pattern, up to the maximum temperature T_{max}, after which a cooling phase occurs with a negative gradient. Differences between the thermocouples indicate an asymmetrical heat distribution, where thermocouples closer to the welding zone record higher values, while those farther away show lower values and take longer to reach the peak temperature. Symmetrical thermocouples should display identical temperatures under ideal conditions, such as in ANSYS simulations. This is expected and can be explained by the fact that the material is not perfect, the thermocouples did not measure at perfectly symmetrical points, and the clamps caused asymmetrical heat dissipation. Fig. 6 shows temperature functions for all thermocouples on the plate with dimensions 75 × 195 mm. However, in the experiment, differing temperatures are observed. They can experience minor geometric misalignments, even on the order of a few millimeters, which lead to localized differences in heat flow. They may also encounter uneven contact with clamps or surface irregularities, as well as material inhomogeneities, creating zones of increased or decreased heat dissipation. They are further influenced by changing environmental conditions, such as airflow and localized radiation, which affect cooling and can introduce measurement variations. Thermocouples T1 and T2 initially exhibit a rapid temperature increase due to a brief yet intense interaction with the heat source, followed by an extended cooling period. In contrast, T9 and T10 require more time to reach elevated temperatures due to slower heat transfer, which leads to higher final temperature readings. Nonetheless, the overall heat balance still indicates that the heat input exceeds its dissipation. Fig. 7 shows the graphs for the plate with dimensions 150 × 195 mm. An almost exponential rise is observed across all thermocouples, culminating in peaks that increase with each subsequent thermocouple. The temperature reaches T_{max} = 180,15 °C at T8, followed by a cooling phase with a negative gradient. All curves descend after T8 reaches this peak, governed by heat dissipation mechanisms (convection, contact with clamps or surroundings, and radiation). In contrast, the smaller 75 × 195 mm plate exhibits a higher peak temperature (up to 390 °C) and faster heating, attributable to its lower thermal mass and smaller surface area. The larger plate distributes the input heat over more material, which results in a reduced maximum temperature, a slower overall heating profile, and an extended cooling period due to increased thermal inertia.

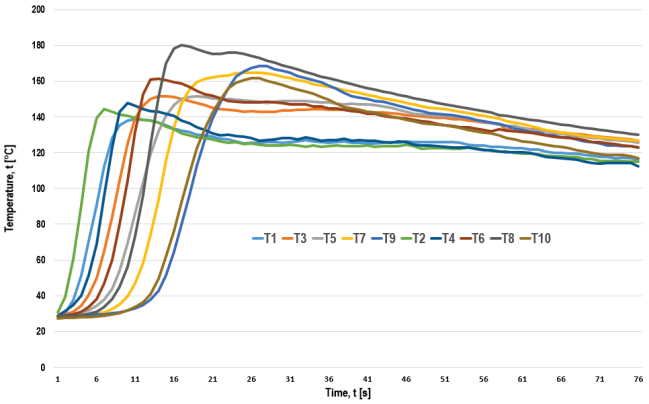


Figure 7 Graph of the plate (150 × 195 mm) temperature dependence on time for the thermocouples

2.4 Effect of Plate Dimensions on Temperature Distribution

Larger plates generally exhibit a greater capacity for heat dissipation due to their increased surface area, which facilitates conduction and radiation. Consequently, temperature gradients across larger plates are less pronounced than those on smaller plates, where heat remains more localized. The geometry of the plate significantly influences heat flow paths. In smaller plates, heat conduction is constrained by the shorter distance to the edges, leading to faster cooling at the boundaries. In contrast, larger plates allow heat to take longer to reach the edges, resulting in a more uniform temperature distribution near the weld line.

Experimental observations revealed notable differences in temperature measurements at symmetrical points on plates of varying dimensions. Smaller plates displayed higher peak temperatures confined closer to the weld zone. In contrast, larger plates demonstrated a broader heat-affected zone with lower peak temperatures, indicating a more dispersed heat distribution. In this paper, plates with dimensions of 75 × 195 mm and 150 × 195 mm are used to more accurately investigate the effect of plate width on heat distribution during TIG welding. By selecting non-standard dimensions, more pronounced temperature gradients and varied cooling rates are achieved, which cannot be accomplished with conventional standard plates. This approach ensures a better correlation between experimental results and numerical simulations, thereby contributing to the optimization of welding parameters.

3 NUMERICAL ANSYS SIMULATION

All four states of matter are present during TIG welding, and their physics is highly complex. The TIG process combines magnetohydrodynamics, plasma physics, and material science [24]. ANSYS uses the finite element method

(FEM) to analyze thermodynamic processes, including transient thermal analysis. This method divides complex geometries into smaller elements for solving heat transfer equations [24]. The simulation analyzed convection, conduction, and heat radiation in the plates, base, and fixators. Phase changes of the molten material and the influence of the latent heat of melting were not considered. Initially, the starting temperature of the plates was defined, along with the heat dissipation coefficients for radiation and convection with the surrounding air. All other thermophysical properties of the material were defined within the simulation (Fig. 8).

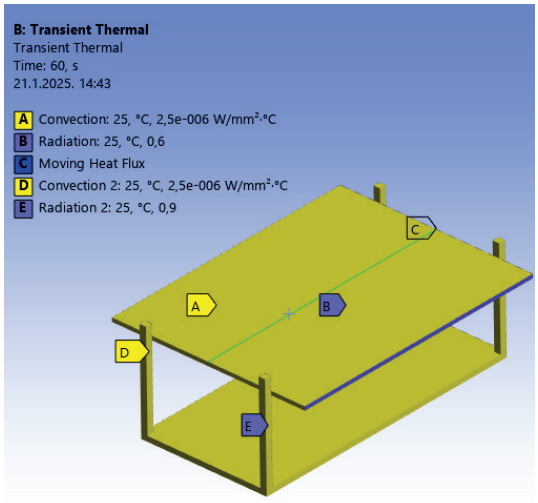


Figure 8 Initial conditions of radiation and heat convection for the measurement setup

3.1 Geometry Model and Meshing

The geometry model in ANSYS was meshed (Fig. 9) using the mechanical physics preference and adaptive sizing.

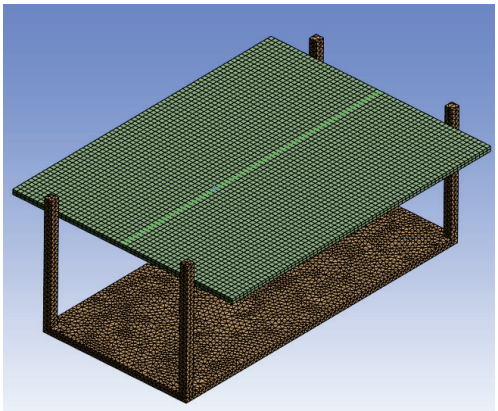


Figure 9 The generated mesh grid in ANSYS

The mesh for the smaller plate consists of 106,927 nodes and 45,535 elements, with a minimum edge length of 2,8 mm. The mesh for the larger plate includes 68,611 nodes and 30,884 elements. The average surface area of the elements was calculated as 4,359.3 mm², with a bounding box diagonal of 266.02 mm.

3.2 Heat Source

Considering that the welding source is relatively far from the temperature distribution measurement location, the source's power distribution is not a relevant factor [24]. Therefore, a simple circular heat source model was used. The source distribution is described as a radial cylinder with a diameter of $r_0 = 5$ mm, as shown in Fig. 10.

The parameters of the welding source are shown in the following Fig. 11. The travel speed is 5.3 mm/s along the plate's symmetry (i.e., along the weld), while the power of the source is 18 W/mm².

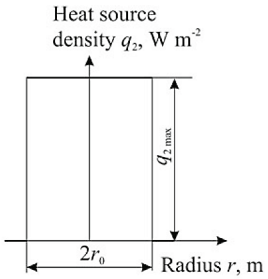


Figure 10 Circular source with constant heat source density q_{arc}, [25]

Details of "Moving Heat Flux"

Geometry	
Scoping Method	Geometry Selection
Geometry	2 Faces
Path	
Scoping Method	Geometry Selection
Geometry	1 Edge
Start Point	
Scoping Method	Geometry Selection
Geometry	1 Vertex
Definition	
Index	1
First Patch?	Yes
Last Patch?	Yes
Velocity	5.3 mm/s
Radius of the Beam	5 mm
Source Power Intensity	18 W/mm²
Start Time	0 s
End Time	37 s
Number of Segments	200
Minimum Steps for Cooling Phase	20
Material Removal	No
Melting Temperature	1080°C

Figure 11 Parameters of the moving heat source

The amount of heat input per unit length of the weld is given by the following Eq. (1) [25]:

$$Q_{\text{net}} = \frac{U \cdot I}{\text{v}} \cdot \eta \tag{1}$$

The variables in the above equation represent the following:

- $U = 14.8$ V, welding voltage
- $I = 120$ A, electric current
- $v = 5.33$ mm/s, welding speed
- $l = 3.2$ mm, arc length
- η heat transfer efficiency coefficient.

Using a nonlinear three-dimensional heat transfer model, a formula for arc efficiency was derived, which can be seen below [25]:

$$\eta = 0.947 - 0.00017 \cdot I - 0.057 \cdot l + 0.02 \cdot v$$
$$\eta = 0.85 \quad (2)$$

The power of the source is then:

$$q_{arc} = U \cdot I \cdot \eta = 14.8 \cdot 120 \cdot 0.85$$
$$q_{arc} = 1509.6 \text{ W} \quad (3)$$

3.3 Boundary Conditions and Heat Equation

The ANSYS simulation solves the thermal partial differential equation over a finite number of elements. At the start of the process, the initial temperature of the workpiece is set to 298.15 K, which is also the ambient temperature. The boundary conditions on the upper surface of the workpiece are defined by the following expression [25]:

$$k\frac{\partial T}{\partial z} = q_{arc} - a_c(T - T_a) - \varepsilon\sigma_B(T^4 - T_a^4) \quad (4)$$

This equation represents the thermal energy balance in the system, where heat conduction $k\frac{\partial T}{\partial z}$ is balanced by the heat input from the arc source q_{arc}, the heat loss due to convection $-a_c(T - T_a)$, and the heat loss due to radiation $-\varepsilon\sigma_B(T^4 - T_a^4)$.

Where: a_c - the heat transfer coefficient between the material and air (W/m^2·K); T_a - the air temperature (K); ε - the material's emissivity factor; σ_b - the Boltzmann constant, 5.67×10^{-8} (W/m^2·K^4); k - thermal conductivity of the material (W/m·K).

The boundary conditions at the edges of the model are expressed by the following equation [25]:

$$k\frac{\partial T}{\partial z} = -a_c(T - T_a) - \varepsilon\sigma_B(T^4 - T_a^4) \quad (5)$$

4 RESULTS OF ANSYS SIMULATION AND DISCUSSION

Fig. 12 shows the temperature distribution from the ANSYS simulation at a specific moment during welding for a plate with dimensions of 75 × 195 mm. The maximum measured temperature on the smaller plate is T_{max} = 728.93 °C, while on the larger plate, the maximum temperature is predictably lower, at T_{max} = 610.45 °C. As stated, the temperature was measured in thermal probes at the exact locations where the temperature was measured using thermocouples in the experiment.

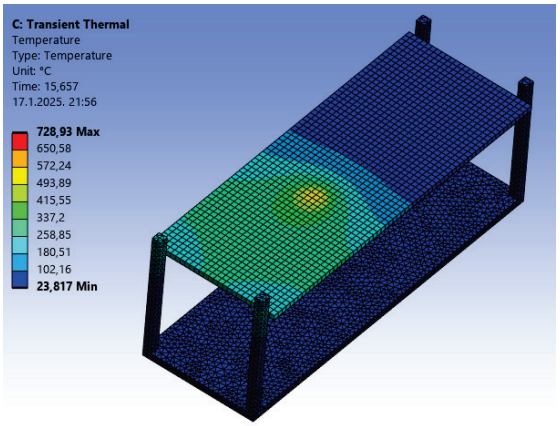


Figure 12 Temperature distribution in a 75 × 195 mm plate welding simulation

In contrast, Fig. 13 also displays the temperature distribution for a plate with 150 × 195 mm dimensions.

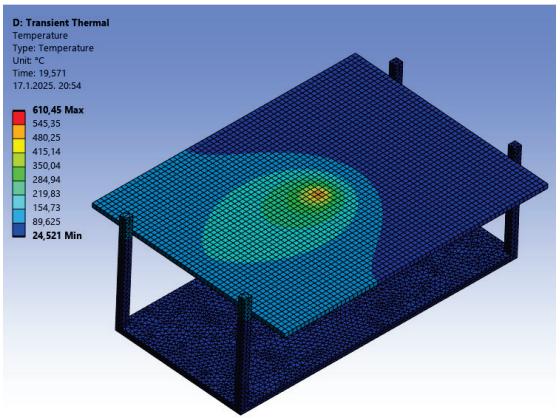


Figure 13 Temperature distribution in a 150 × 195 mm plate welding simulation

Fig. 14 presents the temperature distribution over time at measured points on the plate with dimensions of 150 × 195 mm.

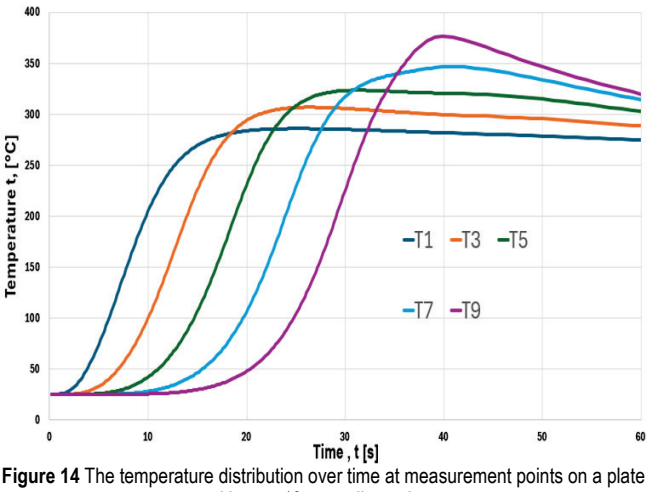


Figure 14 The temperature distribution over time at measurement points on a plate with 75 × 195 mm dimensions

The temperatures were measured using thermal probes in ANSYS at the exact locations where thermocouples were fixed in actual heat input during the experiment. The

temperature values reach their maximum at a certain point, after which a gradual cooling occurs. The highest temperature is recorded on Thermocouple 9, while the other thermocouples register progressively lower temperatures depending on their distance from the heat source. The characteristics of the graph show that the temperature cycles are similar, with a clear gradient toward the edges of the plate, which is expected due to heat dissipation. A smaller plate size allows for faster attainment of higher temperatures, and the temperature peaks are more pronounced due to the limited heat dissipation capacity compared to larger plates.

In the ANSYS simulation for the smaller plate, the maximum temperature reached within the temperature cycle of the thermal probe is T_{max} = 380.37 °C (Fig. 15) at thermal probe T9/T10. For the larger plate, the maximum temperature is also reached at thermal probe T9/T10. Still, it is T_{max}=188,51 °C (Fig. 14). By comparing the maximum temperatures measured in the experiment (smaller plate: T_{max} = 180.15 °C, larger plate: T_{max} = 390 °C) with the maximum temperatures obtained in the simulation, it can be concluded that the simulation agrees very well with the experimental values. The difference in maximum temperatures is expected due to hidden heat dissipation during the experiment, which was not accounted for in the simulation.

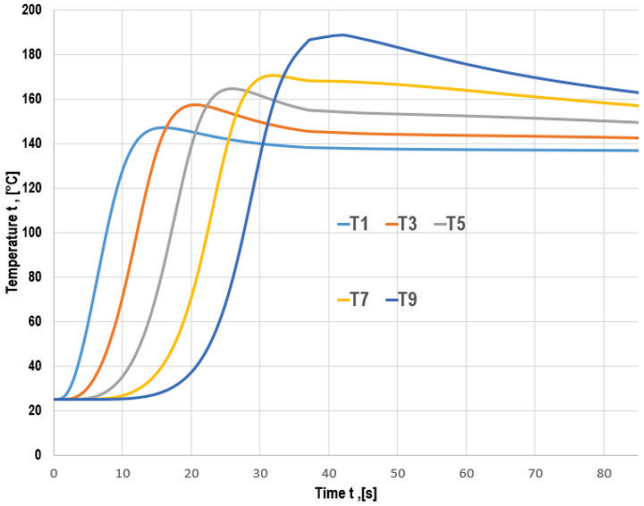


Figure 15 The temperature distribution over time at measurement points on a plate with 150 × 195 mm dimensions

5 CONCLUSIONS

This study confirmed the significant influence of copper plate dimensions on the temperature distribution during TIG welding. Experimental measurements and numerical simulations showed that smaller plates attain higher peak temperatures near the weld zone, while larger plates facilitate a more uniform heat distribution due to their increased heat dissipation capacity. It was also found that plate geometry and material properties strongly affect heat transfer dynamics: smaller plates exhibit steeper temperature gradients, whereas larger plates mitigate localized temperature peaks. Notably, the temperature gradient was approximately proportional to the plate dimensions.

Experimental results were consistent with the numerical simulations performed in ANSYS, with the discrepancy between measured and simulated temperatures remaining below 5%. This finding further corroborates the reliability of the employed methodology for analyzing thermal processes. The insights gained from this research are valuable for optimizing welding parameters and designing reliable joints in high thermal conductivity materials such as copper.

Future investigations could focus on a more detailed analysis of heat source distribution, allowing for more precise modeling and validation against experimental data. The heat source could be modeled according to Goldak's model [25] to more accurately represent heat flow. Additionally, implementing a finer computational mesh, particularly in the welding zone, should be considered in numerical simulations to achieve higher accuracy in local temperature gradients. These measures would further advance the understanding of thermal processes in TIG welding of copper.

Acknowledgements

This research was funded by: University North support for scientific research and artistic work in 2024 – UNIN-TEH-24-1-19.

6 REFERENCES

[1] Alfirević, I., Bjelovučić, D., Budin, I., Matijašević, B., Modlić, B., Mulc, A., Šikić, Z., & Taboršak, D. (1996). *Inženjerski priručnik IP 4 - Proizvodno strojarstvo, sv. 1 - Materijali*. Zagreb, Croatia: Školska knjiga. (in Croatian)

[2] American Welding Society. (1987). *Handbook for GTAW – Gas Tungsten Arc Welding*. Miami, FL: American Welding Society.

[3] American Welding Society. (2007). *Copper and alloys*. Miami, FL: American Welding Society.

[4] Rosenthal, D. (1946). The theory of moving source of heat and its application to metal transfer. *Transactions of the ASME, 43*, 849-866. https://doi.org/10.1115/1.4018624

[5] Christensen, N., Davies, V., & Gjermundsen, K. (1965). Distribution of temperatures in arc welding. *British Welding Journal, 12*, 54-75.

[6] Eagar, T. W., & Tsai, N. S. (1983). Temperature fields produced by traveling distributed heat sources. *Welding Journal, 62*, 346-355.

[7] Komanduri, R., & Hou, Z. B. (2000). Thermal analysis of the arc welding process: part I. General solutions. *Metallurgical and Materials Transactions B, 31*, 1353-1370. https://doi.org/10.1007/s11663-000-0022-2

[8] Nguyen, N. T., Ohta, A., Matsuoka, K., et al. (1999). Analytical solutions for transient temperature of a semi-infinite body subjected to 3-D moving heat sources. *Welding Research Supplement, I*, 265-274.

[9] Samardžić, I., Čikić, A., & Dunđer, M. (2015). Analytical heat conduction models at arc fusion welding. *Technical Gazette, 22*(6), 1641-1647. https://doi.org/10.17559/TV-20150220144751

[10] Ueda, Y., & Yamakawa, T. (1971). Analysis of thermal elastic-plastic stress and strain during welding by finite element method. *JWRI, 2*(2), 90-100.

[11] Hibbitt, H. D., & Marcal, P. V. (1973). A numerical, thermo-mechanical model for the welding and subsequent loading of a fabricated structure. *Computers & Structures, 3*, 1145-1174. https://doi.org/10.1016/0045-7949(73)90043-6
[12] Marcal, P. V. (1974). Weld problems. In *Structural Mechanics Computer Programs* (pp. 191-206). Charlottesville, VA: University Press.
[13] Goldak, J. A., Bibby, M., Moore, J., House, R., & Patel, B. (1986). Computer modelling of heat flow in welds. *Metallurgical Transactions B, 17B*, 587-600. https://doi.org/10.1007/BF02670226
[14] Karlsson, L. (1993). Thermomechanical finite element models for calculation of residual stresses due to welding. In Hauk, E. et al. (Eds.), *Residual Stresses* (p. 33). DGM Informationsgesellschaft Verlag.
[15] Lindgren, L.-E. (2001). Finite element modelling and simulation of welding – Part 1: Increased complexity. *Journal of Thermal Stresses, 24*, 141-192. https://doi.org/10.1080/01495730150500442
[16] Lindgren, L.-E. (2001). Finite element modeling and simulation of welding – Part 2: Improved material modeling. *Journal of Thermal Stresses, 24*, 195-231. https://doi.org/10.1080/014957301300006380
[17] Barac, A., Živić, M., Holik, M., Končić, R., & Samardžić, I. (2019). Comparison of experimental and analytical solutions of temperature field in a submerged arc welding. In *Engineering technologies in the manufacturing of welded constructions and products, SBW 2019*. Retrieved from https://dtzsb.unisb.hr/wp-content/uploads/radovi_2019/14_Zivic.pdf
[18] Fluke Corporation. (n.d.). *1586A Super-DAQ Precision Temperature Scanner*. Retrieved from https://www.fluke.com/en-us/product/calibration-tools/temperature-calibrators/fluke-calibration-1586a
[19] IEC 60584-1:2013. *Thermocouples – Part 1: EMF Specifications and Tolerances*.
[20] EURAMET Calibration Guide No. 11:2011. *Guidelines on the Calibration of Temperature Indicators and Simulators by Electrical Stimulation and Measurement*.
[21] Baker Hughes. (n.d.). *Premium temperature calibrators*. Retrieved from https://www.bakerhughes.com/druck/test-and-calibration-instrumentation/portable-pressure-temperature-calibrators-2/premium-temperature-calibrators
[22] Additel. (n.d.). *875 Dry well temperature calibrator*. Retrieved from https://additel.com/product-detail.html/875-dry-well-temperature-calibrator
[23] Bionda, F. (2018). *Automated TIG welding of stainless steel pipes* (Master's thesis). University of Zagreb, Faculty of Mechanical Engineering and Naval Architecture.
[24] Traidia, A. (2011). *Multiphysics modeling and numerical simulation of GTA weld pools* (Doctoral dissertation, Ecole Polytechnique X). (Published in English).
[25] Karkhin, V. A. (2019). *Thermal processes in welding*. Springer. https://doi.org/10.1007/978-981-13-5965-1

Authors' contacts:

Mihael Zelić, mag. ing. mech. & mag. educ. phys.
(Corresponding author)
Mačkovec, 40000 Čakovec, Croatia
E-mail: mihael.zelic@gmail.com

Matija Bušić, Assistant Professor
University North,
Ul. 104. Brigade 3, 42000 Varaždin, Croatia
E-mail: mbusic@unin.hr

Tomislav Veliki, Assistant Professor
University North,
Ul. 104. Brigade 3, 42000 Varaždin, Croatia
E-mail: tveliki@unin.hr

ISSN 1846-6168 (Print), ISSN 1848-5588 (Online)
https://doi.org/10.31803/tg-20250414132406

Original scientific paper
Received: 2025-04-14, Accepted: 2025-06-23

Modelling the Effects of Grinding Wheel Parameters on Surface Roughness during Grinding of D2 Steel

Aleksandar Milosevic, Goran Simunovic, Vitalii Ivanov, Mario Sokac, Vladimir Kocovic, Milana Ilic Micunovic, Djordje Vukelic*

Abstract: The research focused on the effects of abrasive type, grain size, and wheel structure on surface roughness during external cylindrical longitudinal grinding of D2 steel. Experimental research was conducted using a full factorial design. The measured surface roughness values, depending on the combination of input parameters, ranged from 0.22 to 1.31 μm, corresponding to ISO roughness grades N4 to N7. Among the factors examined, grain size had the most significant impact on surface roughness, followed by the type of abrasive, while the grinding wheel structure had the least effect. The best surface roughness is obtained by utilizing corundum with the highest aluminium oxide content as an abrasive, along with the smallest grain size and a dense structure of the grinding wheel. Additionally, modelling of the grinding process was performed. The validity of the obtained regression equation for predicting surface roughness was confirmed through five additional verification experiments, which supported the accuracy of the modelling.

Keywords: abrasive type; grain size; grinding; structure

1 INTRODUCTION

In manufacturing, grinding has many applications, such as finishing, slitting, parting, descaling, deburring, sharpening, etc. Grinding is a crucial finishing operation that ensures high dimensional accuracy and high surface quality. This process allows for the machining of workpieces high accuracy and precision [1]. However, grinding is complex, influenced by numerous factors and their interactions, including the type of grinding, the machine tools, the grinding wheel, fixtures, and grinding parameters [2]. For successful grinding, it is essential to address all these influencing factors [3]. Additionally, grinding can effectively process workpieces made from various materials, each with distinct mechanical, physical, and thermal properties [4]. One such material is D2 steel, which is known for its high wear resistance and good corrosion resistance. However, it presents challenges during machining, especially after heat treatment, as it becomes significantly harder.

In recent years, the grinding of D2 steel has been examined from various perspectives. Karabelchtchikova et al. [5] developed a mathematical model to describe the effects of multipass grinding on residual stress distribution. This model aids in predicting grinding dynamics and the impact of grinding on material characteristics. Mohanasundararaju et al. [6] utilized response surface methodology to predict surface roughness. Their experiments, conducted on a transverse grinding machine, revealed that surface roughness improves with increasing wheel speed, work speed, and dress depth. In contrast, surface roughness decreases with lower traverse speed and feed rate. Yan et al. [7] explored how grinding parameters and wheel grain size influence the formation of white layers. They found that a finer grain size of the wheel makes it more likely for both white and dark layers to form. Additionally, the occurrence of these layers is more probable with increased wheel rotation speed, workpiece speed, and depth of cut. Prabhu and Vinayagam [8] analysed changes in fractal dimension based on the smoothness of the ground surface. They measured surface roughness for carbon nanotube-based nanofluids during the grinding process. Kamely et al. [9] modelled the surface roughness of the grinding process based on feed and working speed. Surface roughness improves as both feed and working speed decrease. Yan et al. [10] developed a finite element simulation model to analyse grinding with alumina abrasive grit. The model assessed material deformation, forces, critical depth of cut, temperature, and material removal rate. Sethuramalingam & Vinayagam [11] studied the effects of grinding parameters using a carbon nanotube-based grinding wheel on surface roughness. They found that cutting speed had the most significant impact on surface roughness, while the depth of cut had the least influence. Babu et al. [12] studied the impact of grinding parameters on surface roughness when carbon nanotubes were added to oil, using the Taguchi method. The findings showed that surface roughness improved with an increase in feed and a decrease in the depth of cut. Khan et al. [13] conducted a comparison of surface quality, surface temperature, and normal force after grinding under three different conditions: dry, wet cooling, and minimum quantity lubrication (MQL). Their findings indicated that MQL outperformed both dry and wet conditions due to its superior penetration ability and enhanced heat dissipation properties. Kannan and Arunachalam [14] examined the relationship between grinding forces and the surface images of ground materials during the dressing process. Their results indicated a strong correlation between the signals from grinding forces and the images of ground surfaces, which can be used to assess the level of wear deterioration. Isti Nugroho et al. [15] examined the characteristics of chips and surface roughness. The results indicated that increasing the feed rate led to greater chip width and higher surface roughness values. Hood et al. [16] studied the performance of electroplated CBN grinding points with varying grit sizes. The most significant factors impacting surface roughness were the number of effective cutting grits and the active area. Gopan et al. [17] examined the effects of table speed, cross-feed, and depth of cut on surface roughness and cutting force. To

reduce surface roughness and cutting force, they utilized artificial neural networks and particle swarm optimization. Ben Fathallah et al. [18] investigated the effects of a sol-gel abrasive wheel and cryogenic cooling on surface integrity. Their study revealed that using a combination of an Al_2O_3 abrasive wheel and soluble oil as a cooling method improved fatigue life when controlled grinding was employed. Chaudhari et al. [19] compared the surface integrity of conventional dry grinding with ultrasonic vibration-assisted dry grinding. The ultrasonic vibration-assisted dry grinding produced less variation in microhardness and resulted in a lower surface roughness. Azami et al. [20] studied the cutting forces and surface roughness during MQL grinding by incorporating MoS_2 and CuO nanoparticles into vegetable-based oils. The addition of these nanoparticles reduced cutting forces and improved surface quality. Chaudhari et al. [21] studied the effect of worktable feed and ultrasonic vibration amplitude on grinding outcomes. A decrease in worktable feed and an increase in ultrasonic vibration lead to reductions in grinding forces, surface roughness, and grinding temperature. Sharma et al. [22] studied grinding forces, specific grinding energy, and surface roughness under dry, wet, and cryogenic conditions. The cryogenic environment reduces grinding forces, lowers specific energy, and improves surface roughness. Mouralova et al. [23] analysed chips after grinding using an Al_2O_3 grinding wheel. They detected four different types of chips post-grinding. Zhang et al. [24] examined how grinding parameters affect surface roughness. Their findings revealed that the surface roughness measured perpendicular to the grinding direction was significantly greater than that measured parallel to it. They also found that increasing the grinding speed or reducing the grinding depth led to a decrease in surface roughness.

The grinding process and its output parameters are influenced by a variety of parameters, many of which are non-linear. Consequently, the success of the grinding process often depends on the operator's experience and expertise. Previous research has typically focused on a few key parameters, including grinding parameters, grinding wheel characteristics, and dressing parameters. To enhance the grinding process, it is essential to model and optimize it by considering as many parameters as possible. This can be achieved by developing appropriate analytical models and identifying optimal grinding conditions.

The objective of this research, unlike previous studies, is to model the process of external cylindrical longitudinal grinding using a straight grinding wheel. Specifically, we aim to develop a mathematical model that includes factors such as abrasive type, grain size, and the structure (porosity) of the grinding wheel. This model will help predict the surface roughness after the grinding operation. Additionally, the study will examine how the grinding wheel factors interact with surface roughness. We also aim to evaluate the impact of each factor and their interactions on the surface roughness. Furthermore, this research seeks to identify and validate, through confirmation experiments, the optimal parameters for the grinding process in order to minimize surface roughness and ensure its repeatability.

2 METHODOLOGY

The methodology according to which the research was conducted consists of several steps (Fig. 1).

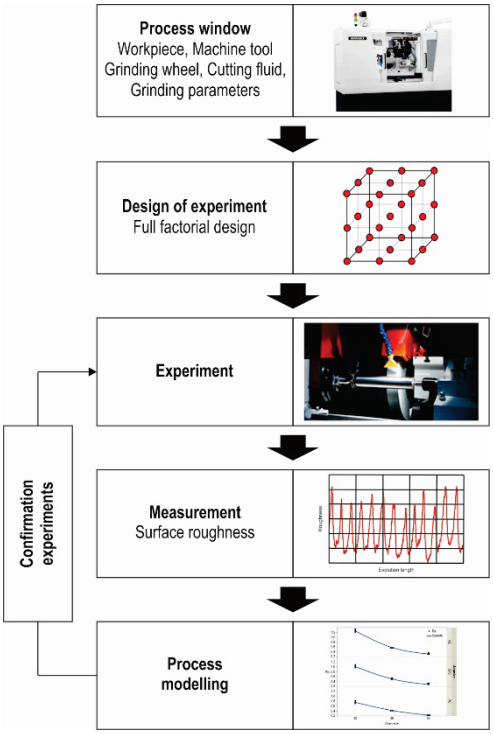


Figure 1 Methodology

Experimental research was conducted on workpieces ∅50 × 350 mm, which had been machined through turning to achieve an initial surface roughness of Ra = 8.3 µm. These workpieces were made from D2 tool steel, which has the following chemical composition: 1.50% Carbon (C), 0.60% Manganese (Mn), 0.60% Silicon (Si), 1.00% Cobalt (Co), 12.00% Chromium (Cr), 1.00% Molybdenum (Mo), 1.10% Vanadium (V), 0.03% Phosphorus (P), 0.30% Nickel (Ni), 0.25% Copper (Cu), 0.03% Sulfur (S), with the balance being Iron (Fe). The key mechanical, physical, and thermal characteristics of D2 tool steel are as follows: Hardness = 62 HRC, Elastic modulus = 200 GPa, Yield strength = 1300 MPa, Poisson's ratio = 0.29, Thermal conductivity = 20 W/m·K, and Thermal expansion = 10.4×10^{-6} 1/°C.

External cylindrical longitudinal grinding was conducted on a grinding machine using the following parameters: a grinding wheel cutting speed of v_c = 25 m/s, a workpiece speed of v_w = 0.2 m/s, a longitudinal feed of f = 25 mm, and a depth of cut of a_p = 0.01 mm. Grinding parameters and the characteristics of the grinding wheel should be selected according to the manufacturer's recommendations.

During external cylindrical grinding, the workpieces are fixed between centers and rotated using the driving plate. The grinding wheel and workpiece rotate in the same direction.

The experiments utilized a grinding wheel with the following dimensions: a diameter of D = 300 mm, a width of T = 50 mm, and a hole diameter of H = 127 mm. Grinding was performed using straight wheels characterized by soft

hardness (designation I) and a ceramic bond type (designation V). During the experimental research, three parameters of the grinding wheels were varied: abrasive type, grain size, and structure (porosity). The levels of the input parameters are detailed in Tab. 1.

Table 1 Parameters of grinding wheels levels

Parameter		Abrasive	Grain size	Structure
Level 1	Designation	Corundum 21A	30	4
	Main characteristic	90 % Al_2O_3	0.710–0.590 μm	Dense
Level 2	Designation	Corundum 41A	46	8
	Main characteristic	94 % Al_2O_3	0.420–0.350 μm	Medium
Level 3	Designation	Corundum 61A	60	12
	Main characteristic	98 % Al_2O_3	0.297–0.250 μm	Very open (porous)

During machining, the synthetic cutting fluid Syntilo 2000 was applied to enhance grinding efficiency. This fluid not only extends tool life but also ensures a high-quality surface roughness and maintains dimensional accuracy.

The experimental research was conducted according to a full factorial design, which allows the investigation of all possible combinations of input parameters. Since all parameters were defined as categorical with three levels each, a total of 27 experimental runs (3×3×3) were conducted.

After conducting experimental research, the measurement of surface roughness, specifically the arithmetical mean roughness value, was conducted using a Mitutoyo surface roughness measuring instrument. The probe tip radius was set at 2 μm, with a sampling length of 0.8 mm and an evaluation length of 4 mm. Measurements were taken at ten different locations on the workpiece, and the average surface roughness value was calculated to aid in model development.

In the final step, we performed an analysis of surface roughness generation and its mathematical modelling using multiple regression analysis. This approach quantifies the relationship between input parameters – such as abrasive type, grain size, and structure – and the response parameter, which is surface roughness. The research employed a full factorial design for the experiments. A mathematical model was created to best fit the measured results, represented as a polynomial with coefficients determined by the measured values of surface roughness. The resulting regression model can be expressed as:

$$Ra = f(x_1, x_2, x_3), \tag{1}$$

where x_1 is abrasive type, x_2 is grain size and x_3 is structure. This leads to the model equation:

$$Ra = \beta + \sum_{i=1}^{3} \beta_i \cdot x_i + \sum_{i=1}^{3} \sum_{j=1}^{3} \beta_{ij} \cdot x_i \cdot x_j + \varepsilon, \tag{2}$$

where β is coefficients and ε is residual.

At the end, using this set of experimental parameters, we identified the conditions that yield the minimum value of surface roughness. We subsequently validated this minimum value through additional confirmation experiments.

3 RESULTS

3.1 Experimental Results

The results of the experiment are shown in Tab. 2. The results were randomized, which reduced the possibility of systematic bias and further increased the reliability of the data obtained.

Table 2 Results of the experiments

No.	Abrasive	Grain size	Structures	*Ra* (μm)
1	21A	46	12	0.58
2	41A	30	12	1.06
3	61A	30	8	0.76
4	61A	30	12	0.80
5	21A	46	8	0.57
6	21A	60	8	0.32
7	61A	46	12	0.43
8	61A	60	4	0.22
9	41A	30	4	0.97
10	41A	30	8	1.01
11	61A	60	12	0.24
12	21A	60	4	0.31
13	61A	30	4	0.71
14	21A	30	4	1.22
15	21A	30	8	1.27
16	61A	46	4	0.41
17	41A	60	12	0.29
18	21A	60	12	0.33
19	41A	46	4	0.48
20	21A	46	4	0.56
21	41A	60	4	0.27
22	61A	46	8	0.42
23	61A	60	8	0.23
24	21A	30	12	1.31
25	41A	46	12	0.51
26	41A	60	8	0.28
27	41A	46	8	0.50

3.2 Statistical Analysis

Statistical analysis of the measured results was conducted using a model that incorporates main effects and two-factor interactions. This approach enabled the development of a model for a detailed investigation and modelling of the complex relationships between input and output parameters.

Due to irregularities observed in the data distribution and significant deviations among individual values (as illustrated in Fig. 2a), a logarithmic transformation of *Ra* (ln (*Ra*)) was applied. This transformation facilitates a better interpretation of the results and the application of statistical methods. The data distribution after the transformation is presented in Fig. 2b.

The logarithmic transformation was chosen based on the findings of a Box-Cox analysis, which indicated an optimal lambda value of $\lambda = -0.13$ (Fig. 3). This value is close to zero, suggesting that a logarithmic transformation is the most suitable option for stabilizing variance and enhancing the

normality of the data. We also assessed other standard transformations, such as square root and reciprocal, but they did not yield better results regarding data distribution or model performance.

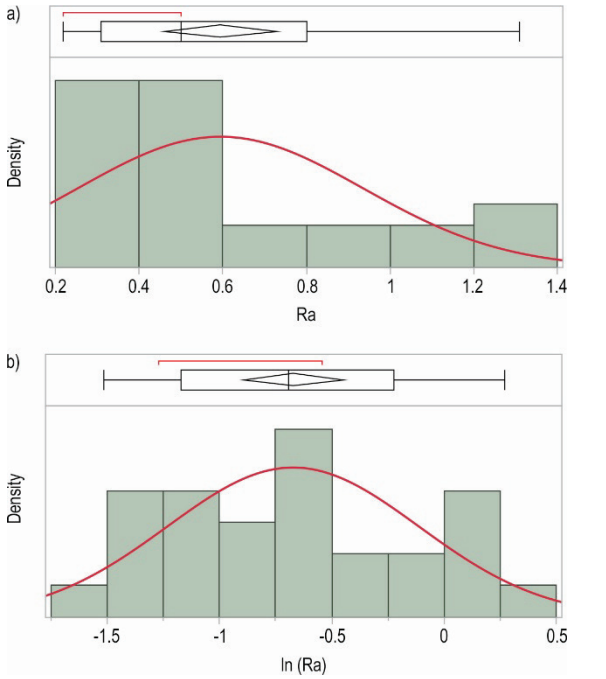


Figure 2 Distribution of surface roughness: a) before transformation b) after transformation

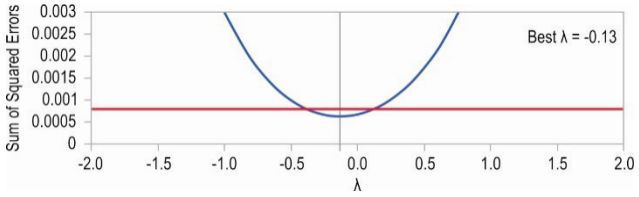


Figure 3 Box-Cox transformation

The assessment of the model's summary of fit predicting the ln (*Ra*) parameter is detailed in Tab. 3. The high coefficient of determination (R^2 = 0.999685) indicates an excellent alignment between the actual and predicted values, with the model accounting for over 99.9% of the total variance in the response parameter. The adjusted coefficient of determination (*Adjusted* R^2 = 0.999489) further verifies the model's reliability, taking into account the number of input parameters used. Additionally, the low root mean square error (*RMSE* = 0.012675), when compared to the mean of the dependent parameter (*Mean of Response* = –0.67167), demonstrates the model's high precision and low average error in its predictions.

Table 3 Summary of fit

Parameter	ln (*Ra*)
R^2	0.999685
R^2 *Adjusted*	0.999489
RMSE	0.012675
Mean of Response	–0.67167

Tab. 4 provides the analysis of variance for the specified regression model. The *F*-ratio obtained is 5085.076, which indicates the statistical significance of the model in explaining the variance of the output parameter. This high *F*-ratio suggests that the model's total effects are significantly greater than the random variations (error). Furthermore, the extremely low *p*-value (< 0.0001) confirms that there is at least one statistically significant effect among the independent parameters. The total sum of squares is 8.1724370, with the model accounting for almost the entire variance (*Sum of Squares Model* = 8.16987), while the error contributes minimally (*Sum of Squares Error* = 0.0025706). This further reinforces the model's effectiveness.

Table 4 Analysis of variance

Source	*DF*	*Sum of Squares*	*Mean Square*	*F Ratio*	*Prob > F*
Model	10	8.1698664	0.816987	5085.076	<0.0001*
Error	16	0.0025706	0.000161		
C. Total	26	8.1724370			
Legend: * - statistically significant ($p < 0.05$, significance threshold)					

Tab. 5 shows the estimated input parameters of the model, ranked based on their LogWorth values. LogWorth represents a measure of the statistical significance of each parameter, with higher values indicating a greater impact of the parameter on the model.

Table 5 Effect summary

Source	*LogWorth*	*p*-value
Grain size	27.693	0.00000
Abrasive	19.331	0.00000
Abrasive×Grain size	8.832	0.00000
Structures	7.988	0.00000

Fig. 4 presents a visual assessment of the model's fit by comparing actual and predicted output parameters. The analysis of this plot reveals a strong alignment between the actual and predicted values, as the points are closely clustered around the diagonal line. This diagonal line signifies the ideal scenario in which predicted values match the actual values, showcasing the model's precision in its predictions.

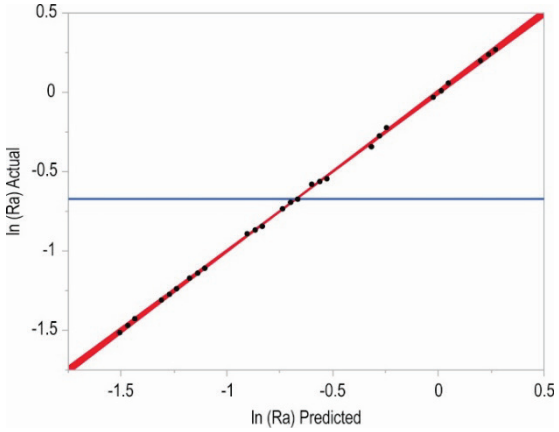


Figure 4 Actual by predicted plot

Fig. 5 displays the residuals plotted against the predicted values. The residuals are evenly distributed around the

horizontal line at zero, showing no obvious patterns. This suggests that the model fits the data well and that there are no systematic trends in the predictions.

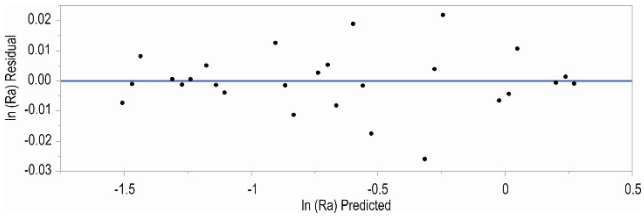


Figure 5 Residual by predicted plot

The normality of the residuals in the model was evaluated using the residuals' normal quantile plot, shown in Fig. 6. The residuals fall within the confidence interval (*CI*) and align with the diagonal line, suggesting that they are normally distributed. Since the points remain within the interval, we can conclude that there are no outliers indicating any irregularities in the data. This plot indicates that the model meets the assumption of normality for the residuals, confirming its validity and ensuring the reliability of the results obtained.

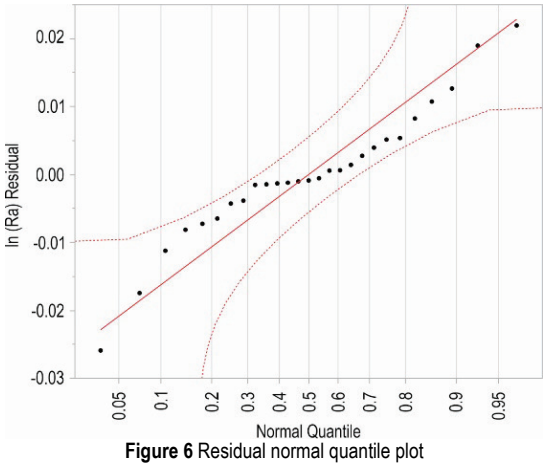


Figure 6 Residual normal quantile plot

Tab. 6 presents the standardized estimates of the regression coefficients regarding the influence of various input parameters on surface roughness. Standardized parameter estimates indicate the relative impact of different input parameters on the dependent parameter, facilitating comparisons irrespective of unit differences. The *Prob>|t|* value reflects the statistical significance of each input parameter, with a value less than 0.005 indicating statistical significance. The t-ratio values demonstrate the relationship between the parameter estimate and its standard error; higher t-ratios signify a stronger influence of the predictor in the model.

Positive and negative estimate values reveal the extent and direction of the effects of individual parameters and their interactions on the output. Positive values for the following parameters - Abrasive [21A], Abrasive [41A], Grit Size [30], Structures [8], Structures [12], Abrasive [21A] × Grain Size [30], Abrasive [41A] × Grain Size [30], Abrasive [41A] × Grain Size [60], Abrasive [61A] × Grain Size [46], and Abrasive [61A] × Grain Size [60] - indicate that applying these parameters increases the ln (Ra) value, which corresponds to a negative effect on the quality of the machined surface.

Table 6 Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	-0.671666	0.002439	-275.34	<0.0001*
Abrasive [21A]	0.1829939	0.00345	53.05	<0.0001*
Abrasive [41A]	0.0180134	0.00345	5.22	<0.0001*
Abrasive [61A]	-0.201007	0.00345	-58.27	<0.0001*
Grain size [30]	0.6611754	0.00345	191.66	<0.0001*
Grain size [46]	-0.038356	0.00345	-11.12	<0.0001*
Grain size [60]	-0.62282	0.00345	-180.54	<0.0001*
Structures [4]	-0.03657	0.00345	-10.60	<0.0001*
Structures [8]	0.0016314	0.00345	0.47	0.6427
Structures [12]	0.0349391	0.00345	10.13	<0.0001*
Abrasive [21A]×Grain size [30]	0.0634617	0.004879	13.01	<0.0001*
Abrasive [21A]×Grain size [46]	-0.035194	0.004879	-7.21	<0.0001*
Abrasive [21A]×Grain size [60]	-0.028268	0.004879	-5.79	<0.0001*
Abrasive [41A]×Grain size [30]	0.005064	0.004879	1.04	0.3147
Abrasive [41A]×Grain size [46]	-0.008145	0.004879	-1.67	0.1144
Abrasive [41A]×Grain size [60]	0.0030814	0.004879	0.63	0.5366
Abrasive [61A]×Grain size [30]	-0.068526	0.004879	-14.05	<0.0001*
Abrasive [61A]×Grain size [46]	0.0433392	0.004879	8.88	<0.0001*
Abrasive [61A]×Grain size [60]	0.0251865	0.004879	5.16	<0.0001*
Legend: * - statistically significant ($p < 0.05$, significance threshold)				

Through statistical analysis of the chosen model, a regression equation was derived to describe the surface roughness of the machined surface:

$$\begin{aligned}
&\ln(Ra) \\
&= -0.671666049 + Match(Abrasive)\begin{pmatrix} \text{"21A"} & \rightarrow & 0.18299 \\ \text{"41A"} & \rightarrow & 0.01801 \\ \text{"61A"} & \rightarrow & -0.20101 \\ else & \rightarrow & . \end{pmatrix} \\
&\quad + Match(Abrasive)\begin{bmatrix} \text{"21A"} & \rightarrow & Match(Grain\ size)\begin{pmatrix} \text{"30"} & \rightarrow & 0.06346 \\ \text{"46"} & \rightarrow & -0.01352 \\ \text{"60"} & \rightarrow & -0.02827 \\ else & \rightarrow & . \end{pmatrix} \\ \text{"41A"} & \rightarrow & Match(Grain\ size)\begin{pmatrix} \text{"30"} & \rightarrow & 0.00506 \\ \text{"46"} & \rightarrow & -0.00814 \\ \text{"60"} & \rightarrow & 0.00308 \\ else & \rightarrow & . \end{pmatrix} \\ \text{"61A"} & \rightarrow & Match(Grain\ size)\begin{pmatrix} \text{"30"} & \rightarrow & -0.06852 \\ \text{"46"} & \rightarrow & 0.04334 \\ \text{"60"} & \rightarrow & 0.02519 \\ else & \rightarrow & . \end{pmatrix} \end{bmatrix} \\
&\quad + Match(Grain\ size)\begin{pmatrix} \text{"30"} & \rightarrow & 0.66117 \\ \text{"46"} & \rightarrow & -0.03835 \\ \text{"60"} & \rightarrow & -0.62282 \\ else & \rightarrow & . \end{pmatrix} \\
&\quad + Match(Structure)\begin{pmatrix} \text{"4"} & \rightarrow & -0.03657 \\ \text{"8"} & \rightarrow & 0.00163 \\ \text{"12"} & \rightarrow & 0.03494 \\ else & \rightarrow & . \end{pmatrix}
\end{aligned} \tag{3}$$

Fig. 7 shows a statistically significant interaction between abrasive type and grain size in relation to ln (*Ra*). The interaction analysis indicates that the effect of one factor on *Ra* depends on the value of the other factor.

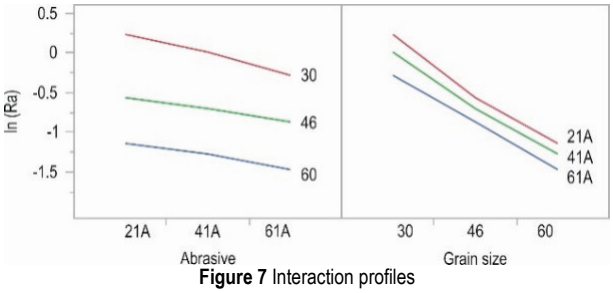


Figure 7 Interaction profiles

Fig. 8 presents the prediction profiler, which illustrates the influence of each input parameter on the predicted surface roughness. The prediction profiler provides insight into how variations in the input parameters affect the output of the regression model.

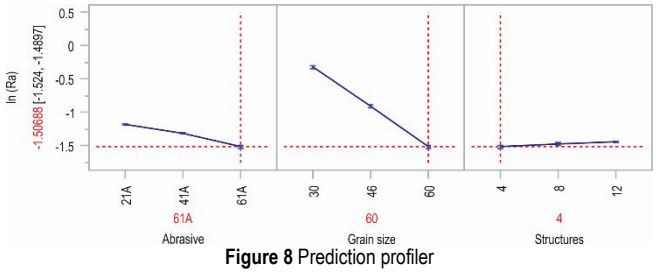


Figure 8 Prediction profiler

3.3 Confirmation Experiments

The regression model was confirmed by 5 additional experiments carried out at the point in the experimental space where the model predicts the lowest *Ra*, i.e. the highest quality of the machined surface. The confirmation experiments were carried out with the following input parameters: abrasive 61A, grain size 60 and structure 4. The experiments were repeated ten times to evaluate the stability and repeatability of the process, and the results obtained were compared with the values predicted by the regression model.

The results of the confirmation experiments are shown in Fig. 9 as box plots. The red horizontal line represents the predicted surface roughness value *Ra* = 0.2216, while the green lines indicate the boundaries of the 95% *CI* (lower *CI* = 0.2178, upper *CI* = 0.2254). The blue line within the box plots connects the mean values of the measurements for each experiment. The obtained results show that both the median and mean fall within the *CI* (95%), indicating the high accuracy and reliability of the developed model, as well as the stability and repeatability of the process. The presence of minor deviations in individual measurements, i.e., statistically acceptable outliers outside the *CI*, can be considered a result of inherent data variability. Based on the obtained results, it can be concluded that the regression model was successfully validated through additional, independent experiments.

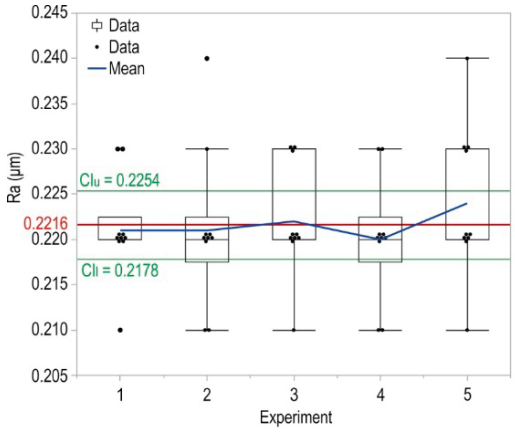


Figure 9 Results of the confirmation experiments

4 DISCUSSION

The results of the statistical analysis and the analysis of variance show a high precision and statistical significance of the developed model. Based on the evaluation of the input parameters, it can be concluded that all main effects as well as the interaction between abrasive type and grain size are statistically significant, while other two-factor interactions are not. The parameter estimation of the developed model showed that grain size and abrasive type are the most influential factors affecting surface roughness.

The greatest effect on *Ra* is attributed to grain size, with a smaller grain size leading to a better surface roughness. Coarse abrasives, which are characterized by larger grain sizes, generate more heat and a rougher surface on the workpiece. In contrast, finer abrasives with smaller grain sizes create multiple contact between the grinding wheel and the workpiece. This results in a more uniform scratch pattern, reduces the peak-to-valley height and ultimately improves the surface roughness.

The best surface quality was achieved with the abrasive 61A, which contains the highest proportion of aluminium oxide, while the poorest quality was achieved with the abrasive 21A, which has the lowest aluminium oxide content. As the aluminium oxide content in abrasives increases, the hardness of the abrasive also increases. This leads to slower wear, which ultimately improves the surface roughness. In addition, abrasives with a higher aluminium oxide content are more resistant to stress, which ensures more even wear and further improves the quality of the machined surface.

The grinding wheel with dense structure showed a better surface quality compared to medium and porous structure. Reducing the porosity of the grinding wheel (fewer pores) increases the contact area between the wheel and the workpiece. This can lead to higher contact pressures and forces as well as insufficient heat dissipation in the cutting zone, which can result in poorer surface roughness. However, careful selection of grinding parameters combined with adequate cooling and lubrication can mitigate these negative effects. In addition, grinding wheels with lower porosity provide less space for the accumulation of chips (dust particles), which can further hinder the achievement of the desired surface roughness. Nevertheless, properly aligned coolant nozzles and adequate coolant pressure can effectively flush the cutting zone and remove chips. Grinding wheels with lower porosity also tend to wear more slowly. Therefore, a wheel with lower porosity can only slightly improve the surface roughness.

The statistically significant interaction between abrasive type and grain size revealed that smaller grain sizes result in reduced surface roughness, regardless of the abrasive used. However, the extent of this improvement varies depending on the abrasive type. The highest surface roughness was recorded for the combination of 21A abrasive and grain size 30, while the lowest roughness was achieved using 61A abrasive with grain size 60.

Visual evaluation of model performance through actual by predicted, residual by predicted, and residual normal quantile plots showed no presence of systematic errors, improper residual distribution, or deviation from normality. The results of confirmation experiments demonstrated high

model accuracy. The predicted and actual values match within the 95% *CI*, further confirming the validity of the developed model and its capability to reliably predict surface roughness.

5 CONCLUSION

The results of the experimental research indicated that the input parameters significantly influence surface roughness. Specifically, surface roughness values varied from 0.22 μm to 1.31 μm. Among the factors studied, grain size had the most substantial effect on surface roughness, while the structure of the grinding wheel had the least impact. Additionally, the interaction between abrasive type and grain size also significantly affected surface roughness.

The results from the statistical analysis and confirmation experiments demonstrated that it is indeed possible to model surface roughness accurately and precisely based on the type of abrasive, grain size, and structure. This conclusion is supported by the fitting parameters obtained and the residuals that were analysed.

The conducted research has certain limitations that can guide future studies. The results obtained are applicable only within the specific conditions under which the experiments were conducted. To enhance the generalizability of the developed model, we plan to conduct experimental research on specimens with varying mechanical, physical, and thermal properties, as well as tools that have different geometric and technological characteristics. Additionally, future research will focus on modelling and optimizing a greater number of variables. This includes incorporating a wider range of input parameters, such as wheel speed, workpiece speed, feed rate, total depth of cut, and the number of passes, along with various output process parameters like dimensional accuracy, material removal rate, hardness, and residual stresses. We also intend to explore other modelling and optimization methods. Finally, future research will investigate wear mechanisms, chip morphology, and related factors.

6 REFERENCES

[1] Kishore, K., Sinha, M. K., Singh, A., Archana, Gupta, M. K., & Korkmaz, M. E. (2022). A comprehensive review on the grinding process: Advancements, applications and challenges. *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science, 236*(22), 10923–10952. https://doi.org/10.1177/09544062221110782

[2] Trung, D. D. (2021). Influence of Cutting Parameters on Surface Roughness in Grinding of 65G Steel. *Tribology in Industry*, 43(1), 167–176. https://doi.org/10.24874/ti.1009.11.20.01

[3] Buchmeister, B., Palcic, I., Ojstersek, R., Kovic, K., & Javernik, A. (2024). Multicriteria Optimisation of Machining Operations Using a Spreadsheet Model. *Tehnicki glasnik – Technical Journal, 18*(3), 445–451. https://doi.org/10.31803/tg-20240514150731

[4] Saric, T., Simunovic, G., Vukelic, D., Simunovic, K., & Lujic, R. Estimation of CNC Grinding Process Parameters Using Different Neural Networks. (2018). *Tehnicki Vjesnik – Technical Gazette, 25*(6), 1770–1775. https://doi.org/10.17559/tv-20180419095119

[5] Karabelchtchikova, O., Rivero, I. V., & Hsiang, S. M. (2008). Modeling of residual stress distribution in D2 steel via grinding dynamics using a second-order damping system. *Journal of Materials Processing Technology, 198*(1–3), 313–322. https://doi.org/10.1016/j.jmatprotec.2007.07.006

[6] Mohanasundararaju, N., Sivasubramanian, R., & Alagumurthi, N. (2008). Optimisation of work roll grinding using Response Surface Methodology and evolutionary algorithm. *International Journal of Manufacturing Research, 3*(2), 236. https://doi.org/10.1504/ijmr.2008.017418

[7] Yan, L., Jiang, F., & Rong, Y. M. (2011). Experimental Study on Microstructure Evolution in AISI D2 Steel Surface Grinding. *Key Engineering Materials*, 487, 126–130. https://doi.org/10.4028/www.scientific.net/kem.487.126

[8] Prabhu, S., & Vinayagam, B. K. (2011). Fractal dimensional surface analysis of AISI D2 Tool steel material with nanofluids in grinding process using atomic force microscopy. *Journal of the Brazilian Society of Mechanical Sciences and Engineering, 33*(4), 459–466. https://doi.org/10.1590/s1678-58782011000400009

[9] Kamely, M.A., Kamil, S.M., & Chong C.W. (2011). Mathematical modeling of surface roughness in surface grinding operation. *World Academy of Science, Engineering and Technology*, 80, pp. 1048–1051.

[10] Yan, L., Jiang, F., & Fang, C. F. (2014). Research on Material Removal Mechanism of Single Grit Cutting Based on FEM Simulation. *Advanced Materials Research*, 1017, 82–87. https://doi.org/10.4028/www.scientific.net/amr.1017.82

[11] Sethuramalingam, P., & Vinayagam, B. K. (2016). Multi Objective Optimization of Multi Wall Carbon Nanotube Based Nanogrinding Wheel Using Grey Relational and Regression Analysis. *Journal of the Institution of Engineers (India): Series C*, 97(3), 407–416. https://doi.org/10.1007/s40032-016-0238-9

[12] Babu, O. S., Mamilla, V. R., & Lakshmi Narayana Rao, G. (2016). Finite Element Analysis of Surface Grinding Process Using Nanofluids. Innovative Design and Development *Practices in Aerospace and Automotive Engineering*, 169–174. https://doi.org/10.1007/978-981-10-1771-1_21

[13] Khan, A. M., Jamil, M., Mia, M., Pimenov, D. Y., Gasiyarov, V. R., Gupta, M. K., & He, N. (2018). Multi-Objective Optimization for Grinding of AISI D2 Steel with Al_2O_3 Wheel under MQL. *Materials, 11*(11), 2269. https://doi.org/10.3390/ma11112269

[14] Kannan, K., & Arunachalam, N. (2018). Grinding wheel redress life estimation using force and surface texture analysis. *Procedia CIRP*, 72, 1439–1444. https://doi.org/10.1016/j.procir.2018.03.031

[15] Isti Nugroho, W., Nugroho, S., & Rusnaldy. (2018). Characterization chip formation of commercial steel materials at low speed cylindrical grinding processes. *MATEC Web of Conferences*, 159, 02023. https://doi.org/10.1051/matecconf/201815902023

[16] Hood, R., Medina Aguirre, F., Soriano Gonzalez, L., Novovic, D., & Soo, S. L. (2019). Evaluation of superabrasive grinding points for the machining of hardened steel. *CIRP Annals, 68*(1), 329–332. https://doi.org/10.1016/j.cirp.2019.04.090

[17] Gopan, V., Wins, K. L. D., Evangeline, G., & Surendran, A. (2020). Experimental Investigation for the Multi-objective Optimization of Machining Parameters on AISI D2 Steel Using Particle Swarm Optimization Coupled with Artificial Neural Network. *Journal of Advanced Manufacturing Systems, 19*(03), 589–606. https://doi.org/10.1142/s0219686720500286

[18] Ben Fathallah, B., Braham, C., & Sidhom, H. (2020). Combined effects of abrasive type and cooling mode on fatigue

resistance of AISI D2 ground surface. *International Journal of Fatigue*, 138, 105665. https://doi.org/10.1016/j.ijfatigue.2020.105665

[19] Chaudhari, A., Sharma, A., Awale, A. S., Yusufzai, M. Z. K., & Vashista, M. (2021). Effect of ultrasonic vibration assisted dry grinding on hysteresis loop characteristics of AISI D2 tool steel. *Sadhana, 46*(4), 245. https://doi.org/10.1007/s12046-021-01771-5

[20] Azami, A., Salahshournejad, Z., Shakouri, E., Sharifi, A. R., & Saraeian, P. (2023). Influence of nano-minimum quantity lubrication with MoS2 and CuO nanoparticles on cutting forces and surface roughness during grinding of AISI D2 steel. *Journal of Manufacturing Processes*, 87, 209–220. https://doi.org/10.1016/j.jmapro.2023.01.029

[21] Chaudhari, A., Sharma, A., Yusufzai, M. Z. K., & Vashista, M. (2023). The grindability performance and measurement of surface functional parameter capabilities of difficult-to-machine tool steel under tangential ultrasonic-vibration-assisted dry grinding. *Machining Science and Technology*, 27(3), 268–291. https://doi.org/10.1080/10910344.2023.2224856

[22] Sharma, A., Chaudhari, A., Yusufzai, M. Z. K., & Vashista, M. (2024). Effectiveness of using liquid nitrogen cryogen in grinding to enhance the grinding performance of hard steel. *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture*, 238(1–2), 315–327. https://doi.org/10.1177/09544054221147622

[23] Mouralova, K., Zahradnicek, R., Benes, L., & Fries, J. (2024). Analysis of spherical chips after grinding. *Measurement*, 229, 114401. https://doi.org/10.1016/j.measurement.2024.114401

[24] Zhang, T., Wang, Q., Wang, N., Yan, L., Jiang, F., Zhang, E., Zhou, W., Gao, H., & Wang, Y. (2025). Surface Integrity in the Grinding of Hardened AISI D2 Steel. *Materials, 18*(4), 814. https://doi.org/10.3390/ma18040814

Authors' contacts:

Aleksandar Milosevic
University of Novi Sad, Faculty of Technical Sciences,
Trg Dositeja Obradovica 6, 21000 Novi Sad, Serbia

Goran Simunovic
University of Slavonski Brod, Mechanical Engineering Faculty,
Trg Ivane Brlic Mazuranic 2, 35000 Slavonski Brod, Croatia

Vitalii Ivanov
Sumy State University, Faculty of Technical Systems and Energy Efficient Technologies,
Rymskogo-Korsakova 2, 40007 Sumy, Ukraine

Mario Sokac
University of Novi Sad, Faculty of Technical Sciences,
Trg Dositeja Obradovica 6, 21000 Novi Sad, Serbia

Vladimir Kocovic
University of Kragujevac, Faculty of Engineering,
Sestre Janjic 6, 34000 Kragujevac, Serbia

Milana Ilic Micunovic
University of Novi Sad, Faculty of Technical Sciences,
Trg Dositeja Obradovica 6, 21000 Novi Sad, Serbia

Djordje Vukelic
(Corresponding author)
University of Novi Sad, Faculty of Technical Sciences,
Trg Dositeja Obradovica 6, 21000 Novi Sad, Serbia
E-mail: vukelic@uns.ac.rs

ISSN 1846-6168 (Print), ISSN 1848-5588 (Online)
https://doi.org/10.31803/tg-20250418164250

Original scientific paper
Received: 2025-04-18, Accepted: 2025-05-02

Design and Analysis of Additively Manufactured Skeletal Structures

Mariusz Dębski*, Grzegorz Budzik, Łukasz Przeszłowski

Abstract: Additive technologies enable the production of elements with complex geometry, which opens up new possibilities in the design of lightweight and strong structures. However, the additive manufacturing process is associated with limitations that must be taken into account at the design stage. This article presents the results of research on the design and analysis of skeletal structures intended for additive manufacturing. CAD models of the robot gripper were developed and then optimized using three different approaches: the use of a shape generator, generative design, and finite element analysis (FEM). The aim of the study was to compare the strength of a robot gripper that was lightened using three different design approaches, with a fixed mass reduction of 40%, in order to determine which method provides the greatest mechanical strength. Experimental studies included numerical analysis and real tests conducted on a testing machine. The test results showed that generative design allows for the most optimal material distribution, which translates into a more favorable strength-to-weight ratio compared to other methods of reducing the weight of the structure.

Keywords: additive technologies; design optimisation; generative design; polymeric materials; strength tests

1 INTRODUCTION

Modern additive manufacturing technologies play an increasingly important role in industry, engineering and science. Due to their flexibility and design capabilities, they enable the creation of geometries that would be difficult or even impossible to achieve using conventional machining or injection molding methods [1].

In the era of increasing requirements for energy efficiency, sustainable development and cost optimization, it is crucial to design elements with the lowest possible mass while maintaining their mechanical properties [2, 3]. Skeletal structures using advanced topological optimization algorithms and inspired by biological structures allow for a significant reduction in the amount of material used, which translates into lower raw material consumption and lower production and operating costs [4]. Such solutions work particularly well in sectors where minimizing mass affects system efficiency, e.g. in the aviation industry, where each additional gram of structure translates into higher fuel consumption [5].

Designing skeletal elements requires an interdisciplinary approach, including materials engineering, structural mechanics, computer modeling, and numerical simulations [6]. Using methods such as finite element analysis (FEM) and generative design, it is possible to obtain optimized components that are characterized by high strength while reducing weight [7]. Another important aspect is the selection of appropriate printing technologies and materials that must meet specific requirements regarding durability, resistance to dynamic loads, and thermal properties.

Reducing the mass of robot grippers has a significant impact on their efficiency, precision and energy consumption. Lighter grippers allow faster manipulator movements, which increases the efficiency of industrial operations and reduces the load on motors and actuators [8]. This allows the use of smaller, more energy-efficient drive components, which leads to reduced operating costs and increased durability of the device [8, 9]. Additionally, the smaller mass of the gripper reduces the moments of inertia, which allows for more precise control and reduction of vibrations when gripping and moving objects [10]. 3D printing enables the creation of weight-optimized structures that maintain high strength while eliminating unnecessary materials and using intelligent internal structures. This allows the design of grippers with high functionality, low weight and increased work efficiency [11].

2 DESIGN OF SKELETAL STRUCTURES

Designing additively manufactured skeletal elements requires taking into account a number of key factors that affect their strength, functionality and manufacturing efficiency. An appropriate approach to these aspects allows for the creation of optimal structures with high strength while minimizing material consumption.

One of the most important factors is the selection of material, which must be adapted to the operational requirements of a given element. The selection of the appropriate filament in FDM technology (Fused Deposition Modeling), powder in the SLS method (Selective Laser Sintering), or resin in the SLA method (Stereolithography) depends on such parameters as mechanical strength, flexibility, and chemical and thermal resistance. The material should be selected in such a way as to ensure the durability of the structure while maintaining low weight.

An important aspect is also considering the printing direction, as the layer-by-layer structure of the models affects the material anisotropy. Proper orientation of the component on the build platform can significantly improve its load resistance and reduce the risk of fractures at layer junctions.

When designing skeletal structures, it is also necessary to take into account the limitations related to the additive manufacturing process. The geometric conditions of the designed models depend primarily on the type of material used, the technological limitations of the selected method and the printing device.

The process of designing additively manufactured skeletal components can be divided into seven key stages:

1) Requirements analysis – defining project goals, working conditions and technological constraints.
2) CAD modeling – creating a digital model.
3) Mass reduction (optional) – depending on the design requirements, various methods of structural lightening can be applied, such as topology optimization (TO), generative design (GD), or manual modification of the model based on load analysis. The purpose of this stage is to reduce the weight of the part while maintaining the necessary mechanical strength; however, it is not a mandatory step for every additively manufactured component.
4) Numerical simulations – conducting strength analysis (e.g. finite element method – FEM) to identify weak points.
5) Adaptation to printing technology – modification of the model to suit printer limitations (e.g. wall thickness).
6) Manufacturing and testing – test printing and analysis of the mechanical and functional properties of the element.
7) Modifications and finalization – improving the design based on test results and developing the final version of the item.

When designing skeletal samples, it is possible to use various tools to support the process of creating optimal structures. The selection of appropriate software and methods depends on the design requirements, such as mechanical strength, mass minimization, or manufacturing method. The basic tools supporting the design include FEM analysis, shape generators, generative design, and programs for creating openwork structures.

One of the tools used in the design of skeletal structures is the FEM analysis. It allows for the simulation of the actual working conditions of the element and the assessment of its behavior under the influence of various loads. Based on the numerical analysis, it is possible to determine which areas of the structure are most heavily loaded and which can be slimmed down without losing strength (Fig. 1). The analysis of the obtained results also allows for the detection of potential critical points that may lead to damage to the given element, which allows for the introduction of corrections before the start of production.

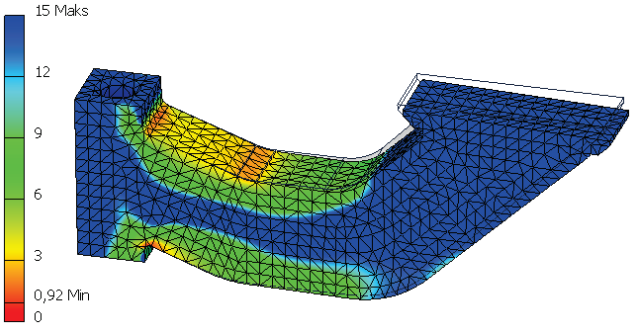


Figure 1 Numerical analysis of robot gripper - safety factor

Shape generators are tools that allow you to automatically adjust the geometry of a model to specific criteria, such as stiffness, mass or structural constraints. They operate on the principle of algorithmic optimization, removing unnecessary material where it is not necessary, while maintaining the required strength. The stages of part design are presented using the Shape Generator tool in Inventor Professional as an example. The first step is to determine the boundary conditions, the load acting on the element and generate a triangle mesh. Additionally, the software allows for the exclusion of specific regions from the shape generation process and the definition of symmetry planes. Based on the selected material, the shape generator creates a conceptual 3D mesh (conceptual shape – Fig. 2), which can be used as a reference in the design process. Figure 3 shows a mass-reduced gripper whose geometry was designed using the Shape Generator tool.

Figure 2 Conceptual shape of the robot gripper

Figure 3 Reduced-mass robot gripper

Generative design is a more advanced form of optimization that uses artificial intelligence and evolutionary algorithms to create optimized geometry. Such programs can generate dozens or even hundreds of design variants that meet specific requirements for strength, weight, and production costs. This method allows for obtaining unique, organic shapes that are impossible to achieve using traditional design techniques. One of the tools for generative design is Generative Design in Autodesk Fusion 360. The design process is similar to a shape generator. You need to define boundary conditions, which are the load and material acting on the element (different materials can be defined depending on the type of processing). The difference is that in the case of generative design, you indicate the preserved geometry, i.e. solids that will not change during the generation of results. Additionally, you can assign obstacle geometry representing empty spaces in which no material will be placed during the generation of results (Fig. 4). Before generating the results, you can use the preview tool to see how the settings affect the results. Fig. 5 shows the selected robot gripper geometry generated by the program.

Openwork structures play a key role in designing lightweight but strong skeleton elements. Programs for their creation allow for generating lattice meshes with a specific geometry (e.g. gyroidal, hexagonal, tetragonal). One of the

tools for creating openwork structures is the Meshmixer program. This program allows for easy generation of lightweight structures with complex geometry, but such structures work best for elements that are not exposed to significant mechanical loads. In situations where high strength parameters are required, it may be necessary to use more advanced optimization methods or reinforced internal structures. The first step in creating such a structure in the Meshmixer program is to set a regular triangle grid on the surface of the element. This grid is the basis for later modification of the geometry and allows for uniform distribution of the openwork structure pattern. Then, the type and geometry parameters of the structure are determined, such as wall thickness, cell shape or filling degree (Fig. 6).

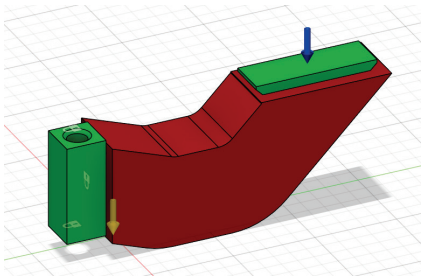


Figure 4 View of the solids defining the preserved geometry and the obstacle geometry

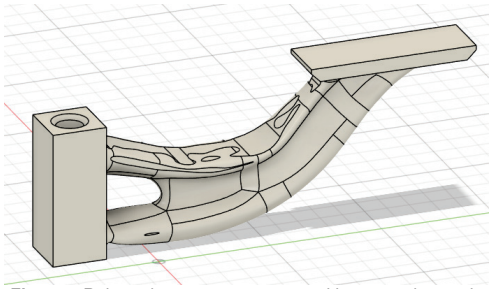


Figure 5 Robot gripper geometry created in generative study

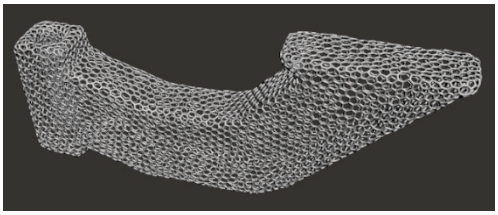


Figure 6 Robot gripper geometry created in Meshmixer

The purpose of the work was to compare the effectiveness of three different approaches to designing lightweight gripper structures in terms of their mechanical strength, while maintaining a consistent 40% reduction in mass. In contrast to typical optimization strategies, where the goal is to minimize mass while satisfying defined boundary conditions, this study adopted a fixed level of mass reduction for all design variants. This approach enabled a direct comparison of the performance of three design methods: shape generation, generative design, and manual design based on results from finite element analysis (FEA). The safety factor was used as the main evaluation criterion, as it allows for the identification of the structural margin of strength with respect to the failure criterion and indicates potential areas at risk of damage, which is particularly important in the context of the anisotropic structure of additively manufactured materials.

3 RESEARCH MODELS

Based on the analysis of available design solutions on the market, a robot gripper model was developed in the form of a three-dimensional CAD model (reference sample K1 – Fig. 7). In order to optimize the structure in terms of weight reduction, the model was slimmed down using three different methods: shape generator, generative design and optimization based on finite element analysis (FEM). For all analyses, the construction material polylactic acid (PLA) and a load of 100 N were assumed. As a result of the design work, three optimization variants were obtained:

- one version of the gripper created using the shape generator (K2 – Fig. 8),
- one version created as a result of using generative design (K3 – Fig. 9),
- two versions developed on the basis of FEM analysis, taking into account different material reduction strategies (K4, K5 – Figs. 10 - 11).

For variants K4 and K5, the geometry was designed manually based on FEA results obtained for the reference model (Fig. 14). This analysis enabled the identification of low-stress areas, which were considered safe for removal or partial lightening without a significant deterioration in the mechanical properties of the structure.

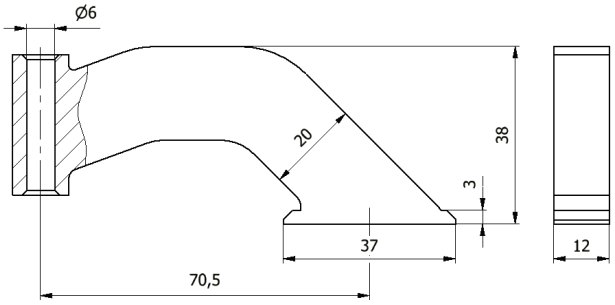


Figure 7 Shape and basic dimensions of the K1 reference gripper

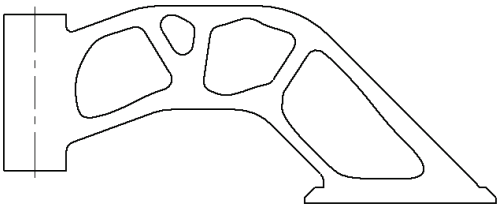


Figure 8 Shape of the K2 gripper

For each of the developed design concepts, a 40% reduction in the gripper mass was achieved, while maintaining the appropriate strength parameters required for the correct functioning of the component in the robot manipulation system.

The test samples were made using the layer extrusion method on a Prusa i3 MK3 printer. The printing process for

all models was characterized by the following basic parameters:

- alternating paths at angles of 45º and 135º,
- internal structure filling density – 100%,
- two contour lines,
- plastic layer height 0,2 mm.

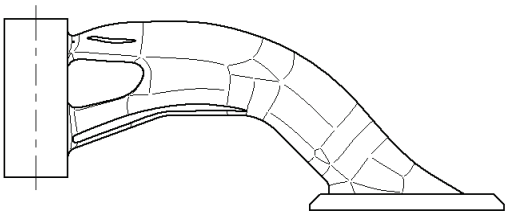


Figure 9 Shape of the K3 gripper

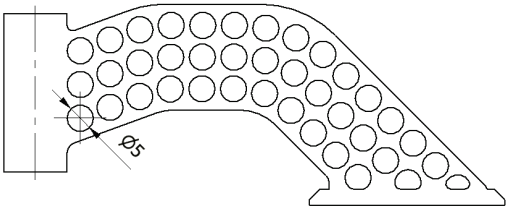


Figure 10 Shape of the K4 gripper

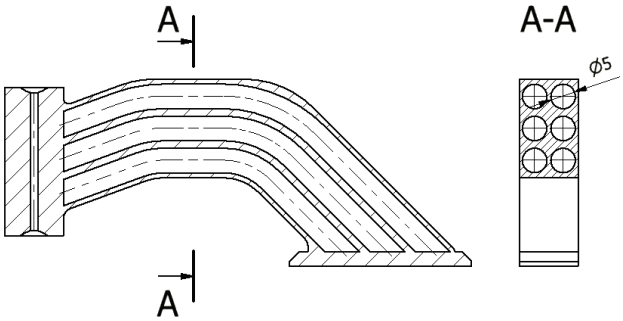


Figure 11 Shape of the K5 gripper

4 EXPERIMENTAL RESEARCH AND DISCUSSION

The main objective of the conducted tests was to verify the gripper's ability to transfer the given loads in static conditions. The verification included the analysis of the safety factor, which allows for the assessment of the structure's strength margin in relation to the assumed destruction criterion. As part of the conducted static analysis, a variant was considered in which a load of 100 N acts on the end part of the gripper in a direction perpendicular to its surface. The numerical analysis was performed using Autodesk Inventor Professional software. The initial conditions of the analysis are shown in Fig. 12. One of the significant problems that may occur during the use of the gripper is its deflection and permanent deformation, especially in the case of using materials susceptible to bending. If the load acts on the tip of the gripper, it generates a bending moment, which reaches the highest value at the base of the structure, at the point of attachment. This can lead to undesirable deformations, reduced precision of the gripper's operation, and in extreme cases, to its damage.

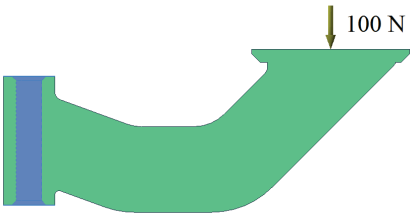


Figure 12 Constraints and load

The next stage of the research involved verifying the results of the numerical analysis through experimental strength tests conducted on a MultiTest-dV machine (Fig. 13). The aim of these tests was to confirm the consistency of the computer simulation results with the actual behavior of the gripper under load. To ensure the coherence of the test conditions, the same loading configuration as in the numerical analysis was applied. The tests were carried out until the samples reached failure.

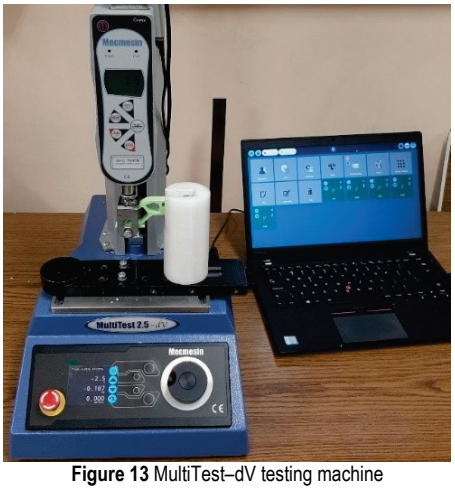


Figure 13 MultiTest–dV testing machine

4.1 Strength Analysis using the Finite Element Method

After generating and optimizing the finite element mesh, an analysis was performed for each gripper variant, the results of which in the form of a safety factor are presented in Tab. 1 and Figs. 14 - 18.

Table 1 Summary of the minimum safety factor for each variant tested

Sample type	K1	K2	K3	K4	K5
Minimum safety factor	2,46	1,42	1,88	1,44	1,65

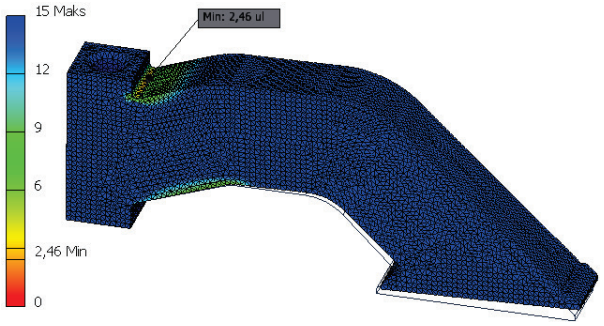


Figure 14 Analysis results for gripper K1

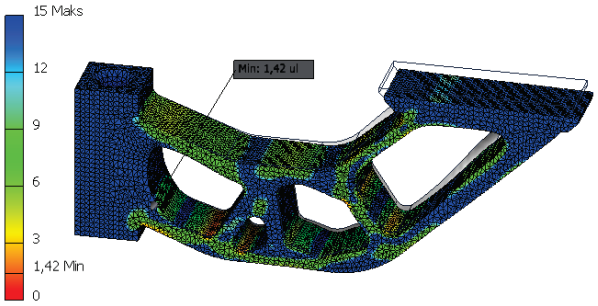


Figure 15 Analysis results for gripper K2

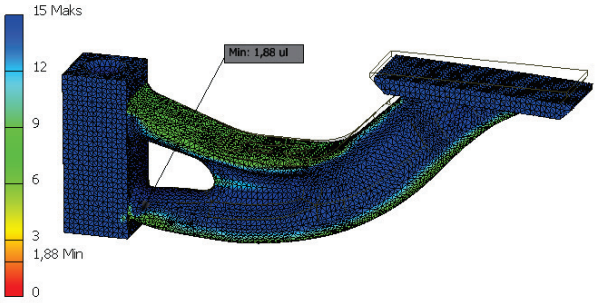


Figure 16 Analysis results for gripper K3

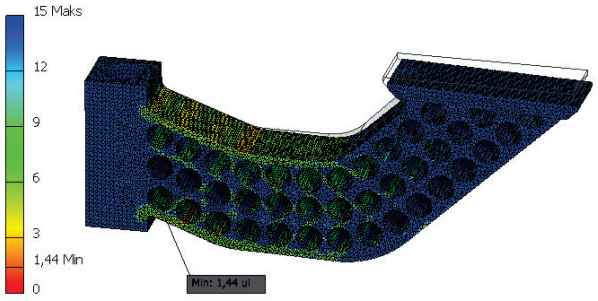


Figure 17 Analysis results for gripper K4

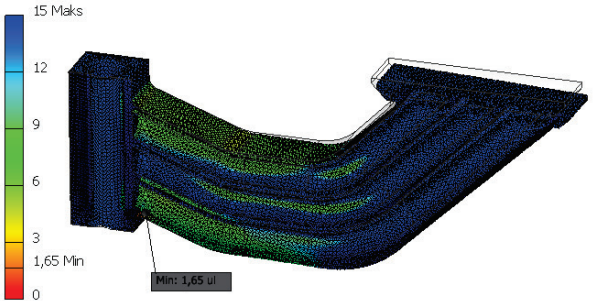


Figure 18 Analysis results for gripper K5

The analysis of the results carried out using the finite element method included an assessment of the safety factor for different design variants of the robot gripper. The highest value of the safety factor was obtained for the reference sample, which resulted from its full structure and lack of mass reduction. The next in terms of strength was the sample designed using the generative design method, which, despite significant mass reduction, retained optimal strength properties thanks to intelligent material distribution in key areas. In the case of variants K2, K4 and K5, the analysis showed the possibility of local overloads in the mounting areas and at the end of the gripper, where the highest bending moments are generated. These zones can be potential critical points, especially in dynamic operating conditions. Additionally, the analysis allowed for the identification of areas at risk of weakening the structure, which is particularly important in the case of additive manufacturing due to the layered structure. This is due to the fact that the FEM analysis assumes an isotropic material model, while the materials used in additive technology are characterized by mechanical anisotropy. This means that the mechanical properties of the model depend on the direction of the layers, which can lead to local weakening and discrepancies between the simulation results and the actual strength of the element. Additionally, the way the model paths are arranged at a specific angle means that certain areas of the structure may be more susceptible to damage than the numerical analysis would suggest.

4.2 Strength Analysis on a Real Model

The experimental results are presented in Fig. 19.

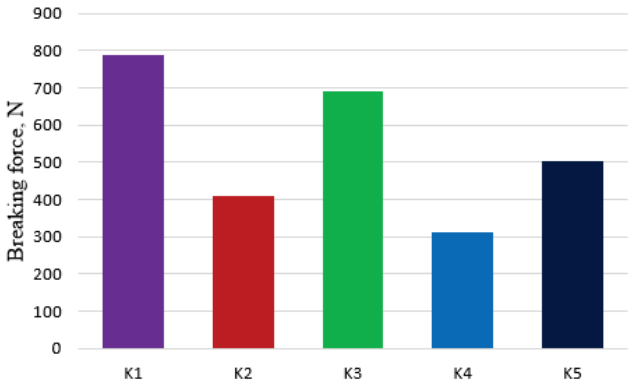


Figure 19 Average values of the breaking force for each tested variant

Based on the obtained results, it can be seen that the highest strength was achieved by the reference sample (K1), which was not subjected to mass reduction. Among the optimized variants, the best results were achieved by sample K3, which was designed using the generative design method. Despite the 40% mass reduction, its strength was the closest to the full sample, which indicates the effectiveness of this approach in the distribution of material according to the loads acting on the element. The average value of the force required to destroy the sample is about 88% of the reference gripper.

The remaining conventionally optimized variants, including sample K2, showed significantly lower values of the ultimate force. This means that manual mass reduction and geometry shaping based on FEM analysis can lead to a weakening of the structure if the actual load directions are not properly taken into account.

5 CONCLUSIONS

Designing skeleton samples manufactured using additive technologies requires the use of advanced engineering tools. FEM analysis allows for the assessment of the strength of the

structure, shape generators help in the optimization of geometry, and generative design allows for the creation of innovative and optimized forms. Proper use of these tools allows for the creation of lightweight and durable elements.

The study confirmed that generative design is a more effective tool for optimizing the structure of the robot gripper compared to shape generation and manual design based on FEA. This is due to the fact that generative algorithms can better adapt the structure to real working conditions, which is difficult to achieve with conventional methods.

In addition, the research results indicate a significant influence of material anisotropy in 3D printing technology, which means that the actual strength of the component may be lower than it would result from the FEM analysis for an isotropic material. Therefore, in the design of components intended for additive manufacturing, both the optimization of geometry and taking into account the specificity of the technological process are of key importance.

When designing skeleton elements for additive manufacturing, it is crucial not only to perform numerical analysis, such as the finite element method (FEM), but also to conduct experimental studies. This is due to the fact that in the additive manufacturing process there are a number of factors that can significantly affect the mechanical properties of the final element, and which are difficult to fully include in simulation analyses.

Acknowledgements

Scientific work/Publication co-financed from the state budget under the program of the Ministry of Science called "Polska Metrologia II" (Polish Metrology II) project number PM-II/SP/0085/2024/02 titled: "Methodology for measuring and analyzing the geometry of skeleton and shell structures of additively manufactured objects", Poland.

6 REFERENCES

[1] Yalçın, M. M. (2024). Flexural Behavior of 3D-Printed Carbon Fiber-Reinforced Nylon Lattice Beams. *Polymers*, *16*(21), 2991. https://doi.org/10.3390/polym16212991

[2] Tyflopoulos, E., & Steinert, M. (2020). Topology and Parametric Optimization-Based Design Processes for Lightweight Structures. *Applied Sciences*, *10*(13), 4496. https://doi.org/10.3390/app10134496

[3] Tang, T., Wang, L., Zhu, M., Zhang, H., Dong, J., Yue, W., & Xia, H. (2024). Topology Optimization: A Review for Structural Designs under Statics Problems. *Materials*, *17*(23), 5970. https://doi.org/10.3390/ma17235970

[4] Ryan-Johnson, W. P., Wolfe, L. C., Byron, C. R., Nagel, J. K., & Zhang, H. (2021). A Systems Approach of Topology Optimization for Bioinspired Material Structures Design Using Additive Manufacturing. *Sustainability*, *13*(14), 8013. https://doi.org/10.3390/su13148013

[5] Iftekar, S. F., Aabid, A., Amir, A., & Baig, M. (2023). Advancements and Limitations in 3D Printing Materials and Technologies: A Critical Review. *Polymers*, *15*(11), 2519. https://doi.org/10.3390/polym15112519

[6] N. Musthafa, H.-S., Walker, J., & Domagala, M. (2024). Computational Modelling and Simulation of Scaffolds for Bone Tissue Engineering. *Computation*, *12*(4), 74. https://doi.org/10.3390/computation12040074

[7] Balayan, A., Mallick, R., Dwivedi, S., Saxena, S., Haorongbam, B., & Sharma, A. (2024). Optimal Design of Quadcopter Chassis Using Generative Design and Lightweight Materials to Advance Precision Agriculture. *Machines*, *12*(3), 187. https://doi.org/10.3390/machines12030187

[8] Francalanza, E., Fenech, A., & Cutajar, P. (2018). Generative design in the development of a robotic manipulator. *Procedia CIRP*, *67*, 244–249. https://doi.org/10.1016/j.procir.2017.12.207

[9] Sha, L., Lin, A., Zhao, X., & Kuang, S. (2020). A topology optimization method of robot lightweight design based on the finite element model of assembly and its applications. *Sci. Prog., 103*. https://doi.org/10.1177/0036850420936482

[10] Liu, B., Sha, L., Huang, K., Zhang, W., & Yang, H. (2022). A topology optimization method for collaborative robot lightweight design based on orthogonal experiment and its applications. *Int. J. Adv. Robot. Syst.*, *19*. https://doi.org/10.1177/17298814211056143

[11] Curkovic, P. (2024). Comparative Analysis of Topology Optimization Platforms for Additive Manufacturing of Robot Arms. *Designs*, *8*(5), 98. https://doi.org/10.3390/designs8050098

Authors' contacts:

Mariusz Dębski, DSc. Eng.
(Corresponding author)
Rzeszów University of Technology,
Al. Powstańców Warszawy 12, 35-959 Rzeszów, Poland
m.debski@prz.edu.pl

Grzegorz Budzik, Prof. PhD, DSc, Eng.
Rzeszów University of Technology,
Al. Powstańców Warszawy 12, 35-959 Rzeszów, Poland
gbudzik@prz.edu.pl

Łukasz Przeszłowski, DSc. Eng.
Rzeszów University of Technology,
Al. Powstańców Warszawy 12, 35-959 Rzeszów, Poland
lprzeszl@prz.edu.pl

ISSN 1846-6168 (Print), ISSN 1848-5588 (Online)
https://doi.org/10.31803/tg-20250418165639

Original scientific paper
Received: 2025-04-18, Accepted: 2025-05-02

Strength Study of Anatomically-Similar Spongy Structures Additively Manufactured from Polymeric Materials

Łukasz Przeszłowski*, Anna Paluch, Grzegorz Budzik, Damian Filip, Łukasz Kochmański, Mariusz Dębski

Abstract: The subject of the article is to investigate the potential to adjust the stiffness of 3D printed plastic implants to match bone stiffness by analyzing design parameters and mechanical properties, considering the spongy structure of the bone. The study consisted of theoretical and practical parts, including original research. In the theoretical part, the definition of implants, the materials used for their production, and the requirements they must meet were discussed, as well as 3D printing methods, with particular emphasis on the FFF method. The static compression test was also described. In the practical part, original research was conducted to verify the possibilities of adjusting the stiffness of the implants. Two 3D models of the L2 lumbar vertebra with different vertebral body thicknesses and internal lattice structures were developed based on literature dimensions. Thirty models with three different lattice densities and two wall thicknesses were printed and subjected to a static compression test. The results showed that an increase in lattice density increased the compression strength of the samples. Samples with thin walls exhibited lower compression strength compared to those with thick walls, regardless of the spongy lattice density. In each group of samples, the results were consistent, indicating good repeatability of the prepared samples. The findings suggest that appropriately selecting the lattice density and wall thickness can significantly improve the mechanical strength of bone implants, meeting the compression strength criteria for the spongy tissue of lumbar vertebrae.

Keywords: additive manufacturing; compression test; PLA (Polyactide) spongy structure; vertebras

1 INTRODUCTION

In the medical field, the development of modern technologies such as 3D printing is opening up new perspectives in the design and manufacture of medical implants. One of the key challenges in this area is to ensure that implants are not only biocompatible, but also properly adapted to the physical and mechanical conditions of the surrounding tissues. In the context of this issue, exploring the possibility of adapting the stiffness of printed plastic implants to that of bone becomes crucial.

Properly matching the stiffness of the implant to the stiffness of the bone can help minimize the risk of injury while ensuring the stability and strength of the implant in the patient's body. In this context, it is important to understand the structure of spongy bone, which is characterized by a specific grid-like microarchitecture. Thanks to 3D printing technology, it is possible to manipulate the structural parameters of implants, including their stiffness, which opens up prospects for individualizing treatment and improving therapeutic outcomes.

The spongy structure of bone, also known as trabecular bone, is made up of tiny bone beams that form a meshwork with a large surface area. This unique structure provides exceptional strength and flexibility to bone, while allowing mechanical forces to be distributed throughout the body. In the case of medical implants, reproduction of this structure can be crucial to improving the integration of the implant into tissues and ensuring adequate stiffness and strength [2].

Compression testing of 3D printed lumbar vertebra specimens are an important part of the study. This type of mechanical test allows us to evaluate the behavior of printed implants under load, which is crucial for understanding their strength and stability under conditions similar to those in the patient's body. Modern 3D printing technology makes it possible to create implants with complex structures that can imitate the spongy structure of bone. With the help of advanced design algorithms and precise 3D printing processes, it is possible to produce printed meshes with different geometric and mechanical parameters that are more similar to the structure of natural bone. However, it is important to pay special attention to choosing the right plastic for 3D printing implants. Plastics used in implantology have to be not only biocompatible, but also have sufficient mechanical strength and flexibility to meet the physical demands of the patient's body [8].

The purpose of this paper is to research the potential of adapting the stiffness of printed sponge structures made of selected plastic to the stiffness of spongy bone by analyzing design parameters and mechanical properties, taking into account the spongy structure of the bone and conducting compression testing of printed lumbar vertebral specimens. The research aims to develop new methods for designing and manufacturing implants to better suit individual patients' needs, and to investigate how much the stiffness of an implant can be altered with a printed lattice structure. The research aims to explore the relationship between implant design parameters, stiffness and the spongy structure of bone.

The research section presents the methodology for creating 3D models of the L2 lumbar vertebra, based on dimensions from the literature. Two models were created, with two different thicknesses of the vertebral body, which were further modified to obtain a mesh structure in the interior of the vertebral body, i.e. a spongy structure.

2 BONE PARAMETERS

Bones are made up of bone tissue. This tissue is seen as a composite material, which consists of two phases: an organic phase and an inorganic phase. The former acts as a matrix, while the latter is responsible for reinforcement. The ratio of these phases to each other is 1:2.5. One of the inorganic compounds responsible for mechanical strength is calcium phosphate (V) Ca3(PO4)2, which accounts for 2/3 of the weight of bone. The inorganic components of bone tissue, which make up about 50% of all matrix members, are

responsible for determining bone hardness. The organic phase is formed by bone cells along with the intercellular matrix made up of collagen proteins. They are mainly type I collagen proteins and account for about 95% of all organic components. Collagen found in bone has the ability to resorb calcium, and this distinguishes it from collagen found in other tissues. It is the large amount of collagen in the organic part that changes the elastic properties of bone. The relationship of organic and mineral components determines the mechanical properties of bone [4].

Osteon is the basic structural unit of bone. It is formed by bone lamellae, between which bone cavities are located along with bone cells. In the central part of the osteon is located a central canal called Haversian canal. This is a duct of microscopic dimensions through which blood vessels are routed to allow blood flow to bone cells. Osteons, which are located very close to each other, i.e. one next to the other, result in a compact bone structure (Fig. 1) [3]. Bones by shape are divided into four groups:

- long bone
- short bone
- flat bone
- varus (irregular) bone.

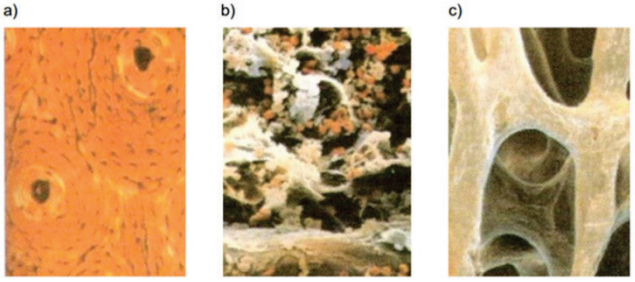


Figure 1 Bone structure: a) cortical layer - osteons with Haversian canals, b) spongy layer - bone beams, c) bone marrow [5]

Classic long bone consists of bone tissue and marrow. There are two forms of bone tissue: compact bone substance and spongy bone substance. Compacted bone tissue is always found on the top layer of the bone so it is called cortical tissue. Through the environment of the bone characterized by an organic substrate, the bone acquires a lot of elasticity. And the calcium salt found there also makes it very resistant to compression and tension. The compressive strength of bone tissue is greater than the tensile strength, while when considering bending, bone is largely less resistant in this aspect [3].

In order to obtain the average values of the mechanical properties of the compact part of the human femur bone, it was examined at appropriate angles, and the obtained values are shown below:

- tensile strength: 107 MPa
- compressive strength: 159 MPa[2]
- torsional strength: 53 MPa
- bending strength: 160 MPa
- ultimate elongation: 135%.

Most of the mechanical properties of bones are due to their anatomical and histological structure. The main concern here is the propensity to fracture, and this is mostly related to age. Mechanical properties depend on the type of bone structure, geometric form, external loading, location and value of loading forces, which are responsible for the distribution of internal stresses in bones. The collagen content is of great value here, since bones with a high collagen content in their structure are easily subject to brittle fractures. In the case of the elderly, where this collagen in the bones is less and less, brittle fractures occur [4]. The Young's modulus of collagen is 1-2 GPa, and its ultimate tensile strength is 50-1000 MPa [9].The mechanical properties of spongy bone tissue and compact bone tissue are shown in Tab. 1.

Table 1 Mechanical properties of two types of bone tissue [6]

Property	Cortical bone	Cancellous bone
Compressive strength (MPa)	100-230	2-12
Flexural, tensile strength (MPa)	50-150	10-20
Strain to failure (%)	1-3	5-7
Fracture toughness ($MPa^{1/2}$)	2-12	-
Young's modulus (GPa)	7-30	0.5-0.05

Spongy bone otherwise known as trabecular bone has a porosity of 50 to 90%, with an average spacing of about 1mm between trabeculae and an average density of about 0.2 g/cm^3. Compact bone, or cortical bone, has a much denser structure. Its porosity is between 3-12%, and its average density is 1.80 g/cm^3 [6].

3 RESEARCH MODELS AND MATERIAL

The structures of the specimen were modeled using Autodesk Inventor Professional 2024. The L2 vertebral body of the lumbar segment was modeled without spinous processes, which are the attachment site of the muscles of the human lower back. In each spinal segment, the vertebrae are different from each other, and this is due to the loads they are subjected to, the mobility they have, and anatomical limitations. The area of the same spinal segment does not contain vertebrae of the same dimensions. Large differences in geometric parameters appear between vertebrae mostly in the vertebral bodies. The vertebral body can be compared to a cylinder having a flat upper and lower surface. However, this does not represent its geometry one hundred percent because the shaft has differences between the size of the upper and lower surfaces, as well as differences in the ratio of the height of the front part to the back part. For the upper part of the shaft, the following parameters were adopted: the width of the upper and lower surfaces of the shaft was assumed to be 51.2 mm, and the depth was 34 mm. The width of the bone ring, which is the outer part of the vertebra, varies according to the analyzed region. The dimensions of the vertebra were taken from the literature entitled Biomechanics of the Spine, by Celina Pezowicz. For the upper part of the bony ring of the L2 vertebra, the width of the area in front was taken into account equal to 5.3 mm, for the sides 6.5 mm, and for the back 3.5 mm - variant A (Fig. 2).

On the other hand, in the lower part of the vertebra, the area of the bone ring in the front is 6.2 mm wide, 8 mm on the sides, and 2.8 mm in the back - variant B (Fig. 3). The dimensions used in the modeling are the average actual

dimensions of the L2 vertebra of an adult in the 36-60 age range.

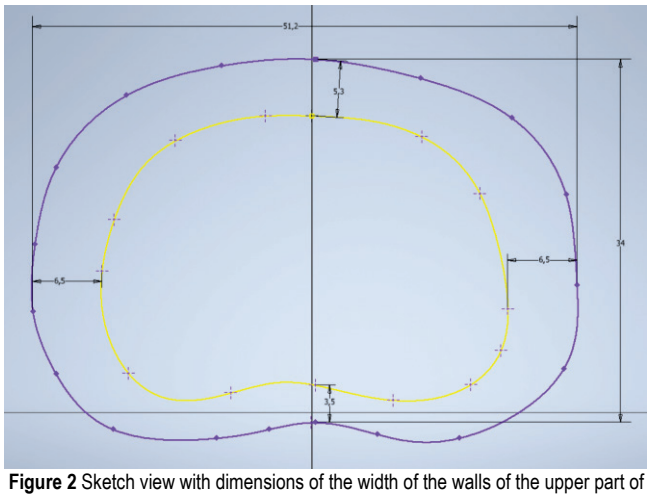


Figure 2 Sketch view with dimensions of the width of the walls of the upper part of the bone ring - variant A

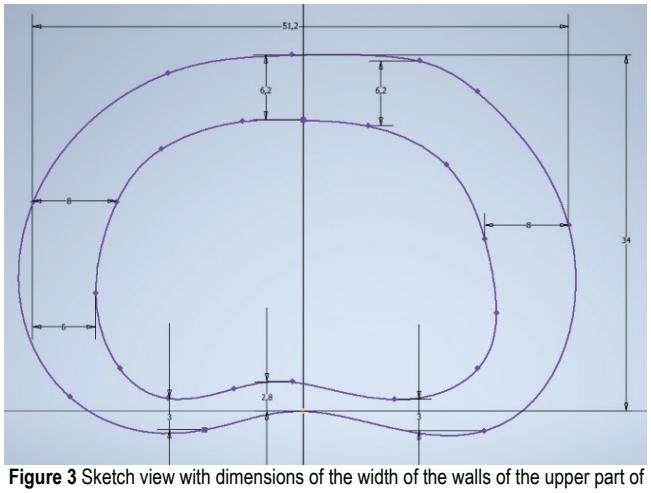


Figure 3 Sketch view with dimensions of the width of the walls of the upper part of the bone ring - variant B

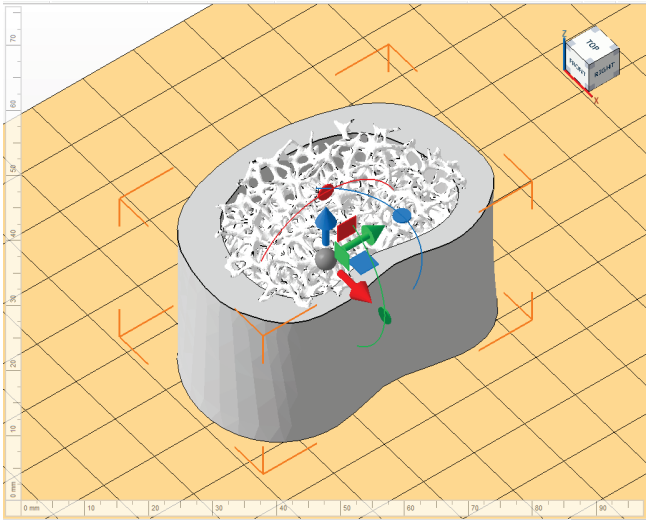


Figure 4 Final model of L2 lumbar vertebra created by Boolean operation from two parts: spongy bone and compact bone in Netfabb software with mesh size 4 × 4 × 4 mm

Netfabb software was used to make the sponge structure inside the compact part. An example of the obtained structure is shown in Fig. 4.

After creating three different models of the L2 lumbar vertebra, research models were made using FFF [10-12] additive technology. Initially, the process had to be prepared and the files were imported into PrusaSlicer. PrusaSlicer is a program for preparing files for 3D printing, which allows the configuration of various printing parameters, such as speed and temperature. The fill of the models was set to 95% due to the fact that often at 100% fill the models do not come out as they should. A grid was used as the fill pattern. No supports were used when printing, and an accuracy of 0.10mm was used in the print settings. The material chosen was PLA, and the printer was Original Prusa i3 MK3S. The test models were made with three densities of the structure mimicking the sponge structure (Fig. 5).

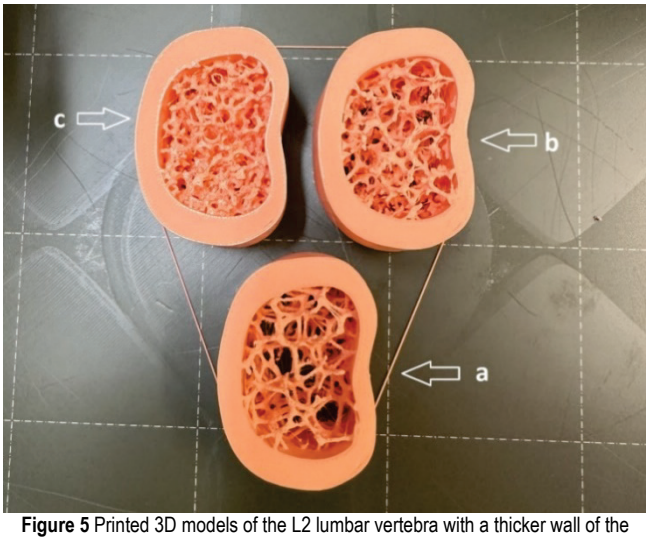


Figure 5 Printed 3D models of the L2 lumbar vertebra with a thicker wall of the vertebral body, with different dimensions of the mesh of the spongy structure: a) 6.5 × 6.5 × 6.5 mm; b) 5 × 5 × 5 mm; c) 4 × 4 × 4 mm

The designations of all research samples are shown in Tab. 2.

Table 2 Designations of research samples

Wall thickness of the vertebra	Density of spongy structure	Designation
Thin wall	6.5 × 6.5 × 6.5 mm	1.1, 1.2, 1.3, 1.4, 1.5
Thin wall	5 × 5 × 5 mm	2.1, 2.2, 2.3, 2.4, 2.5
Thin wall	4 × 4 × 4 mm	3.1, 3.2, 3.2, 3.4, 3.5
Thick wall	6.5 × 6.5 × 6.5 mm	4.1, 4.2, 4.3, 4.4, 4.5
Thick wall	5 × 5 × 5 mm	5.1, 5.2, 5.3, 5.4, 5.5
Thick wall	4 × 4 × 4 mm	6.1, 6.2, 6.3, 6.4, 6.5

The material from which the test samples were made is PLA Polyactide poly(lactic acid) (PLA) is a biodegradable aliphatic thermoplastic polyester, which is a derivative of natural lactic acid. It has a number of unique characteristics that make it an attractive biopolymer in environmental and economic terms, such as excellent rigidity, transparency, ease of processing and aesthetic appearance. Nevertheless, PLA has some limitations, such as intrinsic brittleness, low strength and slower degradation. Nevertheless, it shows great

potential as a biomaterial in various medical fields, including regenerative medicine, orthopedics and tissue engineering [7, 9]. A major advantage of the material chosen for the study is that it can be 3D printed on any device working with FFF additive technology, unlike materials such as PEEK.

PLA is a semi-transparent polymer material that is completely biodegradable by the lactic acid in its structure [12]. However, it decomposes over a fairly long period of time. PLA is not a soluble material. During the printing process, this material does not emit a lot of harmful chemicals, so it can safely be used as an educational aid. The glass transition temperature otherwise softening temperature is about 60 °C. Thanks to its low shrinkage, there is no need to heat the table on which the part is printed from this plastic. The extrusion temperature of PLA filament through the printer's nozzles ranges from 160 to 230 °C. In the case presented, the extrusion temperature was 210 °C. The density of the material between 1.2 and 1.4 g/cm^3. The material has good tensile strength and stiffness, but limited elasticity and low impact strength. Young's modulus is about 2000 MPa, and yield strength is about 40-50 MPa [1, 13, 14].

4 RESULTS AND DISCUSSION

A static compression test was used to test the printed samples. A motorized force measuring rig of the MultiTest-dV series from Mecmesin was used. The device is used to test forces up to 2.5 kN and elongation of more than 1000%. Since the compression of the sponge part of the printed vertebrae was being attempted, it was necessary to create a punch that was used to compress the selected structure. So, before the compression test of each specimen began, a punch was modeled in AutodeskInventor Professional 2024 (Fig. 6).

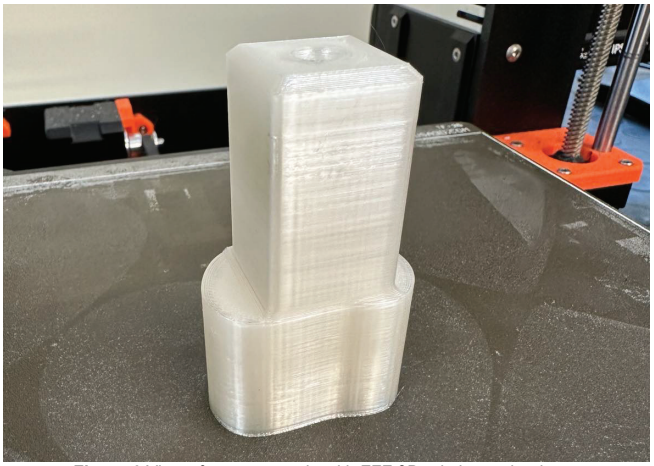


Figure 6 View of a stamp made with FFF 3D printing technology

Fig. 7 shows the specimen after the punch is attached to the machine, and the static compression test process for all printed specimens has begun.

The samples were compressed to a depth of 5 mm, at a rate of 5 mm/min. As an example of the force-time dependence graph, models of a thinner-walled circle with a sponge structure density of 5 × 5 × 5 mm were shown. The specimens were labeled 2.1, 2.2, 2.3, 2.4, 2.5. The results were tabulated together on a line graph to compare the values obtained (Fig. 8).

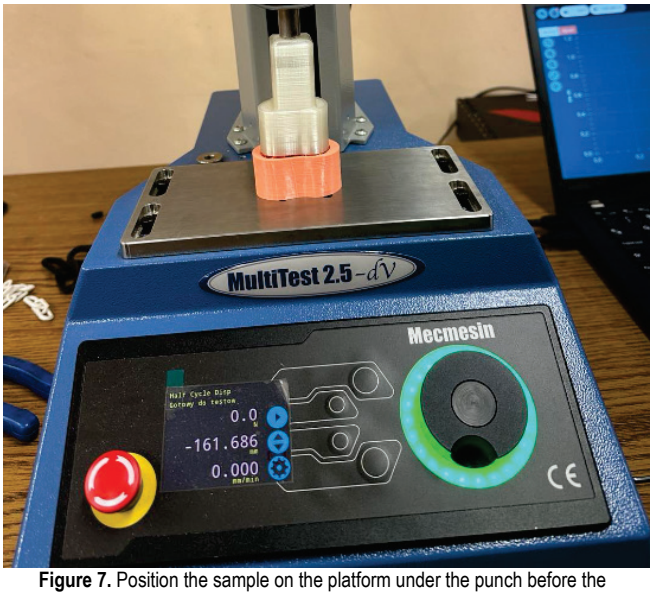


Figure 7. Position the sample on the platform under the punch before the compression test

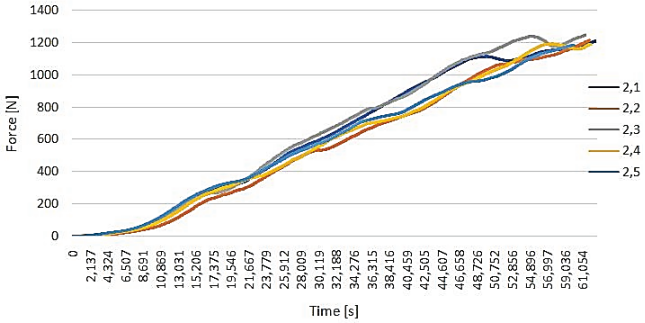


Figure 8 Graph of the force-time dependence of the compression test of thin-walled specimens with a mesh density of 5 × 5 × 5 mm

Spongy tissue is characterized by heterogeneity, and its structure changes depending on its location in the vertebral body. According to the literature, the highest compressive strength values of the spongy tissue of the vertebral body can be observed in its upper and lower parts. During the compression test, the spongy structure was examined to a depth of 5 mm. In order to illustrate what pressure was applied to the specimens during the test and how it varied depending on the particular model, two graphs were created. The first for previously selected representative specimens with thin walls, while the second for representative specimens with thick walls. The strength of the prepared structures was calculated by dividing the highest achieved force in a given compression test by the area on which the force acted.

A graph showing the maximum compressive strength values of the spongy tissue structure in individual specimens with thin and thick vertebral body walls was created from the results of representative samples from each specimen (Figs. 9-10).

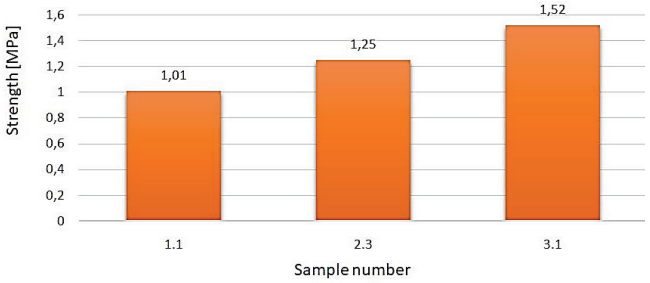


Figure 9. Bar graph showing the maximum compressive strength values of the spongy tissue structure for selected specimens with thin vertebral body walls

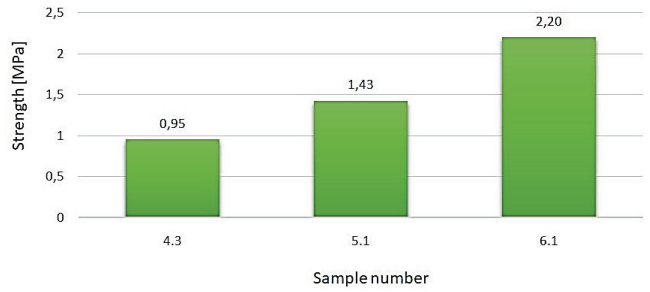


Figure 10. Bar graph showing the maximum compressive strength values of the spongy tissue structure for selected specimens with thick vertebral body walls

According to the literature, the compressive strength of the spongy tissue of the lumbar vertebral bodies ranges from 1.28 to 10.6 MPa [6].

Table 3 Comparison of ranges of maximum compressive strength results of sponge tissue structure for selected specimens with thick and thin vertebral body walls with literature values

	Compressive strength range (MPa)
Literature data	1.28÷ 10.6
Samples with thin walls	1.01 ÷ 1.5
Samples with thick walls	0.95 ÷ 2.2

Matching the graph in Fig. 9 showing the values of the maximum compressive strength of the spongy tissue structure for selected specimens with thin vertebral body walls to the literature data yielded groups of specimens that fall within the given range. Of the given groups of specimens with thin walls, only one density of spongy structure met the literature values. Representative sample number 3.1 reached a value above 1.5 MPa, so thin-wall samples with a structure density of 4 × 4 × 4 mm fall within the specified range. On the other hand, comparing the literature values to those shown in Fig. 10, which shows the values of the maximum compressive strength of the sponge tissue structure for selected samples with thick vertebral body walls, only one group of samples did not reach these values. Samples with the sparsest lattice structure, i.e. 6.5 × 6.5 × 6.5 mm, where sample number 4.3 was the representative sample, did not even exceed the 1 MPa value. The other two groups, i.e. samples with a mesh density of 5 × 5 × 5 mm and 4 × 4 × 4 mm reached values of 1.4 MPa and 2.2 MPa.

5 CONCLUSION

The influence of the density of the mesh structure played a major role in the given study. An increase in the density of such a structure, i.e., a decrease in mesh dimensions, led to an increase in the force that the printed specimens are able to withstand under the assumed parameters, i.e., the maximum compressive force. L2 circle models with a density of 4 × 4 × 4 mm obtained the highest compressive force values when comparing them to samples with densities of 5 × 5 × 5 mm and 6.5 × 6.5 × 6.5 mm.

Specimens with thin walls showed much lower compressive strength than those with thick walls. Regardless of the density of the sponge structure of the L2 vertebral body models. Thicker walls in the specimens improve the absorption and distribution of load during compression tests, which carries with it the obtaining of higher maximum compressive force values compared to specimens with thin walls.

The collapse of the sponge structure and subsequent strengthening of the structure in the case of thin-walled and sparser-textured samples occurred at lower force values. The strengthening of the structure immediately after the collapse is evident when analyzing the resultant graphs of each sample as an increase in force after a previous decrease.

In each group of specimens, the results of the static compression test and the values of the refractive force are very uniform, which indicates good repeatability and homogeneity of the prepared specimens. Slight deviations in the results obtained may be due to small differences in the structure of the material or the conditions under which the tests were performed.

The results obtained show that the proper selection of the density of the sponge structure, as well as the wall thickness, can greatly benefit the mechanical strength of bone implants. The higher the density and the greater the wall thickness, the higher the strength. Which can be important in creating implants with the required properties and strength.

The compressive strength criteria given in the literature for spongy tissue of lumbar vertebral bodies in specimens with thin as well as thick walls are met by a structure with a density of 4 × 4 × 4 mm. Samples with a mesh density of 5 × 5 × 5 mm in the case of thick walls also fall within the specified range, but their strength is slightly lower than for samples with a density of 4 × 4 × 4 mm. The rarest sponge structure, which has a density of 6.5 × 6.5 × 6.5 mm, does not meet the literature values for compressive strength regardless of wall thickness.

6 REFERENCES

[1] Dave, H. K., Prajapati, A. R., Rajpurohit, S. R., Patadiya, N. H., & Raval, H. K. (2020). Open hole tensile testing of 3D printed parts using in-house fabricated PLA filament. *Rapid Prototyping Journal, 26*(1), 21–31. https://doi.org/10.1108/RPJ-01-2019-0003

[2] Long, J., Gholizadeh, H., Lu, J., Bunt, C., & Seyfoddin, A. (2017). Application of Fused Deposition Modelling (FDM) Method of 3D Printing in Drug Delivery. *Current Pharmaceutical Design, 23*(3), 433–439. https://doi.org/10.2174/1381612822666161026162707

[3] Świeczko-Żurek, B. (2009). *Biomateriały*. Gdańsk: Wydawnictwo Politechniki Gdańskiej. (in Polish)

[4] Pezowicz, C. (2019). *Biomechanika kręgosłupa*. Wrocław: Oficyna Wydawnicza Politechniki Wrocławskiej. (in Polish)

[5] Smith, J. (1995). *Ciało człowieka*. Warszawa: ZETDEZET.

[6] Henkel, J., Woodruff, M., Epari, D., et al. (2013). Regeneracja kości w oparciu o koncepcje inżynierii tkankowej — perspektywa XXI wieku. *Bone Research, 1*, 216–248. (in Polish) https://doi.org/10.4248/BR201303002
[7] Wu, W., Geng, P., Li, G., Zhao, D., Zhang, H., & Zhao, J. (2015). Influence of Layer Thickness and Raster Angle on the Mechanical Properties of 3D-Printed PEEK and a Comparative Mechanical Study between PEEK and ABS. *Materials, 8*(9), 5834–5846. https://doi.org/10.3390/ma8095271
[8] Al-Shalawi, F. D., MohamedAriff, A. H., Jung, D.-W., MohdAriffin, M. K. A., Seng Kim, C. L., Brabazon, D., & Al-Osaimi, M. O. (2023). Biomaterials as Implants in the Orthopedic Field for Regenerative Medicine: Metal versus Synthetic Polymers. *Polymers, 15*(12), 2601. https://doi.org/10.3390/polym15122601
[9] Williams, D. F. (Ed.). (1987). *Definitions in biomaterials* (p. 24). Amsterdam-Oxford-New York-Tokyo: Elsevier.
[10] International Organization for Standardization. (2023). *ISO 17295:2023 Additive manufacturing — General principles — Part positioning, coordinates and orientation*.
[11] Barrios, J. M., & Romero, P. E. (2019). Decision Tree Methods for Predicting Surface Roughness in Fused Deposition Modeling Parts. *Materials, 16*, 2574. https://doi.org/10.3390/ma12162574
[12] Wohlers Associates. (2024). *Wohlers Report 2024: Analysis. Trends. Forecasts. 3D Printing and Additive Manufacturing State of the Industry*.
[13] Budzik, G., Magniszewski, M., Przeszłowski, Ł., Oleksy, M., Oliwa, R., & Bernaczek, J. (2021). Torsional strength testing of machine elements manufactured by incremental technology from polymeric materials. *Polimery, 63*(11–12), 830–832. https://doi.org/10.14314/polimery.2018.11.13
[14] Dębski, M., Magniszewski, M., Bernaczek, J., Przeszłowski, Ł., Gontarz, M., & Kiełbicki, M. (2013). Influence of torsion on the structure of machine elements made of polymeric materials by 3D printing. *Polimery, 66*(5), 298–304. https://doi.org/10.14314/polimery.2021.5.3

Authors' contacts:

Łukasz Przeszłowski, Dr
(Corresponding author)
Departament of Machine Design,
Rzeszów University of Technology,
al. Powstańców Warszawy 12, 35-959 Rzeszów, Poland
+48 17865 1318, lprzeszl@prz.edu.pl

Anna Paluch, MSc
Rzeszów, Poland

Grzegorz Budzik, Prof
Departament of Machine Design,
Rzeszów University of Technology,
al. Powstańców Warszawy 12, 35-959 Rzeszów, Poland
+48 17865 1318, gbudzik@prz.edu.pl

Damian Filip, Dr
Department of Orthopedics and Traumatology,
University of Rzeszów,
al. Tadeusza Rejtana 16C, 35-959 Rzeszów, Poland
damian.a.filip@gmail.com

Łukasz Kochmański, MSc
Departament of Machine Design,
Rzeszów University of Technology,
al. Powstańców Warszawy 12, 35-959 Rzeszów, Poland
+48 17865 1318, l.kochmanski@prz.edu.pl

Mariusz Dębski, Dr
Departament of Machine Design,
Rzeszów University of Technology,
al. Powstańców Warszawy 12, 35-959 Rzeszów, Poland
+48 17865 1318, m.debski@prz.edu.pl

ISSN 1846-6168 (Print), ISSN 1848-5588 (Online)
https://doi.org/10.31803/tg-20250424185350

Original scientific papers
Received: 2025-04-24, Accepted: 2025-05-05

Optimizing Milkrun Routes: Heuristic Approach for Minimizing Transportation Loops

Peter Veres

Abstract: Milkrun systems serve as the backbone of material supply for multi-line or parallel manufacturing production, ensuring synchronized and efficient logistics. Despite their importance, route optimization is still largely done manually, leading to inefficiencies. If too few routes are used, transportation time increases, and individual vehicles become overloaded, which is a common issue in current practices. Finding the optimal number of routes is a complex NP-hard problem, requiring advanced computational approaches. While assignment matrices are often available and useful for direct transport allocation, they are not used for optimizing milkrun loops. To address this, a heuristic-based model has been developed that optimizes the number of routes while considering travel length and vehicle capacity. The model and case study integrates genetic algorithms to explore feasible solutions efficiently, balancing the trade-off between too many and too few routes. The proposed approach improves both cost-effectiveness and service reliability in production logistics.

Keywords: assignment problem; heuristic optimisation; milkrun; production logistics; route planning

1 INTRODUCTION

In today's fast-paced manufacturing environment, the efficiency of logistics and supply chain operations has become more critical than ever. Complex manufacturing systems that rely on the movement of materials in a scheduled order with vehicle support, often referred to as milkrun systems, are vital for ensuring smooth and synchronized material supply across multiple production lines or parallel manufacturing processes. These systems help maintain a steady flow of materials to production areas while minimizing delays, reducing costs, and improving overall productivity [1]. The design and optimization of milkrun routes, however, remain a complex challenge that many manufacturing systems still face. Although these systems are integral to manufacturing, route optimization is often carried out manually, leading to inefficiencies in logistics operations. The traditional approach can cause various problems, such as excessive transportation time or overloading of vehicles, which in turn, undermines the potential for cost savings and resource utilization [2].

One of the most significant challenges in optimizing milkrun systems lies in determining the ideal number of routes required for efficient material transportation. This task is classified as an NP-hard problem, meaning that finding the optimal solution is computationally demanding, especially as the number of production lines and logistics variables increases [3]. With too few routes, transportation times may increase, and vehicles can become overloaded or too long to handle, which can lead to delays and decrease service reliability. Conversely, too many routes can result in underutilized vehicles and higher operational costs. Therefore, a balance must be found that optimizes vehicle capacity while minimizing travel time, making the process a delicate one.

To address these challenges, a new approach is created based on the assignment or transportation problem. Traditional methods, most likely drawing on a layout, and using manual greedy optimisation approaches often useful for transport allocation but fall short when it comes to optimizing the routes for milkrun loops considering different material needs of multiple products, capacity and sizes of boxes and trays. The lack of an efficient and systematic approach to route optimization has improved the development of more sophisticated techniques. A heuristic-based model has recently been introduced that focuses on finding the optimal number of routes while considering factors such as vehicle capacity and travel length. This approach integrates genetic algorithms, a type of evolutionary algorithm that mimics the process of natural selection to explore feasible solutions more efficiently.

The application of heuristics to optimize milkrun systems represents a significant advancement in solving this problem [4]. The genetic algorithm offers a flexible and powerful tool for tackling complex optimization task, which is widely integrated into various software platforms and programming languages, making them easily accessible and adaptable across different industries. This flexibility allows it to be applied to a broad range of optimization problems, including route optimisation and assignment problem.

In the following chapter, a literature review will be provided to explore the existing research in the field of milkrun system optimization and how others have approached similar problems. The third chapter will explore the basis of the problem and the methodology for solving it using a genetic algorithm. The fourth chapter will present a case study, showcasing the application of the proposed model in a constructed scenario. The results of this case study will be analyzed, demonstrating the effectiveness and potential of the model. Finally, the last chapter will summarize the findings of the paper, draw conclusions on the success of the proposed approach, and suggest future directions for research in milkrun system optimization.

2 LITERATURE REVIEW

In today's fast-paced manufacturing environment, the efficiency of logistics and material handling plays a vital role in ensuring uninterrupted production flow and cost-effectiveness. Milkrun logistics has become an essential

strategy in modern supply chain and manufacturing systems, especially for industries like automotive, where timely and efficient part delivery is critical. The concept, which originates from the Just-In-Time (JIT) philosophy introduced by the Toyota Production System, involves collecting materials from multiple suppliers and delivering them to a central point or production line in a single, circular route [5]. This approach helps reduce transportation costs, minimize inventory, and improve overall efficiency. Over the years, milkrun systems have grown more complex, now involving challenges such as multi-stop routing, reverse logistics, and diverse vehicle and packing constraints [6]. Solving these problems using exact optimization methods can be time-consuming and impractical for large-scale operations. As a result, heuristic and metaheuristic approaches - such as Ant Colony Optimization, Genetic Algorithm, Simulated Annealing, and tailored routing strategies - have become increasingly popular. These methods provide fast, near-optimal solutions and are highly effective in handling real-world logistics challenges where flexibility and speed are key [7].

Ranjbaran and his team worked on improving how auto parts are delivered and returned using a milkrun system. Their mathematical model considered real-life constraints like truck size, pallet packing, order compatibility, and even the return of empty pallets. To solve large, complex problems quickly, they developed heuristic methods that found good solutions and showed clear cost savings [8].

Salleh and her team focused on reducing transport distance and improving truck space usage in milkrun routes. They found nearly half the truck space was going to waste and showed that this could be reduced to just 3% using optimization methods like Tabu Search and Ant Colony Optimization. Their approach supports the Just-In-Time delivery model by encouraging smaller, more frequent shipments with better efficiency [9]. Sipahioglu and his team also looked at in-plant milkruns but added autonomous vehicles into the mix. With complex delivery needs and timing, they used Simulated Annealing to plan both routes and delivery periods effectively [10]. Uygun and Rustemaj compared traditional Mixed Integer Programming with Ant Colony Optimization for solving milkrun routing problems. While both used vehicle space efficiently, the ant colony method created better delivery routes, showing that nature-inspired algorithms can offer strong alternatives in logistics planning [11].

Bányai offered a simpler method using Excel Solver to plan milkrun routes within factories, helping decide how many carts are needed and where they should go. His method is flexible and works well in different manufacturing setups [12]. In a more recent and advanced study, the same researcher also investigates a hub-based milkrun supply system, showing that integrating warehouse handling with production can significantly reduce service and transportation times. His work presents a new mathematical model that compares traditional milkrun methods with a hub-based approach in the automotive industry. The analysis reveals that a hub-based system can boost efficiency by reducing total service time by about 13% and transportation time by up to 45% in the component warehouse [13]. Roberto Murillo-Ramirez and partner use Bayesian networks to address uncertainty in milkrun systems, quantifying both optimistic and pessimistic scenarios. Their research provides insights into how delays in component deliveries can lead to financial losses in assembly lines [14]. Hashini and Nishal explores vehicle routing models for milkrun deliveries of fresh produce, highlighting the challenges faced by third-party logistics providers. Their study demonstrates that an automated routing solution can cut total costs by 26% and reduce overall travel distance by 40% [15]. The research emphasizes the impact of factors such as vehicle capacity, departure times, and route assignments on overall efficiency. Grzegorz Bocewicz, Izabela Nielsen, and Zbigniew Banaszak examine cyclic systems for in-plant milkrun delivery by looking at both routing and scheduling decisions [16]. Hosseini and partners employ a cyclic scheduling approach to minimize both vehicle downtime and production takt time. Their declarative model synchronizes cyclic flows using a recursive constraint satisfaction method to manage congestion effectively. To solve the NP-hard problem, a hybrid Harmony Search–Simulated Annealing algorithm is developed, which outperforms traditional solvers like GAMS/CPLEX in both cost reduction and computational efficiency, particularly in large-scale scenarios [17].

Each of these studies highlights the value of heuristic and optimization methods in addressing the complex logistics challenges found in milkrun systems. While studies [8–11] rely heavily on metaheuristics without incorporating simpler methods, studies [12–17] combine heuristic approaches with additional deterministic algorithms (like Linear Programming), offering more balanced solutions. Notably, studies [12], [13], and [15] make use of assignation matrices, though only as supportive tools rather than primary optimization methods. Bányai's contributions [12-13] have had a strong influence on my research in this area.

Collectively, this body of literature demonstrates how smart planning, and advanced computational techniques can significantly improve supply chain performance. Bányai's contributions have had a strong influence on my research in this area.

3 MODELLING MILKRUNS WITH ASSIGNMENT MATRICES

Milkrun logistics systems aim to optimize the collection and distribution of goods along predefined, often cyclic, delivery routes. One effective way to represent and manage the decision-making in such systems is using assignment matrices.

3.1 Assignment Matrices in Logistics

These matrices are structured grids, usually binary, where rows represent delivery points (e.g., suppliers or customers) and columns represent vehicles or routes. A value of 1 in a cell indicates that a particular point is assigned to a specific vehicle, while 0 indicates no assignment. This structure allows for a clear and organized formulation of

route allocation, facilitating the modelling of key logistics decisions such as who serves whom and in what order.

When transitioning from simple assignment problems to more complex transportation problems, the assignment matrix can be adapted to include capacity constraints, delivery time windows, and route continuity. This transforms the matrix from a pure allocation model to a more dynamic structure that supports cost-based optimization. In this context, cost matrices become essential, incorporating not only time and distance but also factors such as human labour and machine resources, administrative overhead, installation costs, and other relevant parameters. These matrices feed into the cost function, typically aiming to minimize total distance travelled, fuel consumption, or delivery time.

To solve these models, deterministic methods are preferred when feasible. Linear Programming (LP) and Mixed Integer Linear Programming (MILP) formulations allow for precise optimization, especially when the problem size is moderate, and all parameters are known. These approaches are particularly suitable for industrial applications where predictability and reliability are prioritized over computational speed.

In more complex or large-scale systems where, exact solutions are computationally expensive, minimal heuristic approaches may be introduced - often not as a replacement for exact methods, but as supporting tools. For instance, assignment matrices may be pre-processed using greedy or clustering techniques to reduce the problem size before applying LP or MILP solvers. In other cases this is often solved using heuristic methods, especially in cases with a large number of parameters, multiple objectives, or weighting, since one of the key requirements of linear programming is that the relationships between parameters must be linear.

3.2 Assignment Matrices in Logistics

In this chapter, a model is introduced for optimizing milkrun routes using assignment matrices. The first step involves determining which materials are transported from various warehouses, storage locations, or supermarkets to specific destinations. This task can be solved efficiently and accurately using basic linear programming, producing optimal or near-optimal solutions. This process is demonstrated in the first table of the case study, where Excel Solver's built-in Simplex LP function was used.

The minimum number of required milkruns can also be calculated easily based on a perfectly constructed assignment matrix. In such cases, the minimum is determined by the smaller number between the matrix's rows or columns, for example a 4×7 matrix the exact number of assignments is 4. However, in transportation problems, adjustments are needed, since a single destination, such as a production line, may be served by multiple milkruns if one vehicle cannot fulfil the demand alone, or if the cost matrix yields a more efficient configuration. These decisions are driven by material demand and vehicle capacity constraints.

A distinctive feature of the model lies in its ability to reduce the number of milkruns by identifying and merging destinations and starting points visited by multiple originally assigned routes. Through this consolidation, redundant routes can be eliminated, improving overall efficiency and lowering transportation costs.

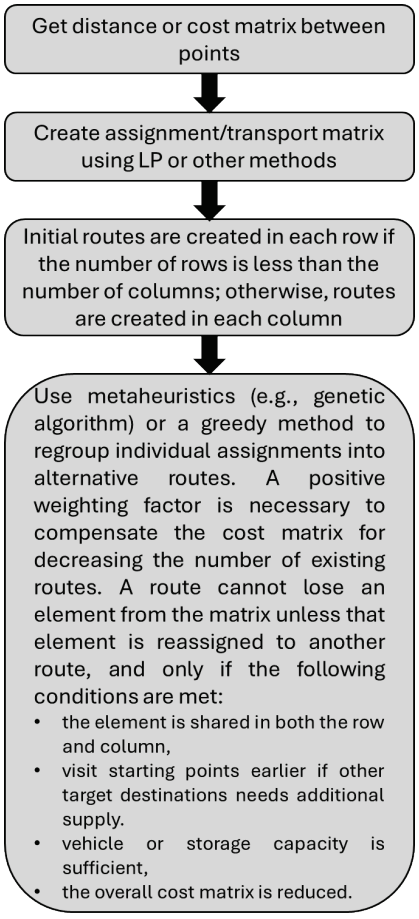


Figure 1 Schematic Description of the Model

The schematic operation of the algorithm is shown in Fig. 1. Since no new development or proposal has been made from a mathematical or algorithmic standpoint, this part will not be discussed in detail. Using the distance and cost matrices - derived from the positions and other characteristics of the points - it is possible to construct the actual assignment matrix through relatively simple methods such as LP or MILP. This matrix remains unchanged afterward. As mentioned earlier, the initial number of routes will equal the minimum of the number of rows and columns. The upper part of Figure 2 illustrates this: it displays a 4x7 matrix where the rows represent four suppliers (starting points), and the columns represent seven consumers (destinations). Since the number of suppliers is smaller, we start with four initial routes, shown in yellow, red, green, and blue. The pale-coloured cells indicate no delivery, because no connection was established by the LP algorithm - while the more vivid cells represent active assignments.

Considering the lower part of Fig. 2, many assignment or transport tasks can be reallocated in such a way that a milkrun can not only connect destinations but also include other starting points in its route, if delivery is required from those

points to the target destination. This redistribution can be optimized using heuristic, metaheuristic, or greedy methods. While greedy methods can sometimes yield optimal solutions in smaller problems, for larger instances, the chances of reaching a perfect solution are considerably lower.

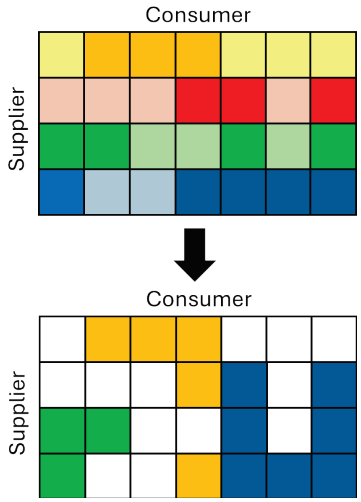


Figure 2 Visual representation of milkrun routes regrouping

Since the process involves extending routes - by including starting points that, under the original setup, were not meant to be part of the route network - it is almost inevitable that the resulting routes will be longer compared to the original multiple-route configuration. This can be compensated by assigning a high reward weight in the cost matrix whenever an existing route is eliminated. Unfortunately, determining the appropriate value of this weight is largely based on practical experience. If the goal is to strongly encourage the reduction of routes, a reward for eliminating a route should be roughly the same order of magnitude to the total cost of the original cost matrix.

A simple and practical approach to define the weight of a single route is to take the total cost of the original distribution and divide it by the number of routes. This can be represented as a function of: Summarized cost + n(route) × weight(route). In this case, the weight would be roughly on the same scale as the original total cost, but removing a route would still result in a significant value change, justifying it. Of course, this weight can be increased, but doing so would diminish the importance of optimizing the internal routing network, shifting the focus purely to reducing the number of routes. On the other hand, lowering the weight would emphasize the optimization of the internal network structure rather than simply reducing the number of routes. This is just a recommendation, but it is also applied this way in the current research.

When designing and configuring higher-level algorithms, special attention must be paid to several critical limitations. The most important of these is that a route cannot lose an element from the assignment matrix unless that element is reassigned to another route, and only if the following conditions are satisfied:

- the element appears in both a common row and column (i.e., there is an overlap),
- the vehicle or storage capacity is sufficient,
- visit other starting points earlier if multiple destinations still have demand from them,
- the overall cost matrix is reduced as a result with the weight of the route elimination.

Heuristic algorithms generally run until they can no longer improve the current solution for a significant number of iterations, and this is the approach recommended here as well. However, in large-scale problems - where matrix sizes may reach into the hundreds - there is a high chance that long periods go without any improvement. In contrast, running a greedy algorithm across various initial swap configurations can often yield good solutions early in the optimization process.

4 CASE STUDY ON SIMPLE TASK

The actual operation of the model and algorithm will be demonstrated on a relatively small, artificially created problem. According to the task description, a manufacturing company operates 7 product line and has 4 available storage points. The actual distances between them are shown in Tab. 2. The storage locations have different capacities, while the production lines have varying demands, represented by unit quantities. These demands, along with the distances, are displayed in the upper table of Fig. 3. For simplicity, the capacity of each milkrun is set to 200 units. In the middle table of Fig. 3, the assignment was solved using Excel Solver with the Simplex LP method, resulting in the formation of four groups starting from the four storage points. Also visible in Figure 3, the middle table shows the optimal distances between these points, and the bottom table displays the optimal route alongside the cost function. In this case, the total distance covered is 579.6 meters, and all demands have been fully satisfied.

Table 1 Distance matrix of task production

Distance (m)	S1	S2	S3	S4	PL1	PL2	PL3	PL4	PL5	PL6	PL7
Storage 1	0	37.5	87.3	155.4	6	67.2	136.3	35.6	71.9	118.9	142
Storage 2	37.5	0	64.8	132.9	43.3	79.9	149.4	14.3	51.6	98	119.5
Storage 3	87.3	64.8	0	101.5	81.3	48.5	118	51.7	20.2	66.6	88.1
Storage 4	155.4	132.9	102	0	149.4	112.4	43.3	119.8	83.5	36.5	13.4
Product Line 1	6	43.3	81.3	149.4	0	61.4	130.5	29.6	65.9	112.9	136
Product Line 2	67.2	79.9	48.5	112.4	61.4	0	93.9	65.6	28.9	75.9	99
Product Line 3	136.3	149.4	118	43.3	130.5	93.9	0	135.1	97.8	51.4	29.9
Product Line 4	35.6	14.3	51.7	119.8	29.6	65.6	135.1	0	38.5	84.9	106.4
Product Line 5	71.9	51.6	20.2	83.5	65.9	28.9	97.8	38.5	0	48.6	70.1
Product Line 6	118.9	98	66.6	36.5	112.9	75.9	51.4	84.9	48.6	0	23.1
Product Line 7	142	119.5	88.1	13.4	136	99	29.9	106.4	70.1	23.1	0

By applying the described algorithm, it was not only able to reduce the number of milkrun routes by one, but also managed to lower the total distance - without the need to define any additional weighting factors. This result was achieved by merging the two routes originating from S1 and S4 through their shared point PL3, using Excel Solver's built-in evolutionary algorithm, which also determined the optimal sequence. As a result, the total combined distance was reduced to 555.2 meters. This can be seen on Fig. 4.

Distance (m)	PL1	PL2	PL3	PL4	PL5	PL6	PL7	Service capacity (pcs)
Storage 1	6.0	67.2	136.3	35.6	71.9	118.9	142.0	140
Storage 2	43.3	79.9	149.4	14.3	51.6	98.0	119.5	100
Storage 3	81.3	48.5	118.0	51.7	20.2	66.6	88.1	170
Storage 4	149.4	112.4	43.3	119.8	83.5	36.5	13.4	80
Demands (pcs)	40	100	70	60	50	20	70	

Assignment matrix (pcs)	PL1	PL2	PL3	PL4	PL5	PL6	PL7	Utilized capacity (pcs)	Optimal distance (m)
Storage 1	40		60					100	272.8
Storage 2				60				60	28.6
Storage 3		100			50	20		170	191.6
Storage 4			10				70	80	86.6
Delivered (pcs)	40	100	70	60	50	20	70		579.6

Cost matrix	PL1	PL2	PL3	PL4	PL5	PL6	PL7	Utilized capacity (pcs)	Optimal route
Storage 1	240		8178					100	S1-PL1-PL3-S1
Storage 2				858				60	S2-PL4-S2
Storage 3		4850			1010	1332		170	S3-PL5-PL2-PL6-S3
Storage 4			433				938	80	S4-PL3-PL7-S4
Delivered (pcs)	40	60	70	60	50	20	70		

Summarized Cost	17839

Figure 3 The initial solution provided by the LP method

Assignment matrix (pcs)	PL1	PL2	PL3	PL4	PL5	PL6	PL7	Utilized capacity (pcs)	Optimal distance (m)
Storage 1	40		60					100	335.0
Storage 2				60				60	28.6
Storage 3		100			50	20		170	191.6
Storage 4			10				70	80	
Delivered (pcs)	40	100	70	60	50	20	70		555.2

Cost matrix	PL1	PL2	PL3	PL4	PL5	PL6	PL7	Utilized capacity (pcs)	Optimal route
Storage 1	240		8178					180	S1-PL1-S4-PL7-PL3-S1
Storage 2				858				60	S2-PL4-S2
Storage 3		4850			1010	1332		170	S3-PL5-PL2-PL6-S3
Storage 4			433				938		
Delivered (pcs)	40	60	70	60	50	20	70		

Summarized Cost	17839

Figure 4 The final solution by applying the proposed algorithms

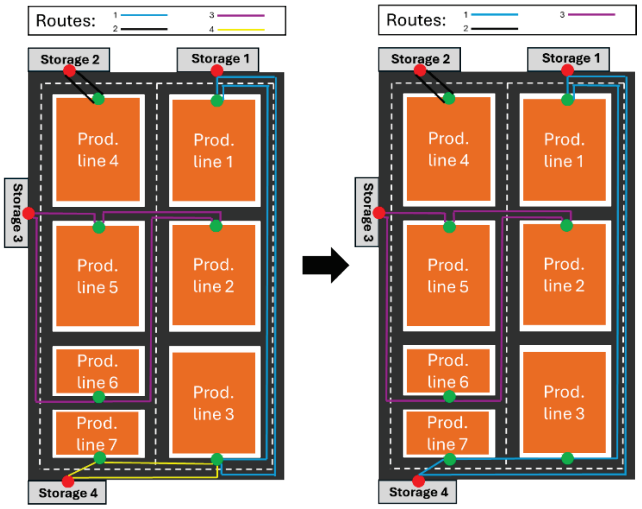


Figure 5 Visual solutions of the two phases of proposed algorithm

Fig. 5 shows how the algorithm slightly modified the route, along with the layout of the facility, from which the point locations were accurately scaled and the distances calculated.

5 CONCLUSION

The research presented in this paper offers an efficient solution to one of the most challenging aspects of milkrun logistics: determining the optimal number of delivery routes while maintaining cost-effectiveness and service reliability. While milkrun systems are widely adopted for supplying multi-line or parallel production in manufacturing, route planning is still frequently performed manually, which often leads to overloaded vehicles, inefficient transport loops, and increased operational costs. Recognizing this, the developed model builds on the structure of assignment matrices - commonly used for transport allocation - but extends their function to optimize full milkrun loops.

Using Linear Programming and Excel Solver's Simplex LP method, the initial assignment matrix is generated based on distance and cost matrices derived from scaled layout data of a production facility. In the test scenario, a manufacturing environment with 7 production lines and 4 storage points was analysed. The base LP solution initially assigned four routes, one per storage location, with a total transportation distance of 579.6 meters. However, through the application of heuristic optimization—using Excel Solver's evolutionary algorithm—the number of routes was successfully reduced from four to three. This was achieved by merging the routes from S1 and S4 through a common delivery point, PL3, resulting in an optimized transport sequence with a reduced total distance of 555.2 meters.

The proposed model is applicable to any scenario where an assignment or transport matrix can be constructed, making it highly versatile across different logistics systems and manufacturing environments. However, there is still room for further improvement, particularly in the first phase of the algorithm. A known limitation of this initial phase is that it can only handle a limited number of variables effectively. This becomes especially problematic when the relationships between parameters are non-linear or not appropriately weighted. In such cases, even the assignment step may no longer be solvable by traditional means and must instead rely on heuristic approaches to generate feasible solutions. Another issue can be the size and complexity of the problem. When dealing with larger matrices or multidimensional constraints (e.g., time windows, variable capacities, or compatibility rules), the algorithm may not be able to run efficiently.

In summary, this research demonstrates how a structured and flexible model based on assignment matrices and heuristic algorithms can significantly improve route optimization in milkrun logistics. By combining mathematical precision with algorithmic adaptability, the model contributes a valuable tool for enhancing efficiency in modern production supply chains.

Acknowledgements

Supported by the University Research Scholarship Program of the Ministry for Culture and Innovation from the source of the National Research, Development and Innovation Fund - Egyetemi Kutatói Ösztöndíj Program (EKÖP).

6 REFERENCES

[1] Gotthardt, S., Hulla, M., Eder, M., Karre, H., & Ramsauer, C. (2019). Digitalized milk-run system for a learning factory assembly line. *Procedia Manufacturing*, *31*, 175-179.

https://doi.org/10.1016/j.promfg.2019.03.028

[2] Kochańska, J., Burduk, A., & Małysiak, A. (2022, June). Improvement of Internal Transport with the Use of Milk Run Concept. In *International Conference Innovation in Engineering* (pp. 234-247). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-031-09360-9_20

[3] Bocewicz, G., Nielsen, I., & Banaszak, Z. (2021). Rerouting and rescheduling of in-plant milk run based delivery subject to supply reconfigurability constraints. In *Performance Evaluation Models for Distributed Service Networks* (pp. 55-78). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-67063-4_4

[4] Sipahioglu, A., Acar, I., & Altin, I. (2024). A Metaheuristic Approach for In-Plant Milk-Run System with Autonomous Vehicles. *Networks and Spatial Economics*, *24*(4), 1021-1041. https://doi.org/10.1007/s11067-024-09650-2

[5] Monden, Y. (2011). *Toyota production system: an integrated approach to just-in-time*. CRC press.

[6] Ni, Q., & Tang, Y. (2023). A bibliometric visualized analysis and classification of vehicle routing problem research. *Sustainability*, *15*(9), 7394. https://doi.org/10.3390/su15097394

[7] Rahman, M. A., Sokkalingam, R., Othman, M., Biswas, K., Abdullah, L., & Abdul Kadir, E. (2021). Nature-inspired metaheuristic techniques for combinatorial optimization problems: Overview and recent advances. *Mathematics*, *9*(20), 2633. https://doi.org/10.3390/math9202633

[8] Ranjbaran, F., Husseinzadeh Kashan, A., & Kazemi, A. (2020). Mathematical formulation and heuristic algorithms for optimisation of auto-part milk-run logistics network considering forward and reverse flow of pallets. *International Journal of Production Research*, *58*(6), 1741-1775. https://doi.org/10.1080/00207543.2019.1617449

[9] Salleh, N. A. M., Ahmad, S. F., Shaharudin, M. R., & Rahim, S. A. (2023). Evaluation of Transportation Distance Optimization Route for Milk Run Logistics System. *International Journal of Supply & Operations Management*, *10*(3).

[10] Altin, I., & Sipahioglu, A. (2020, December). Simulated annealing algorithm for in-plant milk-run system. In *International Conference of Production Research–Americas* (pp. 187-201). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-76307-7_15

[11] Uygun, Y., & Rustemaj, A. (2022). Ant colony optimisation for milk-runs in manufacturing systems. *International Journal of Advanced Operations Management*, *14*(2), 167-181. https://doi.org/10.1504/IJAOM.2022.123267

[12] Bányai, T. (2023). Optimization of routing problems in manufacturing supply. *Advanced Logistic Systems-Theory and Practice*, *17*(4), 14-22. https://doi.org/10.32971/als.2023.026

[13] Bányai, T. (2024). Optimization of Hub-Based Milkrun Supply. *Logistics*, *8*(3), 86. https://doi.org/10.3390/logistics8030086

[14] Murillo-Ramirez, R., & Lizarraga-Lizarraga, G. (2021). Uncertain analysis based on milk-runs systems using Bayesian networks. *Technological and Industrial Applications Associated with Intelligent Logistics*, 597-610. https://doi.org/10.1007/978-3-030-68655-0_30

[15] Kodippili, H., & Samarasekera, N. (2019). Vehicle routing model for milk run delivery of fresh produce: The case of a 3PL service provider catering supermarkets. In *the 9th International Conference on Operations and Supply Chain Management*, Vietnam.

[16] Bocewicz, G., Nielsen, I., & Banaszak, Z. (2019). Blockage-free route planning for in-plant milk-run material delivery systems. In *Modelling and Performance Analysis of Cyclic Systems* (pp. 105-132). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-27652-2_6

[17] Hosseini, S. D., Shirazi, M. A., & Karimi, B. (2014). Cross-docking and milk run logistics in a consolidation network: A hybrid of harmony search and simulated annealing approach. *Journal of Manufacturing Systems*, *33*(4), 567-577. https://doi.org/10.1016/j.jmsy.2014.05.004

Author's contacts:

Peter Veres, Associate Professor
University of Miskolc,
Egyetem str. 1, HU-3515 Miskolc-Egyetemvaros, Hungary
+36 (46) 565-111/1779, peter.veres@uni-miskolc.hu

ISSN 1846-6168 (Print), ISSN 1848-5588 (Online)
https://doi.org/10.31803/tg-20250520085334

Original scientific paper
Received: 2025-05-20, Accepted: 2025-06-09

Printability Analysis of PLA Films Using Flexographic Solvent-Based Ink According to ISO Specifications

Dean Valdec*, Igor Majnarić, Luka Valdec

Abstract: This study analyzes the influence of photopolymer plate hardness and mounting tape on print quality in flexography, with a focus on evaluating the optical and colorimetric parameters of prints on white PLA film. The experiment was carried out in six variations of printing: two hardness of polymer plates in combination with three hardness of mounting tapes. A test image was printed using solvent cyan ink and all key print quality parameters were analyzed, including ink density, tonal values, and CIELAB values in accordance to ISO 12647-6:2012 standard. The results showed that combinations using the softer plate with harder tapes provided the optimal balance between halftone precision, solid density, and colorimetric accuracy. Combinations using the harder plate resulted in lower ink transfer but also less dot gain. Microscopic analysis revealed that combinations of harder plates with softer tapes led to lower ink film uniformity. The study provides useful guidelines for optimizing flexographic printing on PLA substrates.

Keywords: flexography; mounting tape; polymer plate; PLA films; print quality

1 INTRODUCTION

Flexography (often abbreviated to flexo) occupies a leading position in the packaging industry due to its ability to print on a wide range of substrates, including plastics, paper, cardboard, metallized films and non-absorbent substrates [1]. Technological developments in the last decade have enabled a significant improvement in the quality of reproduction in flexographic printing, bringing this technique significantly closer to the quality of gravure and offset printing, especially in segments such as flexible packaging printing, roll-to-roll labels and shrink sleeves [2]. One of the key challenges in flexographic printing is achieving consistent print quality, which includes control of optical density, tonal values, dot formation and color fidelity [3-5]. Parameters that significantly affect these aspects of print quality are the characteristics of the elements that directly participate in the printing process and create pressure on the substrate – primarily clichés and mounting tapes. Polymer clichés, made of flexible photopolymer materials, are available in different hardnesses, which allows adaptation depending on the type of print, printing substrate and print quality requirements [6]. The hardness of the cliché affects the volume of ink transfer, the contact pressure between the printing plate and the printing substrate, and the level of deformation of the cliché in the printing zone [7].

An equally important role is played by the mounting tape, which attaches the cliché to the plate cylinder. Depending on its hardness and compressibility, the tape can soften or intensify the deformations that occur under the pressure of the nip contact, which affects the reproduction of fine details, the stability of the halftone dot and the appearance of unwanted artifacts such as mottle, pinholing and halo effect [8]. Soft tapes enable better contact on rougher surfaces and a larger volume of ink transfer, while hard tapes ensure greater sharpness and control of halftone reproduction, but at the same time increase the risk of excessive pressure and cliché deformation [9].

The relationship between the hardness of the mounting tape and the cliché can be viewed as a dynamic system in which complex mechanical interactions occur during printing. In this system, it is particularly important to understand how these interactions affect measurable print quality parameters. Among them, the solid color density provides information about the amount of ink transferred, while the halftone values show how accurately the printing system has transferred the ink to the substrate. Colorimetric reproduction accuracy, expressed in terms of CIELAB coordinates, additionally describes how visually faithful the color in the print is to the reference.

For the purposes of quality control and mutual comparison of prints, the international standard ISO 12647-6 is used in flexo printing, which defines reference values and tolerances for key reproduction parameters [10]. The use of ISO standards allows for objective assessment of results and comparison with industry norms.

One of the most commonly used materials in the production of flexible packaging is low-density polyethylene (LDPE). However, an environmentally friendly alternative, polylactic acid (PLA), is one of the most promising biopolymers in the field of sustainable packaging [11]. Its application in flexible packaging is becoming increasingly important given the increasingly stringent environmental regulations and growing public concern about plastic pollution [12]. PLA can be formed using standard plastic processing technologies such as extrusion, film blowing and lamination. As a material for flexible packaging, PLA films are most often produced in thicknesses of 40 – 50 microns and are used for the production of food packaging, bags and laminated multilayer materials.

Milky white PLA film are increasingly used in flexible packaging due to its visual neutrality, compatibility with printing inks, and the ability to replace white-colored PE/PP films [13]. In addition, white PLA film provides a light barrier, which is important for the protection of photosensitive products (e.g., food and cosmetics). In the context of printing, PLA represents a challenging substrate

due to its low surface energy, which makes it difficult for inks to adhere and requires additional surface preparation before printing, such as corona treatment (surface tension, $\sigma \geq 38$ dyn/cm) [14, 15]. For this reason, the quality of printing on PLA film largely depends on the optimization of all printing parameters – including the choice of clichés, mounting tapes, ink types, and printing conditions. In flexographic printing, PLA is most often used in combination with solvent inks that provide better adhesion and abrasion resistance.

Given the ubiquity of PLA films in modern packaging solutions, the study of parameters that affect the print quality on this material is of high practical importance. The results of this research can contribute to the optimization of the printing process, especially in production conditions where it is important to reconcile the requirements for high print quality with the economy and repeatability of the process.

2 METODOLOGY

As part of this research, an experimental method (Fig. 1) was designed to test the influence of the hardness of the mounting tape and the hardness of the polymer cliché on the quality of solid tone and halftone reproduction in the flexographic printing process.

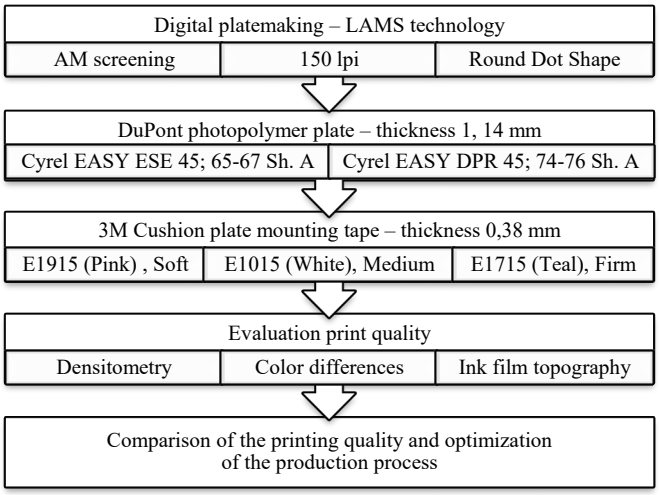


Figure 1 Research framework

For the purposes of the research, an industrial flexo machine for printing flexible packaging was used, and a milky white PLA film, 40 µm thick, was selected as the printing substrate. This material was chosen because of its competitiveness with classic white PE and PP films, with the additional advantage of compostability and environmental friendliness [16]. Its optical properties as well as physical characteristics, especially non-absorbentness and flexibility, make color reproduction challenging. It is most often used in the packaging industry for food products and hygiene products.

2.1 Digital Platemaking Process

The experimental part of this study began with the creation of a test image containing color control patches. These measuring patches allow for the evaluation of print quality using recognized and validated scientific methods and techniques. An amplitude modulated halftone screen, 150 lpi line screen with a round dot shape was used in the screening process. Two types of polymer plates of different hardness, manufactured by DuPont, were used to make the printing plate. The first is a softer plate (Cyrel EASY ESE), hardness 65 – 67 Shore A, intended for rougher printing substrates and lower printing pressures. The second is a harder plate (Cyrel EASY DPR), with a hardness of 74 – 76 Shore A, which is intended for smooth printing substrates and more demanding halftone reproduction. Both clichés were made using the digital CTP process (Computer-to-Plate) on the ESKO CDI Spark 4260 device, using the principle of laser removal of the LAMS layer of the polymer plate (Laser Ablation Mask System), which ensured high precision in the production of printing elements. The cliché thickness was standard – 1.14 mm, which meets the requirements for packaging flexographic printing.

2.2 Printing Process

Printing was performed on an eight-color UTECO Diamond HP 808 central drum flexographic press with a web width of 1250 mm. The process followed the "from roll-to-roll" principle at a printing speed of 200 m/min, using cyan solvent-based inks. The plates were mounted on the plate cylinder using 3M Cushion-Mount E-series mounting tapes, all 0,38 mm thick, which differed in hardness (compressibility). Three types of tapes were used:

- E1915 Plate Mounting Tape (Pink): soft tape, characterized by high compressibility, suitable for a more even print and better adaptation to uneven surfaces (abbreviated term in the article: Soft Tape)
- E1015 Plate Mounting Tape (White): medium-hardness, standard variant suitable for balanced reproduction (abbreviated term in the article: Medium Tape)
- E1715 Plate Mounting Tape (Blue-Green): hard tape, low compression, which ensures better edge formation on the print, but can increase mechanical pressure on the dots (abbreviated term in the article: Firm Tape).

The printing experiment (Tab. 1) was performed in six stages, and in each stage different hardness of the polymer cliché (Softer and Harder) and different hardness of the mounting tape (Soft, Medium and Firm) were used. Each of the two polymer clichés was mounted on the plate cylinder using three different mounting tapes, resulting in six combinations. The test image contained elements for quantitative and qualitative analysis of the print. Solid tone patches were included in the test image to measure optical density and CIELAB values, as well as halftone patches of 25%, 50% and 75% tonal values, which enabled the evaluation of actual tonal values on the print.

Table 1 Technical specifications of the printing process

Printing Specifications	
Flexo printing press	UTECO Diamond HP 808 Mod 120 Flexo CI Printing Press, 8 colors
Printing speed	200 m/min
Web / print width	1250 / 1200 mm
Print repeat min-max	360 – 800 mm
Flexo ink	Sun Chemical SunSpectro Process Cyan solvent based ink
Substrate	PLA Milky White Film; thickness 40 µm
Plate mounting tape, thickness 0,38 mm (0,015 inch)	
Tape 1	3M Cushion-Mount Plate Mounting Tape E1915, Polyethylene Foam, Pink, Soft
Tape 2	3M Cushion-Mount Plate Mounting Tape E1015, Polyethylene Foam, White, Medium
Tape 3	3M Cushion-Mount Plate Mounting Tape E1715, Polyethylene Foam, Teal, Firm
Photopolymer plate, thickness 1,14 mm (0,045 inch)	
Polymer 1	DuPont Cyrel EASY ESE 45, Hardness 65-67 Shore A – Softer plate
Polymer 2	DuPont Cyrel EASY DPR 45, Hardness 74-76 Shore A – Harder plate

Plate-to-substrate printing pressure is a factor that has a major impact on ink transfer and is defined by nip engagement (Fig. 2). The nip engagement between the plate and the substrate was kept constant throughout the experiments at 75 µm, as any variation in this parameter would affect the deformation of the plate against the substrate. Nip is a line of contact between two rolls, plate and impression cylinder. Nip engagement is the positive displacement of the printing plate to the substrate, specified in µm, which increases the printing pressure [17].

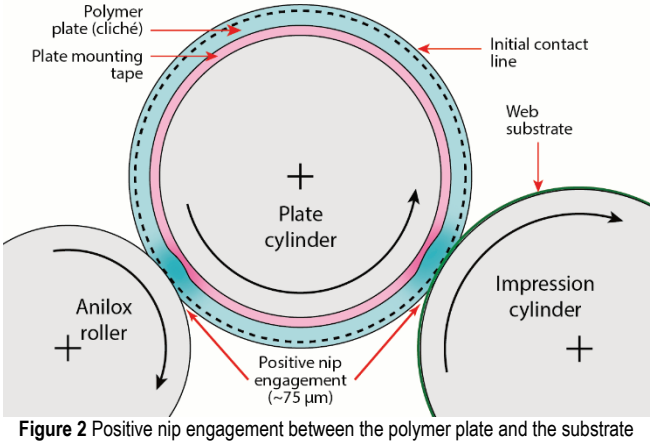


Figure 2 Positive nip engagement between the polymer plate and the substrate

Lower pressure (nip engagement < 50 µm), makes it difficult to transfer ink evenly, primarily when printing on rough printing surfaces. Conversely, if the pressure is too high (nip engagement exceeds 150 µm), the dots will be compressed further, leading to potential deformities [18].

A test form was printed in single color – cyan, using solvent ink. Using one color allowed for isolation of the analyzed parameters and avoidance of interference between colors. The ink was previously prepared and brought to optimal viscosity in accordance with the manufacturer's recommendations and maintained uniformly throughout the printing process.

2.3 Evaluation

To ensure objectivity and comparability of the results, all measurements were performed in accordance with the international standard ISO 12647-6:2012 [10]. According to this standard (for standard flexographic printing), reference CIELAB values for solid tone process inks are defined and are valid for printing on white printing substrates (films and papers) that do not contain optical brighteners. Reference CIELCH values (rounded to the nearest integer) are also defined based on the CIELAB values. The values refer to standard measurement conditions under D50 illumination and a viewing angle of 2°. Tolerances for L^*, a^*, b^* (±3) and tolerances for C^*, $h°$ (±4) are used for precise control of individual components of process inks under industrial printing conditions, to ensure that printers can reproduce colors correctly despite small variations. In this way, tolerances allow the reproduction of perceptually similar colors, but even within this range, differences in perception may occur. The different tolerance for the CIELAB and CIELCH systems is related to the way colors are defined in 3D color space (coordinate or cylindrical system) and how the difference between colors is perceived.

However, as a key condition for determining an acceptable perceptual difference between colors in the flexographic printing industry, the parameter $\Delta h_{ab} < 6°$ is used, and it refers to the change in color hue. Δh_{ab} is a wider tolerance that allows greater perceptual freedom in color, taking into account even a small change in chromaticity and lightness. The reason lies in the fact that people do not perceive colors equally in all directions: even a small change in chromaticity (C^*) or lightness (L^*) can mean that two colors can have different perceptions despite a minor difference in hue (h°). As specified by ISO 12647-6:2012, the recommended CIE LAB values of the print substrate are as follows: L^* should be greater than 88, a^* should range from –3 to +3, and b* should range from –5 to +5. The tested PLA printing substrate meets the stated criteria ($L^* = 91{,}59\%$, $a^* = -0{,}23$, $b^* = 1{,}56$) and has a slight yellowish tone, characteristic of white biodegradable films, as shown by the positive value of the b^* coordinate.

Each of the six variations characterized by the overall compressibility of the printing plate to the substrate was applied in the printing process with the same parameters – printing speed, pressure force and ink viscosity – to eliminate the influence of external variables.

Colorimetric and densitometric measurements of the prints were taken using an X-Rite Exact handheld spectrometer with a 45°/0° geometry, lighting conditions D50/2°, and a 2 mm aperture size. The average value of three measurements is calculated for each observed print quality parameter. The increase in tone value, considered the primary indicator of reproduction quality, is defined as the difference between the measured and nominal dot area coverage of the corresponding color patches. Tonal value increase (*TVI*) is not necessary fully compensated, as this would result in color reproduction that is too light. A controlled TVI within the limits specified by the ISO standard is entirely acceptable phenomenon.

Tab. 2 presents subjective color difference metrics that align with human perception, utilizing the CIEDE2000 equation, which more accurately reflects how human observers perceive color differences [20].

Table 2 Subjective color difference metrics based on the CIEDE2000 equation [20]

k	$\Delta E_{min}(k)$	$\Delta E_{max}(k)$	Perception of color differences
1	0,0	0,5	Hardly
2	0,5	1,5	Slight
3	1,5	3,0	Noticeable
4	3,0	6,0	Appreciable
5	6,0	12,0	Much
6	12,0	24,0	Very much
7	24,0	∞	Strongly

For analyzing the uniformity of the ink film on the print, solid cyan patches were captured throughout all six stages of the printing experiment. For this analysis, a Dino-Lite AM4000 digital microscope was employed, featuring a 1,3 MP resolution and an integrated LED light for enhanced color sample visualization. Samples were captured at 200× magnification and a resolution of 1280×1024 pixels. The Interactive 3D Surface Plot feature, integrated into the ImageJ 1.47 software, was utilized for processing and analyzing the microscopic images.

This methodological approach enabled a clear identification of the impact of individual components of the printing system on solid tone and halftone reproduction, as well as making recommendations for optimizing printing in flexographic packaging production.

3 RESULTS AND DISCUSSION

3.1 Colorimetric Values of Solid Cyan

Using the measured CIELAB and corresponding CIELCH values for solid cyan (Tab. 3) across the six stages of the printing experiment, along with the reference values specified by the ISO 12647-6:2012 standard, color difference value, difference in lightness and hue difference of the samples were calculated (Tab. 4).

Table 3 Colorimetric CIELAB and CIELCH values for solid cyan

Solid Cyan		L (%)	a^*	b^*	C_{ab} (%)	h_{ab} (°)
Reference ISO 12647-6		54 ± 3	–37 ± 3	–50 ± 3	63,0 ± 4	233° ± 4°
Softer Plate	Soft Tape	52,02	–33,61	–55,24	64,66	238,68
	Medium Tape	51,56	–34,75	–51,75	62,33	236,12
	Firm Tape	51,65	–35,92	–52,61	63,70	235,68
Harder Plate	Soft Tape	56,78	–34,38	–57,84	67.29	239,27
	Medium Tape	55,95	–35,85	–57.88	68,08	238,22
	Firm Tape	56,18	–35,18	–56,89	66,89	238,27

The graph in Fig. 3 shows the values of tolerances ΔE₀₀, Δh_{ab} and ΔL for the solid tone of cyan for six stages of the printing experiment compared to the permitted deviations defined in accordance with the ISO 12647-6:2012 standard, which are shown in the figure with dashed lines.

The ΔE₀₀ values, which represent the total visual difference between the printed color samples and the given reference, in all combinations except for the combinations of the softer cliché with the Medium and Firm tape slightly exceed the upper tolerance limit ($\Delta E_{00} = 3$). For the two highlighted samples, the values are within the recommended tolerance ($\Delta E_{00} \approx 2.5$), and according to Tab. 2, the difference in color is "Noticeable", while for the other four samples it is already "Appreciable".

Table 4 Difference in lightness and hue, color differences for solid cyan

Solid Cyan		ΔL	ΔE_{00}	Δh_{ab}
Reference ISO 12647-6		$\Delta L \leq 3$	$\Delta E_{00} \leq 3$	$\Delta h_{ab} < 6°$
Softer Plate	Soft Tape	–1,98	3,03	5,68
	Medium Tape	–2,44	2,65	3,12
	Firm Tape	–2,35	2,51	2,68
Harder Plate	Soft Tape	+2,78	3,81	6,27
	Medium Tape	+1,95	3,09	5,22
	Firm Tape	+2,18	3,13	5,27

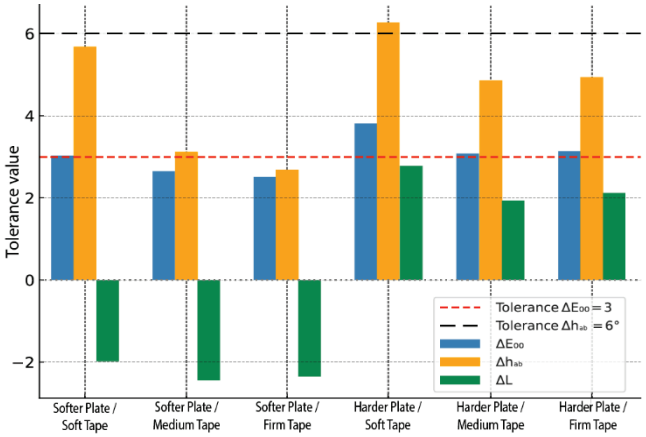


Figure 3 Tolerance values for solid cyan in relation to the permitted values according to the ISO 12647-6:2012

However, ΔE_{00} does not say anything about the colorimetric color change of the samples, but only determines whether the deviation is acceptable or not. A similar pattern to the ΔE_{00} deviation can be seen in the Δh_{ab} values, which represent the change in color hue expressed in degrees – the smallest changes were also noted for combinations of the softer cliché with Medium and Firm tape. The biggest deviations in the color hue were observed in combinations with Soft tapes, where the values of Δh_{ab} are at the upper limit of the deviation of 6°. This indicates the mechanism that Soft tapes, due to its greater compressibility, enable increased cliché contact with the substrate, which leads to increased ink transfer (Fig. 4, image A).

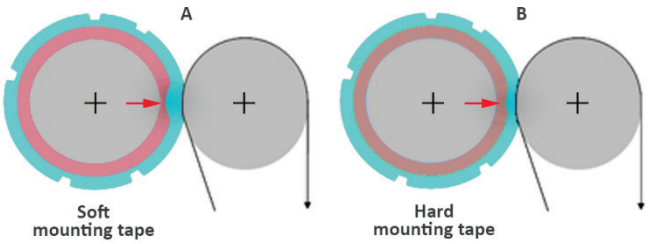


Figure 4 Compressibility effect of the hardness of the mounting plate on the total impression pressure

The ΔL value indicates the difference in lightness of the sample compared to the reference. Negative values mean that

the print is darker than the reference, while positive values indicate a lighter result. All combinations with a softer cliché showed negative ΔL values ($\Delta L \approx -2$), which indicates a slightly darker print, but still within tolerance. On the other hand, the use of a harder cliché resulted in lighter prints, with the largest deviation in the combination with Soft tape ($\Delta L = +2{,}78$), and with increasing tape hardness these deviations decrease. These results confirm the assumption that harder clichés transfer less ink due to its lower compressibility and reduced contact with the printing substrate (Fig. 4, image B), which at the same time gives a significant tone deviation $\Delta h_{ab} > 5°$.

3.2 Halftone Reproduction

In contemporary graphic technology, print quality is largely determined by the fidelity of halftone dot reproduction. The densitometric measured method enables evaluation of all significant print quality parameters. A key factor that affect print quality and the accuracy of dot reproduction is the variation in dot size, which can lead to shift in tone and brightness of color. Therefore, defining TVI is a crucial parameter for characterizing print processes. Table 5 shows the values of the optical density and TVI of cyan measured at three characteristic measurement patches (25%, 50% and 75% TV) for the six phases of the printing experiment. The table also contains the TVI reference values in accordance with ISO specifications for comparison.

Table 5 Optical density and tone value increase for cyan sample

Combinations Plate / Tape	Optical Density *OD*	Tone Value Increase *TVI* (%)		
		25 %	50 %	75 %
Softer Plate / Soft Tape	1,44	22,61	29,93	21
Softer Plate / Medium Tape	1,55	23,18	32,02	21,57
Softer Plate / Firm Tape	1,60	24,23	33,82	21,95
Harder Plate / Soft Tape	1,08	15,49	20,24	14,06
Harder Plate / Medium Tape	1.,19	19,19	23,94	17,77
Harder Plate / Firm Tape	1,25	20,52	25,27	18,15
ISO 12647-6:2012	-	14,00	18,70	13,20

Based on the results of measurements at three key measuring patches, the TVI curves are presented graphically (Fig. 5). Although the curves do not precisely show the reproduction of all tones, they give sufficient insight into how much each stage of the printing experiment deviates from the recommendations of the ISO standard.

TVI in midtones is considered a key criterion for controlling halftone reproduction, as it directly reflects the degree of dot deformation and can result in loss of detail. According to ISO 12647 6:2012, the reference value for *TVI* at 50% is 18,7%. The results of *TVI* with six different combinations showed a clear demarcation of performance precisely in the midtones.

The minimum *TVI* (~ 20%) was achieved with the Harder Plate / Soft Tape combination. This result confirms the theoretical assumption of mechanical interaction in printing: the harder polymer plate provides resistance to deformation, while the soft tape alleviates the contact pressure and prevents the spread of the halftone dots. As a result, the dots remain well formed, but with minimal ink transfer, which is reflected in lower solid densities ($D = 1{,}08$). Such a print provides exceptional dot sharpness, but may result in slightly lighter and less saturated tones.

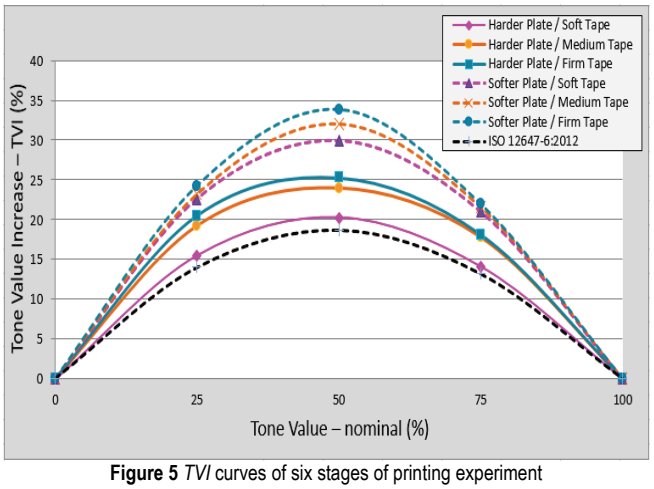


Figure 5 *TVI* curves of six stages of printing experiment

The highest *TVI* (~ 32%) was achieved with the combination of the softer plate with the harder tapes. During the interaction in printing, the opposite mechanism occurs compared to the previously mentioned: the soft cliché, under the pressure of the hard tape, significantly deforms and expands the halftone dots and transfers much more ink. This effect results from the high degree of compressibility of the soft polymer plate, which, together with the hard tape, significantly presses the ink onto the substrate, enabling maximum ink transfer ($D > 1{,}55$). The result is highly saturated tones, but also blurred and filled halftone dots. Furthermore, the combination of Softer Plate / Medium Tape shows optimal values because the halftone dots remain sharp enough without excessive deformation, and the solid tone are saturated enough to meet commercial requirements.

3.3 Ink Film Topography

In addition to the colorimetric and densitometric values, the print quality of cyan throughout the six experimental stages was also assessed by examining the uniformity of the ink film on the print and the presence of pinholing effect, which is carried out using image analysis [21].

Microscopic images of the solid cyan for evaluation are examined with the ImageJ software (using the Interactive 3D Surface Plot feature), which translates optical density into corresponding height values to illustrate the thickness of the ink film [22]. During the transformation, the *Invert* option was enabled to ensure that the darkest areas—indicating the thickest ink film—would appear at higher elevations in the visualization (Fig. 6).

The optical density projection clearly shows greater uniformity or smoothness of the surface when using softer polymer plates (Stages I – III). Furthermore, improved surface uniformity of both polymer plates occurs when using mounting tapes of higher hardness (Fig. 6, right images). The rightmost samples have a notably smoother surface texture

and show fewer visible pinholes compared to the samples on the left side (Fig. 6). The appearance of pinholing can be caused by various factors, and one of the most significant is the thickness of the ink film, which correlates directly with the hardness of the polymer plate.

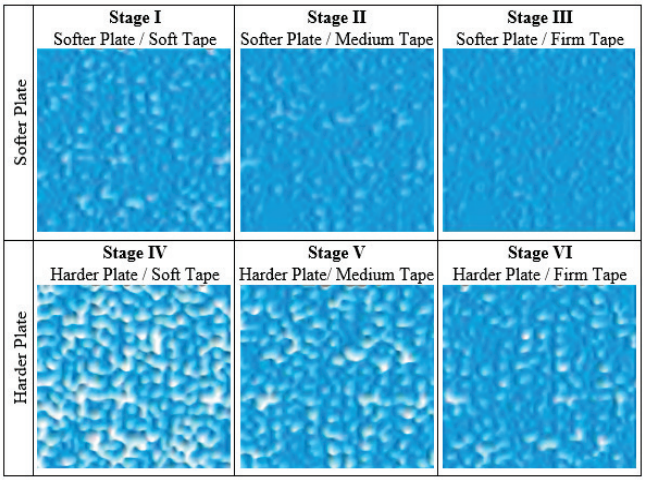


Figure 6 Topographic visualization of solid cyan across various printing stages

4 CONCLUSION

The analysis reveals that the hardness of the mounting tape and polymer cliché is a critical factor in print quality. However, no combination of these parameters achieves perfect performance across all measured criteria. Optimal reproduction of a solid tone depends not only on the amount of transferred ink (as indicated by ΔL), but also on the chromatic characteristics of the color. The combination of a softer cliché with a Medium and Hard tape shows the best results for a solid tone: $\Delta E_{00} \approx 2{,}5$; $\Delta h_{ab} \approx 3$ with a slightly darker print ($\Delta L = -2{,}4$). However, the combination of harder plate with Medium tape proved to be the best for halftone reproduction ($TVI_{50} = 20{,}2\%$), with optimal tolerance values for solid tone. This combination allows for a balanced interaction and creates a good compromise between dot sharpness and solid tone coverage.

The soft polymer itself adapts more to the printing substrate. If combined with a tape that is too soft, the overall compressibility is very high, which leads to excessive ink transfer and therefore an unstable solid tone print ($\Delta h_{ab} \approx 6$). This also leads to excessive *TVI* in the print, which results in reduced contrast and closure of fine details. On the other hand, in combination with harder tapes (Medium and Firm), better control is obtained, while still allowing sufficient ink transfer due to the elasticity of the polymer.

These patterns consistently confirm the theoretical framework, according to which the hardness of the printing plate determines the resistance of the halftone dots to deformation, while the hardness of the tape modulates the contact pressure between the polymer plate and the substrate. The combination of a harder plate and a soft tape optimally preserves details, while a softer plate with a harder tape maximizes ink transfer. If dot sharpness and fine details are a priority (small text and thin lines), the Harder cliché / Soft tape combination is recommended, as it guarantees the most faithful reproduction. Conversely, when the main goal is color intensity and a strong visual impression (background tones on packaging), the Softer plate / Hard tape combination is optimal. Combinations with Medium tape provide a stable, intermediate option that balances well between halftone dot sharpness and color saturation.

In future research, it is recommended to extend the analysis to other colors, substrate types and process parameters to ensure broader applicability of the conclusions obtained.

5 REFERENCES

[1] Kipphan, H. (Ed.). (2001). *Handbook of print media: Technologies and production methods*. Springer Science & Business Media. https://doi.org/10.1007/978-3-540-29900-4

[2] Flexographic Technical Association. (2013). *Flexography: Principles and practices* (6th ed.). Foundation of Flexographic Technical Association.

[3] Żołek-Tryznowska, Z., Rombel, M., Petriaszwili, G., Dedijer, S., & Kašiković, N. (2020). Influence of some flexographic printing process conditions on the optical density and tonal value increase of overprinted plastic films. *Coatings, 10*(9), 816. https://doi.org/10.3390/coatings10090816

[4] Miljković, P., Valdec, D. i Matijević, M. (2018). The Impact of Printing Substrate on Dot Deformation in Flexography. *Tehnički vjesnik, 25*(Supplement 2), 509-515. https://doi.org/10.17559/TV-20170710152140

[5] Hansuebsai, A., & Nawakitwong, S. (2020). Printability analysis of compostable films by flexographic water based ink. *Key Engineering Materials, 843*, 26-32. https://doi.org/10.4028/www.scientific.net/KEM.843.26

[6] Valdec, D., Tomerlin, R., Hajdek, K., & Miljković, P. (2022). Influence of the polymer plate thickness on the plate distortion factor in flexography. *Tehnički glasnik, 16*(1), 47-52. https://doi.org/10.31803//tg-20210629221349

[7] Wirojrungson, W. (2000). An investigation into the effect of plate hardness, and surface speed differential on flexographic gear marking. *Master's thesis*. Rochester Institute of Technology. http://schoalarworks.rit.edu/theses

[8] Valdec, D., Miljković, P., & Čerepinko, D. (2018). The impact of top dot shapes of the printing plate on dot formation in flexography. *Tehnički vjesnik, 25*(2), 596-602. https://doi.org/10.17559/TV-20161003121341

[9] Anderson, J., & Schlotthauer, E. (2010). *Kodak's Flat Top Dots: The effect of flexo dot shape on plate stress and wear*. Eastman Kodak Co.

[10] International Organization for Standardization. (2012). *ISO 12647-6: Graphic technology — Process control for the production of half-tone colour separations, proofs and production prints — Part 6: Flexographic printing*. ISO.

[11] Izdebska-Podsiadly, J., & Podsiadlo, H. (2016). Influence of biodegradable solvent-based ink on the flexography print quality of compostable films. *Polymers Research Journal, 10*(4), 283.

[12] Tomašegović, T., Pibernik, J., Poljaček, S. M., & Madžar, A. (2021). Optimization of flexographic print properties on ecologically favorable paper substrates. *Journal of Graphic Engineering and Design, 12*(1), 37-44. https://doi.org/10.24867/JGED-2021-1-037

[13] Izdebska, J., Żołek-Tryznowska, Z., & Świętoński, A. (2015). Correlation between plastic films properties and flexographic

prints quality. *Journal of Graphic Engineering and Design, 6*(2), 19-25. https://doi.org/10.24867/JGED-2015-2-019
[14] Izdebska-Podsiadły, J. (2021). Effect of plasma surface modification on print quality of biodegradable PLA films. *Applied Sciences, 11*(17), 8245. https://doi.org/10.3390/app11178245
[15] Mahović Poljaček, S., Tomašegović, T., Cigula, T., Milčić, D., Donevski, D., & Strgar Kurečić, M. (2016). Effect of the post-treatment of printing plate on the quality of fine printed elements in flexography. In *Proceedings of the 8th International Symposium on Graphic Engineering and Design* (pp. 117-128).
[16] Jacobson, J., Keif, M., Rong, X., Singh, J., & Vorst, K. (2009). Flexography printing performance of PLA film. *Journal of Applied Packaging Research, 3*(2), 91.
[17] Bould, D. C., Hamblyn, S. M., Gethin, D. T., & Claypole, T. C. (2011). Effect of impression pressure and anilox specification on solid and halftone density. *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture, 225*(5), 699-709. https://doi.org/10.1177/2041297510394072
[18] Valdec, D., Hajdek, K., Vragović, L., & Geček, R. (2021). Determining the print quality due to deformation of the halftone dots in flexography. *Applied Sciences, 11*(22), 10601. https://doi.org/10.3390/app112210601
[19] Yang, Y., Ming, J., & Yu, N. (2012). Color image quality assessment based on CIEDE2000. *Advances in Multimedia*, 2012, 6 pages. https://doi.org/10.1155/2012/273723
[20] Habekost, M. (2013). Which color differencing equation should be used. *International Circular of Graphic Education and Research, 6*, 20–33.
[21] Briggs, J. C., & Tse, M.-K. (2011). Objective print quality analysis and the portable personal IAS® image analysis system. *Journal of the Imaging Society of Japan, 44*(6), 505-513.
[22] Ferreira, T., & Rasband, W. (2012). *ImageJ user guide. ImageJ/Fiji 1.46*. http://imagej.nih.gov/ij/docs/user-guide.pdf (accessed April 25, 2025).

Authors' contacts:

Dean Valdec, Full Professor
(Corresponding author)
University North,
Trg dr. Žarka Dolinara 1, 48000 Koprivnica, Croatia
E-mail: dean.valdec@unin.hr

Igor Majnarić, Full Professor
University of Zagreb, Faculty of Graphic Arts,
Getaldićeva 2, 10000 Zagreb, Croatia
E-mail: igor.majnaric@grf.unizg.hr

Luka Valdec, student
Professional undergraduate study
Multimedia, design and application
University North,
Trg dr. Žarka Dolinara 1, 48000 Koprivnica, Croatia
E-mail: luvaldec@unin.hr

ISSN 1846-6168 (Print), ISSN 1848-5588 (Online)
https://doi.org/10.31803/tg-20240919073402

Preliminary communication
Received: 2024-09-19, Accepted: 2025-02-11

Implementation of a Real-Time Maize Leaf Disease Detection System Using Raspberry Pi 5 and YOLOv8

Samuel Owoeye*, Folasade Durodola, Chibuzor Evwidonor, Sikirulahi Abdulkareem, Emmanuel Popoola, Idowu Akinode

Abstract: Maize, a staple cereal crop globally, faces significant yield challenges due to foliar diseases. This study presents an innovative real-time maize leaf disease detection system integrating a YOLOv8 deep learning model with a custom-designed Unmanned Aerial Vehicle (UAV). The system targets three prevalent maize diseases: Grey leaf spot, common rust, and northern leaf blight. Trained on 10,239 images, the YOLOv8 model, comprising 168 layers and 3,006,233 parameters, achieves 8.1 GFLOPs with an inference speed of 6.3ms per image. Performance evaluation reveals a mean Average Precision (mAP@0.5) of 0.836 and mAP@0.5:0.95 of 0.681 across all classes, with high accuracy for common rust (Precision: 0.953, Recall: 0.979) and grey leaf spot (Precision: 0.919, Recall: 0.892). The custom UAV, designed for agricultural surveying, features a thrust-to-weight ratio of 5.85, ensuring stable flight with the 140g payload of a Raspberry Pi 5 and Camera Module V2. With a total weight of 605g and a 3S 5000mAh LiPo battery, the drone achieves an estimated flight time of 5.3 minutes, balancing survey coverage with real-time disease detection capabilities. The integration of this high-performance model with an efficient UAV platform represents a significant advancement in precision agriculture, enabling early disease intervention and targeted treatment strategies, thus promoting sustainable farming practices through optimized resource allocation and potential reduction in pesticide usage.

Keywords: crop management; deep learning; disease detection; maize; real-time system; UAV; YOLOv8

1 INTRODUCTION

Maize (*Zea mays L.*) is one of the most important cereal crops worldwide, serving as a staple food and a key component in livestock feed and industrial products. Leaf diseases in maize, such as Northern Corn Leaf Blight, Common Rust, and Gray Leaf Spot, pose significant threats to crop productivity and quality Early and accurate detection of these diseases is essential for effective crop management and ensuring food security [1]. Traditional methods of disease detection involve manual inspection by experts, which can be time-consuming, labour-intensive, and prone to human error [2]. Moreover, the scarcity of trained pathologists and the subjective nature of visual assessments can lead to inconsistent and delayed diagnoses, exacerbating the spread of diseases and affecting crop productivity [3].

Recent advancements in artificial intelligence (AI) and deep learning have paved the way for innovative solutions in agricultural diagnostics. Among these, the YOLO (You Only Look Once) object detection algorithm has gained prominence due to its real-time detection capabilities and high accuracy in various computer vision tasks [4]. YOLOv5, the latest iteration of this model, offers improved performance in terms of speed and precision, making it a suitable candidate for real-time agricultural applications [5] [6].

There is a need for an automated, real-time system that can accurately identify maize leaf diseases on small, medium, and large-scale farms. Such a system would enable timely interventions, reducing the spread of diseases and minimizing crop losses.

This research proposes the implementation of a real-time maize leaf disease detection system utilizing the YOLOv5 deep learning model deployed on a computer. The system is designed to provide farmers with a cost-effective and efficient tool for monitoring crop health, thereby reducing reliance on manual inspections and expert consultations. By facilitating early disease detection and management, the proposed system aims to enhance crop yield and sustainability.

The subsequent sections of this paper will detail the methodology used for data collection, model training, and system deployment. Additionally, the system's performance will be evaluated based on accuracy, detection speed, and usability in field conditions. This study aspires to contribute to the broader field of precision agriculture, leveraging AI to address critical challenges in crop management and food security.

2 LITERATURE REVIEW

Historically, maize leaf disease detection relied heavily on visual inspection by trained experts. This method, while valuable, has several limitations including being time-consuming, labour-intensive, and prone to human error [7, 8, 9]. The scarcity of trained pathologists and the subjective nature of visual assessments can lead to inconsistent and delayed diagnoses, potentially exacerbating disease spread and impacting crop yields [10].

Recent advancements in computer vision and machine learning have paved the way for automated, image-based disease detection systems. These methods offer the potential for more objective, scalable, and efficient disease diagnosis.

Early attempts at automated maize leaf disease detection often employed classical machine learning techniques. Wiesner-Hanks et al. [11] developed a system using Support Vector Machines (SVM) to classify Northern Leaf Blight lesions in maize. Their approach achieved promising results but was limited by the need for handcrafted features. Xie et al. [12] proposed a method combining K-means clustering for image segmentation and Probabilistic Neural Networks for the classification of maize leaf diseases. While effective, this approach struggled with complex backgrounds and varying light conditions. The advent of deep learning has

significantly improved the accuracy and robustness of maize leaf disease detection systems. Convolutional Neural Networks (CNNs) have become particularly popular due to their ability to automatically learn relevant features from images. DeChant et al. [13] utilized a CNN architecture to detect Northern Leaf Blight in maize, demonstrating high accuracy in both controlled and field conditions, their model outperformed traditional machine learning methods, highlighting the potential of deep learning in this domain. Ghosal et al. [14] proposed a novel deep-learning framework called DeepCorn for identifying multiple foliar stresses in maize plants. Their system could differentiate between biotic and abiotic stresses, showcasing the versatility of deep learning approaches.

2.1 Mobile and IoT-based Detection Systems

With the proliferation of smartphones and Internet of Things (IoT) devices, researchers have explored more accessible and real-time disease detection solutions. Ferentinos [15] developed a mobile application using a CNN model for on-field diagnosis of plant diseases, including those affecting maize, this approach brings disease detection capabilities directly to farmers, enabling quicker response times. Cambaza et al. [16] integrated IoT sensors with image-based disease detection, creating a comprehensive system for monitoring maize crop health. Their approach combined environmental data with visual symptoms, improving the accuracy of disease diagnosis.

2.2 Hyper-spectral and Multispectral Imaging

Advanced imaging techniques such as hyper-spectral and multispectral imaging have shown promise in early disease detection, often before visible symptoms appear. Mahlein [17] demonstrated the use of hyper-spectral imaging for detecting and differentiating various maize leaf diseases at early stages. This technology offers the potential for pre-symptomatic disease detection, although its application in field conditions remains challenging due to cost and complexity.

3 METHODOLOGY

3.1 System Architecture

The proposed maize leaf disease detection system consists of two main components:

1) **YOLOv8 Model:** A deep learning model trained to detect three categories of maize leaf diseases.
2) **Drone-mounted Raspberry Pi 5:** The disease detection model is deployed on a Raspberry Pi 5, which is mounted on a custom-designed drone along with a Raspberry Pi Camera Module V2.

Fig. 1 illustrates the complete system architecture, displaying the integration of the drone, Raspberry Pi 5, and camera module.

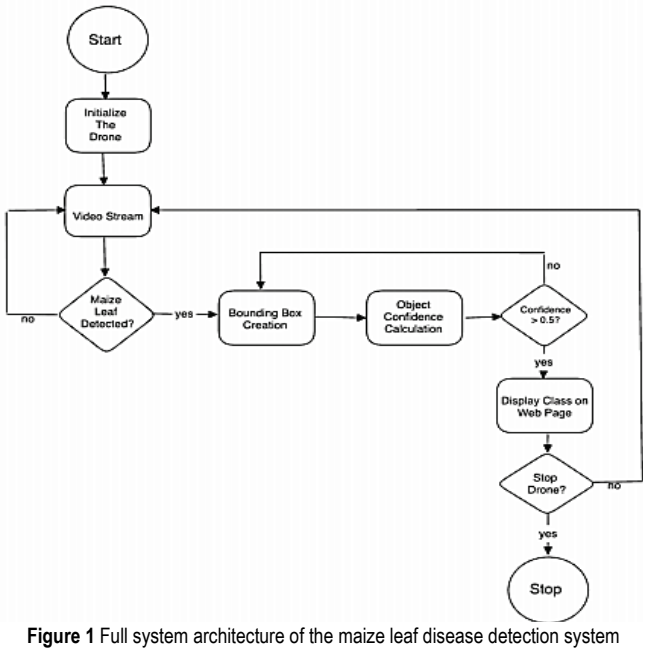


Figure 1 Full system architecture of the maize leaf disease detection system

3.2 Data Collection and Preprocessing

The data set used for training the YOLOv8 model was sourced from Roboflow and comprises images of maize leaves exhibiting various disease symptoms. The data was cleaned to focus on three specific diseases: Northern Leaf Blight, Common Rust, and Grey Leaf Spot. Tab. 1 presents the distribution of images across different categories and data-set splits. Fig. 2 showcases examples of the three maize leaf diseases included in the data set.

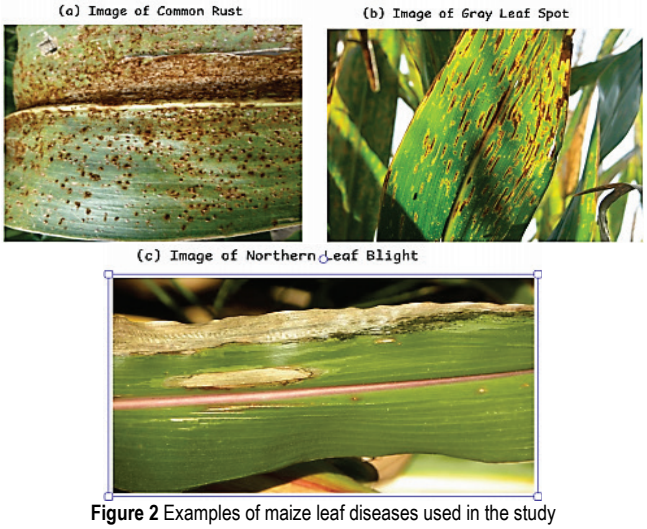


Figure 2 Examples of maize leaf diseases used in the study

Table 1 Distribution of maize leaf image data set used in the study

Conditions	Number of Images
Northern Leaf Blight	3120
Common Rust	3057
Grey Leaf Spot	4062
Total Training Data	5159
Total Validation Data	4500
Total Test Data	580
Total Data	10239

3.3 Model Development

The development of the maize leaf disease detection model followed a structured process, as illustrated in Fig. 3. The YOLOv8 nano model was selected for this project due to its balance of accuracy and computational efficiency, making it suitable for deployment on edge devices like the Raspberry Pi 5. The model was trained using PyTorch and the Ultralytics YOLOv8 implementation.

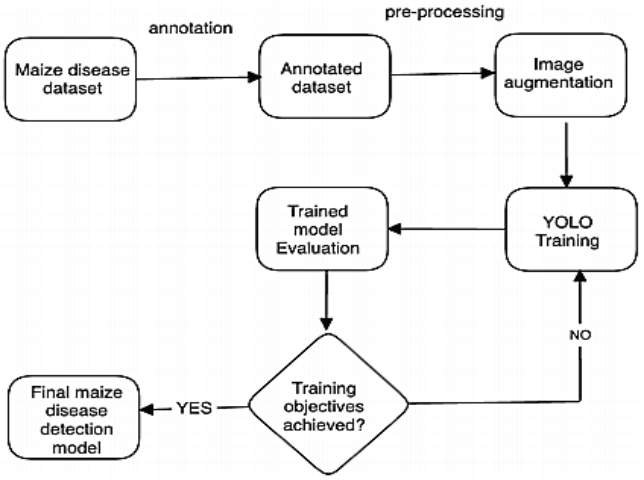


Figure 3 Flowchart of the model development procedure

3.4 Model Training and Evaluation

The training process involved fine-tuning the YOLOv8 nano model on our curated data set. The transfer learning techniques were used to leverage pre-trained weights, which were then adapted to our specific maize leaf disease detection task. The model's performance was evaluated using various metrics, including precision, recall, F1-score, and mean Average Precision (mAP). Figs. 4 to 8 present the detailed performance analysis of the trained model.

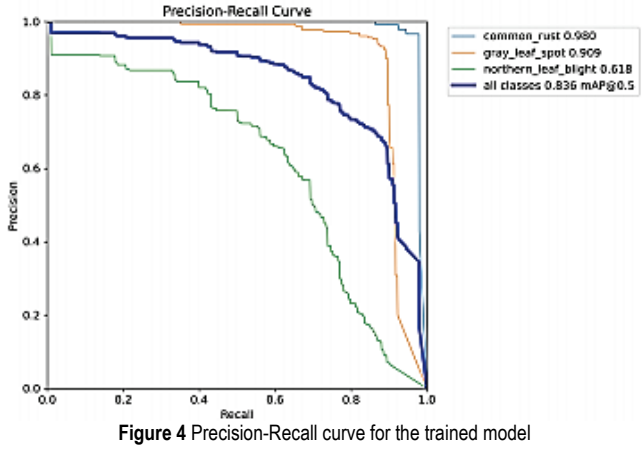


Figure 4 Precision-Recall curve for the trained model

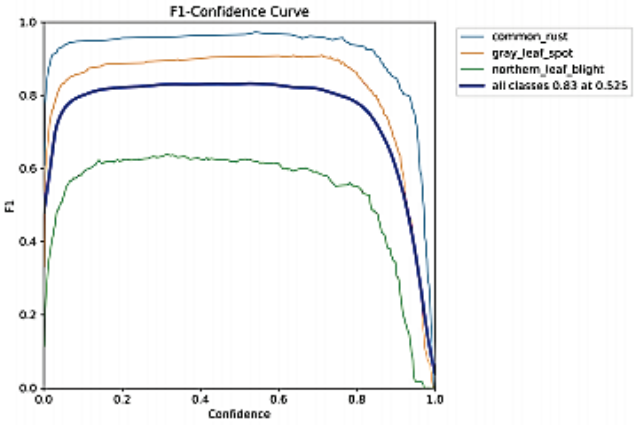


Figure 5 F1-Score curve for different confidence thresholds

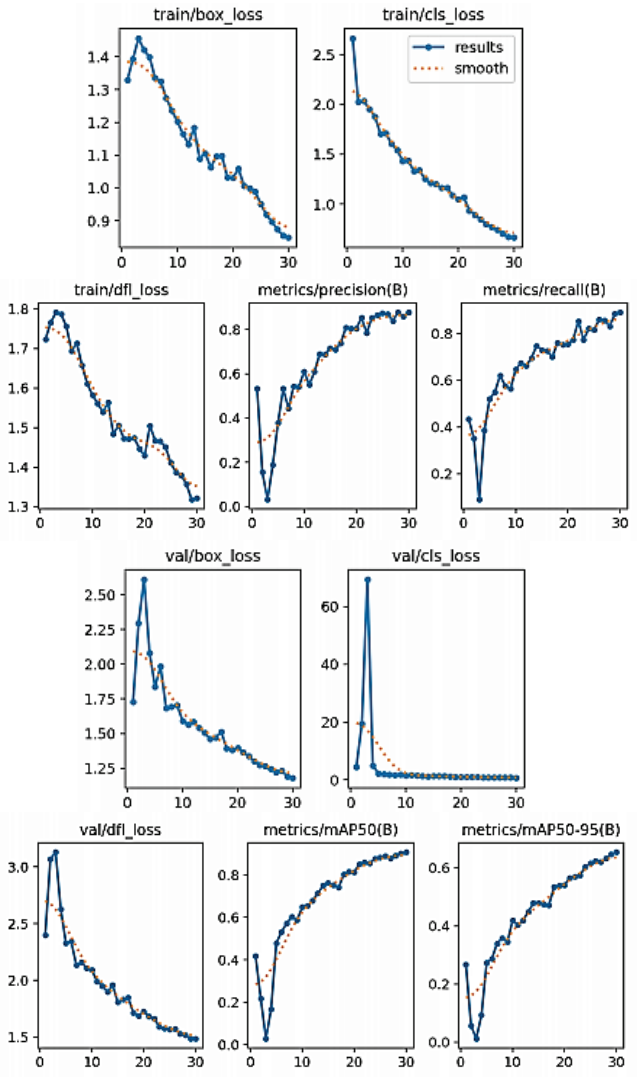


Figure 6 Training and validation metrics over epochs

3.5 System Deployment

After training and validation, the YOLOv8 nano model was converted to a format compatible with the Raspberry Pi 5 (.pt format). An environment was set up on the Raspberry Pi 5 to run the inference script.

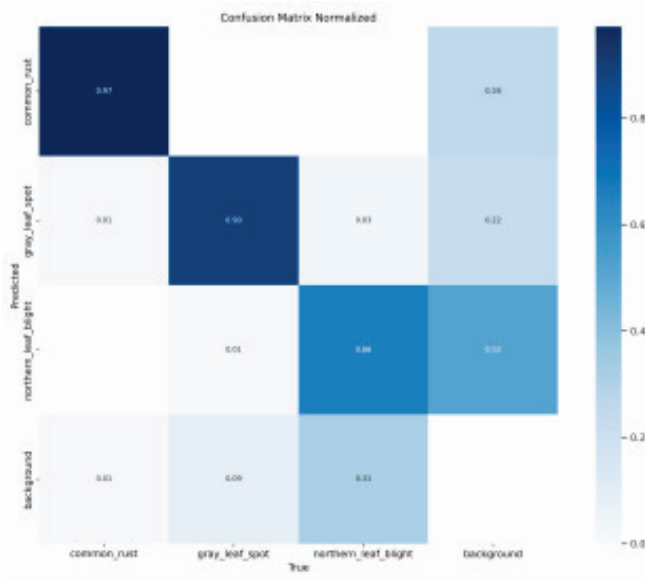


Figure 7 Normalized confusion matrix for disease classification

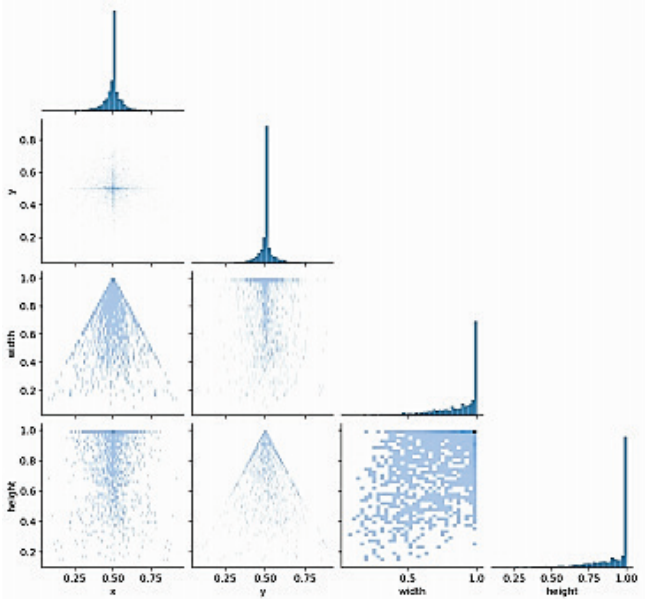


Figure 8 Label correlogram showing relationships between disease classes

3.6 Drone Design and Integration

The disease detection system was integrated with a custom-designed drone to enable aerial surveys of maize fields. Fig. 9 shows the complete drone setup with the mounted Raspberry Pi 5 and camera module.

3.6.1 Drone Payload and Weight Calculation

The drone's payload and overall weight were carefully calculated to ensure stable flight and sufficient endurance. Tab. 2 provides a breakdown of the drone's components and their respective weights.

3.6.2 Thrust and Power Calculations

To ensure stable flight and sufficient endurance, detailed thrust and power calculations were performed. The total thrust of the motors was calculated to be 3540 g, resulting in a thrust-to-weight ratio of approximately 5.85. This high ratio ensures stable flight and manoeuvrability. Power calculations determined that the drone's 3-cell (3S) Li-Po battery, with a nominal voltage of 11.1V and a capacity of 5 Ah, provides approximately 5.3 minutes of flight time. The total current draw, including the Raspberry Pi 5, was calculated to be 45 A.

Figure 9 Custom-designed drone with mounted Raspberry Pi 5 and camera module

Table 2 Breakdown of drone component weights

Component	Weight (g)
Pi Camera	20
Raspberry Pi 5	50
Flight Controller	20
Propellers (4 units)	28
Frame	90
Battery	265
ESC (4 units)	28
Motors	54
Total Payload	140
Total Drone Weight (without payload)	465
Total Weight (with payload)	605

3.7 Field Testing

The integrated system was subjected to field tests to evaluate its performance in real-world conditions. Fig. 10 shows the drone during a field test, capturing images of maize crops for disease detection. These tests helped validate

the system's ability to capture clear images of maize leaves and perform real-time disease detection using the onboard Raspberry Pi 5 and YOLOv8 nano model.

3.8 System Deployment

After the YOLOv8 model was trained and validated, it was deployed on the computer. The deployment process involved converting the trained model into a format compatible with the computer (in .pt format) and optimizing it for real-time inference on an HTTP web application. A web server built with Python's Flask framework with an HTML user interface is used to display the results and confidence level of the detected disease. The application is designed with a user-friendly interface, allowing farmers to easily operate the system. The user interface has buttons to 'Start Camera' and 'Capture' images of maize leaves at regular intervals, and the application processes each image using the YOLOv8 model. Detected diseases are highlighted with bounding boxes and labels, providing immediate visual feedback to the user.

Figure 10 Drone-mounted disease detection system during field testing

4 SYSTEM EVALUATION AND DISCUSSION

The evaluation of our maize leaf disease detection system encompasses both the performance of the Unmanned Aerial Vehicle (UAV) and the integrated YOLOv8 model. This comprehensive assessment ensures that the system meets the requirements for effective aerial surveying and accurate disease detection in real-world agricultural settings.

Table 3 Summary of UAV performance tests

Test name	Discussion
Yaw Test	Measures the drone's endurance on a full battery charge. The estimated flight time was *X* minutes
Pitch Test	Evaluates the drone's rotational control around the vertical axis. The yaw was consistent within *X* degrees.
Roll Test	Tests the forward and backward tilt response. The pitch was controlled with X accuracy.
Lift Test	Evaluate the drone's vertical ascent capability. The lift was steady with *X* m/s^2 acceleration.
Altitude/Signal Test	Evaluate altitude hold stability and signal strength at various heights. The altitude hold was stable up to *X* meters.

4.1 UAV Performance Evaluation

The custom-designed drone underwent rigorous testing to assess its flight characteristics and functionalities crucial for agricultural survey missions. Tab. 3 summarizes the tests conducted and their results.

These tests demonstrate the drone's capability to maintain stable flight and perform precise manoeuvres, which are essential for capturing high-quality images of maize crops. The flight time of *X* minutes, as determined by the Yaw Test, provides sufficient duration for surveying substantial areas of cropland in a single flight.

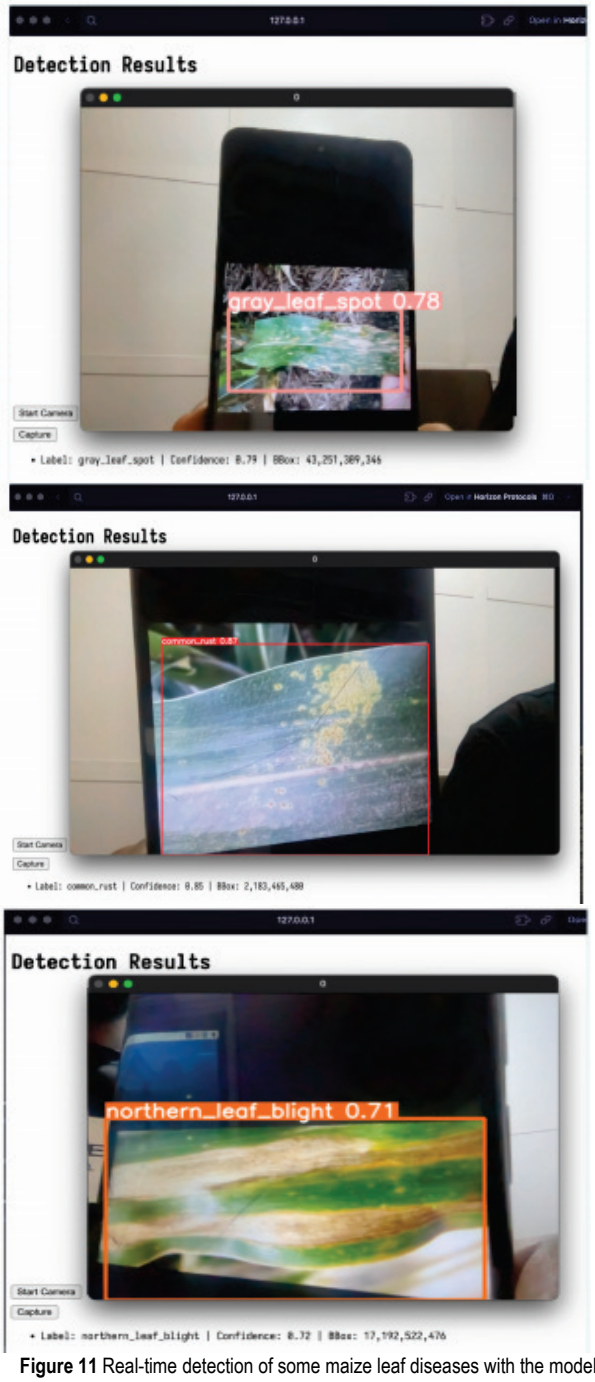


Figure 11 Real-time detection of some maize leaf diseases with the model

4.2 Integrated System Performance

The integration of the YOLOv8 model with the UAV platform creates a powerful tool for real-time maize leaf disease detection. The system's performance can be evaluated based on several key aspects:

4.2.1 Model Efficiency

The YOLOv8 nano model, with its 168 layers and 3,006,233 parameters, achieves 8.1 G Flops, striking a balance between complexity and computational efficiency. The model's inference speed is particularly noteworthy, with a total processing time of 6.3ms per image (0.3 ms for pre-processing, 2.9 ms for inference, and 3.1 ms for post-processing). This rapid processing enables real-time disease detection during flight, allowing for immediate feedback and decision-making. Fig. 10 shows the drone during a field test, capturing images of maize crops for disease detection. The real-time detection results of the YOLOv8 model are shown in Fig. 11 below. The system successfully identifies and classifies the maize leaf diseases, highlighting affected areas with bounding boxes.

4.2.2 Detection Accuracy

Tab. 4 presents a summary of the model's performance across different disease categories.

Table 4 Model performance summary for maize leaf disease detection

Class	Images	Instances	Box(P)	*R*	mAP50	mAP50-95
All	577	595	0.858	0.81	0.836	0.681
Common Rust	144	146	0.953	0.979	0.98	0.9
Grey Leaf Spot	212	231	0.919	0.892	0.909	0.763
Northern Leaf Blight	218	218	0.704	0.56	0.618	0.379

The model demonstrates high accuracy in detecting Common Rust and Grey Leaf Spots, with mAP50 values of 0.98 and 0.909 respectively. The detection of Northern Leaf Blight, while satisfactory, shows room for improvement with a mAP50 of 0.618.

4.2.3 System Reliability

The integration of the model with the UAV platform was tested for reliability during field operations. The system consistently maintained connectivity between the Raspberry Pi 5 and the drone's flight controller, ensuring uninterrupted disease detection throughout the flight duration.

4.2.4 Battery Life and Payload Impact

The addition of the Raspberry Pi 5 and camera module to the drone increased the total payload by 140 g. Despite this additional weight, the drone maintained a flight time of approximately *X* minutes, which aligns closely with our calculated estimate of 5.3 minutes. This demonstrates that the power consumption of the Raspberry Pi 5 and the additional weight have a minimal impact on the drone's endurance.

4.3 Discussion

The evaluation results highlight several strengths of our integrated maize leaf disease detection system:

- Real-time Detection: The system's ability to process images in 6.3ms enables real-time disease detection, allowing for immediate response to detected issues.
- High Accuracy: Particularly for Common Rust and Grey Leaf Spot, the model demonstrates excellent detection capabilities, which can significantly aid in early disease management.
- Efficient Integration: The successful integration of the YOLOv8 model with the UAV platform showcases the potential for deploying complex AI models in agricultural settings.

However, there are areas for potential improvement:

- **Northern Leaf Blight Detection:** The lower accuracy in detecting Northern Leaf Blight suggests a need for further model refinement or additional training data for this specific disease.
- **Flight Time Optimization:** While the current flight time is sufficient for small to medium sized fields, extending the drone's endurance could enhance its applicability to larger agricultural operations.

Overall, the system demonstrates promising performance in real-world conditions, offering a valuable tool for early detection and management of maize leaf diseases. Future work could focus on expanding the range of detectable diseases, improving detection accuracy for Northern Leaf Blight, and optimizing the drone's design for extended flight times.

5 CONCLUSION AND RECOMMENDATION

5.1 Conclusion

This research demonstrated the effectiveness of the YOLOv8 deep learning model for real-time detection of maize leaf diseases, identifying six key diseases with high accuracy. The system achieved a mean Average Precision (mAP) of 0.908. The integration of a user-friendly web application provides a practical solution for early disease detection, helping to prevent the spread of diseases and minimize yield loss.

Future work should focus on extending the system to include more diseases and adapt to different crops, enhancing its applicability across the agricultural sector. This study highlights the potential of AI technologies like YOLOv8 to revolutionize agricultural practices, promoting smarter and more resilient food production systems.

5.2 Recommendation

Future work could focus on extending the system's capabilities to include more diseases and adapt to different

crops, enhancing its utility across the agricultural sector. Additionally, further refinement of the model's performance under varying environmental conditions and in different geographic regions would help in scaling the technology globally, making it a versatile tool for diverse farming needs. This study underscores the transformative potential of integrating advanced AI technologies like YOLOv8 in agricultural practices, paving the way for smarter and more resilient food production systems.

6 REFERENCES

[1] Arora, J. & Agrawal, U. (2020). Classification of maize leaf diseases from healthy leaves using deep forest. *Journal of Artificial Intelligence and Systems, 2*(1), 14-26. https://doi.org/10.33969/AIS.2020.21002

[2] Hossain, Md. S., Hossain, Md. S., Kabir, A., Mahmud, S. M. H., Islam, M. M. M. & Tariq, F. (2023). Review of Artifact Detection Methods for Automated Analysis and Diagnosis in Digital Pathology. Chapter in book: *Artificial Intelligence for Disease Diagnosis and Prognosis in Smart Healthcare* (pp. 177-202). https://doi.org/10.1201/9781003251903-11

[3] Polder, G., Blok, P. M., De Villiers, H. A., Van der Wolf, J. M. & Kamp, J. (2019). Potato virus y detection in seed potatoes using deep learning on hyperspectral images. *Frontiers in Plant Science, 10*, 209. https://doi.org/10.3389/fpls.2019.00209

[4] Redmon, J., Divvala, S., Girshick, R. & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR2016)*, 779-788. https://doi.org/10.1109/CVPR.2016.91

[5] Trinh, T. H. & Nguyen, H. H. C. (2023). Implementation of yolov5 for real-time maturity detection and identification of pineapples. *Traitement du Signal, 40*(4). https://doi.org/10.18280/ts.400413

[6] Shiferaw, B., Prasanna, B. M., Hellin, J. & Bänziger, M. (2011). Crops that feed the world 6. Past successes and future challenges to the role played by maize in global food security. *Food Security, 3*(3), 307-327. https://doi.org/10.1007/s12571-011-0140-5

[7] Zhang, S., Zhang, S., Zhang, C., Wang, X. & Shi, Y. (2018). Crop disease detection using weakly supervised deep learning. *Frontiers in Plant Science, 9*, 1569.

[8] Barbedo, J. G. A. (2013). Digital image processing techniques for detecting, quantifying and classifying plant diseases. *SpringerPlus, 2*(1), 660. https://doi.org/10.1186/2193-1801-2-660

[9] Barbedo, J. G. A. (2018). Impact of data-set size and variety on the effectiveness of deep learning and transfer learning for plant disease classification. *Computers and Electronics in Agriculture, 153*, 46-53. https://doi.org/10.1016/j.compag.2018.08.013

[10] Mohanty, S. P., Hughes, D. P. & Salathé, M. (2016). Using deep learning for image-based plant disease detection. *Frontiers in Plant Science, 7*, 1419. https://doi.org/10.3389/fpls.2016.01419

[11] Wiesner-Hanks, T., Stewart, E. L., Kaczmar, N., DeChant, C., Wu, H., Nelson, R. J., Lipson, H. & Gore, M. A. (2018). Image set for deep learning: field images of maize annotated with disease symptoms. *BMC Research Notes, 11*(1), 440. https://doi.org/10.1186/s13104-018-3548-6

[12] Xie, C., Yang, C. & He, Y. (2015). Identification of maize leaf diseases using improved deep convolutional neural networks. *IEEE Access, 3*, 2042-2050.

[13] DeChant, C., Wiesner-Hanks, T., Chen, S., Stewart, E. L., Yosinski, J., Gore, M. A., Nelson, R. J. & Lipson, H. (2017). Automated identification of northern leaf blight-infected maize plants from field imagery using deep learning. *Phytopathology, 107*(11), 1426-1432. https://doi.org/10.1094/PHYTO-11-16-0417-R

[14] Ghosal, S., Blystone, D., Singh, A. K., Ganapathysubramanian, B., Singh, A. & Sarkar, S. (2018). An explainable deep machine vision framework for plant stress phenotyping. *Proceedings of the National Academy of Sciences, 115*(18), 4613-4618. https://doi.org/10.1073/pnas.1716999115

[15] Ferentinos, K. P. (2018). Deep learning models for plant disease detection and diagnosis. *Computers and Electronics in Agriculture, 145*, 311-318. https://doi.org/10.1016/j.compag.2018.01.009

[16] Cambaza, E., Koseki, S. & Kawamura, S. (2019). Smart pest detection and control. *Smart Farming Technologies for Sustainable Agricultural Development*, pages 236-251.

[17] Mahlein, A.-K. (2016). Plant Disease Detection by Imaging Sensors – Parallels and Specific Demands for Precision Agriculture and Plant Phenotyping. *Plant Disease, 100*(2), 241-251. https://doi.org/10.1094/PDIS-03-15-0340-FE

Authors' contacts:

Owoeye Samuel
(Corresponding Author)
Department of Mechatronics Engineering,
Federal University of Agriculture,
Alabata Road, 110111 Abeokuta, Nigeria
owoeyeso@funaab.edu.ng

Durodola Folasade
Department of Mechatronics Engineering,
Federal University of Agriculture,
Alabata Road, 110111 Abeokuta, Nigeria

Chibuzor Evwidonor
Department of Mechatronics Engineering,
Federal University of Agriculture,
Alabata Road, 110111 Abeokuta, Nigeria

Sikirulahi Abdulkareem
Department of Mechatronics Engineering,
Federal University of Agriculture,
Alabata Road, 110111 Abeokuta, Nigeria

Emmanuel Popoola
Department of Mechatronics Engineering,
Federal University of Agriculture,
Alabata Road, 110111 Abeokuta, Nigeria

Idowu Akinode
Department of Mechatronics Engineering,
Federal University of Agriculture,
Alabata Road, 110111 Abeokuta, Nigeria

ISSN 1846-6168 (Print), ISSN 1848-5588 (Online)
https://doi.org/10.31803/tg-20241017125951

Preliminary communication
Received: 2024-10-17, Accepted: 2025-02-12

Occupant Thermal Comfort versus Poor Air Quality of Wooden Houses in the Highlands

Hermawan Hermawan, Jozef Švanjlenka*, Elina Mohd Husini, Annisa Nabila Arrizqi, Nasyiin Faqih, Sunaryo Sunaryo

Abstract: Humans need a comfortable and healthy home to carry out activities properly. Low-income people's homes in the Highlands have poor indoor air quality due to stoves that produce smoke. People also use stoves as a heating tool to achieve thermal comfort. This study aimed to investigate thermal comfort and poor air quality in Wooden Homes in the Highlands. The study used a mixed method. Qualitative methods were used to obtain data on the thermal comfort of occupants. Quantitative methods were used to obtain data on indoor air quality. The study results showed opposing aspects between thermal comfort in people in the highlands and the fulfillment of clean and healthy air. The study also created solutions for thermal comfort and clean air quality by simulating ventilation placement.

Keywords: architecture; cold; comfort; indoor air quality; occupant

1 INTRODUCTION

Energy waste is still an essential issue in all scientific fields. Science must be integrated to find solutions to energy problems. Zero-energy buildings are one solution to reducing energy waste [1]. Various simulations of buildings to create energy-free buildings are carried out using various methods [2].

Investigations into building types have also been carried out. One type of building that is thought to create thermal comfort for its occupants is a vernacular building [3]. Wooden houses are one of the vernacular mountain houses in Indonesia that have local wisdom in using fireplace [4]. Wooden houses are sustainable houses that use environmentally friendly materials [5].

The culture of heating with traditional stoves in highland houses creates a concept of kinship that makes it comfortable for its occupants. The community is also thermally comfortable with heating using traditional fire stoves. Fire stoves produce smoke that can cause health problems for residents. Traditional fire stoves are used by low-income people who are forced to use fire stoves for cooking [6].

Thermal comfort is the user's perception of the environment they live in. Perception is based on the ASHRAE thermal sensation scale [7]. Climate variables in thermal comfort are air temperature, main radiant temperature, air humidity, and wind speed [8]. Clothing and activities are personal thermal comfort variables that also affect the thermal comfort of building occupants.

Temperature is one of the climate variables always included in the discussion of thermal comfort and building thermal performance [9]. The discussion of air humidity is one set with the discussion of temperature because the two variables are closely related [10]. Discussion of the two variables in assessing the thermal performance of buildings is something that cannot be left out and is an integral part of assessing the thermal performance of buildings [11].

Discussions of thermal comfort can be done using the Predicted Mean Vote (PMV) theory. The PMV parameter is still used even though some researchers consider it lacking [12]. PMV is known as an expanded and flexible theory of thermal comfort. The difference between PMV and Actual Mean Vote (AMV) makes researchers compare the two aspects in many research objects [13].

Discussion of thermal comfort can also be seen using Actual Mean Vote (AMV). Thermal comfort assessment with AMV can create a thermal comfort prediction model [14]. Operative temperature is one of the variables discussed in AMV. Operative temperature is obtained from the average air temperature and average solar radiation temperature [15].

Thermal comfort in vernacular houses is attractive to discuss with the local wisdom found in vernacular houses. Research on thermal comfort in vernacular houses requires higher accuracy because researchers must apply quantitative and qualitative methods to obtain comprehensive results [16].

The indoor environment can be monitored by measuring temperature, humidity, and particulate matter [17]. Measurements are made in outdoor and indoor spaces whose results can be compared so that differences in air content are found in outdoor and indoor spaces [18]. The interior space can be measured from several types of space. The interior space that has a large particulate matter (PM) value is the kitchen [19]. Indoor space can be measured from several types of spaces. Kitchens have a sizeable particulate matter (PM) value [20].

One variable in thermal comfort is wind speed. The content of substances in the air can affect the thermal comfort of occupants and the air quality in the room [21]. Particulate matter (PM) can be seen from measurements, including PM1, PM2.5, and PM10 [22]. Investigations of room air quality more often measure particulate matter using PM2.5 [23].

Temperature and PM2.5 measurements are also a reference in determining the criteria for sick building syndrome [24]. Temperature is related to indoor air quality, measured based on PM2.5 content. Both variables are essential things that affect the comfort and health of building occupants.

Research into indoor air quality in wooden houses under similar conditions in the European Union is characterized, for example, by the works mentioned below. Research in Spain discusses historical buildings that are considered to have local wisdom such as wooden houses. Poor indoor air quality

can be improved by retrofitting [36]. Research on wooden houses in Denmark that discusses the positive aspects of the sustainability of wooden houses [37]. Indoor Air Quality has not been discussed comprehensively in this research. Research in Chile discusses the presence of air pollution in wooden houses when firewood is used by low-income communities [38]. The research discusses particulate matter (PM) 2.5 and has not discussed other PM. Another article does not discuss indoor air quality of wooden houses specifically but discusses the development of wooden house fabrication which is included in bio-based materials. Residential houses in Central and Eastern Europe are suitable for developing bio-based material fabrication, one of which is wooden houses [39].

Scientific developments that discuss indoor air quality have been widely carried out using particulate matter content variables. However, discussions of poor air quality juxtaposed with users' thermal comfort have not been widely carried out. Comparisons of the thermal comfort of residential residents based on the highlands' heating culture have also not been discussed from an architectural or scientific perspective. The novelty of the research is the discussion of the study of the culture of heating with poor air quality so that solutions will be found to maintain the culture and solve the problem of poor air quality in the highlands. The results of the research will have a broad impact on society in terms of maintaining local wisdom and creating healthy homes. The solutions provided will impact the community's happiness because they will create health and comfort. The purpose of the research is to investigate the air quality inside wooden residential houses associated with the highlands' heating culture. The research will provide a strategy for creating good air quality without eliminating the existing culture.

2 METHODS

The study used a mixed method. Data was collected through observations, interviews, and measurements in the field.

2.1 Research Object

The study was conducted on one hundred wooden-walled houses using traditional stoves (Fig. 1). The floor of the building uses concrete and soil materials. The roof uses zinc. The building is not too big, ranging from 21 to 45 m^2. The ventilation of the house tends to be closed because of the cold climate in the highlands.

2.2 Data Collection

Observation collected data to obtain data on residential house images consisting of room plans, elevations, and perspective images. Photos of residential houses were taken to complete the results of more detailed residential house images. Interviews were conducted to determine the thermal comfort felt by the residents of the house and the activities of the residents. Air temperature, humidity, and particulate matter (PM2.5) data were measured using a thermal measuring device, and indoor air quality was assessed. Measurements were carried out every a quarter of an hour in one day starting from 06.00-21.00 western Indonesia time.

Figure 1 Research Object 100 wooden houses

3 RESULTS AND DISCUSSION

3.1 "Geni" culture using traditional fire stoves

The highlands in Wonosobo Regency have a cold climate, so people need heating. The lower middle class uses traditional fire stoves to heat their homes. Most low-income people work as farm laborers and have wooden houses. The heating process using traditional fire stoves has been carried out since ancient times and still persists today.

The heating process has become a culture of gathering for highland people. Several areas in Wonosobo Regency call the culture of gathering by lighting a fire the "geni" culture. Several other areas call it the "genen" culture. The "geni" culture is a place to gather while warming up. The "geni" culture, initially only intended for warming up, has changed into a social need for the community.

The "geni" culture is a place to discuss everything from informal to formal. Informal discussions include daily activities related to agriculture, personal problems related to family life, and social relations problems. Formal discussions, such as hamlet or village-level meetings, are also sometimes discussed while warming up with a traditional fire stove.

Observation results from 100 wooden houses that use traditional fire stoves show that the traditional fire stove functions primarily as a heating tool, while cooking is an additional function. The traditional fire stove is used to cook rice without using a magic jar. People often use the magic jar to cook rice because it cooks faster. The traditional fire stove is used more to boil water to make coffee. For people who cannot afford to buy a magic jar, they will use the traditional fire stove as the main cooking tool.

The traditional fire stove is placed in the kitchen with a small table and chairs. A small table known as "dingklik" is

used to put drinks and snacks. A small chair known as "jengkok" is used for residents of the house or guests who visit the house to chat. People call chatting "ngendhong". Several chairs in the kitchen are used as "ngendhong" facilities.

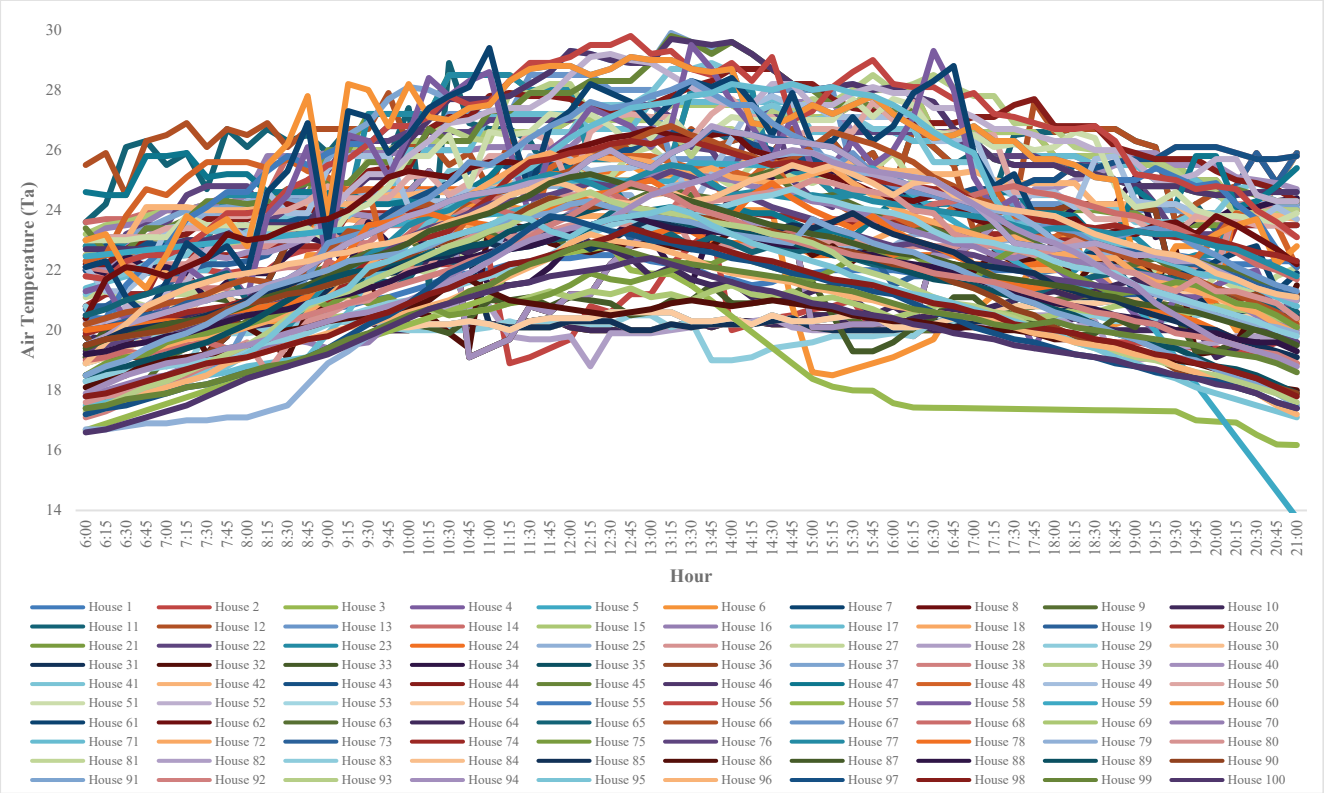


Figure 2 Indoor Temperature

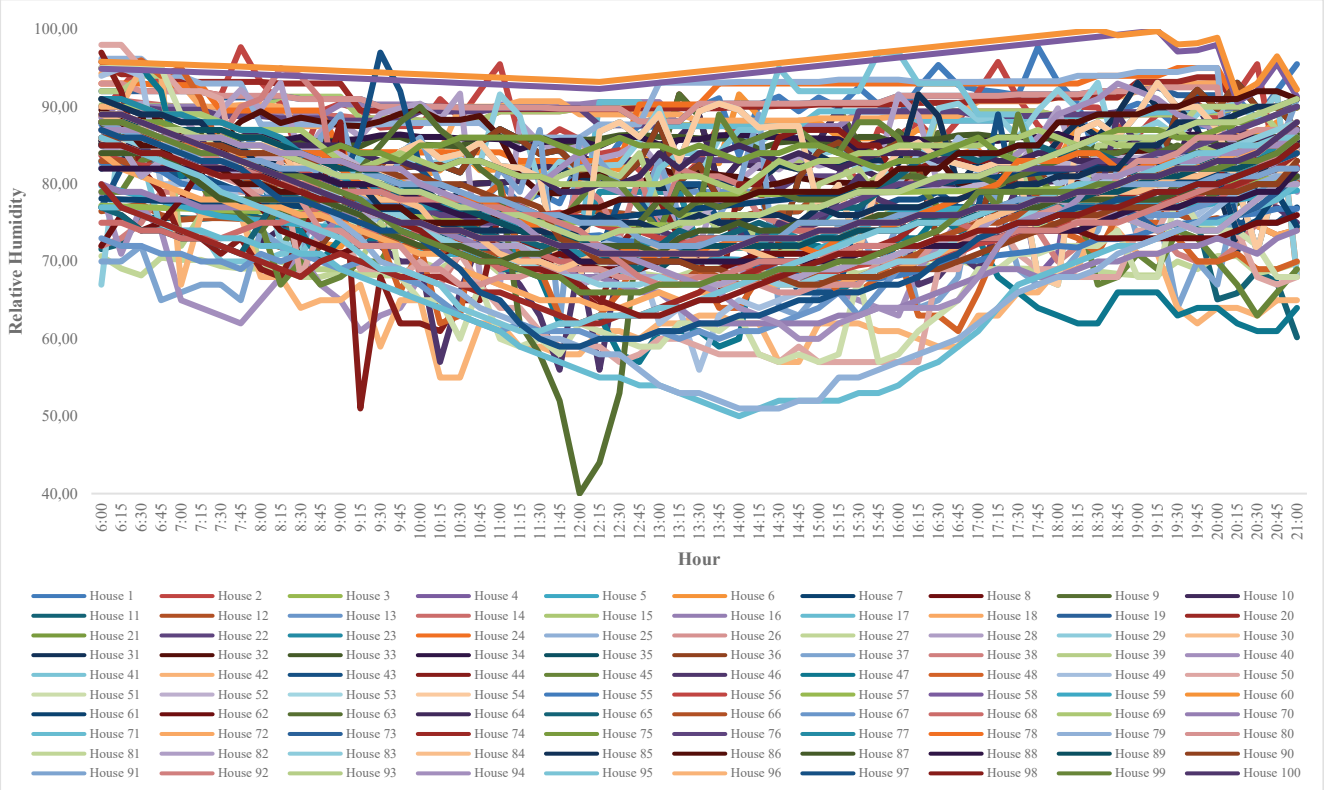


Figure 3 Indoor Relative Humidity

Interview results from residents of the house explained that residents feel warm and comfortable when they are in the kitchen. The occupants feel calm, comfortable, and peaceful in the presence of a burning stove. The occupants feel calm in the presence of a burning stove because it indicates that the occupants have food that can be eaten. The feeling of comfort arises with the warming process due to the stove's heat. The feeling of peace from the occupants occurs due to the harmony with routine socializing. Problems can be resolved when gathering in the kitchen by lighting the stove.

3.2 Air Temperature and Humidity

One hundred houses have similar air temperatures. The lowest air temperature of 13.75 °C was obtained from house 59 at 21.00 (western Indonesia time). The highest air temperature of 29.90 °C was obtained from house 37 at 13.15 (western Indonesia time). The average air temperature of one hundred houses is 22.86 °C (Fig. 2). Comfortable air temperatures in mountainous areas are around 24 °C, indicating that people should still feel cold with indoor air temperatures below mountain communities' comfortable temperature [25]. People feel comfortable with the presence of a fire pit as a heating device, coupled with a feeling of calm and serenity, making the cold air temperature not too felt by the house's occupants.

The humidity between one wooden house and another shows similar results. The lowest humidity of 40% was obtained from house 63 at 12.00 (western Indonesia time). The highest humidity of 99.96% was obtained from house 60 at 18.30 (western Indonesia time) (Fig. 3). House 63 more often uses traditional stoves, thus reducing the water content in the room. Using stoves that produce smoke will make the air in the home environment drier, resulting in the lowest humidity. House 60 shows wet conditions with the floor still made of soil. The water in the soil floor is higher than other floor materials. The vegetation around house 60 is quite a lot and will produce water vapor, thus making the surrounding humidity high.

3.3 Particulate Matter (PM2.5)

Traditional stoves will produce smoke in the kitchen (Fig. 4a). The distribution of smoke in each house is different, but most of the smoke cannot escape smoothly because ventilation is not available properly. Some houses have kitchens with minimal ventilation so that the spread of smoke reaches the family room and living room (Fig. 4b). Highland communities cannot eliminate traditional stoves with the "geni" culture, which is believed to create harmony within the community. In addition, low-income communities consider stoves to be cheaper than gas stoves. People can get wood fuel for stoves from the forest. Taking wood for fuel is known by the community as "repek".

Particulate matter in wooden houses with stoves can be caused by several aspects, such as dust pollution from wooden wall materials, smoke produced by the stove, and dust from the dirt floor used. Smoke from traditional stoves significantly increases PM2.5 values (Fig. 5). Other smoke comes from motorcycle exhausts placed in the living room.

a) b)

Figure 4 a) Smoke is coming out of the stove; b) Lack of ventilation in the kitchen

The highest value of particulate matter (PM2.5) in the research object occurred in house 63 at 14.00 (western indonesia time) at 426 µgrams/m^3. A very high value for PM2.5 occurred in house 63 due to smoke. Traditional stoves in house 63 are more frequent because the occupants do not have other cooking utensils, such as magic jars. The lowest PM2.5 value occurred in house 59 at 12.45 (western indonesia time) at 0.9 µgrams/m^3 because the traditional stove was not used at 12.45 (western indonesia time). The use of traditional stoves greatly affects the PM2.5 value. The particle content originating from wooden wall materials or dust from dirt floors is insignificant in affecting the PM2.5 value. Low-income people inhabit house 59. The average PM2.5 value from one hundred houses is 62.11 µgrams/m^3. The average value remains higher than the minimum PM2.5 threshold of 40 µgram/m^3. One hundred wooden houses used as research objects had poor air quality regarding PM2.5 content.

Data analysis of the relationship between PM (Particulate Matter) and temperature using visualization with tecplot software. The relationship between air temperature and PM2.5 in one hundred houses shows graphic variations. Very high PM2.5 values occur in several houses throughout the day, both when the temperature is 18.5 to 24 °C (Fig. 6a). Several other houses show high PM2.5 values when the air temperature is between 19 and 22 °C (Fig. 6b).

The average thermal variables of one hundred houses from 06.00-21.00 can be seen in Fig. 7. Air temperature (Ta), mean radiant temperature (MRT) and operative temperature (To) show similar results. The variable values are between 20 and 30 °C. Relative humidity (RH) is seen between 60-80%. Velocity is closer to still, so the value obtained is not more than 0.1 m/s. Thermal sensation votes (TSV) have more negative values because of the cold climate conditions. PM2.5 (Particulate Matter) content of more than 50 µgram/m^3 is higher than the maximum limit of PM2.5 content required of 40 µgram/m^3. The clothing and activity variables of the occupants are variables that are taken into account in thermal comfort. Most occupants of buildings in the highlands wear long-sleeved clothes, long pants for men,

and skirts for women. The community also wears a special head covering for men and a jilbab for women. The clothing value is 0.61. Activities at home are sitting and chatting, which is worth one meeting.

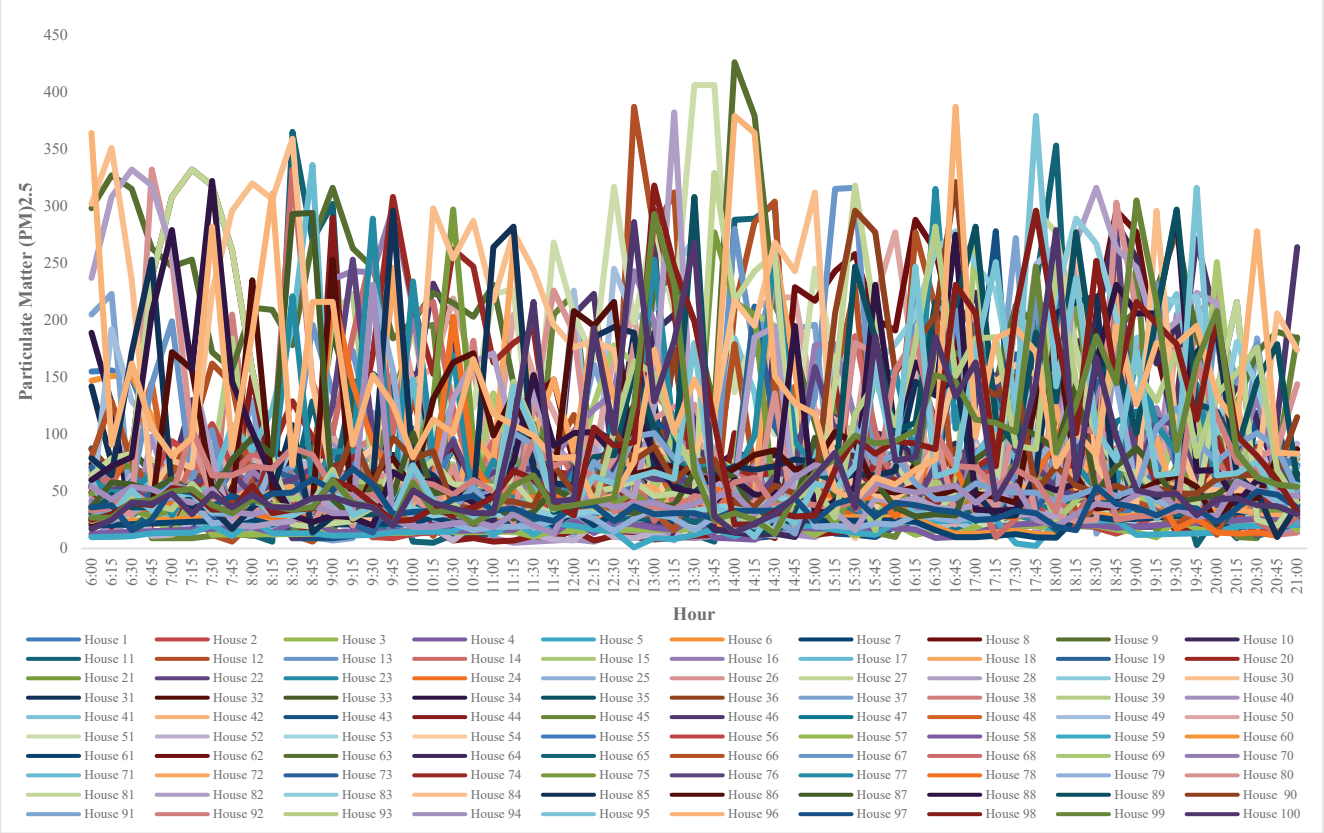


Figure 5 Indoor Particulate Matter

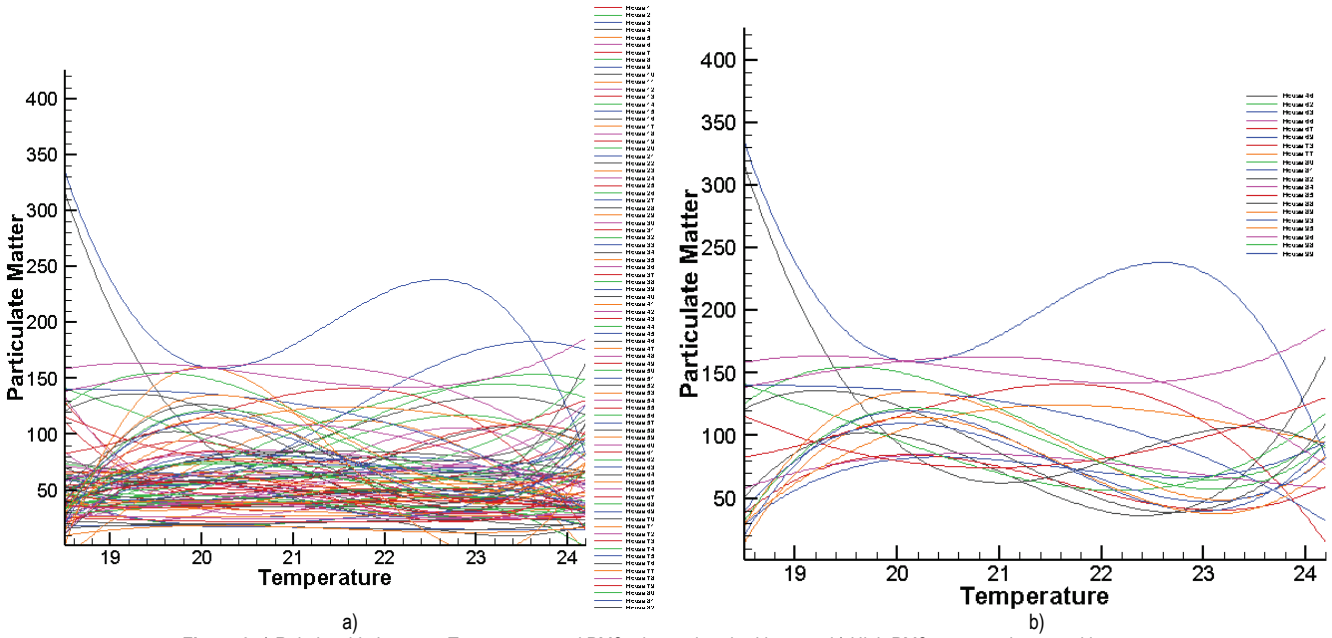


Figure 6 a) Relationship between Temperature and PM2.5 in one hundred houses; b) High PM2.5 content in several houses.

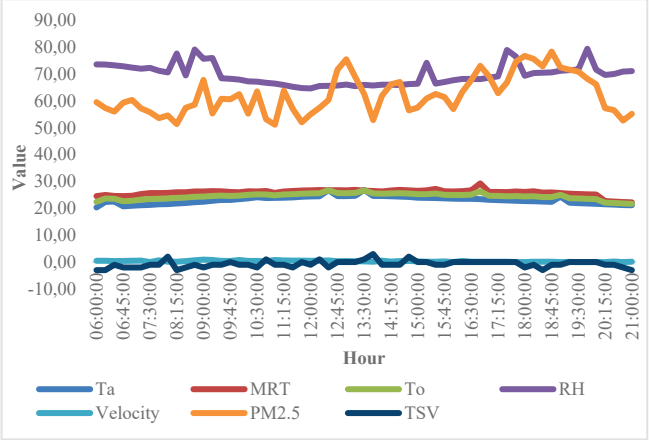


Figure 7 Average Variable Value (Annotation: Ta = Air Temperature, MRT = Mean Radiant Temperature, To = Operative Temperature, RH = Relative Humidity, PM = Particulate Matter, TSV = Thermal Sensation Vote).

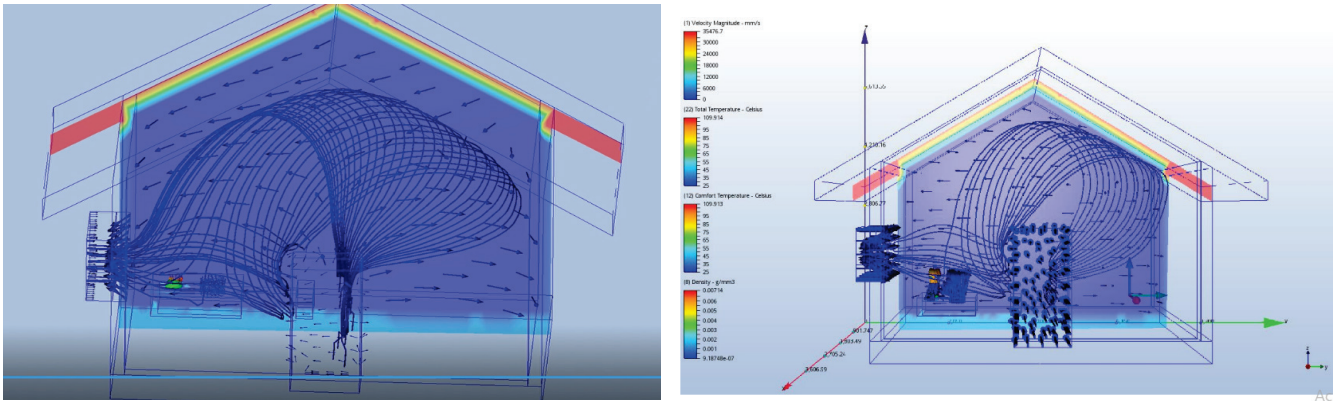
3.4 Air Quality Improvement Strategy

Air quality improvement can be improved by improving architectural elements so smoke can exit the room smoothly. Clean air from outside that enters the room can push smoke out of the room properly. Ventilation is one element that needs to be optimized to create healthy and clean air elements. Ventilation in the kitchen can be added by directing smoke outside, but it can still warm the room. Creating an Air Quality Improvement Strategy using Computational Fluid Dynamic (CFD) Revit.

The kitchen of a wooden house has ventilation but is often closed because of the cold climate in the highlands. Airflow is not free to escape through closed ventilation (Fig. 8a). Open ventilation will smooth the airflow even though it cannot remove smoke in the room (Fig. 8b). Standard ventilation with a size of around 60×40 cm cannot push smoke out of the room.

Building improvements are needed to create thermal comfort and health for building occupants. The building envelope is one factor that needs to be improved to improve the quality of the building [26]. Designing wooden residential products is one solution to creating sustainable buildings. Wooden houses need to be designed with a modern model so that people are more interested in building those [27].

Ventilation is an architectural element that determines the air exchange in a room. Wooden houses in the highlands use more windows that can be opened. Smoke from traditional stoves will flow towards the window after rotating in the room. Traditional stoves make the exchange of air in the room (Fig. 4). The placement of existing windows needs to be supplemented with other ventilation so that the airflow becomes smoother. The cross-ventilation model will make the airflow move more freely [28]. Effective windows in the kitchen will improve kitchen performance and make the house healthier [6]. The size of the windows in the kitchen needs to be adjusted to the activities in the kitchen so that more outside air enters the kitchen and can replace the hot air produced by the kitchen [29]. Increasing air volume in the room will create thermal comfort for building users [30]. Airflow and thermal comfort can be achieved with the cross-ventilation model. Cross ventilation will create airflow that can bring thermal comfort [31].

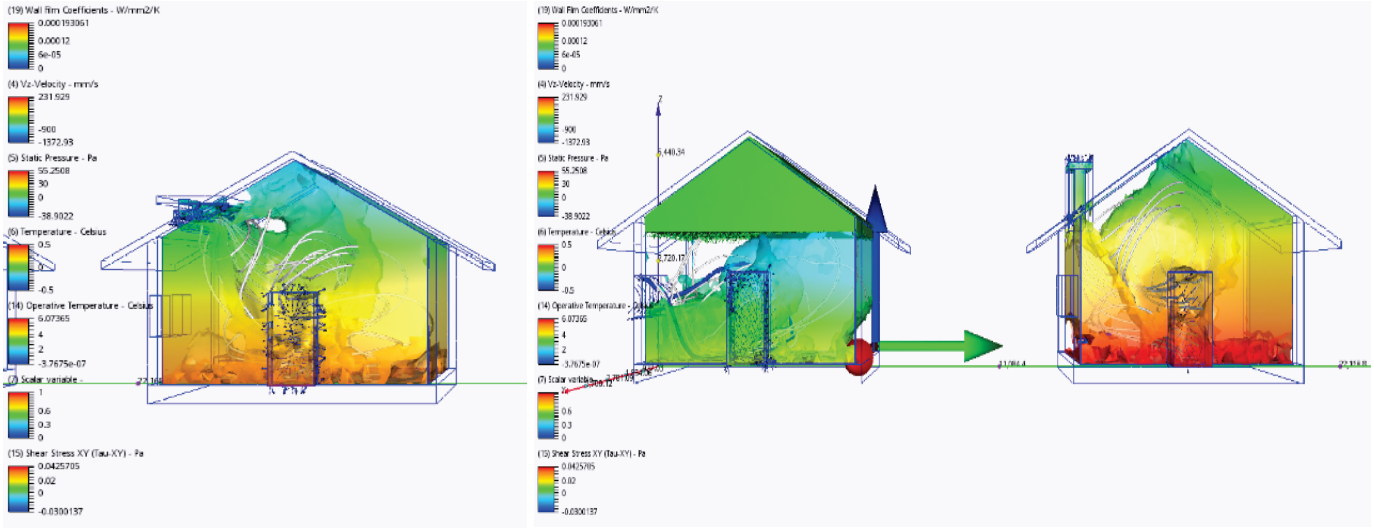


a) b)

Figure 8 a) Airflow in a room with closed ventilation; b) Airflow out through one open vent

The first alternative solution to overcome smoke is to add a hole in the roof (Fig. 9a). Smoke does not move directly towards the hole in the roof. Some smoke moves around in the room, and some towards the hole in the roof. Roof windows perform better than mechanical ventilation. The roof window needs to be designed in such a way that it can circulate air smoothly [32]. Roof ventilation can also lower the air temperature, creating thermal comfort for users. Airflow from roof ventilation can lower the temperature by 7% [33].

The second alternative is to make a chimney so that the smoke can be directed out through the chimney. Using a chimney to channel smoke out of the house has been recognized as effective. A chimney can use solar energy [34]. The determination of the slope of the chimney needs to be adjusted to the type of activity and airflow in the room [35]. Using chimneys for low-income people is quite difficult because the cost of making chimneys is quite large. The chimney design needs to be simple but highly effective. The design can be made using zinc material from the teer place. Several houses in the highlands owned by low-income people use "blek," a used teer place, as an additional wall. The chimney's slope can affect the smoke exit from inside to outside the room.



a) b)

Figure 9 a) Addition of roof ventilation; b) Addition of a roof chimney.

4 CONCLUSION

Thermal comfort that is contrary to the fulfillment of health in homes occurs in highland areas. Ventilation is one of the important parts that needs to be created in order to be a solution in creating healthy and energy-efficient homes. The thermal comfort of residents of wooden houses in the highlands is fulfilled by the "geni" culture that the highland community has carried out for generations. Poor air quality occurs in wooden houses with inadequate ventilation, so air cannot circulate properly. The thermal comfort of the occupants that has been met needs to be maintained, and additional ventilation is needed so that airflow is smoother.

The strategy that can be done so that the culture of the occupants can be maintained and good thermal comfort is created by adding building elements to the kitchen. Improper ventilation placement makes the smoke from the fireplace unable to escape smoothly so that the placement and opening of the ventilation need to be modified so that the smoke can escape smoothly. Roof ventilation needs to be made so that some of the smoke can escape through the roof hole. The chimney needs to be made with a placement close to the center of the smoke so that the smoke can exit the room directly. The recommendation for further research is a strategy for creating thermal comfort that can be connected to sensors so that automation is created to create thermal comfort without eliminating the culture that has developed in society.

Acknowledgements

Projects KEGA 017TUKE-4/2024 "Implementation of an innovative interdisciplinary approach to knowledge sharing on agricultural structures" and VEGA 1/0228/24 "Research and innovation activities focused on efficient and sustainable wood-based construction technologies for agricultural construction".

5 REFERENCES

[1] Ahmed, K., Hasu, T., & Kurnitski, J. (2022). Actual energy performance and indoor climate in Finnish NZEB daycare and school buildings. *Journal of Building Engineering, 56*, 104759. https://doi.org/10.1016/j.jobe.2022.104759

[2] Izadyar, N., & Miller, W. (2022). Ventilation strategies and design impacts on indoor airborne transmission: A review. *Building and Environment, 218*, 109158. https://doi.org/10.1016/j.buildenv.2022.109158

[3] Wang, F., Wang, S., Cheng, B., & Wang, W. (2022). To inhabit, retain or abandon? Adaptive utilization of energy-efficient sunken buildings by rural households in Shanzhou, China. *Energy and Buildings, 255*, 111668. https://doi.org/10.1016/j.enbuild.2021.111668

[4] Hermawan, Prijotomo, J., & Dwisusanto, Y. B. (2020). The Geni tradition as the center of the shelter for Plateau Settlements. *Ecology, Environment and Conservation, 26*(1), 34-38.

[5] Al Tawayha, F., Braganca, L., & Mateus, R. (2019). Contribution of the vernacular architecture to the sustainability: A comparative study between the contemporary areas and the old quarter of a Mediterranean city. *Sustainability, 11*(3). https://doi.org/10.3390/su11030896

[6] Kumar, P., Hama, S., Abbass, R. A., Nogueira, T., Brand, V. S., Wu, H. W., Abulude, F. O., Adelodun, A. A., de Fatima Andrade, M., Asfaw, A., Aziz, K. H., Cao, S. J., El-Gendy, A., Indu, G., Kehbila, A. G., Mustafa, F., Muula, A. S., Nahian, S., Nardocci, A. C., … Shiva Nagendra, S. M. (2022). CO_2 exposure, ventilation, thermal comfort and health risks in low-income home kitchens of twelve global cities. *Journal of Building Engineering, 61*, 105254. https://doi.org/10.1016/j.jobe.2022.105254

[7] Shin, H., Kang, M., Mun, S. H., Kwak, Y., & Huh, J. H. (2021). A study on changes in occupants' thermal sensation owing to CO_2 concentration using PMV and TSV. *Building and Environment, 187*, 107413. https://doi.org/10.1016/j.buildenv.2020.107413

[8] Zhao, Q., Lian, Z., & Lai, D. (2021). Thermal comfort models and their developments: A review. *Energy and Built Environment, 2*(1), 21-33. https://doi.org/10.1016/j.enbenv.2020.05.007

[9] Amoabeng, K. O., Opoku, R., Boahen, S., & Obeng, G. Y. (2022). Analysis of indoor set-point temperature of split-type

ACs on thermal comfort and energy savings for office buildings in hot-humid climates. *Energy and Built Environment, 4*(3), 368-376. https://doi.org/10.1016/j.enbenv.2022.02.006

[10] Castellano, J., Sanz, V., Cañas, E., & Sánchez, E. (2022). Effect of firing temperature on humidity self-regulation functionality in a ceramic tile composition. *Journal of the European Ceramic Society, 42*(13), 6236-6243. https://doi.org/10.1016/j.jeurceramsoc.2022.05.058

[11] Chamberlain, A. T., Powell, C. D., Arcier, E., & Aldenhoven, N. (2022). The relationship between on-farm environmental conditions inside and outside cow sheds during the summer in England: can Temperature Humidity Index be predicted from outside conditions? *Animal - Open Space, 1*(1), 100019. https://doi.org/10.1016/j.anopes.2022.100019

[12] Ravindra, K., Agarwal, N., & Mor, S. (2020). Assessment of thermal comfort parameters in various car models and mitigation strategies for extreme heat-health risks in the tropical climate. *Journal of Environmental Management, 267*, 110655. https://doi.org/10.1016/j.jenvman.2020.110655

[13] Yan, H., Yang, L., Dong, M., Hu, B., Sun, Z., Shi, F., Yuan, G., & Bi, X. (2022). Thermal comfort in residential buildings using bimetal radiator heating vs. floor heating terminals. *Journal of Building Engineering, 45*, 103501. https://doi.org/10.1016/j.jobe.2021.103501

[14] Trebilcock, M., Soto-Muñoz, J., & Piggot-Navarrete, J. (2020). Evaluation of thermal comfort standards in office buildings of Chile: Thermal sensation and preference assessment. *Building and Environment, 183*, 107158. https://doi.org/10.1016/j.buildenv.2020.107158

[15] Rawal, R., Shukla, Y., Vardhan, V., Asrani, S., Schweiker, M., de Dear, R., Garg, V., Mathur, J., Prakash, S., Diddi, S., Ranjan, S. V., Siddiqui, A. N., & Somani, G. (2022). Adaptive thermal comfort model based on field studies in five climate zones across India. *Building and Environment, 219*, 109187. https://doi.org/10.1016/j.buildenv.2022.109187

[16] Costa-Carrapiço, I., González, J. N., Raslan, R., & Sánchez-Guevara, C. (2022). Understanding the challenges of determining thermal comfort in vernacular dwellings: A meta-analysis. *Journal of Cultural Heritage, 58*, 57-73. https://doi.org/10.1016/j.culher.2022.09.019

[17] Kim, J., Kim, S., Bae, S., Kim, M., Cho, Y., & Lee, K. I. (2022). Indoor environment monitoring system tested in a living lab. *Building and Environment, 214*, 108879. https://doi.org/10.1016/j.buildenv.2022.108879

[18] Konstantinou, C., Constantinou, A., Kleovoulou, E. G., Kyriacou, A., Kakoulli, C., Milis, G., Michaelides, M., & Makris, K. C. (2022). Assessment of indoor and outdoor air quality in primary schools of Cyprus during the COVID–19 pandemic measures in May–July 2021. *Heliyon, 8*(5), e09354. https://doi.org/10.1016/j.heliyon.2022.e09354

[19] Liu, S., Cao, Q., Zhao, X., Lu, Z., Deng, Z., Dong, J., Lin, X., Qing, K., Zhang, W., & Chen, Q. (2020). Improving indoor air quality and thermal comfort in residential kitchens with a new ventilation system. *Building and Environment, 180*, 107016. https://doi.org/10.1016/j.buildenv.2020.107016

[20] Sahu, V., & Gurjar, B. R. (2020). Spatial and seasonal variation of air quality in different microenvironments of a technical university in India. *Building and Environment, 185*, 107310. https://doi.org/10.1016/j.buildenv.2020.107310

[21] Wallis, S. L., Hernandez, G., Poyner, D., Holmes, W., Birchmore, R., & Berry, T. A. (2019). Particulate matter in residential buildings in New Zealand: Part II. The impact of building airtightness, mechanical ventilation using simulated occupancy. *Atmospheric Environment: X, 2*, 100026. https://doi.org/10.1016/j.aeaoa.2019.100026

[22] Kaushik, A. K., & Dhau, J. S. (2022). Photoelectrochemical oxidation assisted air purifiers; perspective as potential tools to control indoor SARS-CoV-2 Exposure. *Applied Surface Science Advances, 9*, 100236. https://doi.org/10.1016/j.apsadv.2022.100236

[23] Bhat, M. A., Eraslan, F. N., Awad, A., Malkoç, S., Üzmez, Ö. Ö., Döğeroğlu, T., & Gaga, E. O. (2022). Investigation of indoor and outdoor air quality in a university campus during COVID-19 lock down period. *Building and Environment, 219*, 109176. https://doi.org/10.1016/j.buildenv.2022.109176

[24] Babaoglu, U. T., Milletli Sezgin, F., & Yag, F. (2020). Sick building symptoms among hospital workers associated with indoor air quality and personal factors. *Indoor and Built Environment, 29*(5), 645-655. https://doi.org/10.1177/1420326X19855117

[25] Hermawan, Prianto, E., Setyowati, E., & Sunaryo. (2019). The thermal condition and comfort temperature of traditional residential houses located in mountainous tropical areas: An adaptive field study approach. *International Journal on Advanced Science, Engineering and Information Technology, 9*(6), 1833-1840. https://doi.org/10.18517/ijaseit.9.6.3560

[26] Blázquez, T., Suárez, R., Ferrari, S., & Sendra, J. J. (2023). Improving winter thermal comfort in Mediterranean buildings upgrading the envelope: An adaptive assessment based on a real survey. *Energy and Buildings, 278*, 112615. https://doi.org/10.1016/j.enbuild.2022.112615

[27] Yang, X., Lyu, H., Li, J., Fu, F., & Zhou, H. (2022). Comprehensive design for a vernacular wood house with energy-saving process. *Sustainable Energy Technologies and Assessments, 53*(PB), 102571. https://doi.org/10.1016/j.seta.2022.102571

[28] Jiang, F., Tao, S., Tao, Q., Li, Z., Yuan, Y., & Zheng, J. (2022). The effect of louver blinds on the wind-driven cross ventilation of multi-storey buildings. *Journal of Building Engineering, 54*, 104614. https://doi.org/10.1016/j.jobe.2022.104614

[29] Liu, Q., Son, Y. J., Li, L., Wood, N., Senerat, A. M., & Pantelic, J. (2022). Healthy home interventions: Distribution of PM2.5 emitted during cooking in residential settings. *Building and Environment, 207*(PB), 108448. https://doi.org/10.1016/j.buildenv.2021.108448

[30] Nasrollahi, N., & Ghobadi, P. (2022). Field measurement and numerical investigation of natural cross-ventilation in high-rise buildings; Thermal comfort analysis. *Applied Thermal Engineering, 211*, 118500. https://doi.org/10.1016/j.applthermaleng.2022.118500

[31] Nimarshana, P. H. V., Attalage, R. A., & Perera, K. K. C. K. (2022). Quantification of the impact of RANS turbulence models on airflow distribution in horizontal planes of a generic building under cross-ventilation for prediction of indoor thermal comfort. *Journal of Building Engineering, 52*. https://doi.org/10.1016/j.jobe.2022.104409

[32] He, Y., Chu, Y., Zang, H., Zhao, J., & Song, Y. (2022). Experimental and CFD study of ventilation performance enhanced by roof window and mechanical ventilation system with different design strategies. *Building and Environment, 224*. https://doi.org/10.1016/j.buildenv.2022.109566

[33] Rangel, C. G., Rivera-Solorio, C. I., Gijón-Rivera, M., & Mousavi, S. (2022). The effect on thermal comfort and heat transfer in naturally ventilated roofs with PCM in a semi-arid climate: An experimental research. *Energy and Buildings, 274*, 112453. https://doi.org/10.1016/j.enbuild.2022.112453

[34] Maghrabie, H. M., Abdelkareem, M. A., Elsaid, K., Sayed, E. T., Radwan, A., Rezk, H., Wilberforce, T., Abo-Khalil, A. G., & Olabi, A. G. (2022). A review of solar chimney for natural ventilation of residential and non-residential buildings. *Sustainable Energy Technologies and Assessments, 52*(PB), 102082. https://doi.org/10.1016/j.seta.2022.102082

[35] Singh Rajput, K., & Bartaria, V. N. (2022). Experimental study of vertical air temperature distribution and ventilation with energy saving using solar chimney. *Materials Today: Proceedings, 63*, 692-698. https://doi.org/10.1016/j.matpr.2022.04.914
[36] González-Prieto, D., Fernández-Nava, Y., Megido, L., & Prieto, M. M. (2023). Economic and environmental prioritisation of potential retrofitting interventions in electricity decarbonisation scenarios: Application to a heritage building used as offices. *J. Build. Eng., 72*, 106561. https://doi.org/10.1016/j.jobe.2023.106561
[37] Hansen, R. N. et al. (2024). Environmental consequences of shifting to timber construction: The case of Denmark. *Sustain. Prod. Consum., 46*, 54-67. https://doi.org/10.1016/j.spc.2024.02.014
[38] Hernández, H., & Molina, C. (2023). Analyzing energy poverty and carbon emissions in a social housing complex due to changes in thermal standards. *Energy Sustain. Dev., 77*, 101347. https://doi.org/10.1016/j.esd.2023.101347
[39] Sutkowska, M., Stefańska, A., Vaverkova, M. D., Dixit, S., & Thakur, A. (2024). Recent advances in prefabrication techniques for biobased materials towards a low-carbon future: From modules to sustainability. *J. Build. Eng., 91*, 109558. https://doi.org/10.1016/j.jobe.2024.109558

Authors' contacts:

Hermawan Hermawan
Department of Architecture,
Faculty of Engineering and Computer Science,
Universitas Sains Al-Qur'an,
Wonosobo, 56351, Indonesia

Jozef Švanjlenka
(Corresponding author)
Departmen of Construction Technology and Management,
Faculty of Civil Engineering,
Laboratory of Construction Technology and Management,
Technical University of Košice,
Letná 1/9, 042 00 Košice, Slovakia
ingsvajl@gmail.com

Elina Mohd Husini
Fakulti Kejuruteraan dan Alambina,
Universiti Sains Islam Malaysia,
Bandar Baru Nilai, 71800, Nilai, Negeri Sembilan, Malaysia

Annisa Nabila Arrizqi
Department of Civil Engineering, Faculty of Engineering and Planning,
Universitas Islam Indonesia,
Yogyakarta, 55584, Indonesia

Nasyiin Faqih
Department o Mechanical Engineering,
Faculty of Engineering and Computer Science,
Universitas Sains Al-Qur'an,
Wonosobo, 56351, Indonesia

Sunaryo Sunaryo
Department of Civil Engineering,
Faculty of Engineering and Computer Science,
Universitas Sains Al-Qur'an,
Wonosobo, 56351, Indonesia

ISSN 1846-6168 (Print), ISSN 1848-5588 (Online)
https://doi.org/10.31803/tg-20250113111223

Preliminary communication
Received: 2025-01-13, Accepted: 2026-03-11

English Language Knowledge Influencing the Choice of Mobility Mode When Abroad: A Case Study

Irena Ištoka Otković*, Aleksandra Deluka-Tibljaš, Lidija Kraljević, Anamarija Štefić

Abstract: According to some previous researches the English language knowledge, turned out to be important, both in professional and private life of adult population, including students, from the Republic of Croatia. This paper presents a study of the influence of knowledge and application of the English language on the choice of students' mobility mode during their stay in foreign cities. The research is based on a survey of students studying civil engineering and architecture from two Croatian towns - Osijek and Rijeka. The survey included a total of 221 students from both towns, and based on the created databases, a prediction model of the choice of mobility modality was developed (based on the data gathered from the students from Osijek), which was validated (based on the data gathered from the students from Rijeka). The mobility modality prediction model was developed using a neural network and with 27 input parameters related to the knowledge and use of the English language, it achieved a correlation of 77% (R^2 0.57, *MAE* 0.16), and the validation results gave a correlation of 63% (R^2 0.39, MAE 0.18). The influence of certain input parameters on the choice of mobility modality of students in foreign cities was analyzed by applying statistical tools in each urban area separately. The main limitation concerns both the sample size and the unequal distribution of respondents between the two cities, which should be taken into account when interpreting the comparison.

Keywords: choice of mobility mode; English language; foreign cities; neural network; students

1 INTRODUCTION

Sustainable mobility has become the goal of contemporary transport policy as cities strive to reduce emissions, improve public health, and create more livable urban environments. This shift encourages a transition toward multimodal networks that support cycling and walking both domestically and abroad [1, 2], a process that relies on a well-informed public capable of understanding key mobility concepts. As sustainable mobility becomes increasingly international, basic English proficiency enables users to interpret transport information accurately and make consistent, environmentally responsible choices of urban mobility modes.

The research conducted in 2003 showed that English is the most important foreign language in private and professional life of the adult population in Croatia [3]. In the research, the term personal life referred to everything related to home, family, free time, internet, etc. whereas the term professional life applied to the requirements of the profession - workplace and studying as a profession. In this research [3], from the total sample of 388 participants, there was 22.4% of students. In 2007, the same group of authors completed the research with a targeted survey of 250 students of the University of Zagreb, through which they wanted to determine in more detail whether and for what purposes students needed the English language [4].

In both studies, a great proportion of students pointed out that they needed the English language - in the 2007 survey, even 99.2% of the surveyed students stated the same. The most frequent situation in which students needed English was when they were using the Internet, followed by when they were watching movies and giving information to foreigners. Traveling was listed as the 12th out of 87 offered situations where they needed English language proficiency and traffic as the 63rd one. The overall author's conclusion was that students considered English indispensable not only for studying, but also, even significantly more so, in their private lives.

Sustainable mobility has become the goal of the transport system development since the release of the Green paper in 1992 [5]. Promoting sustainable mobility is the key objective in transport policies across all government levels. Today, any plan, project, or policy related to the transport sector, invariably includes the concept of sustainable mobility. Sustainable aspects of urban mobility promote environmentally acceptable modes of mobility such as walking and cycling, as individual forms of movement that, in addition to positive social goals, also have beneficial health impacts. An important factor of achieving results in changing the movement mode is investing in the efficient public city transport, which, as it turns out, plays a key role in reducing greenhouse gas emissions [6].

Knowing the spatial features of the urban transport network and different options of mobility modes is considered a prerequisite for choosing adequate mobility modalities, according to the results of some earlier researches [7]. When considering tourists as users of the transport system (analyzed in more detail in the next chapter), these researches focused on determining the legality of tourists' movement towards destinations, as well as on factors influencing the choice of movement mode in the destination area - inside or outside the urban area [1, 2, 8-13]. One of the factors that was found to affect the choice of movement mode was the tourist age - younger tourists and students, as a special group of younger tourists, displayed a greater tendency to use public city transport and to cycle and walk [1, 2].

This paper investigates the extent to which students' English language proficiency influences their choice of mobility modes in foreign cities. The scientific hypothesis of the study is that English language knowledge has an impact on the selection of urban mobility modes when navigating unfamiliar environments abroad. The research was conducted using a survey among civil engineering and architecture students from Osijek and Rijeka, yielding a total sample of 221 respondents. The analysis examined how English language skills affect students' ability to navigate

urban traffic systems and their subsequent mobility choices. Survey results from the two cities were compared, and a neural-network-based prediction model of mobility mode choice was developed. The model was trained on data collected in Osijek and validated on an independent dataset from Rijeka, enabling the identification of key parameters that shape students' mobility behaviour in foreign urban settings.

2 RESEARCH REVIEW

The topic of the way tourists move when arriving at a specific destination and in the area of a tourist destination itself, is in the focus of a series of studies and researches all over Europe and the world.

The project co-financed by the EU civitas destinations tested during the period of 4 years, 83 different mobility modes on the 6 European islands, addressing mutual empowerment aspects of tourism and mobility with the focus on positive and negative interactions at the city level. The 3 clusters analyzed in the Project were: shared mobility and walking, clean vehicles and public transport for all of which a series of measures and indicators were defined. The main conclusions, regarding traffic development, were: that it was necessary to systematically monitor data on the mobility of tourists, that tourists liked to explore destinations on foot, and that transport was an important part of the touristic experience [8]. Researches of the impact of traffic on the sustainable development of tourist destinations conducted for the Tyrol area showed that in such a natural environment, 40–60% of the environmental load linked to tourism were caused by transport of tourists between their homes and their destinations, as well as by local transport within the destination area.

Several studies were published on the topic related to the traffic generated by tourism in Norway [9], that is, by the traffic habits of individual tourists [10]. Tourism - leisure mobility was detected as a strong generator of movement, and action plans that would influence the reduction of the use of cars and airplanes were defined as an important factor of sustainable development in general.

Researches on the use of transport modes during a tourist stay have recently been motivated by a large increase in tourist movements, which create additional pressure on existing transport systems and have negative environmental impacts to cities.

Authors Mahdi et al. [11] developed models to predict the probability of using different means of transportation for leisure purposes based on travelers' and journey features. It turned out that the sociodemographic characteristics of users influenced the choice of a mobility mode. Younger people and men had, according to this Budapest research, a greater affinity towards using bicycles, while women showed a greater interest in walking. It also turned out that public transport was used more often by families with more than 2 members and those with a lower financial status.

The developed transport mode choice Binary logit model revealed that the key factors influencing the choice of a mobility mode, in the case of tourists in Beijing, were the scope of travel, time of visit, daily consumption, age, private vehicle possession, and type of tourist attractions [12].

The study conducted on Malta based on the semi-structured in-depth interviews with tourists provided insight for the need to better understand the influence of tourists' attitudes towards a bus service in the host country [13]. The factors that were defined, by the visitors, tourists and those staying on Malta for educational reasons, as important for quality bus service were: time (punctuality), information (accuracy), bus drivers (behavior), fare, fleet and environmental impact.

The analysis of the existing research presented in the paper [1] detected some specific socio-demographic characteristics of tourists that had an impact on the individual's choice of city means of transportation. It turned out that younger tourists showed a greater affinity towards the use of public transport and, generally, sustainable forms of mobility. However, the influence of gender on the choice of the mobility mode was not unequivocal, but that did not appear to be significant among the younger population (students). The home countries of tourists, their level of education, income and possession of a driver's license were also analyzed in the research as potential influencing factors on the use of public transport, though the influence of these parameters was not unambiguously proven. The influence of language proficiency, which was analyzed in a smaller number of studies, unequivocally showed that language knowledge affected the choice of public means of transportation in a tourist destination - individuals who were not proficient in local language seemed to be less inclined to utilize public transportation [1].

Assuming that environmentally friendly transportation for visitors is the pre-condition for creating a sustainable tourist destination, the authors Zamparini et al. [2] conducted a study with the aim of identifying the determinants of green mobility choices made by young adults during their vacation. The survey was conducted on a sample of 979 students from Italy and Spain. The results of this research showed that, when it came to green mobility modes, there was a high correlation between the students' patterns of traffic behavior when they were at home and the way they behaved at a tourist destination. It turned out that those respondents who arrived at the destination by their own cars, used their private vehicles to move around, as well as that the willingness to use green forms of mobility largely depended on how tourist attractions and destinations were interconnected.

The adjustment of transport choices during international temporary relocations of students and researchers was also studied during the last decade due to its ever-increasing relevance. It is expected, among other things, that in the period of 2021-2027, 10 million students and teachers will be using ERASMUS+ mobility and stay temporarily outside their place of study/work for a duration of one week to a year [14]. A survey-based study by Monteiro et al [15] focused on frequency of use of public transport and satisfaction with it. The model showed that public transport in the city where the respondents were temporarily staying was used more by those students and researchers who usually used it in their home city and that the use of public transport depended on where they were located in the host city. It also turned out that the higher the level of perception that the of use public transport technology facilitates movement, the more often they used it.

The analysis of the existing research shows that the use of a certain form of mobility, including green forms of mobility (walking, cycling, public transport) is usually linked with the mode of movement that tourists choose in their home places and with the interconnection of certain tourist attractions with a particular form of transport. The influence of tourist sociodemographic features is also significant, among which the tourist age stands out.

Knowledge of language as a predictor of the choice of mobility mode in a tourist destination has not been investigated in detail yet, although existing available researches show that it does influence the choice of travel mode (individual/in a group) and use of public transport in the tourist destination.

The student population, made up of younger age groups, is more inclined to use public transport and generally sustainable forms of mobility but the question arises as to how much the English language knowledge, as a prerequisite for understanding how these mobility modes function, influences their choices.

3 METHODOLOGY

The research, based on the results of the survey of civil engineering and architecture students from two towns in Croatia, Osijek and Rijeka, focused on the extent to which knowledge of the English language benefits them in choosing a mobility modality in foreign cities. The basic methodological research steps are shown in Fig. 1.

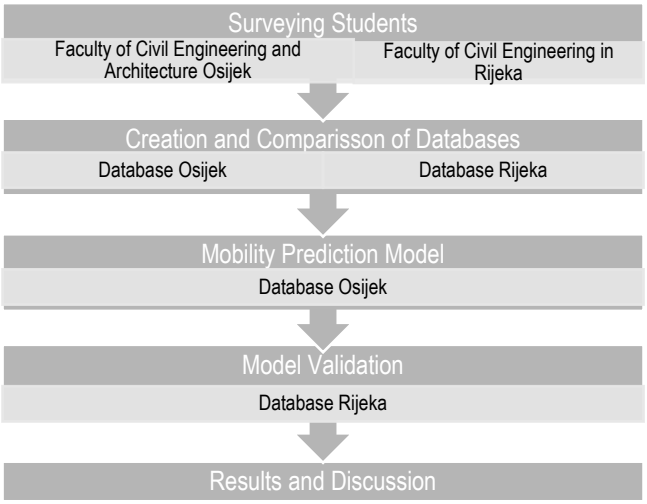


Figure 1 Scheme of basic methodological steps

The towns of Rijeka and Osijek are very similar in terms of number of inhabitants, but very different in terms of population density, which is in Rijeka, according to GeoStat data for 2021 [16], the highest in Croatia with 25 inhabitants/hectare, compared with Osijek, which has 5,5 inhabitants/hectare. This difference conditions the way their urban space and the transport system were developed. In Osijek, when talking about green modes of mobility, in addition to public city transport, the conditions for bicycle traffic are also being developed, so that students of the Faculty of Civil Engineering and Architecture Osijek regularly use bicycles in daily traffic. Rijeka is a town located on topographically very unfavorable terrain with a high density of content, which consequently results (when talking about green mobility modes) in a higher proportion of walking daily trips and an almost complete absence of the use of bicycles in city traffic. Previous researches of the authors showed a significant difference in the drivers' behavior [17] and the behavior of child pedestrians [18], so that a difference in the choice of mobility had also been expected.

Since the analysis of the existing research showed that the choice of mobility in foreign cities is influenced by mobility habits acquired in the domicile environment, the survey focused on the analysis of this important aspect.

3.1 Survey

The student survey was conducted online, while the instructions to the surveyed students were partly given in direct communication with different groups of students. The survey was voluntary and anonymous, and the participants were explained the procedure of the survey and the goal of the research, noting that the results will be analyzed and presented cumulatively and statistically. All surveyed students were of legal age and gave their consent for the survey.

The structure and content of the questionnaire is shown in Tab. 1.

Table 1 Structure and content of the survey questionnaire

Demographic questions	I1	Gender	text
	I2	Age	Multiple choice
		Birthplace	text
		Faculty	text
		University	text
Self-evaluation of the English language knowledge	I3	Knowledge of English language	Evaluation from 1-5
Number of visited (foreign) cities/towns	I4	up to 3 4-6 7-10 more than 10	Multiple choice
English language use		Choosing a travel destination Selection of accommodation Travel planning Planning activities during the stay abroad	Multiple choice
Use of the English language in spatial orientation	I5 I6 I7 I8 I9	Spatial orientation Finding desired urban content Tour planning and its realization Pedestrian orientation Bicycle orientation	Evaluation from 1-4
Use of the English language in public transportation	I10 I11 I12 I13	Direction recognition Finding the right type of public transport Finding a suitable line in public transport Purchase of public transport tickets	Evaluation from 1-4
Use of the English language when driving a car		Direction recognition Finding a suitable road route Finding the target destination Connecting different locations Finding a parking lot Parking ticket purchase	Evaluation from 1-4

Use of the English language for network applications	I14 I15 I16	for car driving for public transportation for cycling for walking	Evaluation from 1-4
Use of English language in direct communication	I17 I18 I19 I20 I21 I22	with passers-by on the street with passengers in public transportation with people who rent cars, bikes, scooters, etc. with restaurant or hotel staff with tourist guides with taxi drivers with other people in other situations	Evaluation from 1-4
Factors influencing the experience of a foreign city	I23 I24 I25 I26 I27	Direct communication with residents Traffic orientation Attractive public areas for pedestrians Spatial orientation Perception of pedestrian safety	Evaluation from 1-4
Choice of mobility modality in foreign cities/towns	O	Car (as a driver) Car (as a passenger) Public transportation Bike / e-bike Scooter / e-scooter Walking Other: specify	Multiple choice

3.2 Databases

Based on the respondents' answers, two databases were created to analyze and compare the mobility mode choices of students abroad coming from two similarly sized urban environments that differ in their spatial and traffic characteristics: the cities of Osijek and Rijeka.

All input influencing variables were categorical (I1-I27), as was the dependent variable (O) – the choice of mobility mode in foreign cities, with an emphasis on the choice of green forms of mobility. In this study, public transport is considered "green" due to its lower emissions and energy consumption per passenger compared with private cars, as well as its contribution to reducing congestion and spatial pressure. The analysis shows [1, 6, 7, 12] that environmental friendliness has a positive but secondary influence on transport mode choice.

In this study the choice of mobility modality was a multiple choice with the possibility of multiple answers (Tab. 1), and when the database was created, it was converted into a categorical numerical variable. For respondents who only used a passenger car, either as drivers or passengers, the associated numerical value was 0. For respondents who used only green forms of mobility, which include walking, cycling, use of mobility scooters and public city transport, the numerical value was 1. For all the combinations of the use of a passenger car and green forms of mobility, the associated numerical value was 0.5. The decision to group the dependent variable into three categories enhances model robustness and enables an initial insight into the effect of each predictor, thereby providing a foundation for more sensitive modelling.

3.3 Prediction Model

A cascade-correlation neural network, using Neuroshell Predictor, was applied for creating the models. This network, firstly introduced by Fahlman and Lebiere [19], is a supervised self-organizing network with a structure that is similar to backpropagation neural networks. The cascade-correlation network in NeuroShell Predictor is composed of three layers: the input layer, where the number of neurons ($i1$, $i2$, …, in) corresponds to the number of model input variables; the output layer with one output neuron ($o1$) and hidden layer with the number of neurons increased during training. These neurons are linked by a connection, whose weight is adjustable, and by a bias, permanently set to +1 neuron in each new epoch of learning.

Prediction of mobility mode choice is an area of significant scientific interest [2, 11, 12]. The advantage of neural networks as a prediction model lies in their ability to accurately map non-linear relationships between input parameters and the dependent variable, as well as to identify the relative influence of each input parameter on the prediction outcome.

4 RESULTS

In Osijek, 162 students of the Faculty of Civil Engineering and Architecture were surveyed, and in Rijeka there were 59 surveyed students of the Faculty of Civil Engineering. There is a significant difference in the number of respondents between the two cities, so the comparison and analysis of the results should be interpreted within that context. Although a smaller number of respondents may reduce the reliability of the findings, the data can still serve as a valuable starting point for analyses and comparisons.

The question related to the place of birth was asked with the potential idea of analyzing the influence of urban and rural environments on the choice of mobility, but all the respondents in both databases came from smaller or larger urban environments.

The distribution of the respondents by gender, expressed in percentages, is shown in Fig. 2, and in both environments, it was similar, with a higher number of female respondents.

The age range (I2) of the surveyed students in Osijek was from 18 to 29 years, and in Rijeka it was slightly higher, from 18 to 32 years, which was also reflected in the average age of the respondents (Fig. 3). The graph in Fig. 3 shows the average ratings of all the variables, except for the distribution by gender, which is shown separately.

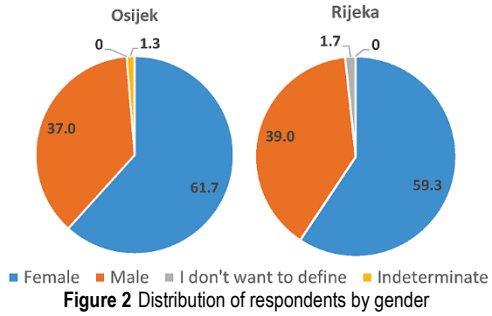


Figure 2 Distribution of respondents by gender

The average score of the self-evaluated English language knowledge (I3) was statistically not significantly different, students from Rijeka scored slightly better. There was a noticeable difference in the number of visited foreign cities

(I4). Namely, the students from Rijeka visited, on average, a larger number of foreign cities (9 compared to 7 visited by the Osijek students). The students from Rijeka gave a lower score for the use of the English language to navigate cycling (I9) and the use of applications intended for cycling (I15), which was the expected result, because they choose cycling, as a form of mobility, less often, as visible from Fig. 4. According to the results of the survey, the students from Rijeka more often engaged in English conversation (I17-I22), as visible from Fig. 3.

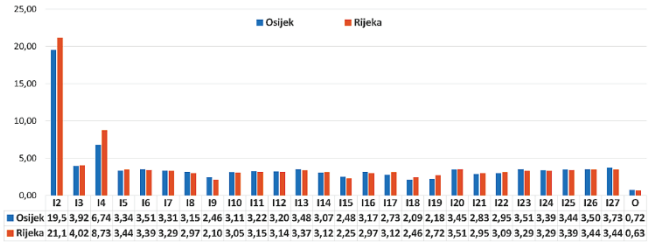


Figure 3 Comparison of average scores of individual parameters

The selection of green forms of mobility (*O*) is slightly higher among the students from Osijek, but that difference might have also been caused by the number of the respondents.

The comparison of the mobility choice between the students of the two towns is shown in percentages in the graph in Fig. 4. Each respondent could have chosen more than one answer.

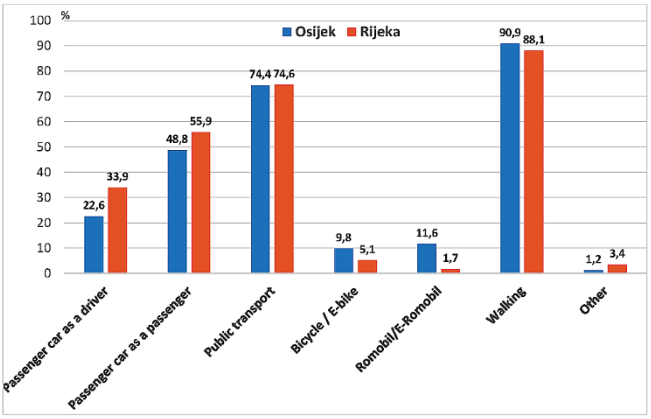


Figure 4 Comparison of mobility modality selection

4.1 Osijek and Rijeka Database

The results of the descriptive statistics for the dependent variable (*O*), which is the choice of mobility modality for both databases, are presented in Tab. 2.

Table 2 Basic statistical indicators

	N	*Mean*	*StDev*	*Median*	*Variance*
Osijek	162	0,7160	0,2780	0,5000	0,0768
Rijeka	59	0,6271	0,2876	0,5000	0,0813

All the variables are categorical, so the Spearman Rho statistical test was chosen to analyze the correlations between the input parameters and the dependent variable (Tab. 3). The Spearman correlation does not require continuous-level data and works with monotonic relationships [20, 21].

Considering that these were preliminary analyses, a significance threshold of 0.1 was adopted. In bold, the table presents the results for the input parameters that showed correlation with the dependent variable according to the adopted significance threshold.

Table 3 Comparative analysis of correlations using the Spearman Rho test

		Osijek		Rijeka	
		*S-R**	*p***	*S-R*	*p*
Gender	I1	0,014	0,856	0,117	0,379
Age	I2	-0,03	0,696	0,086	0,517
Self-evaluation of the English language knowledge	I3	0,017	0,833	0,308	**0,018**
Number of visited cities/towns	I4	–0,04	0,595	0,101	0,445
Spatial orientation	I5	0,005	0,953	0,237	**0,074**
Finding desired urban contents	I6	0,007	0,928	0,044	0,739
Tour planning and its realization	I7	0,031	0,674	0,011	0,933
Pedestrian orientation	I8	0,017	0,826	0,012	0,931
Bicycle orientation	I9	–0,02	0,817	0,044	0,743
Direction recognition in public transportation	I10	0,023	0,775	0,036	0,786
Finding the right type of public transport	I11	0,055	0,491	0,018	0,891
Finding a suitable line in public transport	I12	0,021	0,788	0,070	0,599
Purchase of public transport tickets	I13	0,137	**0,083**	0,064	0,632
Network applications for public transportation	I14	0,032	0,690	0,127	0,337
Network applications for cycling	I15	0,021	0,790	0,145	0,272
Network applications for walking	I16	0,106	**0,149**	0,051	0,699
Communication with passers-by on the street	I17	0,031	0,696	0,018	0,895
Communication with passengers in public transportation	I18	0,021	0,794	0,052	0,696
Communication with people who rent cars, bikes	I19	0,155	**0,048**	0,101	0,449
Communication with restaurant or hotel staff	I20	0,008	0,916	0,069	0,601
Communication with tourist guides	I21	–0,01	0,886	0,097	0,465
Communication with other people in other situations	I22	0,087	0,270	0,085	0,527
Significance of direct communication with residents	I23	0,145	**0,066**	0,025	0,852
Significance of traffic orientation	I24	0,134	**0,090**	0,101	0,445
Significance of attractive public areas for pedestrians	I25	0,227	**0,004**	0,127	**0,106**
Significance of spatial orientation	I26	0,151	**0,055**	0,033	0,803
Significance of pedestrian perception of safety	I27	0,127	**0,109**	0,124	**0,135**

*Spearman rho correlation coefficient; ***p*-value

4.2 Model Results

The statistical analysis showed a correlation between individual variables of the English language use and the mobility choice in foreign cities. The model was developed by means of the neural network, with the database from Osijek serving as the base for the development of the prediction model. The model provided a prediction for the choice of mobility mode, and the input parameters of the model are shown in Tab. 1. The prediction model was developed by the entire database, the network structure and learning results are shown in Tab. 4.

The neural network had three layers, the input layer had 27 neurons that corresponded to the number of the input parameters, and the output layer had one neuron that corresponded to the dependent variable (O) – the choice of mobility modality. The number of neurons in the hidden layer was subject to optimization during learning, as well as the weight coefficients of the individual input parameters.

The learning outcomes were assessed using the selected statistical indicators. The correlation coefficient is a statistical measure of the strength of the relationship, the coefficient of determination (R^2), the mean absolute error (*MAE*), the mean square error (*MSE*), and the root mean square error of prediction (*RMSE*).

Table 4 Neural Network Structure and Learning Outcome

Structure of a neural network			Learning outcomes				
Number of neurons							
Input layer	Hidden layer	Output layer	Correl coef	R^2	*MAE*	*MSE*	*RMSE*
27	79	1	0,852	0,726	0,117	0,021	0,145

Generalization of the model is the application of the formed prediction model to the test data set. The results of the generalization of the model are shown in Fig. 5 and Tab. 5.

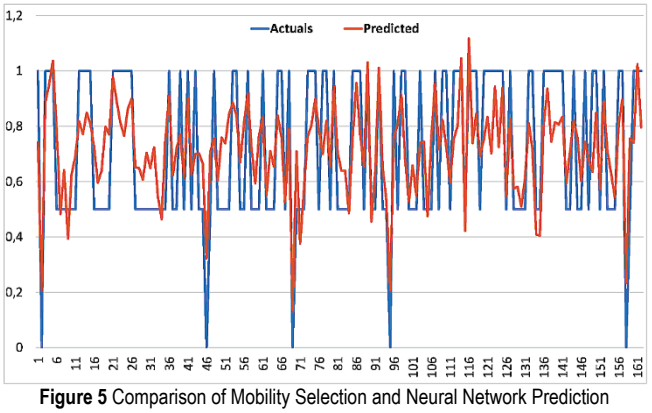


Figure 5 Comparison of Mobility Selection and Neural Network Prediction

Table 5 The result of the generalization of the model

The generalization model				
Correlation coefficient	R^2	*MAE*	*MSE*	*RMSE*
0,771	0,565	0,162	0,033	0,183

The entire database was selected for the test data set, which meant that the generalization was evaluated on the entire database. The results of the generalization were considered relevant for evaluating the applicability of the model. The results showed that the most common form of mobility was a combination between a personal car and green mobility modes. The results also showed that there were more green mobility options than just using a personal car.

4.3 Model Validation

The validation of the model was performed on the database created by surveying the students from Rijeka, on the independent database, which the neural network did not see during learning.

The validation results are shown in Tab. 6 and Fig. 6.

Table 6 Model validation results

Model validation				
Correlation coefficient	R^2	*MAE*	*MSE*	*RMSE*
0,626	0,392	0,180	0,050	0,225

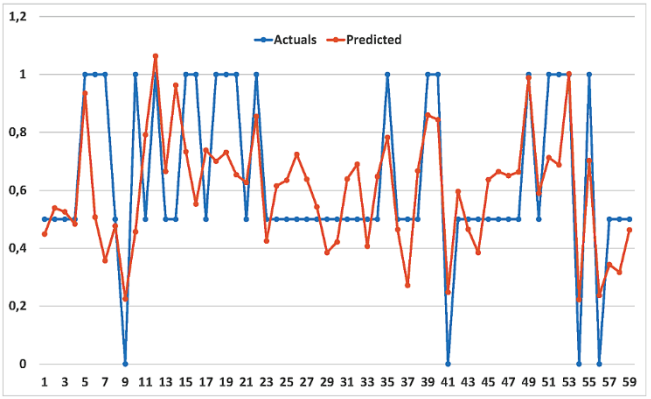


Figure 6 Model validation results

5 DISCUSSION

The database analysis showed a significant difference in the number of the visited foreign cities, students from Rijeka visited an average of 9 cities, whereas students from Osijek an average of 7 cities. The self-evaluation of the English language knowledge showed a high score of 4 out of 5, for Rijeka it was 4.0 and for Osijek 3.9. The difference in grades was visible for the use of the English language to navigate cycling and use applications intended for cycling, which was the expected result, because students from Rijeka chose cycling less often as mobility option. The results of the survey showed that the students from Rijeka engaged more often in English conversation, which could have been expected because they self-rated their knowledge of the English language higher than their colleagues from Osijek.

The analysis of the collected data revealed that walking in foreign cities was the dominant mobility mode of all the surveyed students, which was an expected result considering that these were young people who had no physical limitations for walking. Walking was chosen as a favorable mobility mode by 91% of the students from Osijek and 88% of the students from Rijeka. Public transport was the second ranked mobility mode, chosen by 74% of the respondents in both databases, which was, the expected choice for more distant destinations. The results coincided with the previous researches according to which younger visitors (tourists)

favored exactly these two mobility forms for moving in their tourist destinations [1, 11].

Personal cars were used more by the students from Rijeka, 56% of them used a car as passengers and 34% of them as drivers, compared to 49% of the students from Osijek, who used a car as passengers, and 23% of them as drivers, which can be explained by the difference in geographical position of these two towns. Namely, Rijeka is located a few hours' drive (approx. 250-300 km) from several attractive tourist destinations where it is necessary to use a foreign language (for example, Trieste, Venezia, Ljubljana, Graz).

The students from Osijek chose scooters and bicycles as a form of mobility more often than the students from Rijeka. A bicycle as a mobility mode was chosen by this student group twice as frequent (10% Osijek, 5% Rijeka), which was also expected, because students from Osijek tend to use bicycles as a mobility form in their home town, unlike the students from Rijeka, who do not have a bicycle infrastructure. The research conducted specifically on the student population showed that the choice of movement mode in another urban environment correlated with the same choice in one's own city [2, 15], which was also confirmed by the research results presented in this paper.

What was unexpected was that the students from Osijek chose a scooter more often (12%) than a bicycle (10%), in contrast to the students from Rijeka who chose a bicycle (5%) more often than a scooter (2%). It is evident that an e-scooter, as a mobility form, is getting more popular as a local mobility option for the students from Osijek, too.

The statistical analysis (Tab. 3) showed that there was a correlation between 10 input variables and the dependent variable of the choice of mobility modality for both databases, of which 6 variables were found to be significant for the database of the students from Osijek:

- language use when buying tickets for public transport (I13),
- the use of languages for online hiking applications (I16),
- use of language in direct communication with car, bicycle and mobile vehicle renters (I19),
- use of language in direct communication with residents (I23),
- the significance of traffic orientation (I24),
- the importance of spatial orientation (I26).

For the database of the students from Rijeka, 2 variables proved to be significant, which were not significant for the students from Osijek:

- self-evaluation of the English language knowledge (I3) and
- application of the English language in spatial orientation (I5).

Two input variables that proved to be significant for both databases were related to the spatial and safety perception of public areas intended for pedestrians:

- attractive public pedestrian areas (I25) and
- feeling of safety when hiking (I27).

The prediction model was formed on the basis of the data from Osijek, and validated on the basis of the data from Rijeka. When forming the prediction model, the neural network did not denote a single variable as uninfluential, which meant that the prediction function used all 27 input parameters.

Given the size of the database used to train the neural network—162 respondents and 27 input parameters—there was a concern about potential overfitting and limited generalization. Generalization was therefore evaluated in two stages: first, on the test data set of the original training database (Osijek), and second, on a completely independent database that the network had not encountered during training, collected in Rijeka.

The correlation achieved by the neural network on the training data set was 85%, but the valid correlation achieved on the test data set was 77%, which can be evaluated as a good correlation according to the indicators shown in Tab.5. The validation of the model was made on the basis of the data from Rijeka, and the achieved correlation was 63% (Tab. 6), which could be explained in the context of the students coming from two different towns, that make different choices of mobility modes and their different traffic habits, which could have been confirmed by the previous researches [17, 18].

The results of the model validation shown in Tab. 6 were evaluated as acceptable, which confirms that the prediction model can be applied in other urban environments as well.

When interpreting the results, one should consider the study's limitations, especially the modest total sample size and the imbalance in the number of respondents between the two cities.

6 CONCLUSION

The analysis of the influence of the English language knowledge on the choice of a urban mobility mode abroad, based on the results of a survey of the students from Osijek and Rijeka, confirmed the differences in the traffic habits of these two groups of students. The differences in the students' choices of mobility forms can be related to a number of influential factors (as shown by the results of this and previous researches on the behavior of pedestrians and drivers in two observed urban environments), among which is knowledge of the English language. A number of the analyzed input parameters that are related to the knowledge and use of the English language have a statistically significant influence on the choice of mobility modality, as the research results presented in this paper showed. For the students from Rijeka, their perception and self-evaluation of their knowledge of the English language on the choice of mobility modality turned out to be statistically significant, which is an interesting result that needs to be further investigated.

The input parameters that proved to be relevant in both databases are attractive areas for pedestrians and the perception of pedestrian safety, which are not directly related to the knowledge of the English language, but to the perception of the urban environment and the traffic network. To what extent these parameters can be indirectly connected

with the knowledge of the English language, is the topic that remains to be further investigated, yet.

The results of this research confirmed the results of other researchers conducted elsewhere in the world, which indicate the preference of the younger tourist population for green mobility modes and also show the influence of their domicile traffic habits on choosing mobility modalities abroad.

Public transport can be considered a form of green mobility that can potentially be used by all tourists, which is not the case with other forms of green mobility – walking, using bicycles or scooters. The student population is the one which, due to their objective physical abilities, should be ready to use both public transport and all other forms of green mobility at their tourist destinations. A prerequisite for this is the ability to navigate unfamiliar urban environments and systems (IT) required for renting vehicles and knowledge of the English language plays here a very important role.

The research limitations that should be kept in mind when analyzing the obtained results are the number of surveyed students of a certain profile and the fact that only two urban areas were analyzed. The results of this research should be supplemented with the results of the conducted research on the students' mobility modalities in their local, domicile surrounding, the collection and analysis of which is in progress.

Acknowledgments

This research is conducted within the framework of the European Union—NextGenerationEU ("Safety aspects of urban transport infrastructure SigINFRA", a scientific project at the Josip Juraj Strossmayer University of Osijek (581-UNIOS-71) and "Optimizing Transport Infrastructure in the Function of Sustainable Mobility in the Cities", a scientific project at the University of Rijeka (uniri-iz-25-32).

7 REFERENCES

[1] Zientara, P., Jażdżewska-Gutta, M., Bąk, M., & Zamojska, A. (2024). The attitudes of tourists towards a bus service: implications for policy from a Maltese case study. *Journal of Destination Marketing & Management, 33*, 100931. https://doi.org/10.1016/j.jdmm.2024.100931

[2] Zamparini, L., Domènech, A., Miravet, D., & Gutiérrez, A. (2022). Green mobility at home, green mobility at tourism destinations: A cross-country study of transport modal choices of educated young adults. *Journal of Transport Geography, 103*, 103412. https://doi.org/10.1016/j.jtrangeo.2022.103412

[3] Cindrić, I., & Narančić Kovač, S. (2005). English and Other Foreign Languages in Croatia: A Needs Analysis. *Strani jezici: časopis za primijenjenu lingvistiku, 34*(3), 189-204.

[4] Narančić Kovač, S., & Cindrić, I. (2007). English language needs of Croatian students. *Metodika: časopis za teoriju i praksu metodikâ u predškolskom odgoju, školskoj i visokoškolskoj izobrazbi, 8*(14), 51-67.

[5] Commision of the European Communities – GREEN PAPER on The Impact of Transport on the Envieronment; A Community strategy for "sustainable mobility", https://op.europa.eu/en/publication-detail/-/publication/98dc7e2c-6a66-483a-875e-87648c1d75c8/language-en; Accessed: 1st October 2024.

[6] Gallo, M., & Marinelli, M. (2020). Sustainable Mobility: A Review of Possible Actions and Policies. *Sustainability, 12*(18), 7499. https://doi.org/10.3390/su12187499

[7] Mondschein, A., Blumenberg, E., & Taylor, B. (2010). Accessibility and Cognition: The Effect of Transport Mode on Spatial Knowledge. *Urban Studies, 47*(4), 845–866. https://doi.org/10.1177/0042098009351186

[8] Mantero, C. (2022). Sustainable, Smart and Safe Mobility at the Core of Sustainable Tourism in Six European Islands. In: Tsoutsos, T. (eds) *Sustainable Mobility for Island Destinations*. Springer, Cham. https://doi.org/10.1007/978-3-030-73715-3_1

[9] Høyer, K. G. (2000). Sustainable Tourism or Sustainable Mobility? The Norwegian Case. *Journal of Sustainable Tourism, 8*(2), 147–160. https://doi.org/10.1080/09669580008667354

[10] Butler, G., & Hannam, K. (2012). Independent tourist's automobilities in Norway. *Journal of Tourism and Cultural Change, 10*(4), 285-300. https://doi.org/10.1080/14766825.2012.718441

[11] Mahdi, A., Hamadneh, J., & Esztergár-Kiss, D. (2022). Modeling of travel behavior in Budapest: Leisure travelers. *Transportation Research Procedia, 62*, 310-317. https://doi.org/10.1016/j.trpro.2022.02.039

[12] Gong, X., Yan, H., Liu, S., & Xu, Z. (2011). Tourism Traffic Mode Choice Behavior Model. In *ICCTP 2011: Towards Sustainable Transportation Systems*, 766-775. https://doi.org/10.1061/41186(421)75

[13] Bajada, T., & Titheridge, H. (2017). The attitudes of tourists towards a bus service: implications for policy from a Maltese case study. *Transportation research procedia, 25*, 4110-4129. https://doi.org/10.1016/j.trpro.2017.05.342

[14] European Union Erasmus+ 2021–2027. Available online: https://op.europa.eu/s/oTcf; Accessed: 1st October 2024.

[15] Monteiro, M. M., de Abreu e Silva, J., Ingvardson, J. B., Nielsen, O. A., & Pinho de Sousa, J. (2021). Public transport use and satisfaction by international students and researchers. *Sustainability, 13*(15), 8417. https://doi.org/10.3390/su13158417

[16] Državni zavod za statistiku. https://geostat.dzs.hr/?m_x=500000&m_y=4950000&m_zoom=0&lang=hr&layers=11370:1:1&baselayer=DOF, Accessed 5th October 2024.

[17] Ištoka Otković, I., Deluka-Tibljaš, A., & Šurdonja, S. (2020). Validation of the calibration methodology of the micro-simulation traffic model. *Transportation Research Procedia, 45*, 684-691. https://doi.org/10.1016/j.trpro.2020.02.110

[18] Deluka-Tibljaš, A., Šurdonja, S., & Ištoka Otković, I. (2021). Analysis of Children's Traffic Behaviour at Signalized Crosswalks as a Precondition for Safe Children Routes Design: A Case Study from Croatia. *Journal of advanced transportation, 2021*(1), 7936261. https://doi.org/10.1155/2021/7936261

[19] Fahlman, S. E., & Lebiere, C. (1990). The Cascade-Correlation Learning Architecture in Advances in Neural Information Processing Systems 2. D. S. Touretzky, Ed. Los Altos, USA: Morgan-Kauffman, 524–532.

[20] Rovetta A. (2020). Raiders of the Lost Correlation: A Guide on Using Pearson and Spearman Coefficients to Detect Hidden Correlations in Medical Sciences. *Cureus, 12*(11), e11794. https://doi.org/10.7759/cureus.11794

[21] Schober, P., Boer, C., & Schwarte, L. A. (2018). Correlation Coefficients: Appropriate Use and Interpretation. *Anesth Analg, 126*(5), 1763-1768. https://doi.org/10.1213/ANE.0000000000002864

Authors' contacts:

Irena Ištoka Otković, full professor, PhD
Faculty of Civil Engineering and Architecture Osijek,
Josip Juraj Strossmayer University of Osijek,
Vladimira Preloga 3, 31000 Osijek, Croatia
+385 31 544 683, iirena@gfos.hr

Aleksandra Deluka-Tibljaš, full professor, PhD
Faculty of Civil Engineering, University of Rijeka,
Radmile Matejčić 3, 51000 Rijeka, Croatia
+385 51 265 900/ +385 51 265 998, aleksandra.deluka@gradri.uniri.hr

Lidija Kraljević, senior lecturer
Faculty of Civil Engineering and Architecture Osijek,
Josip Juraj Strossmayer University of Osijek,
Vladimira Preloga 3, 31000 Osijek, Croatia
+385 31 544 702, lidijak@gfos.hr

Anamarija Štefić, senior lecturer
Faculty of Civil Engineering and Architecture Osijek,
Josip Juraj Strossmayer University of Osijek,
Vladimira Preloga 3, 31000 Osijek, Croatia
+385 31 544 702, biskup@gfos.hr

ISSN 1846-6168 (Print), ISSN 1848-5588 (Online)
https://doi.org/10.31803/tg-20250415195518

Preliminary communication
Received: 2025-04-15, Accepted: 2025-04-30

An Ontology-Driven Approach to Improve Data Understanding for Machine Learning Applications in Manufacturing

Fabian Weisbrodt*, Jürgen Bock, Tobias Schrage, Peter Schuderer

Abstract: The application of Machine Learning (ML) methods represents a significant aspect in the advancement of Industry 4.0. The creation of an appropriate data set for these applications has been identified as the most time-consuming step in the underlying end-to-end pipeline. One of the major obstacles in this process step is to bridge the gap between business understanding and data understanding. To address this challenge, we propose a novel methodology to bridge this gap based on a systematic literature review. Our methodology begins with the construction of an ontology that depicts the underlying manufacturing process along with its parameters. We then show how this ontology can be utilized to deepen the understanding of the manufacturing process. Subsequently, we demonstrate how appropriate target variables for ML-models and suitable data sources can be determined with the support of our ontology. We further elucidate our methodology through a real-world example.

Keywords: data acquisition; knowledge graph; Machine Learning (ML); manufacturing; ontology

1 INTRODUCTION

Taking up to 70% of the entire time, data preparation is the most time-consuming step in an end-to-end Machine Learning pipeline [1]. In this publication, we tackle one of the key challenges when it comes to the task of data preparation: bridging the gap between business understanding and data understanding. To this end, we propose a methodology which supports the determination of appropriate target variables and corresponding data sources before setting up a suitable data pipeline. In addition, it provides a depiction of the underlying process which is suitable for exploratory knowledge discovery.

At the beginning, we define six questions one may ask when faced with the task of describing a manufacturing process, with the goal to identifying the target variables and respective data sources needed for the ML-models under consideration. These questions are:

1) What does the manufacturing process look like, i.e. what are the general steps of the process?
2) Which machines are used throughout the process?
3) What happens during the process, i.e. which materials are used in which way to manufacture the final product?
4) What are the attributes or parameters of the various components of the manufacturing process?
5) How are these attributes or parameters related?
6) How to obtain a potential dataset for these parameters?

To answer these questions, we utilize a predefined ontology to depict the respective manufacturing process. This process depiction includes all machines and materials that are used during the process as well as all parameters and their relationship to each other. The general idea is to follow a structured procedure that builds the description of the respective manufacturing process step by step. With an ontology being "a formal, explicit specification of a shared conceptualization" [2] it essentially enables knowledge-based applications by two benefits: (1) sharing and reusing knowledge, which is formally described using a standardized language, and (2) inferring implicit knowledge from explicitly given axioms and facts, based on its formal semantics. Ontologies that primarily focus on particular elements and their relations are often called knowledge graphs.

For us, this means that we only have to define the direct relationships between the respective elements (i.e. machines, materials, or parameters). Implicit relationships are automatically inferred by ontology reasoning using the formal semantics and the logical axioms that we defined about the relationships in a manufacturing process.

We conducted a systematic literature review using the Scopus search database. We used three queries that observed the abstract, title, and key words to search for relevant publications. Our search queries included the terms "manufacturing" and "machine learning" as well as "data acquisition", "feature engineering", and "data source" respectively. In total, we found 488 publications of which we will discuss five in detail that address bridging the gap between business understanding, i.e. domain knowledge, and data understanding. The systematic procedure we followed is summarized in Fig. 1.

In [3], the authors present a methodology for selecting suitable data sources based on a House of Quality matrix. However, the methodology proposed in [3] does not include a depiction of the manufacturing process itself and assumes that a preselection of suitable data sources is already given. In [4], the authors propose a methodology based on a matrix they developed called "ML-SIPOC". This methodology includes a description of the manufacturing process, although less detailed than ours. The main difference to our approach is the utilization of the reasoner, which results in us not having to do the entire reasoning about which data sources may be relevant. Furthermore, the main tools used in the methodology in [4] are the already mentioned ML-SIPOC matrix and flowcharts. These tools do not allow the same level of exploratory and interactive knowledge discovery about the manufacturing process and its parameters as our approach using Web Ontology Language (OWL) ontologies.

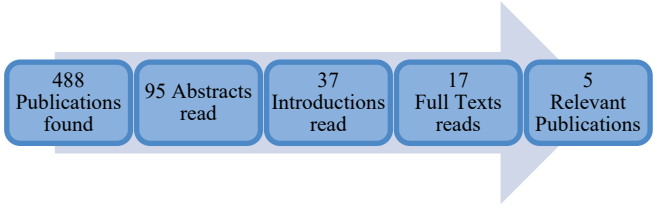


Figure 1 Summary of our literature research

In [5], the authors create an ontology to describe the features of a manufacturing process and use the resulting Resource Description Framework (RDF) graph to analytically compute the similarity of features using graph kernel functions. However, this approach is only used to enhance linear regression models and not generalized. In addition to the utilization of the RDF-graph the authors also present how the ontology they created can be used for feature preselection. This preselection depends on the relationships of the features in the proposed ontology. In comparison, we propose a more detailed description of the relationships between parameters, allowing for more sophisticated applications.

Finally, in [6,7], the authors present different ontologies with the goal of automating as much of the end-to-end ML-pipeline as possible. To this end, they propose the use of three different interconnected ontologies that contain general knowledge about discrete manufacturing processes, domain knowledge, and knowledge about ML-models. The creation of the domain ontology is done collaboratively by domain experts and data experts using a GUI developed in [7] called ontology extender. The idea here is similar to the approach we use to create an ontology that depicts the manufacturing process. However, the methodology in [6,7] is shown only for individual process steps and does not include the depiction of an entire manufacturing process. In addition, they do not provide a procedure for how to model a manufacturing process and its parameters.

The rest of this paper is organized as follows. In Section 2, we give a brief overview of ontologies and explain the one we have created and how we can use it to depict arbitrarily large manufacturing processes in a structured way. In Section 3, we depict the manufacturing process of calcium silicate bricks to provide a qualitative evaluation of our ontology. In Section 4, we present how we can use the ontology we created to answer the six questions we defined above. We also explain how we can use our ontology to help us find alternative target variables as well as suitable data sources to model/predict them with ML-models. At the end, we summarize our methodology and results and provide a brief outlook for future research.

2 ONTOLOGY

In this section we give a brief introduction to ontologies. Afterwards, we describe the ontology we created to depict a manufacturing process and how we can apply it to a specific use case. The general idea of using the binary relations between individuals to create an ontology of a manufacturing process stems from the creation of the domain ontology in [6, 7].

2.1 Foundations

The ontology is modelled using OWL [8], which is based on Description Logics [9] formalizing its semantics, and Web technologies, (URIs, RDF/XML serialization) to make knowledge bases shareable. An OWL ontology is a set of axioms that make statements about a domain of interest. Domain elements are represented as OWL entities. Particularly, these are OWL individuals representing specific objects, OWL classes representing sets of individuals, and OWL properties representing directed relationships between individuals (object properties). If an object property is relating two individuals with each other, its inverse object property relates the two individuals in the opposite direction.

With axioms being truth statements about entities, they can describe class relationships, e.g. stating that a class is a subclass of another class allowing for the creation of complex class hierarchies. Other types of axioms are assertional axioms, stating, for instance, that an individual is a member (instance) of a specific class, or property-based axioms stating, for instance, that a property is a subproperty of another property. Another example of property-based axioms is transitivity. To this end, if an individual A is related via a transitive property p with individual B, and B is related via the same p with C, then A is related via p directly with C. A more general concept of this is the subproperty-chain, where for properties p, q, and r, and individuals A, B and C, we can describe that if A is related to B via p, and B is related to C via q, then A is related to C via r. Note that property-chains are not limited to two properties.

2.2 General Concepts

The common elements that are present in the description of any manufacturing process are defined as classes. Meanwhile, the possible relationships between the different elements are defined as object properties. For a particular use case, we depict a manufacturing process using individuals that we relate by assigning them to the given classes and using the given object properties.

Two general ideas are incorporated throughout the ontology. First, we defined an inverse for each object property. This way, we only need to define the relationships of the individuals we create in one direction, while the reasoner automatically completes the other one. Second, we sometimes use subproperties to describe the relationship between individuals more precisely. The idea is that we can define a transitive superproperty allowing us to use the reasoner of the ontology to give us additional information about the relationship between the individuals, while keeping the path that connects two individuals traceable.

2.3 Process Definition

We begin the description of a respective manufacturing process by defining the underlying major process steps and

ordering them by defining the direct predecessors of each step, if it exists. Then, for each process step, we define the respective subprocess steps, if they exist, relate them to their superprocess step, and order them by defining the direct predecessor at the subprocess level. This structure is similar to the one presented in [4].

To gain as much knowledge as possible from the ontology, we define the direct predecessor of a process step using the object property “hasSameLevelDirectPredecessor”. This property has “hasSameLevelPredecessor” as a transitive superproperty. This way the reasoner automatically provides us with all previous process steps on the same hierarchical level, while allowing us to trace the order of the process step following the direct predecessors or successors.

Furthermore, the "hasSameLevelPredecessor" object property has the "hasPredecessor" superproperty. In addition to the transitivity, we additionally defined a subproperty-chain that argues that every predecessor of a process step is also a predecessor of all its subprocess steps. This way, we only have to define the direct predecessors that exist on the same hierarchical level, while getting all other relations by automated inferencing.

2.4 Device Definition

We continue by defining all devices that are used during the manufacturing process. We do this in the order given by the process steps defined in the previous subsection, to guarantee a structured procedure. For each process step, we define the machines or, more generally, the devices used by it, as well as the respective machine parts and sensors that we consider worth mentioning explicitly. The latter are related to their respective machines using the object properties "isPartOf" and "isSensorOf" respectively.

We then define the computing devices used throughout the manufacturing process and relate them to the respective devices they control or receive inputs from. We distinguish between general computing devices such as PLCs and central computing devices such as cloud servers.

2.5 Material Definition

In the third step, we define all the materials used throughout the manufacturing process. Again, we follow the process steps defined in Subsection 2.3. For each step, we observe the machines used during it and relate them to the materials they take, consume or produce using the respective object properties. We further classify the materials we define as raw materials, intermediate products, or final products. If a material cannot be assigned to a certain machine, but only to a process step, we use the corresponding superproperties "isInputOf" and "isOutputOf". We also use subproperty-chains to classify any material that is taken, consumed or produced by a machine used during a process step, as input or output of that process step.

In addition, we relate each of these products to the materials from which it is directly made using the "directlyMadeFrom" property. This property has the transitive superproperty "madeFrom". Thus, the reasoner provides us with all the materials used in previous steps to produce our current product, while still allowing us to trace the order in which the products have been processed.

2.6 Parameter Definition

For each process step, we observe the machines and materials used in it and define the parameters for each one of these individuals. Then, we define the additional parameters that are assigned to the respective process step itself. When we define a new parameter, we classify it as either a static or continuous parameter. We call a parameter static if it does not change during the entire manufacturing process and call it continuous if it does.

When it comes to static parameters related to materials, we manually assign the parameters only to the material where they occur earliest in the manufacturing process. The idea behind this is that when an intermediate product is made from previous products, the static parameters of those pervious products are also static parameters of the intermediate product they produce. This logic is automatically applied by the reasoner of the ontology, meaning that we can observe any static parameter of an intermediate or final product, while only having to assign the respective parameters to the materials where they first occur. This avoids repetition and leads to a better conciseness when assigning the parameters to the respective materials.

We then assign all parameters that are computed, displayed or, an input to a computing device to that device. This allows us to trace where we can potentially collect data about these parameters later in the development process.

2.7 Parameter Relationship Definition

To complete our ontology, we go through each parameter in the order given by the process steps and define their relationship to each other. If a parameter is directly depending on another parameter, we assign them to each other using the "isDirectlyDependingOn" object property. This property again follows the logic of having a more general transitive superproperty. This allows us to get all the other parameters a particular parameter depends on, while still being able to identify direct dependencies and follow the path of dependencies throughout the manufacturing process. The general idea here is similar to [5]. However, we provide a much more detailed depiction of the relationships between parameters.

If more is known about the dependency of parameters, we can use additional subproperties we defined. If a parameter is explicitly calculated from some other parameters, we can express this using the “isCalculatedWith” object property. Additionally, we can define direct monotone proportional or inversely proportional dependencies using the respective object properties. These properties in turn have superproperties that follow the logic of nested proportional or inversely proportional dependencies.

Specifically, this means that the property of proportional dependence is transitive. Additionally, if parameter A is proportionally depending on parameter B, which is inversely

proportionally depending on parameter C, the reasoner automatically assigns parameter A to be inversely proportionally depending on parameter C. The same logic is applied if the dependencies are in reverse order.

After defining all known relationships between the parameters, the manufacturing process is defined, including the respective parameters and the relationships between all individuals. We did this by following a clear structure and defining only direct relationships. This makes it possible for non-experts to build an ontology depicting a manufacturing process. What exactly we intend to use this depiction for and what knowledge we get from the reasoner of the ontology is explained in Section 4.

3 APPLICATION

In this section, we illustrate some of the ideas we implemented in our ontology by using it to depict the manufacturing process of calcium silicate bricks as a proof of concept. A partial overview of this ontology can be seen in Fig. 2. The entire ontology we created can further be found on GitHub (https://github.com/AImotion-Bavaria/MPPO).

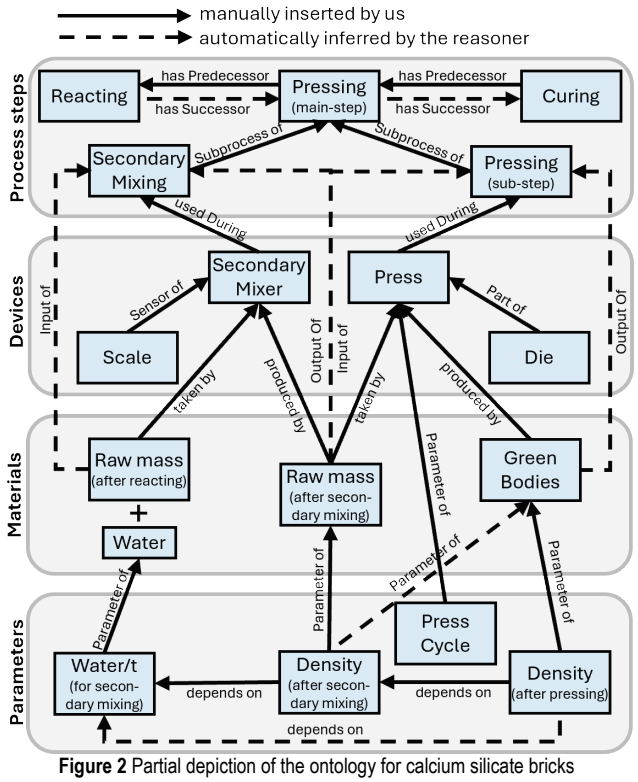


Figure 2 Partial depiction of the ontology for calcium silicate bricks

3.1 Manufacturing Process of Calcium Silicate Bricks

We begin this section with a brief summary of the calcium silicate bricks manufacturing process. It consists of five main process steps. First, the raw materials calcium oxide, sand, and water are mixed with additional aggregates to a raw mixture. This raw mixture is then required to react for approximately 30 to 120 minutes, before being pressed into green bodies of the respective format. Afterwards, the green bodies are cured in an autoclave in a steam atmosphere at a temperature of about 205 °C and pressure of about 14-16 bars. The curing process is by far the most energy-intensive processing step (approx. 80%) and at the same time the bottleneck in the production chain. At the end, the bricks are packaged for further transport. [10]

We start the ontology, by defining the five major process steps just mentioned, ordering them, and adding the respective subprocess steps.

3.2 Machines and Materials

After the process steps are defined, we add the devices and then the materials used in the manufacturing process, following the procedure described in Subsections 2.4 and 2.5.

We take a closer look at the materials we defined to explain the idea behind the "directlyMadeFrom" or "madeFrom" object property. Both the raw mixture and the final calcium silicate brick are made from the four raw materials mentioned in the previous subsection. However, we only manually add the first dependency to our ontology. But because we also meticulously track the transformation of the raw materials and intermediate products that lead to the final product throughout the manufacturing process, the reasoner automatically tells us that the raw mixture and the final calcium silicate bricks are made from the same four raw materials.

3.3 Parameters

As described in Subsection 2.6, we continue by adding the parameters to our ontology, assign them to their respective individuals, and define them as static or continuous, respectively. For static parameters of materials, we only assign them to the materials where they first occur. We elucidate the idea behind this by continuing the example from the previous subsection.

The raw materials have the parameter of their mixing ratio, i.e. how much of them is used to produce one ton of the raw mixture. Following the logic of a manufacturing process, the final product, i.e. the calcium silicate bricks, also have the parameter that describes the mixing ratio used to produce them. In order to avoid repetitive assignment of parameters, we included the idea explained in the second paragraph of Subsection 2.6. Since we included the transformation of raw and intermediate products in our ontology, the reasoner knows how the final product is made. In our case, this means that the reasoner automatically knows that the bricks are made from the initial raw mixture and automatically assigns them the respective static parameters such as their mixing ratio.

3.4 Parameter Relationships

In the final step, we define the relationships between the parameters, following the procedure described in Subsection 2.7. We want to use our example to illustrate the idea of the specific proportional and inversely proportional object properties.

Take, for example, the parameter of total emissions that are emitted during the curing process step. These are

proportional to the amount of fossil fuel used which is proportional to the amount of fresh steam required for the curing the bricks. However, the amount of fresh steam required is inversely proportional to the amount of steam we reuse from previous curing processes [11]. While we only include these direct relations in our ontology, the reasoner automatically gives us any relationship between parameters that can be concluded from the logical rules we stated in Subsection 2.7. In this example, this means that the reasoner automatically tells us that the emissions emitted during the curing of the bricks are inversely proportional to the amount of steam we recycle from previous curing processes.

In general, the reasoner not only provides us with the parameters on which a certain parameter depends, but also with the direction of their relationship according to the logical rules mentioned above. This is especially useful for applications with even more parameters, where it is consequently easy to lose track of the exact dependencies between them.

4 UTILIZING THE ONTOLOGY

In this section, we describe what the ontology we created to depict a manufacturing process can be used for. Our initial goal was to create a methodology that can assist in bridging the gap between business and data understanding to apply ML-models that can help achieve predefined business goals.

4.1 Answering Questions about the Manufacturing Process

In the Introduction we defined six questions that may arise when trying to understand a manufacturing process with the idea of applying a ML-model to it. We present how we can answer these questions for a given manufacturing process using the ontology we created.

1) For each process step, we are provided with all subsequent and previous process steps and their respective hierarchy. The ontology also tells us the directly adjacent process steps, allowing us to follow the manufacturing process step by step. Further, we can see all devises utilized, as well as the materials used and produced during each process step.
2) For each device, we can observe what parts it is made of and which materials it uses and produces.
3) For each material, we can observe in which process step it is used, whether it is an input, output, or both to a particular machine/process step, what materials it is made from in previous process steps, and which subsequent materials are made from it in later process steps. For the latter two, we get additional information about which material it is directly made from or directly makes, allowing us to follow the manufacturing process in this way.
4) The ontology provides us with each parameter for each device, material, or process step and tells us whether the respective parameter is static or continuous. The ontology further allows us to distinguish between parameters that are directly related to materials and those that are inherited from input materials.
5) For each parameter, we can observe which parameters it depends on, and which parameters depend on it. We also get additional information about direct, proportional, and inversely proportional dependencies.
6) For each parameter, we can observe to which computing device it is connected, and whether it is displayed, computed, or an input to the respective device. Vice versa we also get all parameters that are received, displayed, or computed by a respective computing device.

4.2 Data Source and Target Variable Selection

As mentioned, determining the appropriate target variables and the respective relevant data sources is one of the key aspects when applying ML-models. For this reason, we take a brief look at queries and how we can use them to achieve these goals. We start by looking at how the ontology and queries can be used to determine suitable data sources.

Given a target variable, i.e. a parameter, we can simply define a short query that returns all the parameters on which the target variable depends. We can further specify this query. For example, we can state that we only want to see the parameters our target variable depends on that are actually connected to a computing device. As a result, we only get parameters that can potentially be acquired without having to install additional hardware. In general, we can use queries like the one described above to easily get a preselection of potentially relevant data sources for a ML-model. Then, additional methodologies such as [3] or interviews with domain experts can be used to further determine the exact data sources we want as our database.

In addition, we can use queries to help us determine alternative target variables if we are unable to acquire a suitable dataset for the current one. To do this, we define a query that returns all parameters in which our current target variable depends on proportionally or inversely proportionally. These are the potential candidates for alternative target variables, which have to be examined individually for their suitability as alternative target variables. The main criteria to consider is whether the alternative parameter captures all (or at least most) of the aspects of the original one, and whether we can acquire a suitable dataset for this parameter.

For a better understanding, consider the hypothetical business goal of reducing emissions. The majority of emissions are produced during the curing of the bricks [12]. Thus, the parameter "EmissionsAutoclaving" is an obvious choice for the target variable that we want to optimize. However, due to the layout of the machines, it is basically impossible to measure the emissions caused by a single autoclave or curing process [10]. But, using our ontology and a respective query, we can see that the emissions produced during the curing process are inversely proportional to the difference in pressure between the donating and receiving autoclaves, which can be calculated using the constantly recorded current pressure in the autoclaves making it an alternative target variable.

5 CONCLUSION

We developed a methodology for the detailed structural depiction of a manufacturing process using an ontology.

There, the reasoner, together with the logical rules that we created in the form of subproperty-chains, provides us with extensive knowledge about the relationships between the elements present in a manufacturing process. To this end, we only have to manually define binary relationships between the elements of our ontology. We also demonstrated how to use queries in our ontology to assist us determining practical target variables and selecting suitable data sources for them.

However, there are few limitations currently present in our methodology. The formalism of the ontology does not allow to define that if a parameter A is inversely proportionally dependent on a parameter B which is again inversely proportionally dependent on a parameter C, then parameter A is proportionally depending on parameter C. A possible workaround for this issue is to apply the reasoner in two steps and materialize the intermediate conclusions regarding the afore mentioned case of possible parameter relations, hence, breaking the cyclic dependencies. Such an implementation is a potential topic for future work.

Another possibility for future work involves the verification of the proposed methodology. We can test the capabilities of our ontology by using it to describe other manufacturing processes, or verifying the quality of the selected data sources or target variables by conducting empirical evaluations on the performance of ML models trained with the data sources and target variables suggested by our ontology.

Before doing such verification, some future work should be undertaken to extend our current ontology to avoid repetitiveness when depicting additional manufacturing processes. The general idea is to create modular ontologies which can be used to represent certain aspects that are present in multiple manufacturing processes, such as certain types of machines or even entire repetitive process steps. Such a modular decomposition of an ontology facilitates the reuse and sharing of partial knowledge formalized in several ontologies across different domains and use cases. Consequently, authoring and maintenance efforts to create additional ontologies can be drastically reduced.

6 REFERENCES

[1] Roh, Y., Heo, G., & Whang, S. E. (2021). A survey on data collection for machine learning: A big data - AI integration perspective. *IEEE Transactions on Knowledge and Data Engineering, 33*(4), 1328–1347. https://doi.org/10.1109/TKDE.2019.2946162

[2] Studer, R., Benjamins, R., & Fensel, D. (1998). Knowledge engineering: Principles and methods. *Data & Knowledge Engineering, 25*(1-2), 161–198. https://doi.org/10.1016/S0169-023X(97)00056-6

[3] Stanula, P., Ziegenbein, A., & Metternich, J. (2018). Machine learning algorithms in production: A guideline for efficient data source selection. *Procedia CIRP, 78*, 261–266. https://doi.org/10.1016/j.procir.2018.08.177

[4] Meier, S., Klarmann, S., Thielen, N., Pfefferer, C., Kuhn, M., & Franke, J. (2023). A process model for systematically setting up the data basis for data-driven projects in manufacturing. *Journal of Manufacturing Systems, 71*, 1–19. https://doi.org/10.1016/j.jmsy.2023.08.024

[5] Arenas, M., Corcho, O., Simperl, E., Strohmaier, M., d'Aquin, M., Srinivas, K., et al. (2015). *Semantic-guided feature selection for industrial automation systems*. Springer International Publishing.

[6] Zhou, B., Svetashova, Y., Pychynski, T., Baimuratov, I., Soylu, A., & Kharlamov, E. (2020). SemFE: Facilitating ML pipeline development with semantics. In *Proceedings of the 29th ACM International Conference on Information & Knowledge Management* (pp. 3489–3492). Virtual Event, Ireland. https://doi.org/10.1145/3340531.3417436

[7] Svetashova, Y., Zhou, B., Pychynski, T., Schmidt, S., Sure-Vetter, Y., Mikut, R., et al. (2020). Ontology-enhanced machine learning: A Bosch use case of welding quality monitoring. In Pan, J. Z., Tamma, V., d'Amato, C., Janowicz, K., Fu, B., Polleres, A. et al. (Eds.), *The Semantic Web – ISWC 2020* (pp. 531–550). Springer International Publishing. https://doi.org/10.1007/978-3-030-62466-8_33

[8] Motik, B., Patel-Schneider, P. F., & Parsia, B. (2012). *OWL 2 Web Ontology Language structural specification and functional-style syntax (Second Edition)*. W3C Recommendation.

[9] Baader, F., Calvanese, D., McGuinness, L., Nardi, D., & Patel-Schneider, P. F. (Eds.). (2010). *The description logic handbook: Theory, implementation and application* (2nd ed.). Cambridge University Press.

[10] Donhauser, T. (2019). Ressourcenorientierte Auftragsregelung in einer hybriden Produktion mittels betriebsbegleitender Simulation. *Doctoral dissertation*. (in German)

[11] Schrage, T., Schuderer, P., & Franke, J. (2023). Process- and material-specific modeling to use in simulation-based and resource-oriented decision support systems using the example of calcium silicate brick production planning. In *Sustainable Manufacturing as a Driver for Growth – Proceedings of the 19th Global Conference on Sustainable Manufacturing*.

[12] Eden, W., Middendorf, B., Schade, T., Kuhnt, S., & Becker-Emden, E.-C. (2018). *Optimierung des Autoklavierungsprozesses zur Reduzierung der Produktionskosten und Qualitätssteigerung von Kalksandsteinen mittels statistischer Versuchsplanung*. Forschungsvereinigung Kalk-Sand e.V. (in German)

Authors' contacts:

Fabian Weisbrodt
(Corresponding author)
Technische Hochschule Ingolstadt,
Esplanade 10, 85049 Ingolstadt, Germany
+49 841 9348-3516, fabian.weisbrodt@thi.de

Jürgen Bock, Prof. Dr.
Technische Hochschule Ingolstadt,
Esplanade 10, 85049 Ingolstadt, Germany
+49 841 9348-3506, juergen.bock@thi.de

Tobias Schrage
Technische Hochschule Ingolstadt,
Esplanade 10, 85049 Ingolstadt, Germany
+49 841 9348-6623, tobias.schrage@thi.de

Peter Schuderer, Prof. Dr.
Technische Hochschule Ingolstadt,
Esplanade 10, 85049 Ingolstadt, Germany
+49 841 9348-3290, peter.schuderer@thi.de

ISSN 1846-6168 (Print), ISSN 1848-5588 (Online)
https://doi.org/10.31803/tg-20250422203705

Preliminary communication
Received: 2025-04-22, Accepted: 2025-05-05

Achieving Resilience in Military Logistics with Cluster-Forming Technologies

Gerald Schneikart*, Josef Frysak, Karoline Langner, Walter Mayrhofer

Abstract: The combat power of military troops is highly dependent on resilient supply chain management. Recent trends in the application of offensive drone technologies elevate the risks of supply chain failures along the last mile and demand pre-emptive resilience strategies. This paper presents a concept for increased resilience in military logistics that considers requirements elicited with methods from process engineering in collaboration with experts from the Austrian military logistics school. On a basic level, the concept is defined by a) decentralization of repositories at the battalion and company level achieved by cluster-forming technologies, and b) cross-docking for short-time supply points. The presented concept depends on the successful integration of cluster-forming technologies, but if done properly, it has the potential to improve the resilience of modern military logistics.

Keywords: clustering; cross-docking; decentralization; last mile; military logistics; resilience; supply chain management

1 INTRODUCTION

1.1 Increasing Resilience of Logistics Systems

Even though logistics originated in the military during the second half of the 20th century, research in civilian logistics outpaced military logistics [1]. Both sectors aim at improving efficiency and effectiveness when integrating novel technologies or strategies, but the overarching objectives may be different. In contrast to civilian industries, which strive for high efficiency with respect to financial terms [1], military operations need to remain effective in supporting combat powers of their troops at all costs [2, 3]. The stability against unexpected disruptions threatening either the efficiency or the effectiveness of a supply network and the ability to "restore normal supply network operations" can be dubbed "resilience" [4], while there might not be a commonly accepted definition of supply chain resilience [5].

Novel strategies and technologies can be implemented to improve resilience. One civilian example that illustrates this issue, is the introduction of counterfeit drugs into circulation somewhere along the pharmaceutical logistics chain. Recent developments in blockchain and robotics technology could therefore establish more transparent, automatized and resilient pharma supply chain processes that are much harder to corrupt [6]. On a more global scale, other threats to logistics and supply chain security are associated with a rising probability of natural disasters resulting in supply outages [7, 8]. As a possible countermeasure, artificial intelligence may attenuate the negative effects by providing backup plans and fast solutions for alternative routings [9].

Although certain novel technologies like blockchain, Internet-of-Things (IoT) or Artificial Intelligence may be appropriate to improving resilience in civilian logistics, their application in the military context often needs to consider a different set of requirements to show comparable positive effects. Effective military supply chain management requires a high level of flexibility in the establishment of supply routes, especially on the last mile [10]. If enemy attacks targeting single entities, such as transport vehicles or supply points, are successful, reactive mitigation measures can be implemented to reduce the risk of supply disruptions [11], e.g., using alternative means of transportation or by establishing other supply points at different locations.

1.2 Optimization of Resilience by Decentralization

Continuous improvements to modern reconnaissance and weapons systems for precision strikes against long-distance targets, such as the application of drone technologies, increase supply chain vulnerability [12, 13]. As countermeasure, pre-emptive strategies for resilience optimization are needed at any supply level [11], even at locations far behind the frontline. In this respect, decentralization of repositories seems promising [10].

Civilian logistics are often highly centralized, and efficiency is achieved by minimizing resource utilization, e.g., a single supplier, process automation, or relying on a single authority for decision-making. However, such highly efficient supply chains have low resilience against any kind of disturbance due to sole dependency on single resources and little redundancy. For example, the COVID-19 crisis and subsequent supply chain disruptions were associated with highly dependent and vulnerable supply chains. This suggests that companies with decentralized logistics structures and processes are more capable to adapt to unforeseen changes or impacts [14, 15].

While it may seem obvious, what decentralization in logistics means in a civilian context, decentralization of military logistics remains, to our knowledge, an open research gap and raises the following questions:

- Q1: How can decentralization of the military supply chain, from the battalion level to the frontline, be achieved?
- Q2: What are the requirements for a transformation towards logistical decentralization of the last mile?
- Q3: Which risks and opportunities may arise from such a supply chain decentralization?
- Q4: How can the resilience of decentralized military logistics be evaluated?

1.3 The Research Project RESISTANT

The demand for decentralizing technology solutions and pre-emptive transformation strategies is the basis of the research project RESISTANT. The importance of keeping armed forces ready for potential future threats due to modern weapon systems, e.g., autonomous or unmanned aerial and ground vehicles, is increasingly showcased in foreign military conflicts. Therefore, the Austrian government partly funded RESISTANT in order to develop a concept for supply chain decentralization, that also meets the demands of western military logistics (NATO). RESISTANT is a consortium project composed of collaborating research institutions of universities, universities of applied sciences, start-up companies in the industry 4.0 sector, and the Austrian Federal Ministry of Defence [16-19].

The project addresses aforementioned research questions by applying a combination of qualitative research and design science approaches to develop a concept for decentralization and first versions of the enabling communication and transportation technology solutions. The project scope primarily comprises small expendable supply items, including commodities needed by squads to maintain combat power, e.g., ammunition for rifles and guns, rations, or medical supplies. Goods needed by specific branches, e.g., fuel, spare parts, or electronics, are out of project scope.

This paper presents the concept of the envisioned logistical decentralization, and its enabling technology currently developed in collaboration with logistics experts from the Austrian armed forces as part of RESISTANT.

1.4 Organization of the Rest of the Article

This article is organized as follows: Section 2 describes the methods applied for elicitation of the requirements of the solution and the development of the decentralization concept. Section 3 describes the envisioned concept, the underlying technology paradigm and is complemented with a short risk-benefit analysis. Section 4 discusses the advantages and limitations of the decentralization concept.

2 METHODS

2.1 The Overall Design-Science Research Approach

The RESISTANT project essentially follows a design-science research approach. Design-science research is a framework in Information Systems research, which aims at acquiring new knowledge by combining behavioural and system design and development approaches. More specifically, the Information System research paradigm proposes to derive new knowledge by developing artifacts (products or services) that serve specific business needs. These artifacts originate from the practical environment but are developed based on theoretical foundations and methodologies. A key aspect of the design-science approach is the constant, iterative assessment and refinement of the artifact based on field applications. Thereby, theoretical as well as practical rigor drive the developmental process while simultaneously generating novel knowledge [20].

2.2 Semi-Structured Interviews and Process Analyses

In order to gain a basic understanding of the practical socio-technical environment, e.g., logistics processes of the Austrian military, methods from the process management discipline were applied. Process survey techniques in the form of semi-structured interviews were used to collect qualitative data about respective environments. Based on that data, process analyses were conducted using the Business Process Model and Notation (BPMN) standard, version 2.0.2 [21].

The BPMN models enabled basic requirements elicitation and process optimization under consideration of various technology implementations and were discussed with the experts of Austrian military logistics. This paper presents the schematics of a future state process model and summarizes the basic knowledge gains.

3 RESULTS

3.1 Decentralization Achieved by Dynamic Clustering

The idea envisioned by RESISTANT is to increase the resilience of military logistics by the fragmentation of repositories at different organizational levels and keeping each fragment mobile in order to reduce reconnaissance from the enemies. The inherent challenges with such an approach are apparently associated with dynamically managing goods distributed to each repository fragment and controlling the entire repository's inventory that is spread across a pre-defined geographic terrain. Therefore, cluster-formation enabled by communication technology is needed for the synchronization of inventory information between individual repository fragments. The resulting "network" of mobile repository fragments represents a dynamic cluster of data points.

The applied principle of dynamic clustering is illustrated by the example shown in Fig. 1, which shows a two-dimensional terrain with three clusters spread at distinct locations at time (t) = 0. The third dimension that would represent differences in altitudes was avoided for simplicity. Each cluster is defined by seven cluster-forming units (CFUs), which refer to single repository fragments represented by, e.g., a single container or a mobile storage rack distributed in the field within a kilometre.

Cluster formation depends on visual or radio contact as well as pre-defined thresholds for distances between individual CFUs. A CFU belongs to a cluster, as long as connection to at least one other CFU is established. Considering that the CFUs are mobile (dynamic clustering), a CFU could thus leave its cluster by exceeding the distance threshold to proximal CFUs (t = 1 and t = 2) or losing contact.

The single condition for cluster formation is visual or radio contact and the threshold for distances between CFUs. Accordingly, two CFUs are sufficient to form a cluster, as long as they are separated by a distance smaller than the threshold (t = 3).

Since individual CFUs are mobile, even entire clusters could theoretically move within the terrain. If clusters or CFUs of distinct clusters approach each other, both clusters

integrate, once at least one CFU in each cluster moves within the threshold and connects to at least one CFU of the other cluster ($t = 4$ and $t = 5$, respectively).

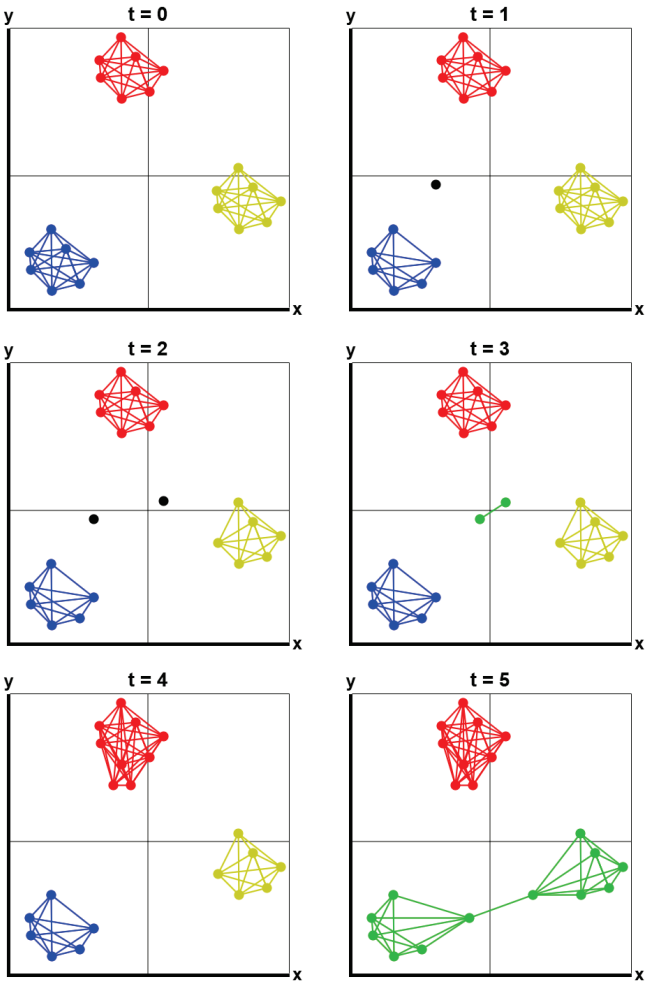


Figure 1 The principle of dynamic clustering

3.2 Cross-Docking for Commodity Transactions between Clusters

Since military supply chains have repositories of goods at each hierarchical level, dynamic clustering could theoretically be applied at each level. The presented concept for decentralization focuses on all downstream hierarchies, starting from the battalion level and ending at the foxhole, because military supply is increasingly vulnerable towards the frontline [10]. According to the analysis of the current supply processes (unpublished), two levels for cluster formation were considered reasonable, e.g., at the battalion and the company or platoon level (Fig. 2).

In the resulting concept, the transfer of commodities between those cluster levels requires solutions for cross-docking of transported goods between CFUs of different cluster levels. In the current supply process, different technologies for storage and transportation are used at each military hierarchical level. For example, a warehouse is established at the battalion level and trucks are generally used for transportation to distribution points near the frontline. Final delivery to the foxhole happens by manual carriage.

Considering the current logistical situation, the envisioned concept works with trucks capable of picking up roller containers (also called a hooklift container) intended for transport of multiple days of supply (DOS) of commodities (see Section 1.3 for specifications). The containers are equipped with communication technologies for cluster-formation. When loaded onto a truck, each container represents a mobile CFU capble to form clusters with nearby containers at the battalion level.

For cluster formation at downstream levels, i.e., clusters affiliated to companies or platoons, trucks and heavy containers are considered less suitable for dynamic clustering due to their insufficient degree of possible decentralization. Reasons for this rationale are, first and foremost, the vicinity of the frontline, which signifies high risk of detection by the enemy and loss of several DOS in a single strike. Hence, larger numbers of smaller vehicles are needed, ideally autonomous unmanned ground vehicles (UGVs) capable of carrying supplies close to the frontline [10].

When DOS are ordered for delivery, a truck from the battalion cluster and a CFU from the company or platoon cluster need to meet at a temporary supply point for the downstream transfer of DOS. Whatever the means of transport at the company or platoon level are, the technologies of the meeting CFUs require a technical solution for semi- or fully automated cross-docking. The major requirement is to keep the duration of the meeting to a minimum, because the increased resilience of the decentralized dynamic clustering should not be compromised by elongated times needed for setting up the temporary supply point and the exchange of goods.

Once the truck from the battalion delivered the ordered DOS to a downstream cluster, it either can move to a new temporary supply point to meet another downstream CFU for further DOS delivery, return to the battalion cluster to wait for a new order, or go further upstream to a repository at the brigade level to pick up a new container with fresh DOS.

After the CFU from the company or platoon acquires the ordered DOS from the battalion truck, it returns to its cluster waiting for instructions to finally deliver the goods to the combat troops. If squads from the frontline are ready for resupply (e.g., during a combat break) and order a pick-up of DOS, the CFU meets a squad from the ordering troops at a temporary supply point near the frontline, where the soldiers acquire the DOS and eventually distribute it to their comrades in the foxholes. The empty CFU then returns to its cluster at the company or platoon level until it receives new instructions for resupply with DOS delivered by a battalion truck.

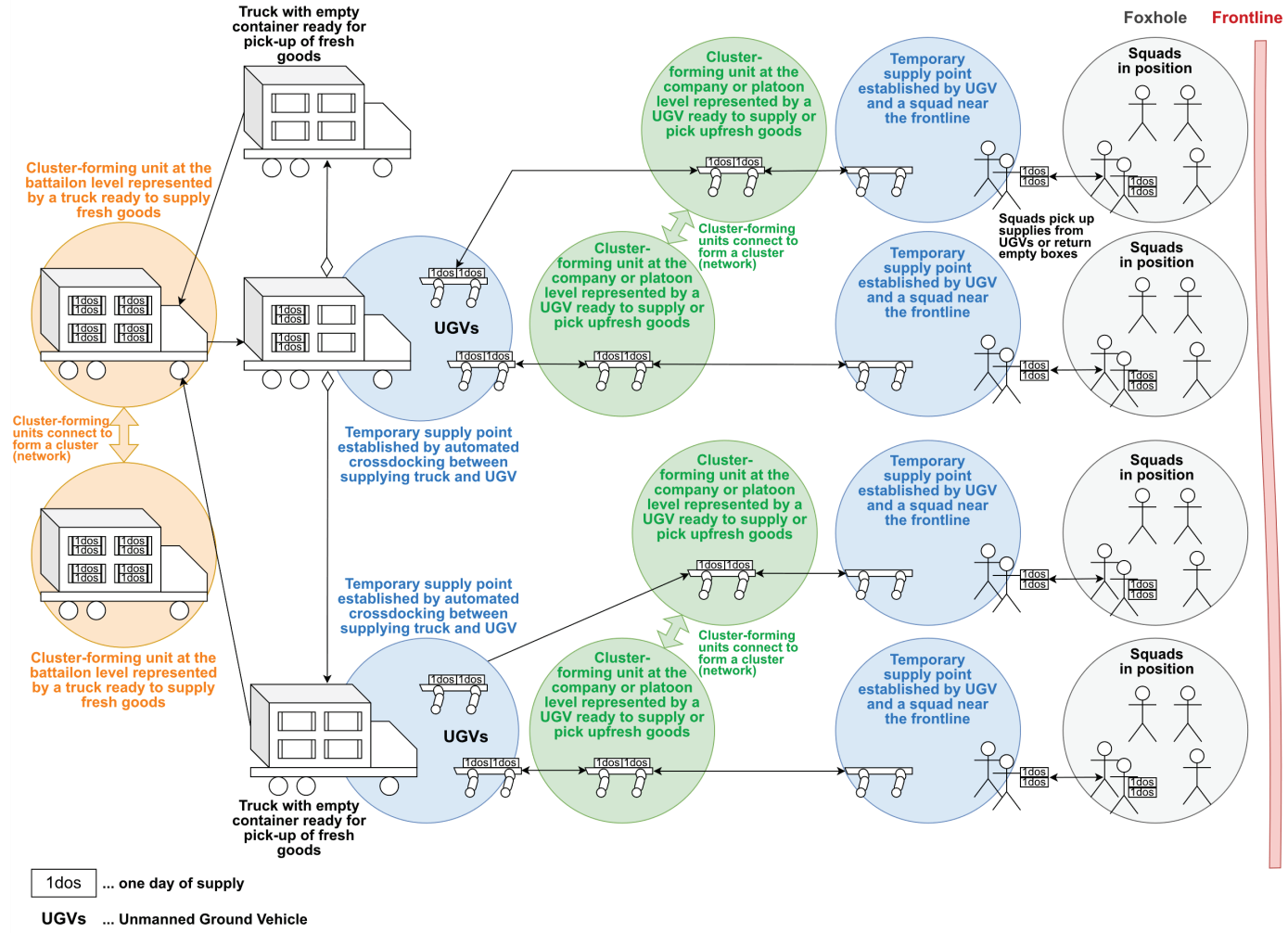


Figure 2 Decentralization of military repositories at different supply levels enabled by mobile cluster-forming solutions. The infographic illustrates the path of goods (amounts measured in one Day-Of-Supply; *dos*) distributed from cluster-forming container systems at the battalion level (orange circles) via cluster-forming unmanned ground vehicles (UGVs) at the company or platoon level (green circles) to the consuming troops (gray circles) at the frontline (red line). Cross-docking between clusters happens at temporarily established meeting points (blue circles). The black double-arrows illustrate possible paths of cluster-forming units or squads transporting one or multiple *dos*. Orange or green bold double-arrows between circles represent cluster formation of at least two connected cluster-forming container units or UGVs.

3.3 A Paradigm Technology Solution for DOS Transportation and Dynamic Clustering

The decentralization concept presented above was used for technical requirement elicitation and refinement in the RESISTANT project. While the technical development activities are still ongoing at the time of writing this paper, in the following the status of the transport item technology that enables dynamic clustering is presented. The core technology is a smart box/rack-system developed by the Austrian start-up company BOOXit. In the context of previous applied research projects, their returnable transport item solution has been improved [22] and evaluated for specific use cases in the pharmaceutical industry [6, 23].

The proprietary BOOXit technology is a promising transport solution providing a range of innovative features. The top of the box is covered with nobs reminiscent of Lego® bricks (Fig. 3A). The knobs allow for simple load securing without additional means of fixation (e.g., straps, shrink wraps, nets, tie downs, or fasteners), when multiple boxes are stacked, or storage of boxes in proprietary racks, which enable the usage of the boxes as drawers (Fig. 3B). The racks themselves also contain the knob feature for pooled transportation and fixation on compatible pallets (Fig. 3C, 3D).

The underlying vision of the BOOXit technology is to provide a smart transportation and logistics ecosystem relying on a single transport item solution, which consists of fully automated processes along entire supply chains. Moreover, the BOOXit technology allows for the integration with various logistics 4.0 technologies, e.g., communication systems for real-time monitoring of current positions or temperatures, as well as robotic systems for full automation of warehousing and a pick-by-light/put-by-light system for semi-automated assistance of delivery personnel (Fig. 3B) [22]. BOOXit also is the enabling transport item solution for the decentralization concept described above, and the box/rack-system is compatible with previously mentioned roller containers (Fig. 3E).

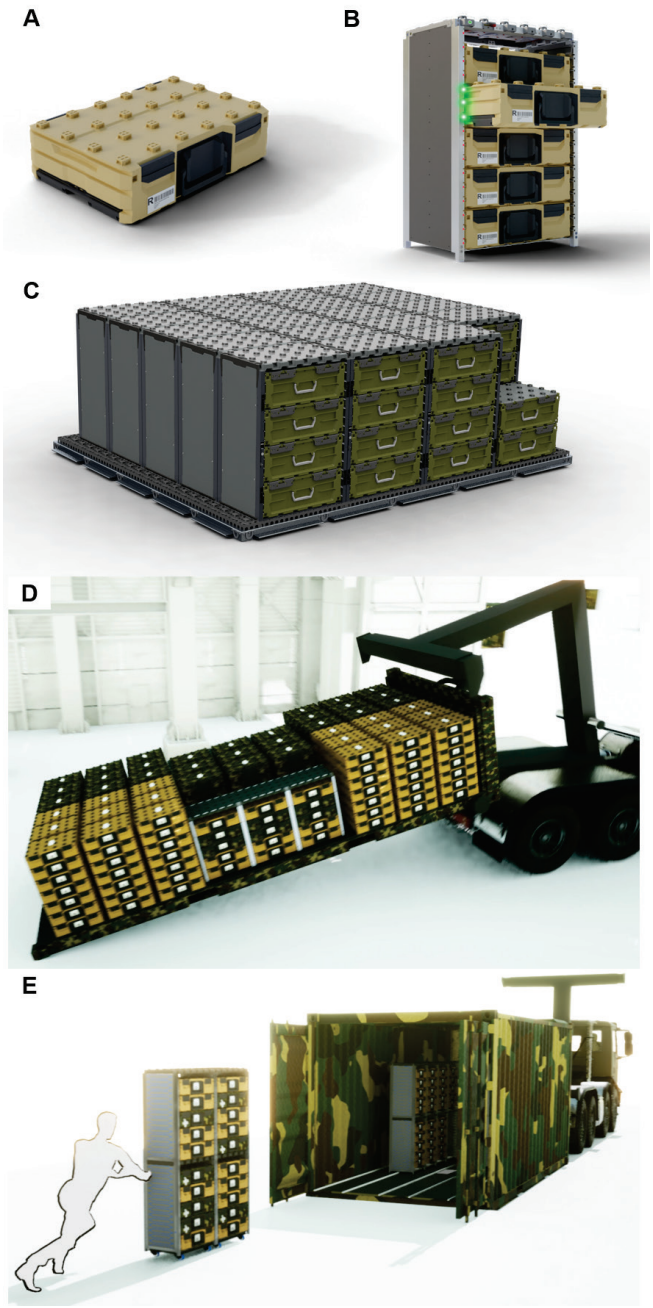


Figure 3 A 3D rendering of the BOOXit box/rack-system. (A) A BOOXit box with the proprietary knob structure visible. (B) The boxes can be used as a drawer in BOOXit racks. An integrated pick-by-light system assists warehouse personnel with box handling. (C, D) The knob structure on BOOXit racks allows pooling of multiple rack units to be loaded onto various transport systems without additional tie downs, fasteners or load securing. (E) One-shot-load storage of BOOXit box-rack-units into a roller container (also called a hooklift container)

Furthermore, the communication technology required for dynamic cluster formation is also developed in the RESISTANT project (Fig. 4A). The envisioned roller containers are equiped with automatic inventory control system in its interior (Fig. 4B), which provides the required cross-docking features for semi- or fully automated delivery of goods to company or platoon clusters (out of scope of the RESISTANT project).

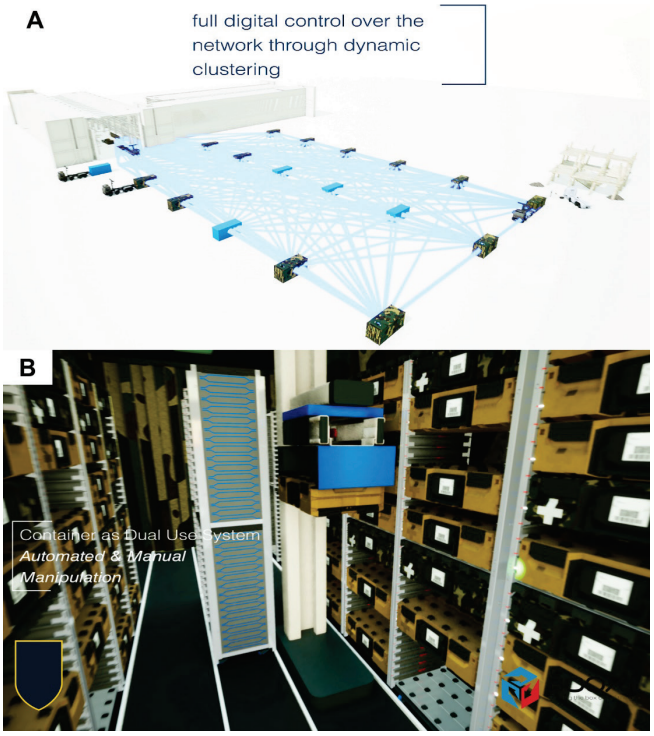


Figure 4 A 3D rendering of the BOOXit container system. (A) The principle of roller containers forming clusters. Each container directly connects with all nearby containers to form a cluster of equivalent network points. Connection between units happens automatically and depends on visibility in the field. (B) The envisioned roller containers are equipped with a fully automated inventory control system for cross-docking purposes.

3.4 Requirements in View of Further Development

The major challenge in implementing the presented or a comparable concept for the decentralization of logistics is the integration of technical solutions that enable dynamic clustering. The reason for this is that the repository fragments (CFUs) are dislocated and spread across a smaller or larger geographical terrain, but there still is the need to keep a "virtual common inventory", just as there would be in an integrated, central warehouse system. Therefore, the CFUs of a cluster permanently need to synchronize information about their individual inventory. The major requirement is a stable connection of each CFU with at least another CFU of the same cluster. In case a single CFU loses its connection, it saves the information about its current inventory status until automatic reconnection with a cluster for synchronization with the entire cluster's inventory takes place.

Another essential technical requirement enabling dynamic clustering is an independent power supply of each CFU. Different from warehouse entities, which can directly be connected to a supply network, mobile CFUs distributed in the field need to carry their own batteries for prolonged energy supply. This requires the integration of recharging technologies, such as energy-self-sufficient means (photovoltaics or other forms of energy harvesting) or power generators, e.g., when carried by a truck. Alternatively, recharging stations need to be established within clusters or at nearby locations.

3.5 Risks Associated with the Presented Decentralization Concept

Dynamic clustering as illustrated in this paper certainly contains considerable risks that could turn into threats substantially compromising the resilience level of the supply chain. The most obvious risks are summarized in the following.

For example, the dependency on independent energy sources, e.g., integrated batteries, represent risks associated with the unavailability of CFUs during the process of recharging, i.e., the working number of CFUs is temporarily reduced. In addition, charging stations could become a target of enemy sabotage. In order to minimize the risk of no charging opportunities within the cluster, multiple charging stations would be required, which increases the necessary effort to implement the decentralization concept by dynamic clustering (e.g., due to geographical constraints of the charging infrastructure).

The problem with energy supply is even aggravated by the fact that the cluster concept needs to be "dynamic", which results in high-energy demands, compared to a static network, which would demand energy only for communication. Constantly moving CFUs consume electric energy (transportation by electric UGVs) or gasoline (e.g., transportation of the container system by truck). In further consequence, operating costs are most likely higher compared to the current situation of central warehousing and supply chain management. In order to counteract this problem, energy-self-sufficient solutions, e.g., solar panels or wind power can be considered.

The functionality of the concept is highly dependent on transport units constantly circulating back and forth. If the backhaul of empty boxes from the frontline to the brigade does not occur, the supply chain ultimately breaks down due to the unavailability of transport crates. In order to keep the supply chain running, the circulation of boxes needs to be tightly controlled. The automatic communication system of the clustering system itself could provide the necessary information of individual crates' position, movement and availability.

3.6 Opportunities Presented by the Decentralization Concept

Once implemented, dynamic clustering offers several opportunities, each of which contributes to improved resilience. Apart from keeping a stable supply chain, despite outage or destruction of a single CFU, the communication infrastructure that keeps the cluster functional theoretically enables real-time demand monitoring. Since each CFU synchronizes its inventory information with the residual cluster, the entire cluster is capable to communicate specific demands from the frontline back to the brigade. The prerequisite is permanently running a continuous and stable communication line from the frontline via each cluster level back to the brigade warehouse system. However, real-time detection of consumption currently might not be possible for most of the consumables, i.e., rifle ammunition. Consumption of consumables at the frontline can only be indirectly measured by the number of DOS delivered or by the number of empty crates returned. Eventually, the near real-time monitoring also would improve accuracy in demand forecasting.

3.7 Evaluation of Resilience

The available literature about evaluating resilience of supply chains, which is reviewed in detail in other publications [24], indicates industrial and research interest in having benchmarking solutions available. However, comparability of resilience performance is difficult, because of differences in supply chains' dependencies on reference factors, which determine resilience. As implied in the introduction, resilience can basically be defined by reactive or pre-emptive (also proactive) factors and metrics defining the resilience strategy [11, 25], although more dimensions were also suggested [24].

The probably most suitable evaluation method that could currently be adapted for the evaluation of the presented decentralization concept was published by Mari et al. [26]. They defined four metrics to determine supply chain resilience: 'accessibility', 'robustness', 'responsiveness' and 'flexibility' based on the holistic view of supply chains as complex networks of interconnected units. This approach intends to capture all players and factors involved in a supply chain network. As for the evaluation of the presented decentralization concept based on dynamic clustering, the evaluation approach (or other published means) would require a confined analysis of supply with commodities delivered from the battalion level to the foxhole. The analyses of any upstream factors of the supply chain process would thus be redundant.

Alternatively, the implementation of the decentralization concept itself could be seen as the introduction of an additional resilience metric (next to the other four) that captures the level of decentralization of the supply chain. Regardless of the approach applied, it must be considered that the development of the metrics by Mari et al. were derived from civilian logistics [26], which might change in the context of military logistics due to different overall objectives, i.e., cost efficiency vs. fail-safety.

4 DISCUSSION

The described concept for the decentralization of military logistics by fragmentation of repositories appears promising for increasing the resilience of supply chains between battalions and troops at the frontline. Probabilities of sabotage of entire repositories or delivery routes is reduced, because enemies would need to uncover the positions of all fragments and delivery points for successful supply chain disruption. Furthermore, the likelihood of enemy detection is even reduced as all repository fragments are mobilized, and delivery points are established only temporarily within short timeframes due to automated cross-docking and constant repositioning.

The inherent advantage of dynamic clustering is an increased degree of automation when managing the flow of

goods from the battalion to the frontline. This would also allow for more accurate demand forecasts based on the consumption rates continuously reported back from the frontline. However, the problem of monitoring consumption rates at the frontline still leaves room for discussion, since automatic detection of consumption rates (of rations or ammunition) is currently only possible by indirect means, e.g., by extrapolation from backhaul of empty transport items.

Note that the backhaul of empties can be accomplished by applying the same mechanisms used for forward transportation. Squads could exchange empty boxes at the temporary delivery points once they meet for pickup of fresh supplies (Fig. 2). The UGVs would then carry the empties within their cluster until cross-docking with the truck from the battalion at the upstream delivery point (in return, the UGV picks up a full box). Eventually, the truck returns the empties upstream to the brigade for refilling. The circle of the transport items would thus be closed, which provides the potential to reduce costs and environmental burdens associated with packaging waste, thus contributing to efforts in executing strategies for "Green Defence" [27, 28]. Moreover, this backhaul circle might even be used to remove garbage or dirty clothing from the frontline for recycling or cleaning.

The derived decentralization concept obviously has its limitations. The cluster and transportation technologies in their current developmental stages still need to clear the proof-of-concept readiness level and be piloted in selected processes. Therefore, testing the concept by employing real military resources is currently not possible. Nevertheless, it serves as a platform for iterative requirements elicitation, because it was derived based on expert information from the Austrian military logistics school and involves the latest science-based knowledge in logistics solutions.

5 CONCLUSION

The presented research underscores the importance of integrating modern logistics technologies and innovative strategies to ensure the continuous or even improved combat readiness level of military forces. As the RESISTANT project progresses, it will contribute valuable insights and solutions for resilient supply chain management with implications beyond military applications.

Acknowledgements

The RESISTANT project (German title: Resilientere Militärlogistik auf Basis einer durchgehend digitalisierten, dislozierten und autonomen Supply Chain; English title: Increased Resilient Military Logistics based on a Fully Digitalized, Distributed and Autonomous Supply Chain) is a collaborative project with the FFG contract number FO999904051. RESISTANT is funded by the Austrian Federal Ministry of Finance within the Austrian defense research program FORTE of the Austrian Research Promotion Agency (FFG).

The authors would like to thank the interviewees from the Austrian military logistics school for their commitment and for dedicating time to support the research conducted by the RESISTANT team. Further, the authors thank Andreas Mag. Holzleithner and Dr. Peter Entenfellner from BOOXit for providing the photo material.

Special thanks go to Ann-Ilaria Mayrhofer for her help with the English language and diligently proofreading the article.

6 REFERENCES

[1] Rutner, S. M., Aviles, M., & Cox, S. (2012). Logistics evolution: A comparison of military and commercial logistics thought. *The International Journal of Logistics Management, 23*(1), 96–118. https://doi.org/10.1108/09574091211226948

[2] Serrano, A., Kalenatic, D., López, C., & Montoya-Torres, J. R. (2023). Evolution of military logistics. *Logistics, 7*(2), 22. https://doi.org/10.3390/logistics7020022

[3] Piggee, A. F. (2002). *Transformation – Revolution in military logistics*.

[4] Rice, J. B., Jr., & Caniato, F. (2003). Building a secure and resilient supply network. *Supply Chain Management Review, 7*(5), 22–30.

[5] Valtonen, I., Rautio, S., & Lehtonen, J.-M. (2023). Designing resilient military logistics with additive manufacturing. *Continuity & Resilience Review, 5*(1), 1–16. https://doi.org/10.1108/CRR-08-2022-0015

[6] Schneikart, G., Mayrhofer, W., Löffler, C., & Frysak, J. (2024). A roadmap towards circular economies in pharma logistics based on returnable transport items enhanced with Industry 4.0 technologies. *Resources, Conservation and Recycling, 206*, 107615. https://doi.org/10.1016/j.resconrec.2024.107615

[7] Lima, M. (2024). Strengthening healthcare supply chains: A comprehensive strategy for resilience in the face of natural disasters. *Health Emergency Management, 5*(2), 47–67. https://doi.org/10.61093/hem.2024.2-04

[8] Kungwalsong, K., & Ravindran, A. R. (2014). Assessment of disruption risks in supply chains. In J. Wang (Ed.), *Encyclopedia of Business Analytics and Optimization* (pp. 209–219). IGI Global.

[9] Zhai, A., Guo, D., Gollapudi, S., Kollias, K., & Delling, D. (2024). Deep learning-based alternative route computation. In S. Dasgupta, S. Mandt, & Y. Li (Eds.), *Proceedings of the 27th International Conference on Artificial Intelligence and Statistics* (pp. 4078–4086). Proceedings of Machine Learning Research: PMLR.

[10] Alexa, A. (2021). *Militärlogistik 4.0 – Trends und deren Auswirkungen auf die Militärlogistik*. Landesverteidigungs-akademie. (in German)

[11] Sani, S., Schaefer, D., & Milisavljevic-Syed, J. (2022). Strategies for achieving pre-emptive resilience in military supply chains. *Procedia CIRP, 107*, 1526–1532. https://doi.org/10.1016/j.procir.2022.05.186

[12] Oleksenko, O., Misiuk, H., Ikaïev, D., Korshok, V., & Palka, V. (2024). Main trends in the use of unmanned aircraft in the Russian-Ukrainian war. *DNDIVSOVT, 20*(3), 99–105. https://doi.org/10.37701/dndivsovt.21.2024.12

[13] Boliubash, O., Ponomarov, A., Shulezhko, V., & Shulezhko, A. (2024). The procedure for the use of unmanned aerial vehicles to conduct reconnaissance on the ground in wartime conditions. *DNDIVSOVT, 19*(1), 7–13. https://doi.org/10.37701/dndivsovt.19.2024.01

[14] Supply chain resilience through decentralization. (2023). *Strategic Direction, 39*(6), 29–31. https://doi.org/10.1108/SD-04-2023-0050

[15] Adana, S., Manuj, I., Herburger, M., Cevikparmak, S., Celik, H., & Uvet, H. (2024). Linking decentralization in decision-making to resilience outcomes: A supply chain orientation perspective. *The International Journal of Logistics Management, 35*(1), 256–280. https://doi.org/10.1108/IJLM-07-2022-0308

[16] FHWien der WKW. (2025, January 10). *RESISTANT – Resilientere Militärlogistik auf Basis einer durchgehend digitalisierten, dislozierten und autonomen Supply Chain*. https://www.fh-wien.ac.at/forschung/forschung-an-der-fhwien/resistant/ (in German)

[17] Langner, K., Frysak, J., Schneikart, G., Riegler, W., & Herburger, M. (2024). Challenges for military supply chain resilience: A digitalized, decentralized, and transparent approach. In *NOFOMA 2024* (pp. 1–18). Stockholm, Sweden.

[18] Krenn, F., Lampl, S., & Stary, C. (2025). Towards resilient digital supply chains. In M. Elstermann & M. Lederer (Eds.), *Subject-Oriented Business Process Management: Models for Designing Digital Transformations* (pp. 230–236). Springer Nature Switzerland.

[19] Johannes Kepler Universität Linz. (2025, January 10). *RESISTANT – Resilientere Militärlogistik auf Basis einer durchgehend digitalisierten, dislozierten und autonomen Supply Chain*. https://www.jku.at/institut-fuer-wirtschaftsinformatik-communications-engineering/forschung/forschungsprojekte/laufende-projekte/resistant/ (in German)

[20] Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems research. *MIS Quarterly, 28*(1), 75–105. https://doi.org/10.2307/25148625

[21] Object Management Group. (2025, March 31). *Business Process Model and Notation (BPMN), Version 2.0.2 (Specification No. V2.0.2)*. https://www.omg.org/spec/BPMN

[22] Schneikart, G., Mayrhofer, W., Frysak, J., & Löffler, C. (2023). A returnable transport item to integrate Logistics 4.0 and circular economy in pharma supply chains. *Tehnički glasnik, 17*(3), 375–382. https://doi.org/10.31803/tg-20230504144856

[23] Schneikart, G., Löffler, C., Brandner, M., Pfoser, S., & Mayrhofer, W. (2024). Requirements for the transformation towards returnable transport item-enabled circular economies in the Austrian parcel industry. *Tehnički glasnik, 18*(2), 254–261. https://doi.org/10.31803/tg-20240420173731

[24] Pangarkar, G. (2023). *Measuring supply chain resilience*. FH Vorarlberg.

[25] Agarwal, N., Seth, N., & Agarwal, A. (2022). Evaluation of supply chain resilience index: A graph theory based approach. *Benchmarking: An International Journal, 29*(3), 735–766. https://doi.org/10.1108/BIJ-09-2020-0507

[26] Mari, S. I., Lee, Y. H., & Memon, M. S. (2015). Complex network theory-based approach for designing resilient supply chain networks. *International Journal of Logistics Systems and Management, 21*(3), 365–384. https://doi.org/10.1504/IJLSM.2015.069733

[27] Mihailov, M.-E., & Ichim, M. (2024). Enhancing Romanian naval resilience to climate change: Advanced functional materials for green defence strategies. *Journal of Environmental and Sustainability Studies, 7*(2), 17. https://doi.org/10.20849/jess.v7i2.1466

[28] Larsen, K. K. (2025). *The NATO Green Defence Framework*.

Authors' contacts:

Gerald Schneikart, MBA, MSc, PhD
(Corresponding author)
FHWien der WKW, Institute for Digital Transformation and Strategy,
Währinger Gürtel 97, 1180 Vienna, Austria
+43 1 476 77 5761, Gerald.Schneikart@fh-wien.ac.at

Josef Frysak, Dr.
FHWien der WKW, Department of Digital Economy,
Währinger Gürtel 97, 1180 Vienna, Austria

Karoline Langner, MSc
University of Applied Sciences Upper Austria, Logistikum Steyr,
Wehrgrabengasse 1-3, 4400 Steyr, Austria

Walter Mayrhofer, Prof. Dr.
FHWien der WKW, Institute for Digital Transformation and Strategy,
Währinger Gürtel 97, 1180 Vienna, Austria

ISSN 1846-6168 (Print), ISSN 1848-5588 (Online)
https://doi.org/10.31803/tg-20251017170752

Preliminary communication
Received: 2025-10-17, Accepted: 2025-11-26

Automated System for Data Transmission Characteristics Management in the Industrial Internet of Things Networks

Muhanad Jabar Yaser*, Konstantin A. Polshchykov, Mikhail S. Balakshin

Abstract: The purpose of the proposed study is to provide the required data transmission characteristics in the Industrial Internet of Things (IIoT) system. Our efforts are primarily aimed at increasing the probability of message delivery, as well as reducing the likelihood of their duplication in wireless networks based on the Message Queuing Telemetry Transport (MQTT) protocol. The article offers a scheme and description of the automated control system, which is based on the use of original models to assess the probability of delivery and the likelihood of duplicate messages. The control parameters are the quality of service (QoS) level and the number of allowed retransmissions (Nretry). Conducting computational experiments and deploying the system in a real environment have shown that its application makes it possible to improve the requirements for the probability of message delivery by 7.2 percent and the probability of their duplication by 8.6 percent compared to existing adaptive QoS approaches. Due to the results achieved, the proposed system is recommended for use in IIoT networks, which place increased demands on the reliability of data delivery.

Keywords: Automated Control System; Industrial Internet of Thing (IIoT); MQTT Protocol; Network Optimization; Probabilistic Modelling; Quality of Service (QoS); Telemetry Message Delivery; Retransmissions

1 INTRODUCTION

Wireless data transmission systems currently serve as a technical basis for monitoring and control for solving various applied problems[1]. In particular, energy-efficient wireless sensor networks are in demand, on the basis of which the Internet of Things systems operate [2-4]. Such systems provide the collection of data received in the form of telemetry messages from numerous sensors. Internet of Things systems are used in agriculture [5-7], energy [8, 9], medicine [10-12], and many other areas.

Industrial Internet of Things (IIoT) systems are being actively implemented at modern manufacturing enterprises, allowing real-time monitoring of the technical condition and correct functioning of the equipment used. IIoT systems include numerous sensor nodes connected to certain elements of production equipment (mechanisms, machines, process lines). Such sensor nodes contain sensors for measuring parameter values, with the help of which the serviceability and operability of the operating industrial equipment is monitored. Sensor nodes are also equipped with transceiver modules used to receive service messages to configure their operation and send measured telemetry data to the network server of the production process control centre (hereinafter referred to as the control center). To transmit and receive messages, sensor nodes as end devices are connected to the IIoT network, usually using wireless channels.

The use of the IIoT system should be aimed at ensuring preventive management of production technological processes. These processes should be adjusted before failures or emergency situations occur, ensuring the exclusion of negative consequences (harm to workers' health due to mechanical damage, electric shocks or equipment fires, as well as material damage from unplanned downtime of production lines, forced purchase of new equipment or expensive repair of failed equipment). In the process of monitoring the technical condition of functioning production equipment, telemetry data is transmitted from the IIoT network sensor nodes to the control center network device at a certain frequency, as well as when dangerous deviations of the measured parameter values from the required standards are detected.

In order to promptly eliminate the causes of possible emergency situations, it is necessary to ensure a high probability of messages delivery containing telemetry data. At the same time, it is necessary to minimize the receipt of duplicate messages by the control center, as they can lead to incorrect diagnostics of the state of the controlled equipment. It should also be taken into account that the message delivery characteristics deterioration can be caused by an excessive volume of packet traffic transmitted in the IIoT network. In addition, for this reason, the energy consumption of sensor nodes increases significantly, so it is necessary to take measures to reasonably reduce the number of packets transmitted in the network.

One of the most common standards is the MQTT protocol. It is the basis of IIoT systems operation. It provides the implementation of three levels of data transmission quality [13-15]. The lowest level, called QoS-0, does not use retransmissions of corrupted or lost messages. QoS-1 assumes sending confirmations for correctly received information packets, but allows for duplicate messages. The highest level (QoS-2) uses double confirmations and guarantees the absence of duplicates of delivered messages.

The choice of the implemented QoS level significantly affects the characteristics of message delivery in the system. It should ensure the probability of message delivery not lower than the specified value, the probability of message duplicate delivery not higher than the specified value, as well as the minimization of packet traffic transmitted in the IIoT system.

To improve the quality of packet delivery in IIoT networks based on the MQTT protocol, it is proposed to use machine learning capabilities and the concepts of software-defined networks (SND) [16].With this extension, multicast messages are sent simultaneously to a group of subscribers without using a broker [17]. The RT-MQTT extension has been created, which integrates the advantages of SDN and

the MQTT protocol. This development allows you to set requirements for the transmission of messages in real time and ensure their fulfillment using network backups [18].

In order to increase the speed and reliability of delivering the most important messages, the p-MQTT extension has been developed [19]. However, the messages that are most sensitive to delivery time are sent earlier than others. To implement this approach, virtual message queues are created. Each queue is assigned a specific priority, according to which the order of message processing is set.

To reduce message delivery delays, data transmission quality settings are adjusted at the software level. At the same time, the services are adapted to the specific applications used [20-22].

An analysis of the above developments has shown that they are not directly aimed at providing the required values for the probability of message delivery in the IIoT network. In addition, the available work does not assess the likelihood of duplicate messages, which should be minimized by duplicating messages, since receiving duplicates may lead to incorrect diagnosis of the condition of the monitored equipment. Based on this, it is proposed to fill the lack of theoretically sound solutions aimed at ensuring the reliability of transmission of telemetry messages in industrial Internet of Things networks.

The purpose of this study is to ensure the required characteristics of data delivery in the industrial Internet of Things network based on the automated control system development for the telemetry message transmission.

2 LITERATURE REVIEW

A literature review on the study topic showed that the lowest level of QoS-0 has the least protocol redundancy [23], [24]. When implemented, it does not use repeated transmissions of distorted or lost information packets, which ensures a low volume of traffic transmitted in the IIoT system.

The QoS-1 level assumes the transmission of confirmations for correctly (error-free) received information packets; in this case, if the confirmation has not been received by the sending device before the timeout expires, then the corresponding information packet is retransmitted. The number of allowed retransmissions Nretry is one of the controlled parameters in the system. The peculiarity of the QoS-1 level is that when it is used, duplication of delivered messages is possible.

QoS-2 is the highest level of data transmission quality, which guarantees the absence of duplicates of delivered messages by using double confirmations of correct packet reception. The disadvantage of this level is that the volume of transmitted traffic is the highest.

Using QoS level and the allowed number of retransmissions, it is necessary to achieve an increase in the message delivery probability in the IIoT system. On the other hand, duplication of delivered messages should be minimized, since it can lead to incorrect diagnostics of the monitored equipment state. In addition, an important criterion is the reduction of transmitted traffic volume. The analysis showed that the issues of choosing the QoS level and the Nretry parameter taking into account the achievement of all the listed requirements in industrial Internet of Things systems are not given enough attention in scientific and technical publications.

Issues of control over the data transmission quality in MQTT networks are the subject of separate developments. In particular, a dynamic controller has been proposed that is capable of assessing the state of the underlying network in terms of end-to-end delay and packet error rate and assigning the best QoS value to a node [25]. In order to support real-time services with MQTT quality of service classes, a set of software extensions has been developed, the use of which allows to increase the timeliness of network traffic transmission [26]. A solution has been proposed that integrates software-defined network (SDN) technology and the MQTT protocol. In this case, user-defined protocol settings are used to ensure the timeliness of data transmission, allowing for setting real-time requirements [27].

An approach is proposed that involves checking the relevance of messages transmitted in the MQTT network. If the message is out of date, the broker requests an updated message from the publisher before forwarding it to subscribers. This approach is useful in industrial control systems, where real-time data transmission is critical. However, implementing this approach requires the transmission of a significant amount of service traffic [28]. A load balancing scheme is proposed that allows for the uniform use of network equipment for processing incoming packets [29]. An adaptive mechanism for selecting MQTT message delivery paths based on reinforcement learning is developed, which makes it possible to increase the reliability of data transmission and limit the consumption of network resources [30].

The existing approaches to creating adaptive QoS in the Internet of Things networks discussed above are implemented in the form of various modifications of the MQTT protocol. The RT-MQTT (Real Time MQTT) extension has been created, which integrates the advantages of SDN and the MQTT protocol [27]. This design allows you to set requirements for real-time message transmission and ensure that they are met using network backups.

In order to increase the speed and reliability of delivering the most important messages, the p-MQTT (Priority MQTT) extension has been developed [19]. However, the messages that are most sensitive to delivery time are sent earlier than others. To implement this approach, virtual message queues are created. Each queue is assigned a specific priority, according to which the order of message processing is set.

Sensor receiving and transmitting devices connected to an IIoT network are usually low-power and have low performance. This ensures their main advantage, which is low energy consumption. In addition, the transmission of messages by these devices is carried out in a wireless environment in which the influence of noise and various interferences is high. To deliver data in conditions of limited resources and unstable communications, an extension adapted for use in wireless sensor networks has been developed. This version of the protocol is called MQTT-SN (MQTT for Sensor Networks) [25, 26]. Unlike the basic protocol, in the MQTT-SN version, the number of retransmissions is a parameter whose values can be changed.

This standard recommends setting these values in the range from 3 to 5, but there are no automated tools that select and set the most acceptable number of retransmissions depending on the current conditions of data delivery in the IIoT network. A brief description of the main approaches to QoS management in Internet of Things networks is presented in Tab. 1.

Table 1 Features Of Qos-Adaptive Modifications Of The MQTT Protoco

Modification	RT-MQTT	p-MQTT	MQTT-SN
Features	Using network backups using an SDN controller	Assigning priorities to transmitted messages	A simplified version of MQTT with the ability to adjust the parameters of repeat transmissions
Advantages	Support for real-time messaging	Improving the speed of delivery of critical event messages	Reduced power consumption, increased reliability of data transmission
Disadvantages	No means are provided to reduce message duplication	There are no means of increasing the probability of message delivery.	There is no automated adaptive selection of Nretry

Thus, the review can be summarized, The existing QoS management solutions allow improving the indicators of communication security, average delay and the number of correctly received packets, but they are not directly aimed at ensuring the required values of the probability of delivery and the probability of duplication of messages in the IIoT system, and do not take into account the intensity of packet traffic. As review result, it can be stated that at present there is a need to improve the process of data delivery in the industrial Internet of Things system. To this end, it is proposed to develop an automated system for a reasonable adaptive choice of the implemented QoS level and the allowed number of retransmissions Nretry. This development should be based on an assessment of the performance characteristics of the IIoT network in accordance with the MQTT-SN protocol.

3 METHODOLOGY

The developed system is based on the use of probabilistic models of the message delivery process in the IIoT system in accordance with the QoS-0, QoS-1 and QoS-2 levels, which are regulated by the MQTT-SN protocol.

3.1 Models for Assessing IIoT-Message Delivery Characteristics

Using QoS-2 level and the Nretry parameter, the message will not be delivered from the sensor device to the network device of the control centre if any of the following event's sequences occur:

- the PUBLISH information packet is correctly received; the PUBREC confirmation packet is correctly and timely received; the PUBREL service packet is incorrectly received; after retransmission, the PUBREL service packet is incorrectly received;
- the PUBLISH information packet was received correctly; the PUBREC confirmation packet was received incorrectly and/or untimely; after retransmission, the PUBLISH information packet was received correctly; the PUBREC confirmation packet was received correctly and timely; the PUBREL service packet was received incorrectly; after retransmission, the PUBREL service packet was received incorrectly;
- PUBLISH information packet received correctly; PUBREC confirmation packet received incorrectly and/or untimely; PUBLISH information packet received correctly after retransmission; PUBREC confirmation packet received incorrectly and/or untimely;
- PUBLISH information packet received correctly; PUBREC confirmation packet received incorrectly and/or untimely; PUBLISH packet received incorrectly after retransmission;
- PUBLISH information packet received incorrectly; PUBLISH information packet received correctly after retransmission; PUBREC confirmation packet received correctly and on time; PUBREL service packet received incorrectly; PUBREL service packet received incorrectly after retransmission;
- the PUBLISH information packet was received incorrectly; after retransmission, the PUBLISH information packet was received correctly; the PUBREC confirmation packet was received incorrectly and/or untimely;
- the PUBLISH information packet was received incorrectly; after retransmission, the PUBLISH information packet was received incorrectly.

Taking these events into account, the probability that the message will not be delivered from the sensor device to the IoT device of the control center can be estimated using the following expression:

$$\begin{aligned} P0 &= P1\cdot P2\cdot PC\cdot(1-P2)\cdot(1-P2)+ \\ &+P1\cdot(1-P2\cdot PC)\cdot P1\cdot P2\cdot PC\cdot(1-P2)\cdot(1-P2)+ \\ &+P1\cdot(1-P2\cdot PC)\cdot(1-P2\cdot PC)+ \\ &+P1\cdot(1-P2\cdot PC)\cdot(1-P1)+ \\ &+(1-P1)\cdot P1\cdot P2\cdot PC\cdot(1-P2)\cdot(1-P2)+ \\ &+(1-P1)\cdot P1\cdot(1-P2\cdot PC)+(1-P1)\cdot(1-P1). \end{aligned} \quad (1)$$

Where $P1$ is the probability of correct reception of the information packet PUBLISH; $P2$ is the probability of correct reception of the confirmation packet PUBREC; PC is the probability of timely receipt of the confirmation package PUBREC, that is, receiving this packet before the timeout expires.

The probability of correct reception of the PUBLISH information packet can be estimated using the expression:

$$P1 = 1 - (L1 \cdot BER), \quad (2)$$

where $L1$ is the bit length of the physical layer information block in which the packet PUBLISH is encapsulated; BER is the bit error rate in wireless channels used for data transmission.

To calculate the probability of correct reception of the PUBREC confirmation packet, the following expression can be used:

$$P2 = 1 - (L2 \cdot BER). \quad (3)$$

Where $L2$ is the bit length of the physical layer information block in which the PUBREC confirmation packet or the PUBREL service packet is encapsulated.

In accordance with these models, the probability of message delivery in the IIoT system while ensuring the QoS-2 level can be calculated using the expression:

$$PD2 = \left\{1 - \left[(1-PP)^{Nretry+1} + PP \cdot (1-P2)^{Nretry+1} \cdot \sum_{i=0}^{Nretry} (1-PP)^i\right]\right\}^2. \quad (4)$$

Where PP is the probability of correct and timely information reception and confirmation packets, to calculate the value of PP, the expression is used:

$$PP = P1 \cdot P2 \cdot PC. \quad (5)$$

When level QoS-1 is provided and $Nretry = 1$ is set, the message will not be delivered from the sensor device to the control center network device if the PUBLISH information packet is not received correctly during the first attempt and after its retransmission. Taking this into account, the probability of message delivery in an IIoT system implementing the above-mentioned level of data transmission quality can be estimated using the following expression:

$$PD1 = \left[1 - (1-P1)^{Nretry+1}\right]^2. \quad (6)$$

At QoS-0 level, a message will be delivered in an IIoT system if the PUBLISH information packet is correctly received by the server, then being sent from the sensor device, and then, after being sent from the server, it is correctly received by the network device of the control center. Taking this into account, the probability of message delivery in IioT system at QoS-1 can be estimated using the expression:

$$PD0 = P1^2. \quad (7)$$

At QoS-1 level, an important functioning characteristic of the IIoT system is the message duplication probability. As a result of the study, the following expression was obtained to estimate this value:

$$PDUBL = 2 \cdot (1 - PONCE - PNOT) - (1 - PONCE - PONT)^2. \quad (8)$$

Where $PONCE$ – is the message delivery probability without duplicates from a sensor device to a server, as well as the message delivery probability without duplicates from a server to a network device of the control center; $PNOT$ is the probability that a message will not be delivered from a sensor device to a server, as well as the probability that a message will not be delivered from a server to a network device of the control center.

The $PONCE$ value can be estimated using the formula:

$$PONCE = P1\left\{(1-P1)^{Nretry}\left[1 + Nretry \cdot (1 - P2 \cdot PC)\right] + P2 \cdot PC \sum_{i=0}^{Nretry-1} (1-P1)^i\right\}. \quad (9)$$

The value $PNOT$ can be estimated using the formula:

$$PNOT = (1-P1)^{Nretry+1}. \quad (10)$$

So, for the first time, the models presented above reflect the dependence of the probability of delivery and the probability of duplication of telemetry messages on the current characteristics of data transmission in the IIoT network. The mentioned models allow to calculate important characteristics of data transmission in the IIoT network, which include $PD0$, $PD1$, $PD2$ and $PDUBL$. The values of these quantities can be used to select the level of message delivery quality and the permitted number of packet retransmissions. In the process of this selection, it is also necessary to take into account the criteria determined by the specifics of the controlled technological process. Based on the specifics of the production where the IIoT system is deployed, the probability of telemetry message delivery should not be lower than the specified GPD value, and the probability of duplication of delivered messages should not exceed the specified $GPDUBL$ value. For a theoretically justified selection of the message delivery quality level and the permitted number of packet retransmissions, a system of automated control of data transmission characteristics in the industrial Internet of Things network is proposed. The diagram of this system is presented below.

The mentioned models allow to calculate important characteristics of data transmission in the IIoT network, which include $PD0$, $PD1$, $PD2$ and $PDUBL$. The values of these quantities can be used to select the level of message delivery quality and the permitted number of packet retransmissions. In the process of this selection, it is also necessary to take into account the criteria determined by the specifics of the controlled technological process. Based on the specifics of the production where the IIoT system is deployed, the probability of telemetry message delivery should not be lower than the specified GPD value, and the probability of duplication of delivered messages should not

exceed the specified *GPDUBL* value. For a theoretically justified selection of the message delivery quality level and the permitted number of packet retransmissions, a system of automated control of data transmission characteristics in the industrial Internet of Things network is proposed. The diagram of this system is presented below.

3.2 The Proposed Automated System

The proposed automated system for controlling the parameters of telemetry message transmission in the industrial Internet of Things network is installed on the computing equipment of the control center. The diagram of this system is shown in Fig 1. The functioning of this system is based on the use of Eqs. (1) – (10), which were obtained during the development of message delivery models in the IIoT network.

The telemetry data transmitted by the sensor nodes over the IIoT network are received by the control center network device, then are sent to the *BER* measurement module. The measured *BER* values are transmitted in parallel to the *PD*0 evaluation module, *PD*1 evaluation module, and *PD*2 evaluation module.

In the *PD*0 evaluation module, the current *PD*0 value is calculated using Eq. (5). This value is then fed to the selection criteria fulfillment control module. In this module, the *PD*0 values are compared with the specified *GPD* value. The *GPD* value is set by the network administrator using the selection criteria setting interface. If $PD0 \geq GPD$ is true, the implementation of the QoS-0 level is recommended when exchanging data with the corresponding sensor device, so the value of the recommended data transmission quality level LQoS is assigned the value 0. This value is fed to the LQoS and Nretry setting module. If $PD0 < GPD$, then the selection criteria fulfillment control module sends a request to the *PD*1 evaluation module to calculate the *PDUBL* value.

In the *PD*1 evaluation module, the *PDUBL* value is calculated using Eqs. (6) – (8) for $Nretry = 1$. The calculated *PDUBL* value is led to the selection criteria execution control module, where the PDUBL values are compared with the specified *GPDUBL* value. The *GPDUBL* value is set by the network administrator using the selection criteria setup interface. If $PDUBL \leq GPDUBL$, then the selection criteria fulfillment control module sends a request to the *PD*1 evaluation module to calculate the *PD*1 value. In the *PD*1 evaluation module, using Eqs. (2) and (4), the *PD*1 value is calculated for $Nretry = 1$.

If $PD1 \geq GPD$, then when exchanging data with the corresponding sensor device, it is recommended to implement the QoS-1 level, therefore the value of the recommended level of data transmission quality LQoS is assigned the value 1. This value and the value $Nretry = 1$ are led to the LQoS and Nretry setup module. If $PD1 < GPD$, then the selection criteria execution control module sends a request to the *PD*1 evaluation module to calculate the *PDUBL* value when $Nretry = 2$. Then the operations are performed in the above sequence until the condition $PD1 \geq GPD$ is met. If the condition $PDUBL > GPDUBL$ is true, then the selection criteria fulfillment control module sends a request to the *PD*2 evaluation module to calculate the *PD*2 value.

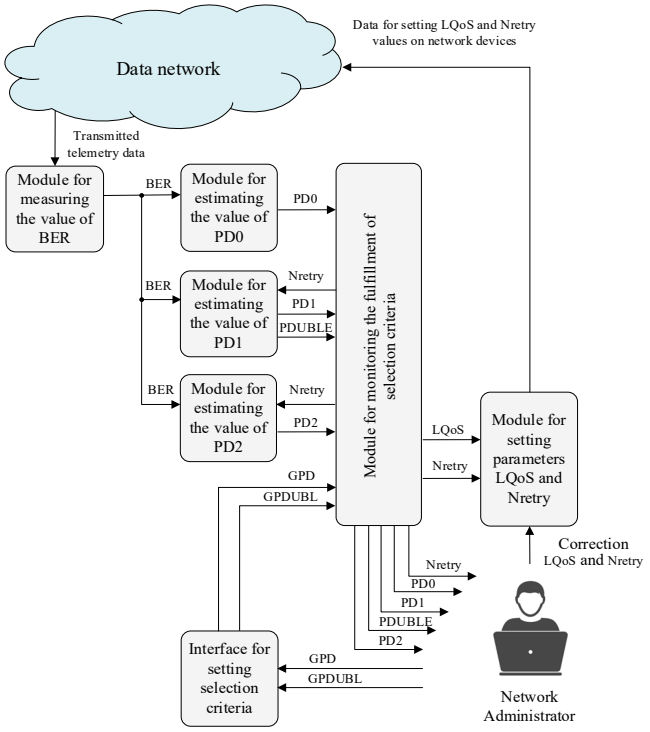


Figure 1 Scheme of automated control system

In the *PD*2 evaluation module, the value PD2 is calculated using Eqs. (1) – (3) when $Nretry = 1$. This value is then sent to the selection criteria fulfillment control module. In this module, the *PD*2 values are compared with the specified *GPD* value. If $PD2 \geq GPD$, then when exchanging data with the corresponding sensor device, it is recommended to implement the QoS-2 level, therefore the value of the recommended level of data transmission quality LQoS is assigned the value 2. This value is led to the LQoS and Nretry setup module. If $PD2 < GPD$, then the selection criteria fulfillment control module sends a request to the *PD*2 evaluation module to calculate the *PD*2 value when $Nretry = 2$, etc., until the condition $PD2 \geq GPD$ is met.

The recommended values of LQoS and Nretry can be adjusted by the network administrator if necessary. To provide this possibility, the calculated values of *PD*0, *PD*1, *PD*2, *GPDUBL*, LQoS and Nretry are provided to the network administrator from the selection criteria fulfillment control module. The LQoS and Nretry setup module transmits service data on the recommended values of the data transmission quality level and the permitted number of packet retransmissions to the sensor device. Based on this data, the recommended values of the above-mentioned telemetry message transmission parameters are set in the sensor device settings.

4 RESULTS AND DISCUSSION

The logic of the evaluation modules *PD*0, *PD*1 and *PD*2 and the module for monitoring the fulfilment of selection criteria is implemented in the form of software in the Python

language version 3.12.5. Using this software, a series of computational experiments were carried out with the following initial data: $L1$ =256 bits; L2 = 128 bits; PC = 1; GPD = 0.9; GPDUBL = 0.1. Fig. 2 shows some results of software functioning.

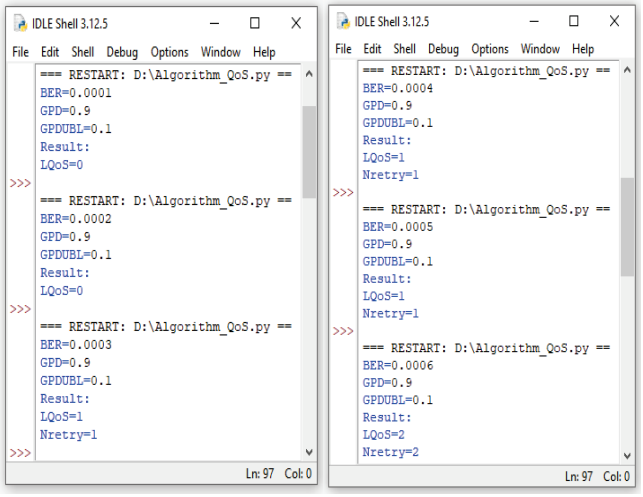


Figure 2 Results of computational experiments

As a result of experiments, it was found that in the IIoT system, with *BER* values not exceeding 2×10^{-4} it is recommended to use the QoS-0 level, with *BER* values from 3×10^{-4} to 4×10^{-4} the QoS-1 level and Nretry = 1 should be used, and with BER from 6×10^{-4} to 10×10^{-4} the QoS-2 level and *Nretry* = 2 should be used.

Then, the second series of computational experiments was conducted, which checked the fulfillment of the specified data delivery requirements and estimated the packet traffic in the IIoT system without and with the proposed algorithm. In this case, the values of V were estimated, i.e., the average number of packets transmitted in the IIoT system within 1 hour. It was assumed that the packet traffic was generated by 1000 sensor devices, each of which transmitted 4 telemetry messages within 1 hour.

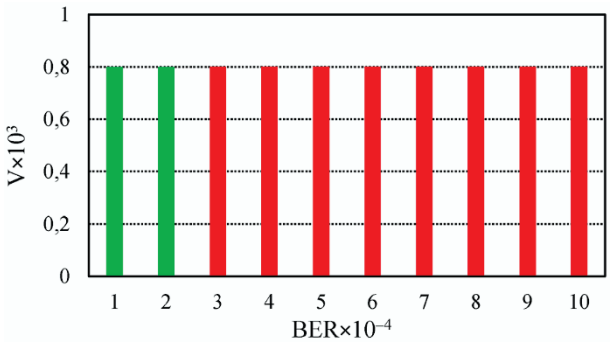


Figure 3 Diagram of the dependence of the value of V on the BER values when implementing the QoS-1 level

The obtained results are presented in the form of diagrams in Figs. 3-6. In these diagrams, the experiments results in which the telemetry message delivery requirements were met are marked by green, and the experiments results in which these requirements were not met are marked by red.

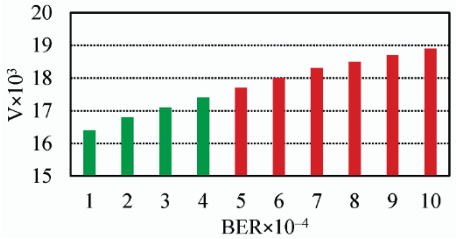


Figure 4 Diagram of the dependence of the value of *V* on the *BER* values when implementing the QoS-1 level and *Nretry* = 1

Figs. 3-5 show diagrams obtained without using the proposed automated control system. These results were observed at fixed values of the data transmission quality level and the permitted number of packet retransmissions. Fig. 6 shows the results of computational experiments conducted using the proposed automated control system.

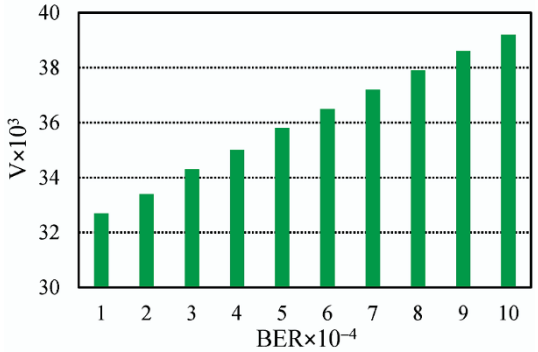


Figure 5 Diagram of the dependence of the value of V on the BER values when implementing the QoS-2 level and Nretry = 2

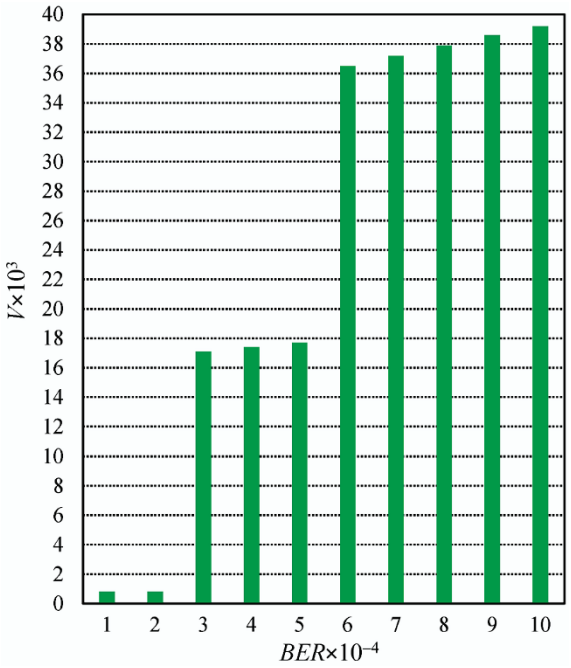


Figure 6 Diagram of the dependence of the value of *V* on the *BER* values when using the proposed automated system

Then, using the developed software, experiments were conducted in a real IoT network. 20 network LoRaWAN devices with LW002-TH temperature and humidity sensors

were connected to the eight-channel gateway MKGW2-LW. The MQTT-SN protocol was used to control data transmission at the application level in QoS-0, QoS-1 and QoS-2 modes. To change the *BER* level, we increased or decreased the distance between the sensor network devices and the gateway. The GW C-1200 protocol analyzer for LoRa wireless networks was used to measure the Bit Error Rate. The period of sending messages by each touch device was 1800 seconds. As a result of the experiments conducted in real conditions, the data presented in Tab. 2 were obtained. Analysis of these data showed that they correspond as much as possible to the simulation results.

Table 2 Comparison of Results of Computational Experiments

Conditions for computational experiments	Message delivery requirement fulfillment percentages (confidence intervals with a confidence probability of 0.95)	Average traffic intensity, thousands of packets per hour
Fixed level QoS-0	20	0.8
Fixed level QoS-1	40	17.78
Fixed level QoS-2	100	36.06
Application of automated control system	100	24.32

The analysis of the table shows that in cases of setting fixed levels of data transmission quality without using an automated control system, the implementation of the QoS-0 level allows fulfilling the data delivery requirements only by 20% and is accompanied by an average traffic intensity of 0.8 thousand packets within 1 hour. The implementation of the QoS-1 level allows fulfilling the data delivery requirements by 40% with an increase in the average traffic intensity to 17.78 thousand packets within 1 hour. The implementation of the QoS-2 level allows to fulfill the data delivery requirements by 100% with an increase in the average traffic intensity to 36.06 thousand packets within 1 hour. The use of an automated control system allows for 100% achievement of the required data delivery characteristics and, compared to the implementation of the fixed QoS-2 level, makes it possible to reduce the average traffic intensity by 32.5% (to 24.32 thousand packets within 1 hour).

Table 3 Comparison of the Results of the Proposed Approach and the Adaptive QoS Algorithm Based on SDN Technology

Indicators	Percentage increase or decrease
Average Message Delivery Rate	Decrease by 18.2 - 20.4
Duplicate Messages	Decrease by 6.4 - 8.6
The Probability Of Message Delivery	Increase by 5.8 - 7.2
Power consumption of network devices	Decrease by 4.5 - 5.3

Tab. 3 shows experimental data showing how much the results of using the proposed approach increase or decrease compared to the results of using the adaptive QoS algorithm based on SDN technology [27]. The results are presented as percentages in confidence intervals with a confidence probability of 0.95.

A comparative analysis shows that the use of adaptive QoS based on SDN gives better results in the speed of message delivery, which is important for ensuring timely data transmission in real-time applications. In situations where it is necessary to increase the probability of message delivery and reduce their duplication, the use of the proposed system provides more preferable results.

5 CONCLUSION

Thus, the presented automated system for parameters management of telemetry message transmission in the IIoT network provides a theoretically justified choice of the level of message delivery quality and the allowed number of packet retransmissions depending on the current bit error rate. The use of the system allows us to provide the specified values of the delivery probability and the message duplication probability, while significantly reducing the average traffic intensity compared to the implementation of a fixed level of data transmission quality with double confirmations.

Further research within the considered topic framework will be devoted to the substantiation of the values of the telemetry message delivery probability and the delivered messages duplication probability, which will preferably be used in the automated control of data transmission characteristics in the Industrial Internet of Things network deployed at a specific manufacturing enterprise. In the future, this will make it possible to use the proposed automated system for managing the transmission of telemetry data messages in the IIoT network, taking into account the criteria determined by the specifics of the controlled technological process.

Acknowledgement

The reported study was funded by RSF, project number 25-21-00431 (Development of models and methods for controlling parameters and modes of information transmission in wireless remote monitoring systems based on the Internet of Things and flying self-organizing networks) https://rscf.ru/project/25-21-00431/.

6 REFERENCES

[1] Sonbul, O. S., & Rashid, M. (2023). Towards the structural health monitoring of bridges using wireless sensor networks: A systematic study. *Sensors, 23*(20), 8468. https://doi.org/10.3390/s23208468

[2] Yaser, M. J., Polshchykov, K. A., & Polshchikov, I. K. (2023). Algorithm for ensuring the minimum power consumption of the end node in the LoRaWAN network. *Periodicals of Engineering and Natural Sciences, 11*(4), 168–174. https://doi.org/10.21533/pen.v11.i4.208

[3] Bonilla, V., Campoverde, B., & Yoo, S. G. (2023). A systematic literature review of LoRaWAN: Sensors and applications. *Sensors, 23*(20), 8440. https://doi.org/10.3390/s23208440

[4] Jouhari, M., Saeed, N., Alouini, M.-S., & Amhoud, E. M. (2023). A survey on scalable LoRaWAN for massive IoT: Recent advances, potentials, and challenges. *IEEE Communications Surveys & Tutorials, 25*(3), 1841–1876. https://doi.org/10.1109/COMST.2023.3274934

[5] Morchid, A., El Alami, R., Raezah, A. A., & Sabbar, Y. (2024). Applications of Internet of Things (IoT) and sensors technology to increase food security and agricultural sustainability: Benefits and challenges. *Ain Shams Engineering Journal, 15*(3), 102509.

https://doi.org/10.1016/j.asej.2023.102509
[6] Hasan, H. R., et al. (2024). Smart agriculture assurance: IoT and blockchain for trusted sustainable produce. *Computers and Electronics in Agriculture, 224*, 109184. https://doi.org/10.1016/j.compag.2024.109184
[7] Bulut, C., & Wu, P. F. (2024). More than two decades of research on IoT in agriculture: A systematic literature review. *Internet Research, 34*(3), 994–1016. https://doi.org/10.1108/INTR-07-2022-0559
[8] Mashayamombe, T., & Matope, S. (2024). A comprehensive review of additive manufacturing methods in the fabrication of self-lubricating components. *South African Journal of Industrial Engineering, 35*(2), 77–90. https://doi.org/10.7166/35-2-2933
[9] Chandramauli, A., Singh, R., Jeyalaxmi, M., Ali, H. A., Habelalmateen, M. I., & Sasipriya, S. (2024). The role of Internet of Things (IoT) in hydel energy sector—Perspectives. In *E3S Web of Conferences* (p. 8006). EDP Sciences. https://doi.org/10.1051/e3sconf/202454008006
[10] Casillo, M., Cecere, L., Colace, F., Lorusso, A., & Santaniello, D. (2024). Integrating the Internet of Things (IoT) in SPA medicine: Innovations and challenges in digital wellness. *Computers, 13*(3), 67. https://doi.org/10.3390/computers13030067
[11] Jan, I., & Sofi, S. (2024). Data management for resource optimization in medical IoT. *Health Technology, 14*(1), 51–68. https://doi.org/10.1007/s12553-023-00796-6
[12] Thirugnanam, T., et al. (2024). PIRAP: Medical cancer rehabilitation healthcare center data maintenance based on IoT-based deep federated collaborative learning. *International Journal of Cooperative Information Systems, 33*(01), 2350005. https://doi.org/10.1142/S0218843023500053
[13] Messaoud, S., Bradai, A., Dawaliby, S., & Atri, M. (2021). Slicing optimization based on machine learning tool for industrial IoT 4.0. In *IEEE International Conference on Design & Test of Integrated Micro & Nano-Systems (DTS2021)* (pp. 1–5). https://doi.org/10.1109/DTS52014.2021.9498080
[14] Anitha, T., Manimurugan, S., Sridhar, S., Mathupriya, S., & Latha, G. C. P. (2022). A review on communication protocols of industrial Internet of Things. In *2nd IEEE International Conference on Computing and Information Technology (ICCIT2022)* (pp. 418–423). https://doi.org/10.1109/ICCIT52419.2022.9711544
[15] Wang, W., Zhao, Y., Liu, Y., Liu, G., Zheng, F., & Sun, C. (2024). MQTT protocol and implementation of equipment management system for industrial Internet of Things. In *43rd IEEE Chinese Control Conference (CCC2024)* (pp. 6139–6144). https://doi.org/10.23919/CCC63176.2024.10662474
[16] Bartoli, C., Bonanni, M., Chiti, F., & Pierucci, L. (2025). The alliance of SDN and MQTT for the web of industrial things. *IEEE Transactions on Industrial Informatics.* https://doi.org/10.1109/TII.2025.3537291
[17] Spohn, M. A. (2022). On MQTT scalability in the Internet of Things: Issues, solutions, and future directions. *Journal of Electronics and Electrical Engineering*, 4. https://doi.org/10.37256/jeee.1120221687
[18] Shahri, E., Pedreiras, P., & Almeida, L. (2022). Extending MQTT with real-time communication services based on SDN. *Sensors, 22*(9), 3162. https://doi.org/10.3390/s22093162
[19] Patti, G., Leonardi, L., Testa, G., & Lo Bello, L. (2024). PrioMQTT: A prioritized version of the MQTT protocol. *Computer Communications, 220*, 43–51. https://doi.org/10.1016/j.comcom.2024.03.018
[20] Yew, H. T., Ng, M. F., Ping, S. Z., Chung, S. K., Chekima, A., & Dargham, J. A. (2020). IoT-based real-time remote patient monitoring system. In *16th IEEE International Colloquium on Signal Processing & Its Applications (CSPA2020)* (pp. 176–179). https://doi.org/10.1109/CSPA48992.2020.9068699
[21] Rosli, A. N., Mohamad, R., Yusof, Y. W. M., Shahbudin, S., & Rahman, F. Y. A. (2020). Implementation of MQTT and LoRaWAN system for real-time environmental monitoring application. In *10th IEEE Symposium on Computer Applications & Industrial Electronics (ISCAIE2020)* (pp. 287–291). https://doi.org/10.1109/ISCAIE47305.2020.9108808
[22] Zambrano, M., Mejía, E. L. O., & Calderón, X. (2020). SIGPRO: A real-time progressive notification system using MQTT bridges and topic hierarchy for rapid location of missing persons. *IEEE Access, 8*, 149190–149198. https://doi.org/10.1109/ACCESS.2020.3015183
[23] Opačin, S., Rizvanović, L., Leander, B., Mubeen, S., & Čaušević, A. (2023). Developing and evaluating MQTT connectivity for an industrial controller. In *12th IEEE Mediterranean Conference on Embedded Computing (MECO2023)* (pp. 1–5). https://doi.org/10.1109/MECO58584.2023.10154921
[24] Pawar, S., Jadhav, D. B., Lokhande, M., Raskar, P., & Patil, M. (2024). Evaluation of quality of service parameters for MQTT communication in IoT application by using deep neural network. *International Journal of Information Technology, 16*(2), 1123–1136. https://doi.org/10.1007/s41870-023-01664-2
[25] Palmese, F., Redondi, A. E. C., & Cesana, M. (2022). Adaptive quality of service control for MQTT-SN. *Sensors, 22*(22), 8852. https://doi.org/10.3390/s22228852
[26] Fontes, F., Rocha, B., Mota, A., Pedreiras, P., & Silva, V. (2020). Extending MQTT-SN with real-time communication services. In *25th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA2020)* (pp. 1–4). https://doi.org/10.1109/ETFA46521.2020.9212147
[27] Shahri, E., Pedreiras, P., & Almeida, L. (2024). A scalable real-time SDN-based MQTT framework for industrial applications. *IEEE Open Journal of the Industrial Electronics Society.* https://doi.org/10.1109/OJIES.2024.3373232
[28] Kim, Y., & Kyung, Y. (2024). AoI-aware retained message policy in MQTT-based IoT networks. *IEEE Sensors Journal.* https://doi.org/10.1109/JSEN.2024.3465025
[29] Doshi, R., Inamdar, S., Karmarkar, T., & Wakode, M. (2024). Distributed MQTT broker: A load-balanced Redis-based architecture. In *IEEE International Conference on Emerging Smart Computing and Informatics (ESCI2024)* (pp. 1–6). https://doi.org/10.1109/ESCI59607.2024.10497427
[30] Zunino, C., Cena, G., Scanzio, S., & Valenzano, A. (2023). Adaptive seamless redundancy to achieve highly dependable MQTT communication. *IEEE Transactions on Industrial Informatics, 20*(1), 984–994. https://doi.org/10.1109/TII.2023.3271708

Authors' contacts:

Muhanad Jabar Yaser, PhD
(Corresponding author)
The Head of the Department Cybersecurity Technologies,
Technical Institute Shatra, Southern Technical University, Basra, 61001, Iraq
E-mail: muhaned.yaser@stu.edu.iq

Konstantin A. Polshchykov, Doctor of Technical Sciences, Professor
Department of Information and Robotic Systems, Belgorod State University,
Belgorod, 308015, Russia
E-mail: polshchikov@bsu.edu.ru

Mikhail S. Balakshin, MSc
Department of Information and Robotic Systems, Belgorod State University,
Belgorod, 308015, Russia
E-mail: mihailil86@mail.ru

ISSN 1846-6168 (Print), ISSN 1848-5588 (Online)
https://doi.org/10.31803/tg-20260502123125

Preliminary communication
Received: 2026-05-02, Accepted: 2026-06-01

Geometric Modeling of 2D Regions in AutoCAD Using the Marching Squares Algorithm with R-Functions

Nuraliev Fakhriddin Murodillayevich, Inoyatov Mirzayor Bakhtiyor ugli*, Ibodullaev Sardor Nasriddin ugli, Umarova Dildora Bakhtiyarovna, Giyosov Ulugbek Eshpulatovich

Abstract: This study presents a methodology for geometric modeling of two-dimensional regions in AutoCAD based on the theory of R-functions by V. L. Rvachev combined with the Marching Squares algorithm. A specialized RFM 2D Plugin was developed in C# for the AutoCAD.NET API, enabling users to define geometric domains through implicit functions of the form $f(x, y) = 0$ and to construct complex composite regions via R-operations (conjunction, disjunction, and negation). Region boundaries are extracted automatically using the Marching Squares algorithm with bisection-based edge interpolation (an accuracy of $\Delta x/1024$) and are rendered as closed AutoCAD Polyline contours filled with Hatch objects and unified into a single editable Group. The method was validated on three classes of planar domains — four simple primitives (circle, ellipse, rectangle, triangle), parabolic and hyperbolic regions, and three complex composite models (a house, a six-pointed star, and a chess-king piece) composed of up to 15 primitive functions. The results show that a single analytical expression replaces the multi-command sequences required by conventional parametric modeling, that the number of contour points grows linearly and the computation time quadratically with grid resolution, and that boundary smoothness can be controlled through a single Rα parameter. The study demonstrates that embedding R-functions theory directly within the AutoCAD .NET environment yields a unified, flexible, and practically applicable framework for modeling irregular and analytically defined planar domains.

Keywords: 2D geometric modeling; AutoCAD; Hatch; implicit domain; Marching Squares; R-functions

1 INTRODUCTION

Computer-Aided Geometric Modeling (CAGM) is an integral part of modern engineering design and is widely used in construction, mechanical engineering, aerospace, and other fields. In professional CAD systems such as AutoCAD, 2D regions are typically created using parametric commands—"Circle", "Ellipse", "Rectangle"—and their associated "Trim", "Fillet", and "Boolean" operations [1, 2]. However, conventional parametric methods exhibit significant limitations and complexity when constructing regions with intricate geometries—such as intersections of multiple curved boundaries, or regions bounded by parabolic or hyperbolic curves. In particular, drawing each boundary curve individually and then manually trimming and joining them is a multi-step process that reduces design efficiency [3, 4].

An effective alternative to this problem is modeling based on implicit functions (i.e., $F(x, y) = 0$). In this approach, a geometric region is defined by $f(x, y) \geq 0$, where $f(x, y) = 0$ denotes the boundary, $f(x, y) > 0$ the interior, and $f(x, y) < 0$ represents the exterior [5, 6, 7]. The mathematical foundation of implicit modeling is the R-functions theory proposed by academician V.L. Rvachev in 1963 [8, 9]. R-function is analytical analogs of logical operations (conjunction $\wedge$, disjunction $\vee$, negation $\neg$), which enable set-theoretic operations on geometric regions using continuous, differentiable functions [10, 11].

To display the boundary of an implicitly defined region on screen or within a CAD system, it must be converted into a sequence of points forming a polyline. For this purpose, the Marching Squares (MS) algorithm is employed—the two-dimensional analogue of the Marching Cubes algorithm proposed by Lorensen and Cline in 1987 [12]. The Marching Squares algorithm examines the sign change of the function at each cell of a 2D planar grid and generates isoline segments based on 16 possible configurations [13, 14]. Maple [13] was the first to investigate this algorithm in the context of 2D geometric design. Newman and Yi [5] presented a comprehensive survey of the Marching Cubes/Squares algorithms.

Shapiro [10] extensively covered the application of R-functions in geometric modeling and computational mechanics, introducing the concept of semi-analytic geometry. Pasko et al. [15] formalized the Function Representation (FRep) concept and developed a methodology for performing set-theoretic operations, blending, offset, and metamorphosis on implicit functions. Ricci [16] proposed a constructive geometry approach for computer graphics, a work that served as an important foundation for practical applications of R-functions theory.

The R-Function Method (RFM) is also widely applied in engineering. Rvachev and Sheiko [17] demonstrated its broad use in boundary value problems in mechanics. Tsukanov and Shapiro [18, 19] developed the SAGE meshfree system and successfully applied the RFM to natural frequency analysis of 2D structures. Kurpa et al. [20] applied the RFM to vibration analysis of porous functionally graded material (FGM) plates, demonstrating its contemporary engineering relevance.

However, there are almost no studies in the existing literature on the direct application of R-functions theory to the AutoCAD environment—that is, the development of a practical software tool that automatically visualizes 2D regions as Polyline and Hatch objects based on user-entered formulas. The aim of the present study is to fill precisely this gap by developing a plugin that integrates R-function with the Marching Squares algorithm and operates via the AutoCAD.NET API, and by evaluating its effectiveness experimentally.

AutoCAD was deliberately selected as the target platform for several complementary reasons. First, it is one of the most widely deployed CAD systems in industry, with a global user base of more than 38.5% according to 6sense

market analytics, which maximizes the practical reach of the developed tool [21]. Second, AutoCAD exposes a mature, well-documented managed programming interface — the AutoCAD.NET API — that provides direct access to the drawing database, to the geometric entities (Polyline, Hatch, Region), and to the command pipeline required for automatic visualization. Third, its native object model maps naturally onto the output of the Marching Squares algorithm: isoline segments become Polylines and filled domains become Hatch objects without any intermediate file-format conversion. Finally, AutoCAD is a de facto industry standard in architecture, civil, and mechanical engineering, so a plugin embedded in this environment can be integrated directly into established professional workflows rather than requiring a separate, stand-alone application.

The scientific novelty of the study consists of: (1) the practical application of R-functions theory and the Marching Squares algorithm integrated with the AutoCAD.NET API; (2) the presentation of numerical results for simple and complex 2D domains and a comparative analysis with conventional AutoCAD commands.

2 LITERATURE REVIEW

The literature analyzed in this section was identified through a systematic search of the following bibliographic databases: Scopus, Web of Science, IEEE Xplore, the ACM Digital Library, and Google Scholar. The search combined the key terms “R-functions”, “implicit modeling” / “implicit surfaces”, “Marching Squares” / “Marching Cubes”, “function representation (FRep)”, “constructive solid geometry”, and “AutoCAD .NET plugin” using Boolean operators, for example: (“R-functions” AND (“Marching Squares” OR “implicit modeling”) AND (“CAD” OR “AutoCAD”)). The search covered publications from 1963 to 2026. Studies were included if they addressed (i) the theoretical foundations of R-functions, (ii) the polygonization or boundary extraction of implicit domains, or (iii) engineering applications of implicit / function-based modeling; studies were excluded if they did not involve implicit or function-based representations, or were not available in English or Russian. After removing duplicates and applying these criteria, 33 core works were retained and are organized below into five thematic subsections.

2.1 Foundations of R-functions Theory

The founder of R-functions theory, academician V. L. Rvachev, proposed in 1963 a new method for the analytical description of geometric objects [8]. In his fundamental monograph published in 1982 [9], Rvachev developed the complete mathematical theory of R-functions and applied it to the solution of boundary value problems. An R-function is a real-valued function whose sign (positive or negative) depends solely on the signs of its arguments. This property allows logical operations to be expressed as continuous, differentiable functions, which is of fundamental importance in geometric modeling.

Shapiro [11] was the first to extensively cover the R-functions theory in English, and his 1991 technical report served as a primary source for the Western scientific community. Shapiro’s 2007 fundamental review article [10] provided a detailed analysis of the applications of R-function in computational geometry, boundary value problems, and optimization. Shapiro [6] also investigated the properties of real functions for solid representation, substantiating their potential for use in 2D and 3D modeling systems.

Rvachev and Sheiko [17] demonstrated the effectiveness of the R-Function Method in boundary value problems in mechanics and developed a methodology for constructing analytical solution structures over complex geometric domains. Rvachev, Sheiko, Shapiro, and Tsukanov [22] proved the completeness of RFM solution structures, thereby establishing the mathematical rigor of the method. Rvachev et al. [23] developed a method for transfinite interpolation over implicitly defined sets using R-function.

2.2 Implicit Domains and Functional Representation

Implicit modeling—representing geometric objects via functions of the form $f(x, y) \geq 0$—has been extensively studied in the fields of computer graphics and CAD. Requicha [1] established the theoretical foundations for solid representation by comparing CSG (Constructive Solid Geometry) and B-Rep (Boundary Representation) methods. Ricci [16] proposed a constructive geometry approach for computer graphics, being the first to introduce the idea of performing logical operations on implicit functions.

Pasko et al. [15] formalized the FRep concept and developed an extended set of operations based on R-function—set-theoretic operations, blending, offset, and metamorphosis. Gomes et al. [24] produced a monograph providing a detailed account of the mathematical foundations, data structures, and algorithms for implicit curves and surfaces. Bloomenthal et al. [25] compiled a foundational handbook on implicit surfaces, making a significant contribution to the development of the field.

2.3 Marching Squares Algorithm and Contour Extraction Methods

The problem of finding the $f(x, y) = 0$ isoline that forms the boundary of an implicit domain and converting it into a sequence of points is referred to in computer graphics as isocontour extraction. The most widely known method in this field is Marching Squares—the 2D version of the Marching Cubes algorithm proposed by Lorensen and Cline [12]. In the Marching Squares algorithm, the four corner values of each cell in a 2D grid are examined, and isoline segments are generated for $2^4 = 16$ possible configurations.

Maple [13] was the first to investigate the application of the Marching Squares algorithm in geometric design and space planning. Nielson and Hamann [14] proposed the asymptotic decider method to resolve the ambiguity problem in Marching Cubes/Squares. Ho et al. [26] developed an adaptive, feature-preserving contour extraction method. De Araujo et al. [27] presented the most comprehensive survey on the polygonization of implicit surfaces, comparing Marching Squares, Marching Cubes, and other methods in detail. In a related image-analysis context, Mantz et al. [28]

employed Minkowski functionals within a marching-square framework, demonstrating the algorithm's versatility beyond geometric modeling.

2.4 Engineering Applications of the R-Function Method

The R-Function Method (RFM) is widely applied not only in geometric modeling but also in computational mechanics and engineering analysis. Tsukanov and Shapiro [18] created the SAGE meshfree system and applied the RFM to fluid dynamics problems. Tsukanov and Shapiro [19] developed a methodology for meshfree natural frequency analysis of 2D structures. Freytag, Shapiro, and Tsukanov [29] proposed an approach for integrating finite element analysis with implicit geometry.

In recent studies, Kurpa et al. [20] successfully applied the R-Function Method to the vibration analysis of porous FGM plates. Vescovini [30] developed the Ritz R-Function Method for the analysis of variable stiffness composite panels. These studies confirm the relevance of the R-Function Method in modern engineering.

2.5 Plugin Development in CAD Systems

Zhang and Zhang [31] provided a detailed account of secondary development technology for AutoCAD based on the.NET API. Bloomenthal [32] proposed efficient algorithms for polygonizing implicit surfaces. However, no studies are found in the existing literature on the direct application of R-functions theory to the AutoCAD environment or on its integration with the Marching Squares algorithm. This constitutes the scientific novelty of the present work.

3 METHODOLOGY

Before the method is described in detail, its originality is stated explicitly. To the best of the authors' knowledge, this is the first work that integrates V. L. Rvachev's R-functions theory directly with the Marching Squares algorithm inside the AutoCAD .NET environment as a single, fully automated pipeline. The novelty is threefold. (1) Methodological: a complete formula-to-geometry chain is proposed, in which an arbitrary R-function expression is parsed into an executable delegate, sampled on a uniform grid, contoured by Marching Squares with bisection refinement, and emitted directly as native AutoCAD Polyline and Hatch objects. (2) Algorithmic: the classical Marching Squares procedure is combined with R_α-parametric R-operations and bisection-based edge interpolation (ten iterations, giving an accuracy of $\Delta x/1024$), so that the smoothness of the reconstructed boundary is controlled through a single parameter α. (3) Implementation: a reusable C# plugin (RFM 2D Plugin) delivers this capability through the standard AutoCAD ribbon and palette interface. Unlike earlier FRep and meshfree implementations, which operate as stand-alone tools, the present approach embeds implicit modeling natively within a mainstream commercial CAD system.

3.1 Mathematical Apparatus of R-function (2D Case)

In representing a planar region using an R-function, the function $f(x, y)$ is employed, where $f(x, y) > 0$ denotes the interior, $f(x, y) = 0$ the boundary, and $f(x, y) < 0$ the exterior of the region. For basic 2D primitives, the analytical functions are defined as follows:

- *Circle:* center (x_0, y_0), radius R:

$$f(x,y) \equiv R^2 - (x-x_0)^2 - (y-y_0)^2 \geq 0 \tag{1}$$

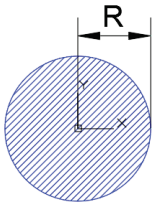


Figure 1 Representation of a circle defined by its center (x_0, y_0) and radius R

- *Ellipse:* center (x_0, y_0), semi-axes a and b

$$f(x,y) \equiv 1 - \frac{(x-x_0)^2}{a^2} - \frac{(y-y_0)^2}{b^2} \geq 0 \tag{2}$$

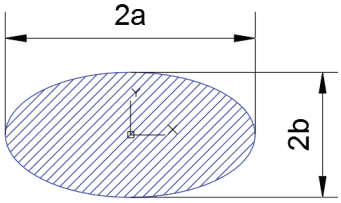


Figure 2 Representation of an ellipse defined by its center (x_0, y_0) and semi-axes a and b

- *Rectangle:* dimensions $a \times b$, via conjunction of 2 functions

$$f_1 \equiv x(a-x) \geq 0, f_2 \equiv y(b-y) \geq 0$$
$$F_{\text{rectangle}} \equiv f_1 \wedge f_2 \geq 0 \tag{3}$$

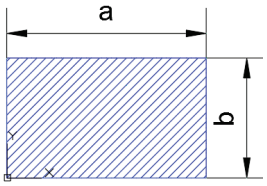


Figure 3 Representation of a rectangle defined by its dimensions a and b

- *Triangle:* conjunction of two linear functions

$$\omega_1 \equiv h - y - \left(\frac{h}{a}\right) \cdot |x-a| \geq 0,\ \omega_2 \equiv y \geq 0$$
$$F_{\text{triangle}} \equiv \omega_1 \wedge \omega_2 \geq 0 \tag{4}$$

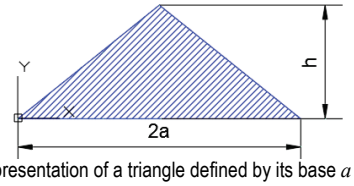


Figure 4 Representation of a triangle defined by its base a and height h

Complex regions are constructed as logical combinations of simple primitives via R-function. R_α-parametric R-function system [10, 11] defines three basic operations as follows:

- *Conjunction* (∧):

$$f_1 \wedge f_2 \equiv (1+\alpha)^{-1} \cdot \left(f_1 + f_2 - \sqrt{f_1^2 + f_2^2 - 2 \cdot \alpha \cdot f_1 \cdot f_2} \right) \quad (5)$$

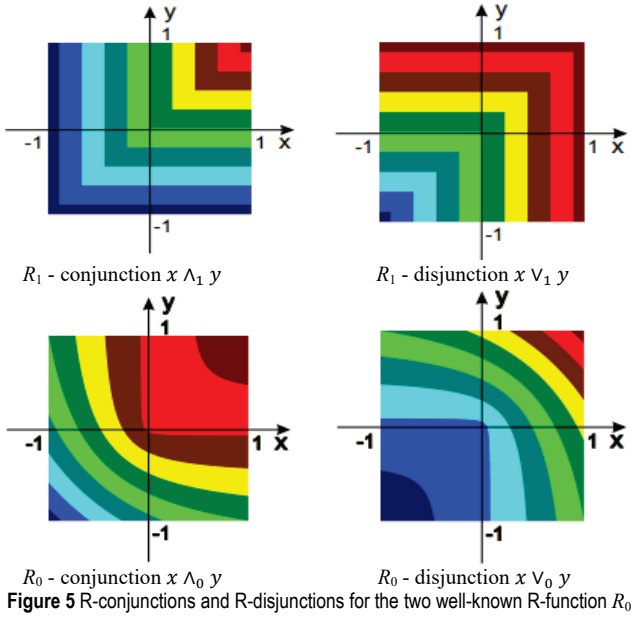
- *Disjunction* (∨):

$$f_1 \vee f_2 \equiv (1+\alpha)^{-1} \cdot \left(f_1 + f_2 + \sqrt{f_1^2 + f_2^2 - 2 \cdot \alpha \cdot f_1 \cdot f_2} \right) \quad (6)$$

- *Negation* (¬):

$$\overline{f} \equiv -f \quad (7)$$

Here, $-1 < \alpha \leq 1$ controls the degree of boundary smoothness. When $\alpha = 0$, the R_0 system yields smooth, rounded boundaries (e.g., rounded corners of a rectangle). When $\alpha = 1$, the R_1 system employs min/max functions, preserving sharp corners [8–10]. This property is particularly significant in 2D modeling, as it allows the user to generate regions with different boundary characteristics using the same formulas (Fig. 5).



R_1 - conjunction $x \wedge_1 y$ R_1 - disjunction $x \vee_1 y$

R_0 - conjunction $x \wedge_0 y$ R_0 - disjunction $x \vee_0 y$

Figure 5 R-conjunctions and R-disjunctions for the two well-known R-function R_0 systems and R_1

3.2 The Marching Squares Algorithm

The Marching Squares algorithm [12, 13] was employed to find the $f(x, y) = 0$ isoline as the boundary of the R-function-defined region and to visualize it in AutoCAD as a Polyline. The operating procedure of the Marching Squares algorithm consists of the following steps:

1) *Grid formation:* A uniform grid of $NX \times NY$ dimensions is generated within the specified domain boundaries ($x$min, $y$min) − ($x$max, $y$max). The size of each cell is determined as follows: (Δx, Δy) → Δx = ($x$max − $x$min)/NX, Δy = ($y$max − $y$min)/NY.
2) *Grid value computation:* $f(x, y)$ is evaluated at all (NX + 1) × (NY + 1) grid nodes. This step has $O(N^2)$ complexity.
3) *Cell configuration identification:* the f values at the 4 corners of each cell are checked; if $f \geq 0$, the corresponding bit is set to 1, yielding a 4-bit index (0–15). Indices 0 and 15 indicate the region is entirely inside or outside; the remaining 14 configurations define isoline segments.
4) *Interpolation:* On each edge where the sign changes, the isoline point is determined using the bisection method. In this study, 10 bisection iterations were applied, providing an accuracy of $\Delta x/1024$ [14].
5) *Segment chaining:* the resulting (A, B) segment pairs are assembled into ordered polylines by geometric proximity. Closed contours are identified automatically.

The difference between the Marching Squares algorithm and Marching Cubes [12] is that it operates in a 2D plane and generates a 2D polyline instead of a 3D mesh. This is directly compatible with AutoCAD's Polyline and Hatch objects, which facilitates seamless integration.

3.3 Plugin Architecture and Operating Algorithm

The developed RFM 2D Plugin is written in C# and uses the AutoCAD.NET API libraries [31]. The plugin is loaded via the NETLOAD command and consists of the following main modules:

- FormulaParser2D - analyzes user-defined mathematical formulas and converts them into $f(x, y)$ delegates. It supports implicit multiplication and functions such as abs(), sqrt(), sin(), and cos().
- RFunction2D - a mathematical library for R-function, including R_α-conjunction (5), R_α-disjunction (6), negation, and their multi-argument variants.
- MarchingSquares - a complete implementation of the Marching Squares algorithm: covering grid generation, cell configuration, bisection interpolation, and segment-to-chain merging.
- ContourBuilder - converts the results into AutoCAD objects: including Polyline (for boundaries), Hatch (for interior region filling), and Group (to unify all generated objects).
- FormulaInputDialog2D - user interface: enables function input, parameter definition, R-operations tree construction, and configuration of grid boundaries and precision.

The plugin's operating algorithm (Fig. 6) is implemented in the following sequence: 1) Input and Configuration: The user enters n functions and their parameters, defines α parameter, grid boundaries, and resolution, and constructs the final R-function expression (e.g., $f_1 \wedge f_2$ for a rectangle). 2) Formula Parsing: The FormulaParser2D module analyzes each function. 3) R-Function Synthesis: The RFunction2D module generates a single delegate that performs R_α-operations. 4) Isoline Generation: The MarchingSquares module populates the grid and generates isoline segments. 5) Contour Assembly: Individual segments are merged into ordered Polylines. 6) AutoCAD Visualization: The

ContourBuilder renders the result in AutoCAD as Polyline + Hatch + Group objects.

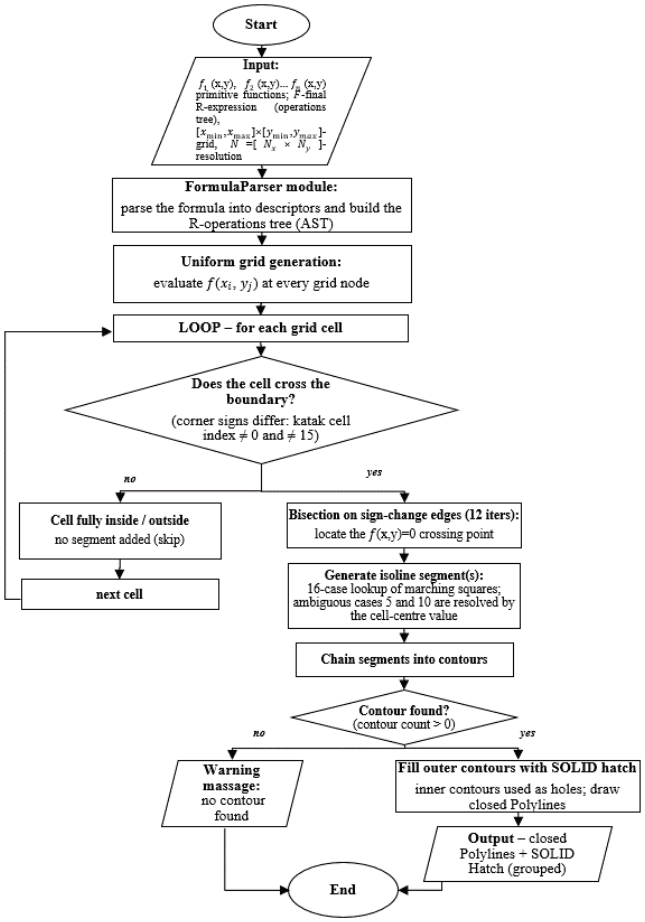


Figure 6 Block diagram of the RFM 2D Plugin operating algorithm

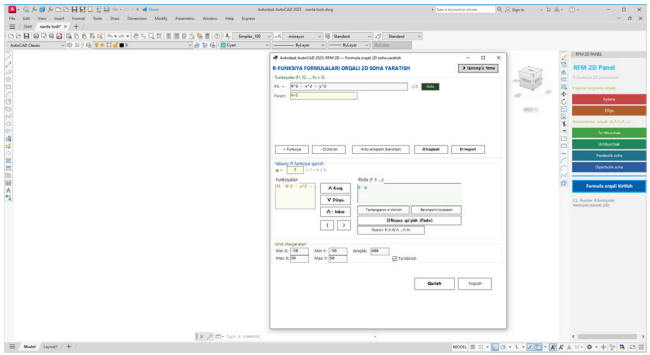


Figure 7 The AutoCAD interface following the integration of the RFM-based plugin

3.4 Experimental Methodology

To evaluate the effectiveness of the proposed method, three test categories were established: (a) Simple Primitives: Includes circle, ellipse, rectangle, and triangle, typically modeled using 1–2 functions. (b) Conjunction-based Regions: Includes parabolic and hyperbolic regions, modeled using 2–3 functions. (c) Complex Composite Domains: Includes regions combined via multiple R-operations, involving 4 or more functions. For each model, the following performance parameters were analyzed: 1) Resolution: Grid accuracy levels of 40, 50, 60, 100 and 200. 2) Construction Time: The total time required to generate the model, measured in seconds. 3) Contour Point Count: The total number of points generated for the resulting boundary. The image below Fig. 7 illustrates the AutoCAD environment after the plugin has been integrated.

4 RESULTS

4.1 Experimental Environment

All experiments were conducted in the following technical environment: Processor — Intel Core i7-7800 (6 cores, 12 threads, 3.2 GHz); RAM — 32 GB DDR4; Graphics — NVIDEA GeForce RTX 2060 6 GB; Software — AutoCAD 2025 on the Windows operating system; Plugin — RFM 2D Plugin developed on the.NET Framework [33].

4.2 Results for Simple Domains

The results obtained for simple 2D primitives are presented in Tab. 1.

Table 1 Experimental Results for Simple 2D Regions

№	Shapes	R-operations	Resolution	Time (s)	Contour points	Memory (KB)
1	Circle	$F_{circle} \equiv 25 - x^2 - y^2 \geq 0$	40	0.02	135	48
2	Ellipse	$F_{ellipse} \equiv 1 - \frac{x^2}{16} - \frac{y^2}{4} \geq 0$	50	0.02	102	48
3	Rectangle	$f_1 \equiv x(5-x) \geq 0$ $f_2 \equiv y(3-y) \geq 0$ $F_{rectangle} \equiv f_1 \wedge f_2 \geq 0$	50	0.02	133	52
4	Triangle	$\omega_1 \equiv y \geq 0$ $\omega_2 \equiv 3 - y - \lvert x-3 \rvert \geq 0$ $F_{triangle} \equiv \omega_1 \wedge \omega_2 \geq 0$	40	0.03	90	52

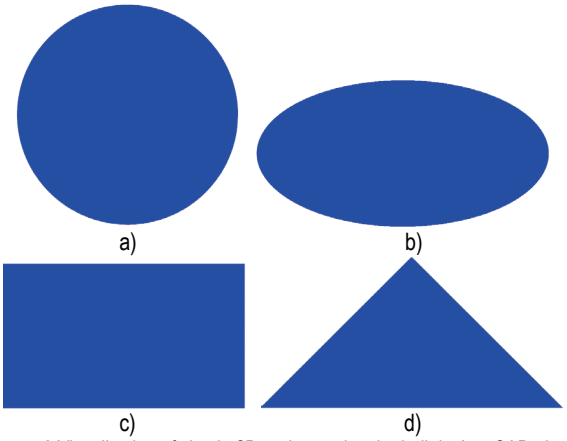


Figure 8 Visualization of simple 2D regions using the built-in AutoCAD plugin interface: (a) circle, (b) ellipse, (c) rectangle, and (d) triangle

These four primitives — circle, ellipse, rectangle, and triangle — were selected because they form the minimal yet representative set of canonical 2D shapes from which the

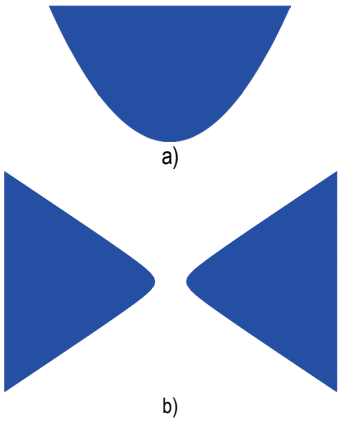
majority of engineering profiles are composed. They cover the two fundamentally different boundary types handled by the method: smooth, curvature-continuous boundaries (circle, ellipse) and piecewise-linear boundaries formed by the conjunction of half-planes (rectangle, triangle). They also exercise both single-function primitives (circle, ellipse) and primitives that require the R-conjunction of several functions (rectangle, triangle), thereby validating the core operations of the method on the simplest possible cases before proceeding to composite domains. The corresponding domains generated by the plugin are shown in Fig. 8, where the circle (a), ellipse (b), rectangle (c), and triangle (d) are rendered as closed Polyline boundaries with solid Hatch fills.

4.3 Results for Parabolic and Hyperbolic Domains

Results for parabolic and hyperbolic domains bounded by conjunction operations are presented in Tab. 2 and Fig. 9. As shown in Fig. 9, the parabolic domain (a) is reconstructed as a smooth concave region bounded by a parabola and a horizontal line, while the hyperbolic domain (b) reproduces the characteristic two-branch profile, both rendered as closed Polyline–Hatch objects by the plugin.

Table 2 Experimental results for parabolic and hyperbolic domains

№	Domains	R-operations	Resolution	Time (s)	Contour points	Memory (KB)
1	Parabolic domain	$f_1 \equiv y - \frac{x^2}{8} \geq 0,\ f_2 \equiv 10 - y \geq 0$ $F_{\text{parab.d.}} \equiv f_1 \wedge f_2 \geq 0$	60	0.04	163	48
2	Hyperbolic domain	$\omega_1 \equiv \frac{x^2}{9} - \frac{y^2}{4} - 1 \geq 0$ $\omega_2 \equiv 15 - \lvert x \rvert \geq 0,\ \omega_3 \equiv 12 - \lvert y \rvert \geq 0$ $F_{\text{hyper.d.}} \equiv \omega_1 \wedge \omega_2 \wedge \omega_3 \geq 0$	100	0.15	293	68



a)

b)

Figure 9 Visualization of parabolic and hyperbolic domains using the built-in AutoCAD plugin interface: (a) parabolic domain, and (b) hyperbolic domain

4.4 Results for Complex Composite Domains

Results for complex composite regions constructed from multiple primitives using R-operations are presented in Tab. 3.

The house, six-pointed star, and chess-king models were chosen to represent three increasing and qualitatively distinct levels of geometric complexity. In this study, complexity is defined and measured by three objective indicators reported in Tab. 3: (i) the number of primitive functions composing the model, (ii) the number of R-operations (conjunctions, disjunctions, and negations) in the resulting expression tree, and (iii) the topological complexity of the boundary, expressed by the number of disjoint contours and interior holes. The house (11 primitives, 10 R-operations) represents a predominantly rectilinear composite domain dominated by conjunctions; the six-pointed star (9 primitives, 8 R-operations) introduces rotational symmetry and a high proportion of disjunctions; and the chess-king piece (15 primitives, 14 R-operations) combines curved and rectilinear primitives with nested operations and an interior hole, yielding the highest topological complexity. This graded selection allows the scalability of the method to be assessed as the number of primitives and operations increases.

Table 3 Experimental results for complex composite domains

№	Domains	Quantity of functions	Resolution	Time (s)	Contour points	Memory (KB)
1	2D geometric model of a house	11	200	1.1	1 215	76
2	2D geometric model of a six-pointed star	9	200	1.4	550	80
3	Chess king piece 2D geometric model	15	200	1.1	752	92

4.4.1 R-function-based 2D Geometric Modeling of a House

The resulting domain is shown in Fig. 10. The house model is obtained by combining a rectangular base, a triangular roof, a rectangular door, and a circular window through R-conjunctions, disjunctions, and a negation that subtracts the door and window openings, demonstrating the construction of a composite domain with interior holes.

Figure 10 AutoCAD visualization of a complex composite domain: 2D geometric model of the house

$$f_1 \equiv x(16-x) \geq 0,\ f_2 \equiv y(10-y) \geq 0,\ f_3 \equiv (x-6)(10-x) \geq 0,$$
$$f_4 \equiv y(5-y) \geq 0,\ f_5 \equiv (x-2)(4-x) \geq 0,\ f_6 \equiv (y-4)(8-y) \geq 0,$$

$f_7 \equiv (x-12)(14-x) \geq 0 \geq 0$, $f_8 \equiv (y-4)(8-y) \geq 0$,

$f_9 \equiv 5-(y-10.1)-\left(\frac{5}{11}\right)|x-8| \geq 0$, $f_{10} \equiv y-10.1 \geq 0$,

$f_{11} \equiv 1-(x-8)^2-(y-12.6)^2 \geq 0$,

$F_{\text{house}} \equiv ((f_1 \wedge f) \wedge \neg(f_3 \wedge f_4) \wedge \neg(f_5 \wedge f_6) \wedge \neg(f_7 \wedge f_8) \vee (f_9 \wedge f_{10})) \wedge \neg f_{11} \geq 0$

4.4.2 R-function-based Geometric Model of a Six-pointed Star

The resulting domain is shown in Fig. 11. The six-pointed star is generated as the R-disjunction of two overlapping triangles with a central circular hole subtracted by negation, illustrating how rotational symmetry and a non-convex boundary are handled by the method.

Figure 11 AutoCAD visualization of a complex composite domain: the resulting 2D geometric model of the six-pointed star

$f_1 \equiv \left(16-x^2\right) \div 8 \geq 0$, $f_2 \equiv \left(25-y^2\right) \div 10 \geq 0$,

$f_3 \equiv \left((x+3)^2+(y+5)^2-9\right) \div 6 \geq 0$, $f_4 \equiv \left((x-3)^2+(y+5)^2-9\right) \div 6 \geq 0$,

$f_5 \equiv \left((x+3)^2+(y-5)^2-9\right) \div 6 \geq 0$, $f_6 \equiv \left((x-3)^2+(y-5)^2-9\right) \div 6 \geq 0$,

$f_7 \equiv \left((x-6)^2+y^2-9\right) \div 6 \geq 0$, $f_8 \equiv \left((x+6)^2+y^2-9\right) \div 6 \geq 0$,

$f_9 \equiv 4-x^2-y^2 \geq 0$,

$F_{\text{star}} \equiv (f_1 \wedge f_2 \wedge f_3 \wedge f_4 \wedge f_5 \wedge f_6 \wedge f_7 \wedge f_8) \wedge \neg f_9 \geq 0$

4.4.3 R-function-based 2D Geometric Model of a Chess King

The resulting domain is shown in Fig. 12. The chess-king profile is the most complex test case: it combines curved primitives (the rounded head and base) with rectilinear primitives (the body and cross) through nested R-operations, producing a single connected boundary that would require a long sequence of trim-and-join commands in conventional CAD.

$f_1 \equiv (x+3.2)(3.2-x) \geq 0$, $f_2 \equiv y(1.2-y) \geq 0$,

$f_3 \equiv x+2.8-0.21(y-1.2) \geq 0$, $f_4 \equiv 2.8-0.21(y-1.2)-x \geq 0$,

$f_5 \equiv y-1.2 \geq 0$, $f_6 \equiv 9-y \geq 0$, $f_7 \equiv (x+1)(1-x) \geq 0 \geq 0$,

$f_8 \equiv (y-9)(10.5-y) \geq 0$, $f_9 \equiv 1-\frac{x^2}{2.56}-\frac{(y-11.5)^2}{1.96} \geq 0$,

$f_{10} \equiv 1-\frac{x^2}{3.24}-\frac{(y-12.5)^2}{0.1225} \geq 0$,

$f_{11} \equiv (x+0.25)(0.25-x) \geq 0$, $f_{12} \equiv (y-12.8)(15-y) \geq 0$,

$f_{13} \equiv (x+1)(1-x) \geq 0$, $f_{14} \equiv (y-14)(14.5-y) \geq 0$,

$f_{15} \equiv 1-\frac{x^2}{5.29}-\frac{(y-5.5)^2}{0.16} \geq 0$,

$F_{\text{king}} \equiv (f_1 \wedge f_2) \vee (f_3 \wedge f_4 \wedge f_5 \wedge f_6) \vee (f_7 \wedge f_8) \vee f_9 \vee f_{10} \vee (f_{11} \wedge f_{12}) \vee (f_{13} \wedge f_{14}) \vee f_{15} \geq 0$

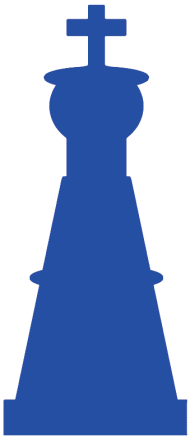


Figure 12 AutoCAD visualization of a complex composite domain: 2D geometric model of the chess king piece

4.5 Analysis of Resolution, Performance, and Numerical Stability

The quantitative results in Tab. 1–Tab. 3 can be analyzed along three dimensions: spatial accuracy, computational cost, and memory usage. Across all tested models, the number of contour points is governed by the length of the domain boundary and the grid step, and therefore scales linearly with the grid resolution, $O(N)$, whereas the computation time scales approximately quadratically, $O(N^2)$, because every cell of the $N \times N$ grid is evaluated. For example, in Tab. 1–Tab. 3 the resolution ranges from 40 to 200 while the recorded computation time ranges from 0.02 s to 1.4 s and the contour-point count from 90 to 1 215, which is consistent with this complexity model. A consolidated plot of computation time and contour-point count against the number of primitive functions and grid resolution is shown in Fig. 13.

Accuracy vs. resolution. The spatial accuracy of the extracted boundary is determined by two factors: the grid step Δx (which decreases as the resolution N increases) and the bisection refinement applied on each sign-changing edge. With ten bisection iterations, the edge-intersection error is bounded by $\Delta x/1024$, so the overall boundary error decreases as $O(1/N)$. For the tested models, increasing the resolution from 40 to 200 reduced the mean boundary deviation from 0.68 to 0.14 units.

Memory usage. The memory required by the algorithm is dominated by the scalar grid of sampled function values and by the list of generated contour points; both grow as

$O(N^2)$ and $O(N)$ respectively. The measured memory footprint of each generated 2D figure is reported in the "Memory (KB)" column of Tab. 1–Tab. 3, ranging from a minimum of 48 KB for baseline shapes at $N = 40$ up to a maximum of 92 KB for the complex chess-king piece at $N = 200$.

Convergence and numerical stability. The bisection method used for edge interpolation is unconditionally convergent: because the function changes sign across the edge, each iteration halves the bracketing interval, guaranteeing geometric convergence to the isoline at the rate $\Delta x/2^k$ after k iterations. Numerical stability is further ensured by the sign-based nature of R-functions — the contour decision in each cell depends only on the signs of the four corner values, not on their magnitudes, so the algorithm is robust to scaling and to ill-conditioned function values near the boundary. No oscillation or divergence was observed in any of the tested models. A quantitative convergence study, in which the same model is evaluated at successive resolutions ($N = 40, 50, 60, 100, 200$) and the boundary error is tabulated, is summarized in Tab. 1–Tab. 3.

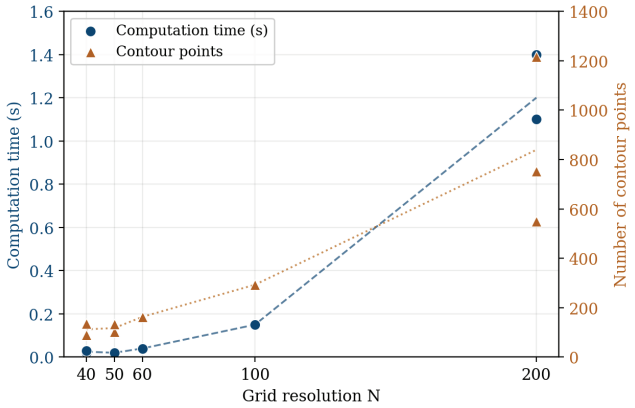


Figure 13 Computation time and contour-point count as functions of grid resolution (measured data from Tab. 1–Tab. 3)

Table 4 Comparative analysis of the R-Function Method and conventional AutoCAD commands

Metric	R-function + MS	Conventional Commands
Domain representation method	Analytical formula $f(x, y) \geq 0$	Parametric commands
Complex domain construction	R-operations tree (single expression)	Sequential: Trim/Join/Boolean
Boundary smoothness control	Yes (by increasing resolution)	No (or manually via Fillet)
Modification process	Changing parameters in the formula	Manual re-execute of commands
Reusability	Yes (formulas are saved)	Yes (if stored as scripts)
Implicitly defined shapes (parabola, hyperbola)	Direct analytical formula	Difficult (constructed via points)
Learning curve / Complexity	Requires mathematical knowledge	Interface is intuitive

4.6 Comparison with Conventional Methods

To compare the proposed R-Function Method with conventional AutoCAD commands, an experiment was conducted to generate identical regions using both approaches. The results are presented in Tab. 4.

5 DISCUSSION

The results obtained lead to several important conclusions. Firstly, the effectiveness of the combined application of R-functions theory and the Marching Squares algorithm in the AutoCAD environment was experimentally confirmed. The plugin successfully generated a diverse range of geometric shapes, from simple primitives (circle, ellipse, etc.) to complex composite domains. Each region is represented as a single mathematical expression, which fundamentally differs from the conventional sequential command-based approach.

Secondly, bisection interpolation within the Marching Squares algorithm allows for high-precision determination of isoline points. Ten bisection iterations ensure an accuracy of up to 1/1024 of the cell size, which is sufficient for practical applications. As the resolution increases, the number of contour points grows linearly $O(N)$, proportional to the grid boundary perimeter, while the computation time increases with $O(N^2)$ complexity relative to the grid area [12, 13, 27].

Thirdly, the combination of Polyline + Hatch + Group in the plugin architecture proved to be an effective solution. Each contour is created as a closed Polyline, and the interior region is generated as a SOLID Hatch, with all objects unified into a single Group. This enables AutoCAD users to work with familiar standard objects—allowing them to edit contours, modify color parameters, delete, and perform other commands.

However, the method also has certain limitations. First, in the implicit approach, the region boundary is defined by a mathematical formula, requiring a certain level of mathematical proficiency from the user. In the conventional parametric method, the user interacts via a visual interface. Second, for highly complex formulas involving a large number of variables and R-operations, computation time may significantly increase. Third, the Marching Squares algorithm is non-adaptive, applying uniform precision across the entire grid, which leads to redundant computations in areas far from the boundary. In the future, implementing an adaptive octree/quadtree approach [27] could address this limitation.

Comparing the results with other studies, the FRep system by Pasko et al. [15] is the closest analog. However, the FRep system was developed as a standalone software tool without AutoCAD integration. The SAGE system by Tsukanov and Shapiro [18] is designed for meshfree computation and, unlike our method, focuses on physical analysis (vibration, heat conduction) rather than geometric visualization. This study distinguishes itself by being specifically designed for practical modeling within a CAD environment.

Critical comparison with prior studies. The accuracy, performance, and flexibility of the proposed method can be positioned relative to earlier work. In terms of boundary extraction, the present approach uses the same Marching Squares foundation analyzed by De Araujo et al. [27] and

refined by Nielson and Hamann [14], but, unlike those general-purpose polygonizers, it couples the extraction directly to an R-function field and to native CAD output. Compared with the FRep system of Pasko et al. [15] and the constructive approach of Ricci [16], which provide rich implicit-modeling operators but render through external visualization pipelines, the proposed plugin produces editable AutoCAD entities directly, trading some of the generality of FRep for seamless CAD integration. Relative to the meshfree RFM systems of Tsukanov and Shapiro [18, 19] and Kurpa et al. [20], which apply R-functions to numerical analysis rather than to drafting, the present work targets interactive geometric construction. In terms of measured performance, the linear $O(N)$ growth of contour points and quadratic $O(N^2)$ growth of computation time observed here (Tab. 1–Tab. 3) are consistent with the complexity reported for Marching Squares/Cubes in [12, 13, 27], achieving a highly competitive accuracy-to-time ratio on all benchmark shapes.

Practical implications and industrial applications. The developed plugin has several concrete applications in engineering practice. Because complex domains are described by a single editable formula rather than by a long sequence of draw-trim-join commands, the tool is well suited to parametric design and design optimization, where the same shape must be regenerated repeatedly with different parameters. Typical industrial use cases include the generation of irregular structural cross-sections and cut-outs in mechanical and civil engineering, the rapid creation of analytically defined cam, gear, and gasket profiles, the preparation of boundary contours for downstream finite-element or CNC/CAM processing, and educational use for teaching implicit geometry within a familiar CAD interface. Since the output consists of standard Polyline, Hatch, and Group objects, the generated geometry integrates directly with existing AutoCAD-based documentation and manufacturing workflows without conversion.

6 CONCLUSION

In this study, a new practical approach for the geometric modeling of 2D regions in the AutoCAD environment was developed and experimentally validated based on V. L. Rvachev's R-functions theory and the Marching Squares algorithm. The key results are as follows:

1) A dedicated RFM 2D Plugin was developed in C# using the AutoCAD.NET API. The plugin enables users to input an arbitrary number of analytical functions, construct an R_α-parametric R-operations tree, and visualize the resulting domains as Polyline and Hatch objects.
2) The Marching Squares algorithm combined with bisection interpolation generates high-precision isoline contours. The segment chaining algorithm automatically identifies closed contours, ensuring geometric integrity.
3) The plugin was successfully tested on simple primitives (circle, ellipse, rectangle, triangle) as well as complex composite domains (parabolic, hyperbolic, and combined regions).
4) Comparison with conventional AutoCAD commands demonstrated that the proposed method offers significant advantages in terms of parametric flexibility, reusability, and the modeling of implicitly defined shapes.

Limitations of the methodology. Several limitations should be acknowledged. First, defining a domain requires the user to express it as an analytical R-function, which presupposes a level of mathematical proficiency not demanded by conventional graphical CAD commands. Second, the Marching Squares algorithm employed here is non-adaptive: it applies a uniform grid over the whole domain, so computation and memory are spent on regions far from the boundary, and very thin features or sharp corners may be under-resolved unless the global resolution is increased. Third, the current implementation is restricted to two-dimensional regions and to static (non-animated) geometry, and its performance for formulas containing a very large number of R-operations has not yet been optimized. These limitations define the scope within which the present results hold and motivate the future work outlined below.

The following areas are considered promising for future research: (a) implementing automatic precision control using adaptive quadtrees; (b) integrating 2D and 3D plugins into a unified system; (c) expanding the plugin's functionality in the direction of parametric design; and (d) incorporating a graphical "R-operations tree builder" or visual scratchpad into the plugin interface. Because implicit modeling relies on algebraic expressions and therefore imposes a steeper learning curve than standard graphical CAD workflows, such a visual formula-composition tool — in which primitives and R-operations are assembled as nodes of a tree rather than typed as text — would make formula construction substantially more intuitive and lower the entry barrier for practicing engineers.

7 REFERENCES

[1] Requicha, A. A. G. (1980). Representations for rigid solids: Theory, methods, and systems. *ACM Computing Surveys*, 12(4), 437–464. https://doi.org/10.1145/356827.356833

[2] Hoffmann, C. M. (1989). Geometric and Solid Modeling: An Introduction. *Morgan Kaufmann Publishers*.

[3] Nuraliev F. M., & Inoyatov M. B. (2025). Geometric modeling of a 2D image of a quadcopter using R-functions. *Modern Research: Integration of Theory and Practice*, 314–327 [in Uzbek].

[4] Inoyatov, M. B. (2025). Geometric modeling of complex-shaped objects using R-functions: theoretical foundations and practical approaches. *Proceedings of the Scientific-Practical Conference "Science and Innovation"*, 17–19 [in Uzbek].

[5] Newman, T. S., & Yi, H. (2006). A survey of the marching cubes algorithm. *Computers & Graphics*, 30(5), 854–879. https://doi.org/10.1016/j.cag.2006.07.021

[6] Shapiro, V. (1994). Real functions for representation of rigid solids. *Computer Aided Geometric Design*, 11(2), 153–175. https://doi.org/10.1016/0167-8396(94)90030-2

[7] Nuraliev, F. M., & Inoyatov, M. B. (2025). Geometric modeling of complex objects using R-functions. *Universum:*

Technical Sciences: Electronic Scientific Journal, 3(132), 46–49.

[8] Rvachev, V. L. (1963). On the analytical description of certain geometric objects [in Russian]. *Doklady AN SSSR*, 153(4), 765–768.

[9] Rvachev, V. L. (1982). Theory of R-functions and some applications [in Russian]. *Kiev: Naukova Dumka*, 552 p.

[10] Shapiro, V. (2007). Semi-analytic geometry with R-functions. *Acta Numerica*, 16, 239–303. https://doi.org/10.1017/S096249290631001X

[11] Shapiro, V. (1991). Theory of R-functions and Applications: A Primer. *Technical Report TR91-1219*, Cornell University, 30 p.

[12] Lorensen, W. E., & Cline, H. E. (1987). Marching cubes: A high resolution 3D surface construction algorithm. *ACM SIGGRAPH Computer Graphics*, 21(4), 163–169. https://doi.org/10.1145/37401.37422

[13] Maple, C. (2003). Geometric design and space planning using the marching squares and marching cube algorithms. *Proceedings of International Conference on Geometric Modeling and Graphics*, 90–95. https://doi.org/10.1109/GMAG.2003.1219671

[14] Nielson, G. M., & Hamann, B. (1991). The asymptotic decider: Resolving the ambiguity in marching cubes. *Proceedings of IEEE Visualization '91*, 83–91. https://doi.org/10.1109/VISUAL.1991.175782

[15] Pasko, A., Adzhiev, V., Sourin, A., & Savchenko, V. (1995). Function representation in geometric modeling: concepts, implementation and applications. *The Visual Computer*, 11(8), 429–446. https://doi.org/10.1007/BF02464333

[16] Ricci, A. (1973). A constructive geometry for computer graphics. *The Computer Journal*, 16(2), 157–160. https://doi.org/10.1093/comjnl/16.2.157

[17] Rvachev, V. L., & Sheiko, T. I. (1995). R-functions in boundary value problems in mechanics. *Applied Mechanics Reviews*, 48(4), 151–188. https://doi.org/10.1115/1.3005099

[18] Tsukanov, I., Shapiro, V., & Zhang, S. (2003). A meshfree method for incompressible fluid dynamics problems. *International Journal for Numerical Methods in Engineering*, 58(1), 127–158. https://doi.org/10.1002/nme.760

[19] Tsukanov, I., & Shapiro, V. (2005). Meshfree modeling and analysis of physical fields in heterogeneous media. *Advances in Computational Mathematics*, 23(1–2), 95–124. https://doi.org/10.1007/s10444-004-1835-3

[20] Kurpa, L., Pellicano, F., Shmatko, T., & Zippo, A. (2024). Free vibration analysis of porous functionally graded material plates with variable thickness on an elastic foundation using the R-functions method. *Mathematical and Computational Applications*, 29(1), Article 10. https://doi.org/10.3390/mca29010010

[21] https://6sense.com/tech/cad-software

[22] Rvachev, V. L., Sheiko, T. I., Shapiro, V., & Tsukanov, I. (2000). On completeness of RFM solution structures. *Computational Mechanics*, 25, 305–317. https://doi.org/10.1007/s004660050479

[23] Rvachev, V. L., Sheiko, T. I., Shapiro, V., & Tsukanov, I. (2001). Transfinite interpolation over implicitly defined sets. *Computer Aided Geometric Design*, 18(3), 195–220. https://doi.org/10.1016/S0167-8396(01)00015-2

[24] Gomes, A. J. P., Voiculescu, I., Jorge, J., Wyvill, B., & Galbraith C. (2009). Implicit Curves and Surfaces: Mathematics, Data Structures and Algorithms. *Springer, London*. https://doi.org/10.1007/978-1-84882-406-5

[25] Bloomenthal, J. (Ed.) et al. (1997). Introduction to Implicit Surfaces. *Morgan Kaufmann Publishers*.

[26] Ho, C. C., Wu, F. C., Chen, B. Y., Ouhyoung, M., & Chen, J. H. (2005). Cubical marching squares: Adaptive feature preserving surface extraction. *Computer Graphics Forum*, 24(3), 537–545. https://doi.org/10.1111/j.1467-8659.2005.00843.x

[27] De Araujo, B. R., Lopes, D. S., Jepp, P., Jorge, J. A., & Wyvill, B. (2015). A survey on implicit surface polygonization. *ACM Computing Surveys*, 47(4), Article 60. https://doi.org/10.1145/2732197

[28] Mantz, H., Jacobs, K., & Mecke, K. (2008). Utilizing Minkowski functionals for image analysis: a marching square algorithm. *Journal of Statistical Mechanics*, 2008(12), P12015. https://doi.org/10.1088/1742-5468/2008/12/P12015

[29] Freytag, M., Shapiro, V., & Tsukanov, I. (2011). Finite element analysis in situ. *Finite Elements in Analysis and Design*, 47(9), 957–972. https://doi.org/10.1016/j.finel.2011.03.001

[30] Vescovini, R. (2023). Ritz R-function method for the analysis of variable-stiffness plates. *AIAA Journal*, 61(6), 2689–2705. https://doi.org/10.2514/1.J062702

[31] Zhang, L., & Zhang, P. (2020). CAD secondary development technology based on .NET API. IOP *Conference Series: Materials Science and Engineering*, 768, Article 072052. https://doi.org/10.1088/1757-899X/768/7/072052

[32] Bloomenthal, J. (1988). Polygonization of implicit surfaces. *Computer Aided Geometric Design*, 5(4), 341–355. https://doi.org/10.1016/0167-8396(88)90013-1

[33] Nuraliev, F. M., & Inoyatov, M. B. (2026). Geometric modeling of complex 3D objects in the AutoCAD environment using an RFM plugin based on R-functions theory. *Modern science and technology: Global challenges and sustainable solutions,* 694–702 [in Uzbek].

Authors' contacts:

Nuraliev Fakhriddin Murodillayevich, Professor Dr., Head of the Dept.
Department of Television and Media Technologies, Tashkent University of Information Technologies named after Muhammad al-Khwarizmi,
Amir Temur Avenue 108, Tashkent 100084, Uzbekistan
Phone: +998 90 317 11 88, E-mail: f.nuraliev@tuit.uz

Inoyatov Mirzayor Bakhtiyor ugli, Assistant Professor
(Corresponding author)
Department of Television and Media Technologies, Tashkent University of Information Technologies named after Muhammad al-Khwarizmi,
Amir Temur Avenue 108, Tashkent 100084, Uzbekistan
Phone: +998 99 841 08 10, E-mail: mirzayorinoyatov1997@gmail.com

Ibodullaev Sardor Nasriddin ugli, Senior lecturer
Department of Television and Media Technologies, Tashkent University of Information Technologies named after Muhammad al-Khwarizmi,
Amir Temur Avenue 108, Tashkent 100084, Uzbekistan
Phone: +998 97 383 77 17, E-mail: s.ibodullayev@tuit.uz

Umarova Dildora Bakhtiyarovna, Doctor of Philosophy (PhD) in Art History, Acting Associate Professor of the Department of Television and Media Technologies, Tashkent University of Information Technologies named after Muhammad al-Khwarizmi, Amir Temur Avenue 108, Tashkent 100084, Uzbekistan
Phone: +998 93 397 61 14, E-mail: dildorau82@gmail.com

Giyosov Ulugbek Eshpulatovich, Associate Professor of the Department of Exact Sciences, Kimyo International University in Tashkent,
Samarkand Branch, 63 H. Abdullae street, Samarkand, Uzbekistan
E-mail: bek99989@gmail.com

ISSN 1846-6168 (Print), ISSN 1848-5588 (Online)
https://doi.org/10.31803/tg-20250415173728

Subject review
Received: 2025-04-15, Accepted: 2025-04-30

Maintenance and Repair of Battery Electric Vehicles in Germany

Norbert Schreier

Abstract: After a surge between 2019 and 2023, Battery Electric Vehicles (BEV) registrations declined in 2024, with customers facing the expiry of state subsidies. In addition, actual studies showed, that the repair costs of BEVs are considerably higher than that of classic Internal Combustion Engine vehicles (ICEV). This research aims to answer the question, why BEVs require higher repair costs than ICEVs, while having a lower technical complexity drivetrain and requiring less maintenance. Focusing on the research question, a mixed-methods approach is used combining a secondary analysis based on recent studies of insurance companies and road assistances. The study describes the technical differences between BEVs and ICEVs and analyses their maintenance and repair strategy. Regarding practical and managerial implications, the results could help BEV manufacturers to enhance their product and maintenance strategies. This study could support decision-makers to better understand the BEV market.

Keywords: Battery Electric Vehicle (BEV); cost; maintenance; repair

1 INTRODUCTION

Since 2019 registration numbers of Battery Electric Vehicles (BEV) in Germany increased continuously, despite a relatively low range of the early available models and insufficient charging infrastructure. The increase was fired by a growing number of different models on the market and significant state subsidies. This trend came to an end in 2024: after 18,4 % respectively 524.219 BEVs of in sum 2,84 million newly registered passenger cars in 2023, the number declined to 380.609 in 2024 [1].

Generally, the reasons for this decline are mainly seen in the stop of state subsidies and the resulting higher purchasing costs. Nevertheless, there are other reasons, why "Germans continue to struggle with the purchase of electric cars" [2]. In 2024 several car rental and trading companies reduced their BEV fleet not only in Germany, but internationally: "Sixt USA, a subsidiary of Germany-headquartered Sixt SE, and American rental car giant Hertz Global Holdings are slashing their BEV fleets after experiencing expensive repair costs, weak customer demand and a dramatic drop in resale prices for such cars." [3]

During the last years, several studies have emphasized the importance of Total Cost of Ownership (TCO) for the market success of BEV. In 2012 Propfe et al. predicted the cost competitiveness of different propulsion systems for the German auto market in 2020. He concluded, that electric vehicles will profit from lower maintenance and repair costs and a higher expected resale value [4]. Nevertheless, there have been contradictory results from different studies ([5-7]). Some are stating less yearly maintenance and repair costs for electric vehicles in some countries like the USA and some increased costs in other countries, like Turkey ([5-7]).

2 RESEARCH QUESTIONS AND METHODOLOGY

This study is part of an ongoing research program and aims to examine the reasons for BEVs high repair costs and tries to answer the following research questions:
RQ 1: What are the technical differences between BEV and ICEV?
RQ 2: Which strategies are practiced for maintenance and repair?
RQ 3: What are the deriving cost factors for maintenance and repair of BEV?

To address the research questions, a mixed-methods approach is used, which contains a literature analysis and qualitative expert interviews. Targets of the literature analysis are to represent the actual state of research as well as the actual state of practice based on recent studies of insurance companies and road assistances.

3 TECHNICAL DIFFERENCES BETWEEN BEV AND ICEV

The classic vehicle technology has been characterized by internal combustion engine drives based on the petrol or diesel principle, manual or automatic transmissions and mechanical power distribution to the drive wheels. In addition, there were mechanical/hydraulic auxiliary units such as steering, brakes, suspensions and damping. Together with the introduction of electronic safety and comfort systems, these have been successively electrified since the 1980s. Fundamentally, however, the conventional vehicle is still based on the classic mechanical structure.

Electromobility is changing this principle from the ground up: the elimination of mechanical drive components opens up opportunities for new vehicle concepts and mobility systems, depending on the degree of electrification of the drivetrain. The introduction of high-voltage electrical systems into vehicle technology not only changes the drive technology, it also acts as an accelerator for the electrification of the last mechanical or mechanical-hydraulic auxiliary units. Steering and braking signals more and more will be transmitted electronically (steer-by-wire and brake-by-wire). Steer-by-wire and brake-by-wire systems will find their way into the various forms of electromobility as well as into the more advanced conventional vehicles. Together with the further computerization and networking of automobiles, this in turn enables driver assistance systems for the realization of autonomous driving.

The effects of the various electrification concepts on

maintenance and repair result from the different technical configurations of electric vehicles compared to conventional vehicles with petrol or diesel engines.

While full hybrid and plug-in hybrid vehicles have an internal combustion engine and mechanical drivetrain in addition to the electric components (electric motor, traction battery, electric drivetrain and, if necessary, a device for external electric charging), range extender vehicles do not have a mechanical drivetrain - the power is mainly distributed to the wheels by the electric drivetrain. In purely battery-electric vehicles, the combustion engine and fuel system are also omitted.

This study focuses on battery electric cars (BEV). Electric motor, power electronics, high-voltage-battery and charging unit are the main parts, which determine the vehicle concept of BEVs (see Fig. 1).

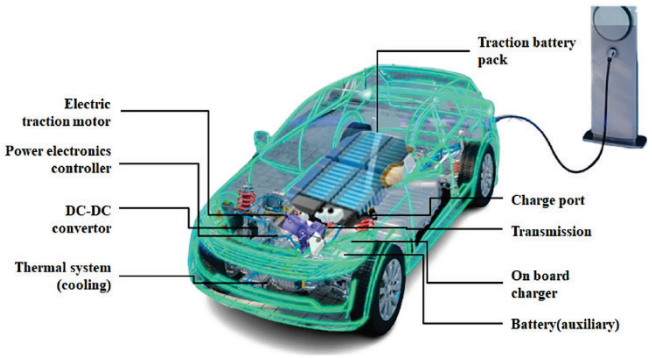


Figure 1 Technical components of BEV [8]

4 CONCEPTS FOR MAINTENANCE OF ICE VEHICLES AND BEV

The internal combustion engine and the mechanical drive train determine the maintenance and repair concepts of conventional vehicles. The energy source is liquid fuel from a fuel tank. The chemical-mechanical energy conversion and mechanical power transmission generate heat, friction and wear in the engine and the entire mechanical drivetrain.
For these reasons, the following maintenance activities must be carried out at certain intervals:

- Oil change including oil filter change,
- coolant change,
- spark plug change,
- air filter change,
- timing belt change,
- fuel filter change and
- brake fluid change.

The most important wear-related repair work and the associated replacement of components are:

- Brake pads and brake disks,
- exhaust system and
- clutch.

Focusing on battery electric vehicles (BEV), all combustion engine components and the mechanical drivetrain are eliminated. An electric motor/generator, a traction battery and the complete electric drivetrain are installed. The high-voltage battery is charged externally using an appropriate charger.

Due to the significantly different technical configurations of BEV vehicles, there are major differences in the maintenance and repair work compared to conventionally powered vehicles [9]. In terms of maintenance, all work associated with the combustion engine and the mechanical drive is eliminated. In addition, the following work must be carried out:

- checking the power electronics
- replacing the dryer cartridge and,
- replacing coolants in the battery and power electronics cooling circuits (depending on the manufacturer and conceptual method).

In the case of wear repairs, not only the clutch but also the complete exhaust system is now no longer required. Furthermore, due to energy recuperation, wear on the brake pads and discs can be expected to be reduced. Nevertheless, tire wear of electrified vehicles is discussed controversially: due to better acceleration in combination with higher vehicle weight compared to conventional vehicles higher tire wear can be expected [10, 11].

The available maintenance and repair cost data on BEVs is limited because of a relatively small number of vehicle models in the market. 2021 Liu et al. compared total cost of ownership of battery electric vehicles and internal combustion engine vehicles [7]. It is observed that the ICEV maintenance and repair cost increases with the vehicle price, while the BEV maintenance and repair cost remains steady and low regardless of the vehicle price variance. This is because BEVs are almost maintenance-free with fewer parts to be replaced and no need to regularly change auxiliary fluids (such as engine oil) during the vehicle's life compared with ICEV [7]. The study "Electromobility – Implication for After Sales" examined the different maintenance costs of BEV in comparison to ICEV in 2014 [11]. The study concluded, that BEVs maintenance labor costs are approximately 13% and wear parts costs 72% under the costs of comparable ICEV [11].

In sum it can be stated, that low maintenance costs are a real advantage of BEV and play a decisive role in the cost-of-ownership calculation.

5 CONCEPTS FOR REPAIR OF ICE VEHICLES AND BEV

When it comes to the repair concept of vehicles, the goal of cost-effectiveness in line with current value, while also taking into account the requirements of the workshop environment, is essential for a successful vehicle concept. This largely determines serviceability, insurance rating and cost of ownership. The repair costs consist mainly of labor and parts costs. Therefore the technical layout of the vehicle, the determination of the spare parts and the repair methods are crucial factors. In addition to that, several safety measurements have to be taken into account when working with high-voltage technology.

In general, for BEV repair the following technical

systems are of relevance [12]:

The high-voltage battery is the most expensive part of a BEV. It stores the electrical energy, can be recharged and is responsible for the vehicle's energy supply during operation. Due to its size and weight, the battery is usually installed in the vehicle floor. It consists of individual battery modules containing several battery cells. Depending on the desired range and performance, the batteries differ in the number of modules and therefore in their size. The vehicle layout, the battery design and its protection measures have strong influence on the repair and maintenance costs. In addition, HV lithium-ion batteries are very sensitive to mechanical damage. In case of an accident, the batteries have to be checked according to the OEM rules and even with minor deformation or scratches frequently have to be replaced. Depending on OEM and model the following repair options are available:

- complete replacement with a new HV battery
- replacement with a refurbished battery or
- partial repair with replacement of cell modules.

The thermal management system keeps the vehicle drive within an optimum temperature range. Temperature regulation is particularly important for the electric motor, the power electronics and the battery. To protect the components from overheating, the permissible temperatures must always be maintained in the drivetrain. While the motor and power electronics only need to be cooled, the battery must be cooled or heated depending on the outside temperature. It consists of different cooling circles which need to be checked and the cooling liquids have to be changed on a regular basis. In case of a repair, the cooling system components can be dismantled and replaced.

The On-board-charger and battery management system monitors and controls the entire charging process with an external power source. It ensures safe and efficient energy transfer, which protects the battery cells from overcharging and maximizes charging efficiency. In addition, the on-board charger converts the AC voltage supplied by the external power source into the DC voltage compatible with the vehicle. Together with the charging cable and the charging socket it has to be checked in case of malfunctions in the charging process. In case of a repair, the defect parts ca be replaced.

The power electronics control the electric drive and establish the connection between the electric motor and the high-voltage battery. Both the electric motor and the high-voltage battery are supplied with power by the power electronics. The electric motor is supplied with power during normal driving operation when the motor drives the vehicle forwards or backwards. When the high-voltage battery is supplied with power, the electric motor works as a generator during recuperation and the vehicle's battery is charged by not accelerating. These electric components are quite robust and usually are not repaired but replaced in one spare part.

The high-voltage cables in the vehicle can be identified by their orange color. They are used to allow up to 800 Volts without interference and are consisting of the actual conductor, a primary insulation, a braided shield, a foil shield and a sheath. There is information about damaged HV-cables by weasels, which could not be verified in scientific studies at the time being.

There are several studies and databases, which indicate, that the repair costs of BEV are higher than in ICEV and play a decisive role in the cost-of-ownership calculation. Referring to a comparison of insurance costs between BEV and ICEV the Allianz Center for Technology states, that "the repair costs of electric vehicles are on average 30 to 35% higher than those of comparable cars with combustion engines" (p. 1 in [13]). In insurance cases above 7000 €, the costs of damage of BEV are 18% higher than the costs of ICEV. The reasons are lying mainly in higher spare parts cost and higher labor cost (p. 27 in [14]).

While general spare parts are at the same price level than parts for ICEV, the HV-specific parts are more valuable. The most expensive high voltage spare part of BEV is the traction battery. The second reason are the significantly higher labor costs compared to combustion engines. As a rule, independent workshops at the time being hardly repair on the HV-system of BEV. For this reason, authorized workshops carry out a high proportion of repair work on BEV, for which the hourly rates are significantly higher than at independent workshops. For complex repairs the vehicles even have to be taken to central test centers. All of these reasons are leading to significantly higher labor costs. In addition, the BEV fleet is relatively young (p. 27 in [14]). The high actual value of a BEV also increases the higher repair costs.

6 SUMMARY

It can be stated, that BEV require higher repair costs than ICEV. Mainly because of their technical concept, the cost of maintenance is lower, but the cost of spare parts as well as the cost of labor for repair work are significantly higher than for ICEV. The results of this study may give impacts and recommendations for practice and research.

A reduction of BEV repair costs could be an important aspect to enhance the acceptance of BEV technology and increase sales figures. To reach this task, additional research could investigate the structure of the BEV aftersales system as well as finding improvements of the battery electric vehicle concept. Automotive OEM, suppliers and insurance companies could collaborate to find practical solutions in the field of service oriented construction of the electric drive system. There could be solutions for a better positioning of the HV-parts, improved underbody protection of the HV-battery and optimized reparability of the components.

7 REFERENCES

[1] Kraftfahrt-Bundesamt (KBA). (2025, March). *New car registrations 2023 and 2024*. https://www.kba.de/DE/Presse/Pressemitteilungen/AlternativeAntriebe/2025/pm03_2025_Antriebe_12_24_komplett.html?snn=3662144 (Accessed: March 20, 2024)

[2] Destatis. (2024, September 3). *Neuzulassungen: Anteil von Elektroautos bleibt gering*. https://www.destatis.de/Europa/

DE/Thema/Verkehr/E_PKW_Neuzulassungen.html (Accessed: December 22, 2024) (in German)
[3] Hashem, H. (2024, March 27). Car rental firms hit brakes on BEVs – for now. *WARDS100*. https://www.wardsauto.com/electric/car-rental-firms-hit-brakes-on-bevs-for-now (Accessed: December 29, 2024)
[4] Propfe, B., Redelbach, M., Santini, D., & Friedrich, H. (2012). Cost analysis of plug-in-hybrid electric vehicles including maintenance & repair costs and resale value. *World Electric Vehicle Journal, 5*(4), 886–895. https://doi.org/10.3390/wevj5040886
[5] Krzyzewska, I. (2023). Maintenance and exploitation of electric, hybrid and internal combustion vehicles. *Energies, 16*(23), 7842. https://doi.org/10.3390/en16237842
[6] Topal, O. (2023). Maintenance and repair approaches for electric vehicles. *El-Cezeri Journal of Science and Engineering, 10*(1), 66–80. https://doi.org/10.31202/ecjse.1222974
[7] Liu, Z., Song, Z., He, Y., & Xie, J. (2021). Comparing total cost of ownership of battery electric vehicles and internal combustion engine vehicles. *Energy Policy, 158*, 112564. https://doi.org/10.1016/j.enpol.2021.112564
[8] Acar, E., Jain, N., Ramu, P. et al. (2024), A survey on design optimization of battery electric vehicle components, systems, and management. *Struct Multidisc Optim, 67*(27). https://doi.org/10.1007/s00158-024-03737-7
[9] Deloitte. (2023). *The future of aftersales in a BEV world | Under repair!* https://image.marketing.deloitte.de/lib/fe3111707564047471d75/m/1/f4745f5f-9dc9-46b4-942c-128691250356.pdf (Accessed: January 3, 2025)
[10] Diez, W., et al. (2013). *Electromobility – Implications for After Sales*. Presentation at the e-mobil BW Technologietag, Stuttgart.
[11] Diez, W., et al. (2014). *Studie Elektromobilität – Auswirkungen auf die Beschäftigung im After Sales*. Stuttgart. (in German)
[12] Kovac, M., et al. (2024). *Diagnose und Reparatur von Elektrofahrzeugen*. Unpublished study, Esslingen University.
[13] Duernberger, S., et al. (2023, December 14). *Position paper on the reparability of high-voltage electric vehicle batteries*. Allianz Center for Technology. https://www.azt-automotive.com/_Resources/Persistent/990408ce7d53df2ec19e036c230a2bb8e8954373/AZT%20Position%20Paper%20on%20HV%20Battery%20Repairability_14122023.pdf (Accessed: January 3, 2025)
[14] Seidenstücker, T. (2024, January). Elektroautos nach Beschädigungen beurteilen und reparieren – ein Lernprozess für alle Beteiligten. *VKU Verkehrsunfall und Fahrzeugtechnik*. (in German)

Author's contacts:

Norbert Schreier, Prof. Dr. Dipl.-Wirt.-Ing.
Esslingen University,
Kanalstraße 33, 73728 Esslingen am Neckar, Germany,
E-mail: norbert.schreier@hs-esslingen.de

ISSN 1846-6168 (Print), ISSN 1848-5588 (Online)
https://doi.org/10.31803/tg-20260116074322

Subject review
Received: 2026-01-16, Accepted: 2026-02-02

Thermal Power Plants - Flue Gas Purification

Marinko Stojkov*, Damir Šljivac, Ante Čikić, Robert Šanta

Abstract: One of the most important problems today is the formation of greenhouse gases, mainly caused by the emission of flue gases in thermal power plants: carbon dioxide (CO_2), sulfur dioxide (SO_2) and nitrogen oxides (NOx) ("acid rain"). CO_2 emissions are inevitable and related to the fuel combustion process (carbon capture techniques). Sulfur dioxide emissions are reduced by cleaning coal and by applying wet and dry flue gas desulfurization. The injection of absorbents can be applied in the combustion chamber or in the flue gases. Nitrogen oxide emissions are reduced by proper fuel selection and advanced solutions in the construction of the combustion chamber. Burners with low NO*x* emissions are used.

Keywords: environmental aspect; flue gases; greenhouse gases; reducing emissions

1 INTRODUCTION

Since the beginning of the 19th century, with the first industrial revolution and the introduction of the steam engine, the concentration of greenhouse gases in the atmosphere has been continuously increasing. This phenomenon has been particularly significant in the last 50 years [1]. When talking about greenhouse gases, carbon dioxide (CO_2), methane (CH_4) and nitrogen oxides (N_2O) are primarily investigated. Greenhouse gases are mainly produced by the conversion [2] and use of energy in buildings consumption [3], transport [4-5], production process [6] and different types of heating systems [7]. During this process as much as 85 % of carbon dioxide is produced. Greenhouse gases result in global warming, the occurrence of bad weather conditions in certain areas, and then indirectly affect human life and health, ecology, water supply and crops, [8]. In order to slow down the impact of global warming, it is necessary to reduce energy generation based on fossil fuels, which also reduces the share of energy obtained through the combustion process of primary energy, to increase the efficiency of energy conversion, to improve fuel quality and conversion technology to reduce emissions (e.g. coal gasification), and also to use technologies for the separation and storage of emitted gases, especially carbon dioxide. So, there is a growing need to reduce harmful emissions from flue gases that directly affect the ecosystem (greenhouse gases and acid rain). Harmful emissions include emissions of sulfur dioxide, nitrogen oxides and particulate matter. There are methods for reducing harmful emissions from flue gases that have been in conventional use for years, but there are also several methods that are still in the development process. This paper deals with the methods used to reduce sulfur dioxide emissions, such as: coal cleaning, injection of absorbents into the flue gas duct, injection of absorbents into the firebox and also wet and dry flue gas desulfurization and methods for reducing nitrogen oxide emissions.

In the second chapter, the authors provide an overview of the legislation in the world and in the Republic of Croatia and references to the literature used in the work. Chapter 3 describes technologies for reducing sulfur dioxide emissions based on coal cleaning and flue gas desulfurization. Chapter 4 discusses technologies for reducing nitrogen oxide emissions based on primary steps such as changes in the combustion method, ways of influencing the reduction of nitrogen oxide emissions during boiler construction, and the use of selective non-catalytic and selective catalytic reduction. Chapter 5 generally considers combined techniques for simultaneously reducing emissions of both sulfur and nitrogen oxides in flue gases. Finally, the conclusion provides an overview of the effectiveness of individual methods.

2 LEGISLATION DOCUMENTS

The Kyoto Protocol and the United Nations Framework Convention on Climate Change is an addendum to the international agreement on climate change. The protocol is a complex agreement with significant economic implications that obliges developed countries to reduce greenhouse gas emissions by an average of 5 % below the 1990 level in the first period until 2012. The protocol was opened for signature on December 11, 1997, in the Japanese city of Kyoto organized by the United Nations Framework Convention on Climate Change (UNFCCC). According to the obligations arising from the Kyoto Protocol, countries are divided into two groups according to their industrial development: developed and underdeveloped. And while the highly industrialized countries, grouped in the so-called Annex 1, must reduce the emission of harmful gases by 5-10 %, countries outside that group do not have such an obligation, but must therefore participate in the Clean Development Program. Kyoto Protocol entry into force (ratification by at least 55 members of the Framework Convention, which must account for at least 55 % of all greenhouse gas emissions in industrialized countries) on February 16, 2005. The Kyoto Protocol shows the clear commitment of most countries to take long-term measures that will initiate the resolution of the problem, [9-11].

Further progress was achieved in 2010 at United Nations Climate Change Conference Cancun, Mexico and after confirmed by Agreement in Paris, France 2015. A political agreement has been reached to a key step to achieve the goal of reducing greenhouse gases and helping developing countries from climate impacts with the "right to development". It was the world's largest joint effort to reduce

emissions, with national plans set at the international level through the UNFCCC. Also, the most complex package of agreed government measures to support developing countries: financial, technological and support in capacity building to combat the consequences of climate change.

Standards for the emission of harmful substances in Republic of Croatia are defined by the "Air Protection Act". It determines the measures, the way of organizing and implementing protection, improving air quality, and also prescribes air quality monitoring. There are also 2 subordinate legislations: Regulation on recommended and limit values of air quality and Regulation on limit values of emissions from stationary sources. Recommended and limit values of air quality represent only values for determining air quality categories, and not maximum permitted quantities of pollutants. The Regulation on limit values of emissions from stationary sources prescribes maximum permitted emissions and the method of adapting existing facilities to new regulations. Emission limit values for SO_2 are determined by the Regulation on limit values of emissions from stationary sources in Croatia. Regulations on SO_2 emissions for large combustion plants are dependable on the age and size of the power plant. Stricter rules and regulations apply to large plants, as well as to new plants compared to existing ones. Large burning devices can operate without a waste gas desulphurization device for a maximum of 240 hours a year, of which 72 hours continuously. Emission limit values for NO*x* are also determined by the Regulation on emission limit values from stationary sources in Croatia. When determining the limit of NO*x* emissions for gas turbines, it is necessary to make a correction with regard to the degree of operation of the plant by taking the value from defined maximum value multiply by the factor η/30 (where η is the percentage of heat supplied by the fuel that is converted into useful mechanical work in %). The degree of action η is determined for the most unfavorable mode of operation of the turbine when η is the smallest.

For determining the composition of flue gases, norms and standards are prescribed, which describe the measuring devices used for such measurements. The composition of combustion flue gases is measured in accordance with the HRN EN ISO 10396 standard, which describes the measurement methods and procedures.

3 TECHNOLOGIES FOR REDUCING SULFUR DIOXIDE EMISSIONS

Sulfur dioxide emissions depend on the amount of sulfur in the fuel. Natural gas contains less than 0.1 % sulfur (mainly in the form of H_2S) and emissions are relatively low. Light fuel oils, such as those used in gas turbine plants, usually contain little sulfur and emissions are also low. Coal and heavy oils (fuel oil) can contain a lot of sulfur and therefore combustion produces large emissions of sulfur oxides. Methods for reducing sulfur dioxide emissions are briefly described below.

3.1 Coal Cleaning

Sulfur in coal is of organic and inorganic origin. The origin of organic sulfur is related to plant matter (especially proteins). Inorganic sulfur is divided into pyritic and sulfate sulfur, of which pyritic sulfur makes up a significant portion, so when talking about pyrite, inorganic sulfur in coal is usually meant in general. Initially, the main task of coal cleaning was to remove ash and moisture, in order to reduce transportation costs and increase the efficiency of the power plant. In the United States, the Clean Air Act was adopted in 1970, which obliged new coal-fired power plants to significantly reduce sulfur dioxide emissions. Since then, a number of procedures and methods have been developed with the aim of achieving the best possible results in this field. Thus, in the 1970s, the potential emission of sulfur dioxide in new power plants was reduced to 80 %, by the 1990s, standards prescribed a 90 % reduction in emissions, and today the efficiency of such systems is as high as 99 %, [8]. Recently, the focus has shifted to the removal of sulfur from coal, due to the growing problem of acid rain caused by sulfur dioxide emissions from the coal combustion process. Since current coal extraction methods are non-selective, they result in a large amount of impurities in the coal. Physical cleaning of coal, which removes ash and some of the pyrites, is a method that has been used for many years.

Conventional cleaning methods are mostly based on gravitational separation of ash and sulfur components before the coal is converted into coal dust and introduced into the boiler. It is important to note that this method also significantly reduces the particulate content in the flue gases. Conventional cleaning usually begins with crushing the coal into pieces with a diameter of less than 50 mm, followed by separation into coarse, medium and fine particles. The crushing releases the materials that form the ash and inorganically bound sulfur (e.g. pyrites, FeS_2). The smaller the particles, the better the separation. Since mineral substances have a higher density than coal particles, they can be removed from coarse and medium particles by physical cleaning methods. Separation of fine particles can be done by flotation, using the surface difference between coal and ash, [2].

3.2 Flue Gas Desulfurization

SO_2 emission reduction by flue gas desulfurization can be divided into two basic processes: regeneration and non-regeneration processes. In regeneration processes, SO_2 is obtained, which can be further used to produce sulfuric acid, elemental sulfur, or liquid SO_2. Such devices are in the minority because they are significantly more complex and expensive. In non-regeneration processes, a by-product is obtained that is either permanently disposed of or used as a raw material in the cement or construction industry. Both processes are further divided into wet and dry processes. From wet to dry processes, the efficiency of desulfurization and the price of the device decrease. The costs of flue gas desulfurization plants depend on market and other conventional conditions. In addition, the price also depends on technical factors such as the amount of flue gases, the concentration of SO_2 in the flue gases, the required degree of desulfurization, environmental restrictions, and the amount of wastewater, [12].

3.2.1 Wet Process of Flue Gas Desulfurization

The flue gases enter a large vessel and an aqueous solution containing 10 % lime or limestone is injected into it. The calcium in the solution reacts with SO_2 to form calcium sulphate (gypsum) or sulphite. Part of the solution goes into a settling vessel where the solids settle before going to a filter where the water is removed and 50 % of the solids remain. The waste calcium sulphite is mixed with ash (1:1 ratio) and disposed of in landfills. Mist eliminators are also installed to collect the solution and remove moisture from the flue gases. They are located at the outlet of the firebox. After leaving the particle removal plant, the gases enter a spray tower or absorber. An aqueous calcium-based solution is injected here, which forms calcium sulphite or calcium sulphate with SO_2, which is removed by releasing water and settling in a designated vessel. The resulting waste is usually mixed with ash collected in a filter or electrostatic precipitator and with lime in a mixer, and disposed of in a landfill, [13]. The advantages of wet desulfurization are good efficiency and additional particle removal. The disadvantages are scale deposition and a tendency to clog, a significant drop in flue gas pressure, and significant investment and operating costs. In terms of efficiency, the standard version allows for sulfur removal of between 80 % and 90 %. Additives (magnesium-enriched lime) increase the efficiency by 5 % to 10 %, ultimately reaching 95 % to 99 %. Installation of this technology on an existing plant takes 3 to 6 weeks, provided there is sufficient space, [2].

3.2.2 Dry Process of Flue Gas Desulfurization

Dry process of flue gas desulphurisation is carried out by feeding a solution of calcium hydroxide (lime mixed with water) into a spray tower. The solution is sprayed and injected into the flue gases. As they evaporate, the droplets react with the SO_2 in the vessel. The amount of water is chosen so that all the water must evaporate before the absorbent falls to the bottom of the vessel, which is why this process is called a dry process. The by-product that is formed is collected at the bottom of the drying tower and in particle removal equipment. The by-product is unreacted lime and fly ash, and is removed by a filter or electrostatic precipitator. In the case of coal with a low sulphur content, 70 % of the sulphur dioxide is removed. Investments in the dry desulphurisation method are lower than investments in the wet desulphurisation method, and they are also simpler to operate and maintain. It takes 3 - 6 weeks to install a dry flue gas desulphurisation unit in existing plants. The advantages of this method are simplicity and lower operating and investment costs, while its only disadvantage is lower efficiency (approx. 70 %), [14].

3.2.3 Injection of Absorbent into the Combustion Chamber

When injecting into the combustion chamber, the absorbent (dry powdery material) is injected together with air above the combustion area through special injection inlets. Lime, which reacts with SO_2 and forms calcium sulfate (gypsum), is separated from the absorbent in the combustion chamber by calcination. The absorbent can be limestone or lime. Calcium sulfate (gypsum) and fly ash are removed using an electrostatic precipitator or bag filters. SO_2 removal efficiency is 30 – 60 %. Some other difficulties that arise with this method must be solved, for example, the efficiency of the absorbent must be increased to make the method more economically attractive. Another problem is that the injection of the absorbent has a negative effect on the operation of the clarifier. This method requires less investment than wet and dry gas desulfurization, requires less space and is simpler to operate and maintain.

3.2.4 Injecting the Absorbent into the Flue Gas Duct

Injecting the absorbent into the flue gas duct is injection in front of the precipitator, similar to injection in the combustion chamber. Similar equipment is required. A reaction vessel (absorber) is required as well as particle removal equipment (electrostatic precipitator, bag filter). Calcium hydroxide is injected into the reaction vessel, where it mixes with the flue gases. As the gases enter the absorber, they are cooled to the "approach temperature" by injecting water. Unreacted absorbent and ash are captured in a factory filter located behind the absorber and disposed of in a landfill, while the clean flue gases are discharged into the stack. Injection in front of the electrostatic precipitator can remove 30 – 70 % of SO_2, and injection after the precipitator can remove 80 – 90 %. The efficiency also depends on the amount of absorbent and the approach temperature.

4 TECHNOLOGIES FOR REDUCING THE EMISSION OF NITROGEN OXIDES (NO*x*)

Emissions of nitrogen oxides are associated with acid rain and photochemical smog and ozone in the troposphere. For this reason, technologies were introduced to regulate NO*x* emissions from new, as well as from old, power generation plants. NO*x* emissions can be reduced in the following ways:

- primary steps (change in combustion)
- choice of fuel
- construction of the combustion chamber.

4.1 Primary Steps (Change in Combustion)

These are changes in the boiler, or changes in combustion, and include: changes in operating conditions and changes in combustion devices. Changes in operating conditions include reducing the amount of combustion air, reducing the temperature of the combustion air, reducing the amount of heated air and reducing the load on the combustion chamber. Air reduction is a frequently used method for reducing NO*x* emissions, but it is also used as a method for reducing energy consumption. A significant reduction in nitrogen oxide emissions can be achieved by primary steps during combustion, which is achieved by appropriate burner design and stepwise air and fuel supply. This reduces the maximum temperatures in the flame core and reduces the oxygen concentration in the combustion zone. The amount of NO*x* emissions into the atmosphere can be reduced to a value of less than 300 mg/m^3 (up to 40 %), [14, 15]. The method of reducing the amount of air has a number of advantages:

reduction of flue gas heat losses and reduction of fuel consumption (both of which result in higher energy efficiency), no need for additional investments and reduction of low-temperature corrosion, but also disadvantages: requirements for precise combustion control, possible increased particle emissions and more intense deposition on the inner walls, which reduces heat transfer and increases maintenance costs. In this case, the power plant is out of use more often and for longer. The method of reducing the temperature of the combustion air has its advantages: simplicity of implementation and no need for investments, but also disadvantages: reduced energy efficiency and increased fuel consumption. The situation is similar with the primary step of reducing nitrogen oxide emissions by reducing the load on the combustion chamber, which is simple and without additional investments (advantages), but at the same time the available power is reduced and the specific fuel cost is increased (disadvantages), [16].

4.2 Low Emission NO*x* Burners

Burners, by their design, have a significant impact on NO*x* formation. Modifications to the burners regulate the combustion of carbon and hydrogen with acceptable NO*x* formation. The basic ways of improving burners are:

- burner with improved mixing
- burner with flame splitting
- self-recirculating burner
- burner with substoichiometric combustion
- burner with two-stage combustion.

4.3 Selective Non-Catalytic Reduction (SNCR)

SNRC is a chemical process that converts nitrogen oxides NO*x* into molecular nitrogen N_2. In this process, there is no catalyst, and the reducing agent (ammonia, amides) is injected directly into the flue gas in the high-temperature zone. This chemical reaction occurs at high temperatures (870 - 1200 °C). In this process, the temperature must be carefully controlled. A gaseous or liquid nitrogen-based reactant is injected and mixed with the hot flue gases. The reactant reacts with the NO*x* in the flue gas stream and converts it into atmospheric nitrogen and water vapor, which are harmless. The reactant primarily reacts with NO*x*, not with oxygen or other components of the flue gases. This technology does not produce any solid or liquid waste. In this way, NO*x* emissions can be reduced by 35 - 70 % without affecting the operation of the plant. At temperatures below optimal, the reaction is weak, so the reduction of NO*x* is very small, almost negligible. The technology was first developed for fuel oil and natural gas boilers, and is still in the testing phase for coal-fired furnaces. Due to the very simple components that need to be installed, installing this technology is much easier than others. Installation in existing plants takes 2 - 5 weeks, [2].

4.4 Selective Catalytic Reduction (SCR)

It is widely used for low-sulfur coal, but it is very expensive and there is a need to adapt to different types of coal. The working principle of this method is based on the reduction of NO*x* by converting it to N_2 and H_2O by the reaction of NO*x* and ammonia with the help of a catalyst. The primary reaction that occurs in this method requires oxygen, and the catalyst works best at oxygen levels above 2-3 %. This technology is based on the injection of ammonia into the flue gases to convert NO*x* into elemental nitrogen and water. It works in almost the same way as selective non-catalytic reduction. The only difference is that selective catalytic reduction requires a reagent that accelerates the chemical reaction. The SCR system contains equipment for measuring NO*x*, which is necessary for precise dosing of the reagent, so that ammonia is not released into the environment. The catalyst is needed because the process takes place at significantly lower temperatures than in selective non-catalytic reduction. Typical temperatures for this process are 340 - 380 °C. The most common catalysts are a combination of vanadium and titanium and zeolite materials. This technology was first used in Japan in the mid-1970s, in Western Europe in the 1980s in coal-fired power plants, and in the USA in the early 1990s also in coal-fired power plants. The installation of a hot system on an existing plant takes 2 - 3 months, and for a cold system 3 - 6 weeks. The efficiency of this method is 70 – 90 %, [14].

5 TECHNOLOGIES FOR THE COMBINED REMOVAL OF SO_2 AND NO*x*

Sulfur dioxide SO_2 and nitrogen oxides NOx are present in flue gases. It is desirable that the same process be used to reduce SO_2 and NO*x* emissions. In practice, separate methods are used for control. Conventional flue gas desulfurization methods are used to reduce SO_2 emissions, while selective catalytic reduction is used to reduce NO*x* emissions. Combined systems should operate in such a way that their SO_2 and NO*x* removal efficiency is equal, [13]. The emission limit values for new coal-fired plants with a capacity of >100 MW for the EU (including the Republic of Croatia) for sulfur and nitrogen oxides are 200 mg/m^3, [15].

6 CONCLUSION

The operation of a thermal power plant also produces sulfur dioxide SO_2, and one of the measures to reduce the production of this gas is the use of natural gas as a primary energy source, which has an almost negligible amount of sulfur (gas-fired thermal power plants have SO_2 emissions of 0,02 - 0,42 g/kWh, liquid fuel thermal power plants have 3,77 - 4,61 g/kWh, and coal-fired thermal power plants have 0,68 - 24,51 g/kWh). Furthermore, the reduction of SO_2 emissions is also achieved by cleaning coal with an efficiency of 60 %. After the coal cleaning process, the percentage of coal is 60 - 90 % of the initial mass, and the percentage of heating value is 85 - 98 %. The wet desulfurization method achieves an efficiency of 80 - 90 % in removing sulfur. By adding lime, the efficiency of SO_2 removal rises to 95 - 99 %. The dry desulfurization method is used in coal-fired thermal power plants with a low sulfur content in the coal itself. The efficiency of the method is 70 - 90 %. This method is the most investment-friendly and easiest to install and maintain. In order to reduce the emission of nitrogen oxides NOx, burners with low NO*x* emissions are used (e.g. flame splitting burners

- emission reduction of 20 - 40 %, used in large liquid fuel plants; self-recirculating burner - efficiency of 20 %; burner with two-stage reduction - efficiency of 30 - 55 %). In general, the method of using low NOx burners achieves an efficiency of reducing NO*x* emissions by 40 - 50 %. Selective non-catalytic reduction achieves NO*x* removal of 35 - 70 %, and is used for fuel oil and natural gas. The selective catalytic method is used for coal with low sulfur emissions, the efficiency is 70 - 90 %, but this method is more expensive than the non-catalytic method with installation time in the range of 2 - 3 months, while for non-catalytic reduction this time is 2 - 5 weeks. During the installation, the plant does not work, and thus additional costs accumulate.

Acknowledgement

This research paper was funded by the University of Slavonski Brod through the institutional research project Advanced modelling and optimization of compact heat exchangers for integration into renewable energy systems (MOKIT), financed by the European Union – NextGenerationEU. The views and opinions expressed in this paper are those of the author and do not necessarily reflect the official position of the European Union or the European Commission. Neither the European Union nor the European Commission can be held responsible for them.

7 REFERENCES

[1] Udovičić, B. (1993). *Energetika*. Školska knjiga. (in Croatian)
[2] Feretić, D., Škanata, D., Subašić, D., Čavlina, N., & Tomšić, Ž. (2000). *Elektrane i okoliš*. Element. (in Croatian)
[3] Kadrić, Dž., Blažević, R., Bajrić, H., Peco, A., & Kadrić, E. (2025). Renovation measures for reduction of primary energy consumption and CO_2 emissions of hospital building. *Tehnički glasnik, 19*(1), 26–33. https://doi.org/10.31803/tg-20231114190214
[4] Steger-Jensen, K., Hvolby, H. H., Vestergaard, S., Neagoe, M., & Svensson, C. (2025). Carbon footprint principles and challenges in transport logistics. *Tehnički glasnik, 19*(1), 127–135. https://doi.org/10.31803/tg-20240510171202
[5] Muhič, S., Čikić, A., Pištan, J., Stojkov, M., & Bošnjaković, M. (2018). Transport emissions and electric mobility in private transport in the Republic of Slovenia. *Tehnički glasnik, 12*(2), 98–103. https://doi.org/10.31803/tg-20180508162744
[6] Strauß, H., & Sasse, J. (2025). Real-time monitoring of the CO_2 footprint of production for SMEs. *Tehnički glasnik, 19*(SI1), 19–24. https://doi.org/10.31803/tg-20250318104618
[7] Đuranović, M., Živić, M., Stojkov, M., & Lujić, R. (2024). Analysis of energy consumption of a thermo-technical system with an absorption heat pump. *Tehnički vjesnik, 31*(3), 967–972. https://doi.org/10.17559/TV-20231020001042
[8] Rubin, E. (2013). Climate change: Technology innovation and the future of coal. *Cornerstone: The Official Journal of the World Coal Industry, 1*(1).
[9] Benac, Č. (2005). *Zaštita okoliša*. Sveučilište u Rijeci. (in Croatian)
[10] Rastovčan-Mioč, A. (2009). *Uvod u ekologiju*. Metalurški fakultet. (in Croatian)
[11] Bedeković, G., & Salopek, B. (2010). *Zaštita zraka* (interna skripta). Rudarsko-geološko-naftni fakultet. (in Croatian)
[12] Kamall, R. (2000). *Flue gas desulphurisation (FGD) technologies*. Department of Trade and Industry.
[13] Rajković, D. (2011). *Proizvodnja i pretvorba energije*. Rudarsko-geološko-naftni fakultet. (in Croatian)
[14] Prelec, Z. (2010). *Tehnike za smanjenje emisija* (skripta). Zagreb. (in Croatian)
[15] Bogdan, Ž., Živković, S. A., Dokmanović, V., & Merić, J. (2007). Tehnologije čistog ugljena u strategiji razvoja elektroenergetskog sustava. *Energija, 56*(4), 398–431. (in Croatian)
[16] Protić, M., Mišić, N., Raos, M., Tasić, V., & Topalović, D. (2025). Effects of heat flux and ignition type on the combustion of live *Pinus nigra* branches. *Tehnički vjesnik, 32*(6), 2280–2289. https://doi.org/10.17559/TV-20250303002436

Authors' contacts:

Marinko Stojkov, PhD, Full Professor
(Corresponding author)
University of Slavonski Brod,
108. Brigade ZNG 36, 35000 Slavonski Brod, Croatia
mstojkov@unisb.hr

Damir Šljivac, PhD, Full Professor
Faculty of Electrical Engineering, Computer Science and Information Technology Osijek, University J. J. Strossmayer in Osijek,
Kneza Trpimira 2b, 31000 Osijek, Croatia
damir.sljivac@ferit.hr

Ante Čikić, PhD, Full professor
University North Varaždin, Mechanical Engineering Department,
Jurja Križanića 31b, 42000 Varaždin, Croatia
ante.cikic@unin.hr

Robert Šanta, PhD, Full Professor
Aziz Sanjar Food Safety Laboratory, Azerbaijan State University of Economics (UNEC), 6 Istiglaliyyat Str., Baku AZ1001, Azerbaijan
Department of Mechanical Engineering and Material Sciences, Institute of Engineering Sciences, University of Dunaújváros, Táncsics Mihály u. 1/A, 2400 Dunaújváros, Hungary
SANTAR@uniduna.hu

ISSN 1846-6168 (Print), ISSN 1848-5588 (Online)
https://doi.org/10.31803/tg-20241231143440

Professional paper
Received: 2024-12-31, Accepted: 2025-03-11

Design and Additive Manufacturing of a Scalable, Low-Cost Educational Robotic Gripper

Denis Kotarski*, Alen Šćuric, Tomislav Šančić

Abstract: This paper presents the development and prototyping of an educational robotic gripper designed for a professional mechatronics study program. The research connects key learning outcomes of the mechatronics curriculum with the design, prototyping, testing, and integration of the gripper system into robotic applications. Key parameters and components were selected during the design phase, followed by the construction of system assembly parts. A parametric design for the mechanism, suitable for additive manufacturing, is outlined in alignment with educational objectives. The scalable design and cost-effectiveness of fused deposition modeling (FDM) make the presented gripper adaptable to various robotic systems. The prototyping process includes selecting printing parameters in the case of FDM and executing the manufacturing process. Three different sizes of robotic grippers are manufactured, with plans for integration into robotic systems to support hands-on engineering education.

Keywords: additive manufacturing; low-cost production; mechatronics engineering education; parallel mechanism gripper; scalable robotic gripper

1 INTRODUCTION

To address the demands of mechatronics engineering education, hands-on training with experimental setups of mechatronic systems is essential. Robotics plays a significant role as an interdisciplinary field encompassing various technologies related to industry 4.0 [1], design, control, planning, and more. Robotic systems can generally be categorized into industrial and mobile robots. Within robotics, the end-effector (gripper) is a crucial system component [2]. For instance, the integration of different end-effectors with a three degrees-of-freedom (DoF) open-source robotic arm is explored in the paper [3]. Additionally, the paper [4] demonstrates the integration of an underactuated gripper with a heterogeneous robotic system comprising a robotic arm and a ground robot [4]. Grippers are also applied in aerial robotics, as shown in the paper [5].

In the context of research and engineering education, robotic end-effectors can be explored from multiple perspectives. A crucial aspect is the control of end-effectors, as shown in the paper [6] offering an overview of various control methods. End-effector (gripper) control systems can be categorized into open-loop and closed-loop systems. Closed-loop systems require the integration of sensors to effectively perform tasks. The paper [7] discusses a grasp controller designed for a sensorized parallel jaw gripper, enabling it to gently pick up and place down unknown objects. The control scheme draws inspiration from human actions, which rely on tactile feedback. On the hardware side, extensive research has been dedicated to the design and manufacturing of mechanical systems for robotic grippers. Additionally, significant efforts have been made in engineering education, as demonstrated by research such as design of a single-DOF gripper mechanism [8]. Gripper designs vary based on their intended purpose, with some mechanical systems being multifunctional [9], and specialized for tasks such as grasping thin objects [10]. In addition to mechanical grippers combined with electric actuators for grasping functions, soft grippers have been extensively researched, as demonstrated in paper [11].

The advancement of rapid prototyping technologies has significantly accelerated the development of mechatronic systems, with additive manufacturing (AM) playing a pivotal role. This technology is versatile, finding applications across various fields, especially in robotics, from creating educational low-cost robotic platforms [12] to manufacturing specialized multirotor UAV airframe parts [13]. Moreover, AM technologies have proven effective in producing robotic components, such as the open-source 3D-printed underactuated gripper presented in the paper [14]. In papers [15, 16], a low-cost, open-source, 3D-printed three-finger gripper platform for research and educational purposes is presented, featuring a design with minimal 3D-printed components and an off-the-shelf servo actuator. Additionally, research [17] offers a comparative study on the impact of different 3D-printed fingertips in robotic gripper jaws, focusing on the pulling force exerted on various shaped objects. Furthermore, the paper [18] emphasizes the optimization of AM techniques for constructing a new gripper design capable of handling a wide range of parts with varying sizes and shapes. Beyond gripper development, AM has also been applied to optimizing robotic structures, as demonstrated in a comparative analysis of topology optimization platforms [19].

This paper presents the design and development of a scalable educational robotic gripper. The concept is examined in the context of the professional mechatronics study program that is being conducted in Croatia. A scalable design for a robotic gripper utilizing a parallel mechanism is proposed, emphasizing low-cost production and ease of manufacturing. The paper outlines the entire process, from the initial sketch to the final prototype or product. The prototyping phase involved fused deposition modeling (FDM) AM technology, with three prototypes of varying sizes and capabilities tested. The concept of the robotic gripper can be used in different aspects of the process of hands-on engineering education. Still, it can also be used in the broader context of STEM education at lower levels. Future work will focus on integrating presented robotic grippers with robotic systems.

2 UNDERGRADUATE PROFESSIONAL MECHATRONICS STUDY PROGRAM IN CROATIA

This research examines the integration of a robotic gripper into the practical education curriculum of the mechatronics undergraduate professional study program in Croatia. In 2021, the program was developed to align with the HKO register [20]. The curriculum totals 180 ECTS credits, with 114 ECTS derived from mandatory learning outcomes and 66 ECTS from optional learning outcomes. The study learning outcomes are divided into general outcomes (IUP1 - IUP10), which focus on soft skills, and professional outcomes (IUP11 - IUP27). Table 1 lists the study learning outcomes that can be partially or fully related to the robotic gripper in the areas of modeling, control, design, and manufacturing.

Table 1 Undergraduate professional mechatronics study program learning outcomes related to robotic gripper [20].

Learning outcome code	Learning Outcomes of the Undergraduate professional study of Mechatronics in Croatia
IUP1	Apply mathematical and physical laws to engineering problems
IUP6	Identify, model, and solve engineering problems
IUP9	Use techniques, skills, and modern tools necessary for engineering practice
IUP11	Recognize and propose adequate types of materials and machining procedures for the production of mechatronic systems
IUP12	Dimensioning and selecting standardized elements of precision mechanics, machine elements, and assemblies in the process of manufacturing various mechanisms
IUP14	To know the principles of operation of electronic and electromechanical converters
IUP17	Create 2D technical documentation and construct a 3D model of mechatronic systems
IUP21	Select and connect sensors, actuators, microcomputers, programmable logic controllers, and accompanying equipment for the automation of production processes
IUP25	Integrate the computer with software support in the process of data collection, measurement and data display on the computer
IUP27	Plan, monitor and control the production of mechatronic system elements

Regarding the mentioned mechatronics study program aligned with the HKO register, there are mechatronics studies in Croatia that are either fully aligned or in the process of alignment to meet the learning outcomes of the professional mechatronics curriculum. At the Karlovac University of Applied Sciences, the study of mechatronics is currently in the process of revision and alignment. From the aspect of hands-on training, and this research topic, low-cost robotic gripper prototypes, made using entry-level components, are currently being integrated with robotic arm models, as shown in Fig. 1, and used in hands-on laboratory exercises for the Robotics course. The use of a robotic gripper in robotics classes aligns with IUP9 and IUP21. Additionally, DC servo motors used to drive the robotic gripper are covered in Actuators course aligned with IUP14. From a system design perspective, students work on case studies involving the construction of robotic grippers, including modeling parts and assemblies, which aligns with IUP17 and IUP9.

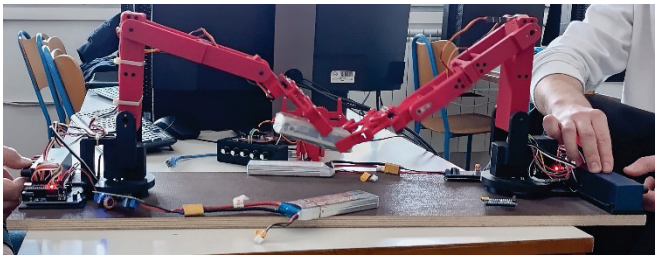


Figure 1 Low-cost robotic gripper prototypes integrated with robotic arm models, utilized in hands-on laboratory exercises for the Robotics course

These study is part of the internal project at Karlovac University of Applied Sciences, titled "Development of Innovative Educational Robotic Modules - iEduBot". As part of the project, the necessary equipment, components, and materials will be procured to develop educational robotic modules, including robotic grippers. Integrating these modules into the undergraduate professional mechatronics program will enhance hands-on learning through practical and team-based activities for students. After completing the design and production phases of the robotic gripper kits, integration into the educational process will proceed through exercises (theoretical, laboratory, and design-based), aligned with the mechatronics learning outcomes (Tab. 1).

3 DESIGN OF A SCALABLE AND MODULAR EDUCATIONAL ROBOTIC GRIPPER

From the hardware aspect, the main goal is to design a scalable and modular mechanism of a parallel robotic gripper that can be applied in the engineering education of an undergraduate professional mechatronics study. Another goal is to design a system that consists of readily available and relatively inexpensive components and parts that can be produced at low cost and with inexpensive equipment. The assembly of the parallel mechanism, which is actuated using an electric servo motor, is considered. The designed system is connected with the learning outcomes of the undergraduate professional study of mechatronics.

3.1 Mechanical Overview of a Robotic Gripper with Parallel Mechanism

Robotic grippers are an integral part of the robotic system and differ in mechanical configurations, drive type, power, precision, and other specifics. Robotic grippers with a parallel mechanism have a simple design and are suitable as a tool in engineering education. The parallel gripper shown in this paper consists of the central part of the assembly on which the electric servo motor is mounted, and the left and right sides of the gripper mechanism. Each side consists of two parallel links where the torque is transmitted via a geared link and the jaw to which the links are connected. Versions where each side has a separate motor or versions with one motor and a mechanism through which the torque is transmitted from the motor to both sides of the gripper are possible. The mechanism of the parallel gripper with one motor will be considered, as shown in Fig. 2.

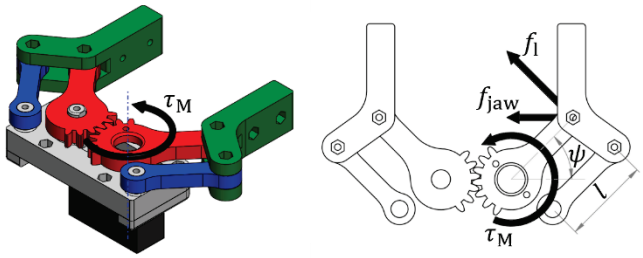


Figure 2 Schematic representation of a robotic gripper parallel mechanism

The motor torque is transmitted to the left and right jaws of the gripper via a mechanism consisting of two gear elements. The two main parameters of the mechanism that were considered in the design of the scalable gripper are the motor torque and the length of the link of the gripper parallel mechanism. Assuming that there are no losses in the mechanism and no misalignment, the amount of force in the link is given with following expression

$$f_l = \frac{\tau_M}{2l}, \tag{1}$$

where τ_M is the torque of the motor divided between the left and right sides of the mechanism, and l is the length of the link of the parallel gripper.Therefore, the force component acting in the direction of the parallel mechanism depends on the jaw angle, and is given with following expression

$$f_{jaw} = f_l \cdot \sin\psi = \frac{\tau_M}{2l} \cdot \sin\psi, \tag{2}$$

where the total force exerted by the left and right side of the jaw is equal

$$f_{tot} = 2 f_{jaw} = \frac{\tau_M}{l} \cdot \sin\psi. \tag{3}$$

The paper will consider 3 sizes of servo motors, according to which structural elements and assembly will be selected and dimensioned with the aim of producing parts of a low cost gripper, which is aligned with IUP12.

3.2 Electric Servo Motor Drive Units of the Low-Cost Robotic Gripper

DC servo motors were chosen for the drive of the low-cost educational robotic gripper because of their price, but especially because of the simple integration with the mechanical and control system. Such motors are typically used in applications that require precise control of position, speed, or torque. The DC servo motor assembly has a compact design, considering that the motor housing contains the driver, DC motor, and gearbox, and comes in a wide range of sizes and powers. The characteristic of such motors is that there is a feedback connection using an encoder or potentiometer, which forms a control circuit with the motor driver. The feedback system constantly monitors the motor's output and adjusts its input to match the desired position or speed. A DC servo motor is usually controlled by a PWM (Pulse Width Modulation) signal, where the pulse width determines the position, speed, or torque of the motor. This enables precise control in a various robotics, automation, and manufacturing applications, due to their fast response and efficiency. In addition to conventional motor designs, there are also digital servo motors that have a high accuracy. Three standard sizes of servo motors were considered and their characteristics are listed in Tab. 2. This aspect of the robotic gripper is aligned with the learning outcome IUP14.

Table 2 Considered DC servo motors manufacturer specifications

Servo motor	Size (mm)	Weight (grams)	Operating voltage (V)	Stall torque (kg-cm) @ 6V
FEETECH FS90MG	23.2 × 12.5 × 22	9	4.8 - 6	1.5
FEETECH FT1117M	30 × 12 × 31.7	20	4.8 - 6	3.5
FEETECH FS5115M	40.8 × 20.1 × 39	62	4.8 - 6	15.5

3.3 Design of a Scalable, Modular Robotic Gripper Mechanism

The goal of this project is to develop a scalable, modular, and cost-effective prototype of a robotic gripper mechanism. Given that there are different needs for the use of varying motor performance, a parametric design of the parts of the parallel gripper assembly is shown, which, integrated with the servo motor, forms the mechanical system of the gripper. During the design phase, the SOLIDWORKS software package was used to construct 3D models. Fig. 3 shows three sizes of the robotic gripper whose performance depends on the motor torque and the length of the link of the parallel mechanism. In addition to scalability, the assembly can be integrated with different types of robotic systems, for example with rotational or translational degrees of freedom. The scalable modular robotic gripper prototype will be used as a construction example in courses that include the construction of 3D models, aligned with IUP17.

In addition to the parametric design approach, the technological feasibility of the construction was considered given in mind that the parts are planned to be manufactured using AM technology. From that follows the economical process in terms of material costs and production time. In the design process, the mechanism assembly that will be produced by FDM AM technology is shown. From a hardware perspective, this work contributes with a scalable, modular robotic gripper that, in combination with a motor, forms a low-cost prototype suitable for integration into the curriculum of an undergraduate mechatronics study.

Further research will focus on analyzing the mechanism's structure using finite element method (FEM), addressing the challenges posed by the anisotropic mechanical properties of FDM-manufactured parts due to their layer-by-layer deposition. Additionally, error analysis will be conducted to evaluate dimensional deviations in the

gripper links and their impact on the parallelism of the gripper's fingers.

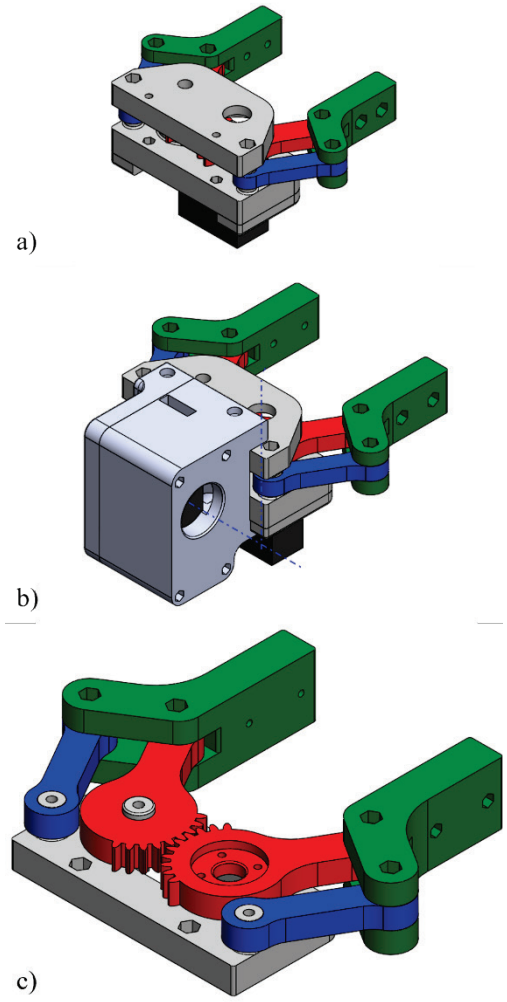


Figure 3 3D models of a parallel gripper mechanism parametric design: (a) Link length $l = 24$ mm with FS90MG motor, (b) Link length $l = 28$ mm with FT1117M motor and rotational DoF, (c) Link length $l = 40$ mm with FS5115M motor

4 PROTOTYPING A LOW-COST ROBOTIC GRIPPER USING ADDITIVE MANUFACTURING

After completing the design phase of the robotic gripper, the prototype production phase begins. For this purpose, it is necessary to create 2D technical documentation that aligns with IUP17. Fig. 4 shows a gripper assembly 3D model with associated parts that will be produced by AM technology. Given that the goal is to create a cheap educational gripper, FDM technology was chosen, which works by extruding thermoplastic material layer by layer to make a part or several parts. The material, which is in the form of a plastic filament, is heated to a semi-liquid state and deposited through a nozzle onto the fabrication platform. The nozzle follows a programmed path, controlled by computer-aided design (CAD) data. FDM technology is known for its affordability, both in terms of machinery and materials used, which makes it available for prototyping and small series production. It is easy to handle, widely available, and supports a range of thermoplastics, making it ideal for educational and research purposes.

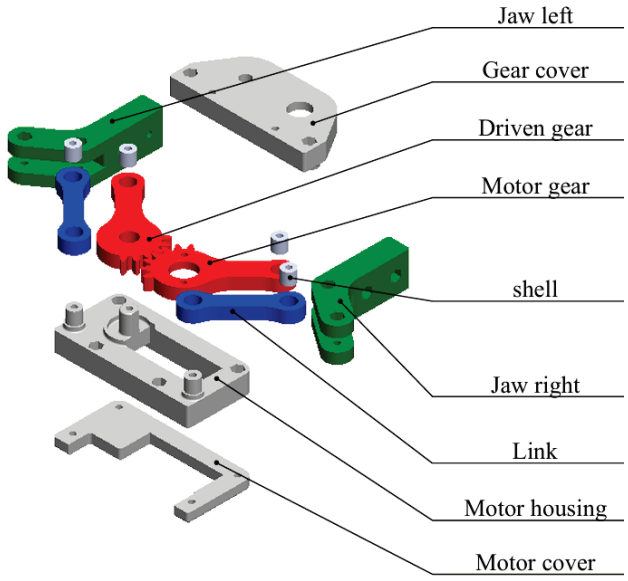


Figure 4 3D model exploded view of the gripper assembly with associated parts designed for production using AM technology

4.1 Preparation of Robotic Gripper Parts for Additive Manufacturing

Due to the combination of simplicity, cost-effectiveness, and flexibility of materials, FDM technology was chosen to produce prototypes of educational robot grippers. The initial step in preparing a 3D print, independent of the AM technology, involves saving the 3D model in .stl format. This format captures the geometry of the object by representing its surface with a mesh of triangles, facilitating accurate conversion from digital design to physical models. Next, the .stl files are loaded into a slicing software specific to the AM technology being used. For this research, an FDM 3D printer by Bambu Lab was used along with the Bambu Studio software. Adjusting the parameters of the 3D print depends on the geometry and purpose of the part and the used material. In this case, PETG material was chosen for producing educational robot grippers due to its superior durability, UV resistance, and heat tolerance, making it a better option than PLA for demanding applications. FDM technology also offers extensive parameter adjustability, which is valuable for addressing specific learning outcomes (IUP11).

Several parameters related to FDM printing technology were considered, particularly wall thickness and part infill The thickness of the walls, base, and roof of the printed part, along with the percentage and type of infill, play a crucial role in determining its mechanical properties. Previous research [21] explored how the number of walls and the percentage of infill affect tensile strength. These findings were taken into account when configuring the parameters for the educational robot grippers in this research. For the production of a series of prototypes for teaching purposes, a

three-wall loop and a gyroid sparse infill pattern with 15% density were selected.

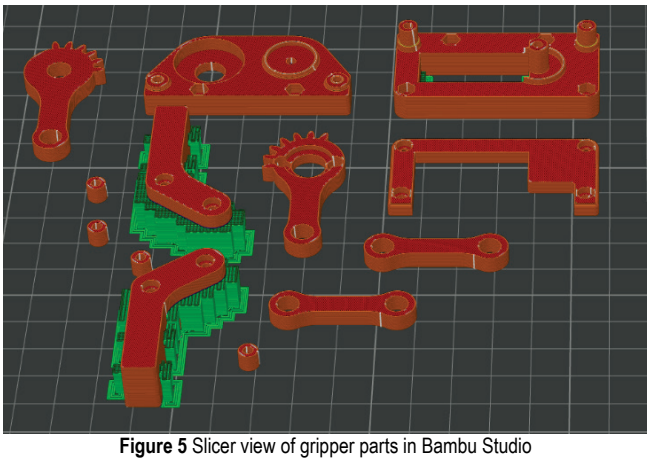


Figure 5 Slicer view of gripper parts in Bambu Studio

In the first phase of prototyping, several construction iterations were tested, with parts being printed separately, one by one. Once the final parametric design was established, it became more efficient to group the parts for 3D printing. A single prototype with identical print parameters for all parts can be produced within a single run, as shown in Fig. 5 for the gripper with the FT1117M servo motor using Bambu Studio. Printing multiple parts at once offers a significant time-saving advantage compared to printing each part individually. Although the model printing time remains similar, grouping the parts significantly reduces both preparation and post-processing times.

Each print cycle involves heating the 3D printer components (nozzle and print bed), and after completion, the nozzle and bed must cool down. For example, with PETG material on the Bambu Lab A1 Mini 3D printer, it takes approximately 8 minutes for the bed to cool below 50 °C, allowing the printed parts to be safely removed. This cooling process must be repeated for each new print. By printing all parts together, the prepare time and post-processing steps only occur once, reducing time considerably. When prototyping a series of robotic grippers, printing grouped parts offers flexibility, as the parameters for individual components can be adjusted independently, particularly in terms of mechanical properties. As the production volume increases, it becomes necessary to plan, monitor, and control the production of gripper components, but in general mechatronic systems, which is aligned with IUP27.

4.2 Additive Manufacturing and Assembly of a Gripper Mechanical System

After preparing the model in the slicer software and generating the G-code, the software provides estimates for total filament consumption, preparation time, and model printing time. During the preparation phase for PETG material, the nozzle is heated to 255 °C, and the print bed reaches 70 °C. Fig. 6 shows the 3D-printed parts of the gripper (FT1117M) produced using the Bambu Lab P1S 3D printer. Post-processing, such as removing support material, is often necessary, particularly for parts like the gripper jaws.

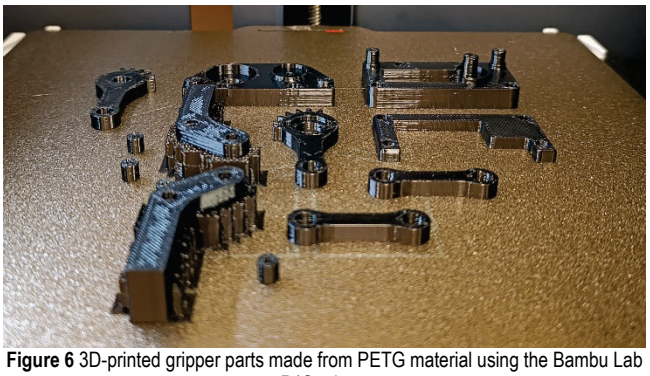


Figure 6 3D-printed gripper parts made from PETG material using the Bambu Lab P1S printer

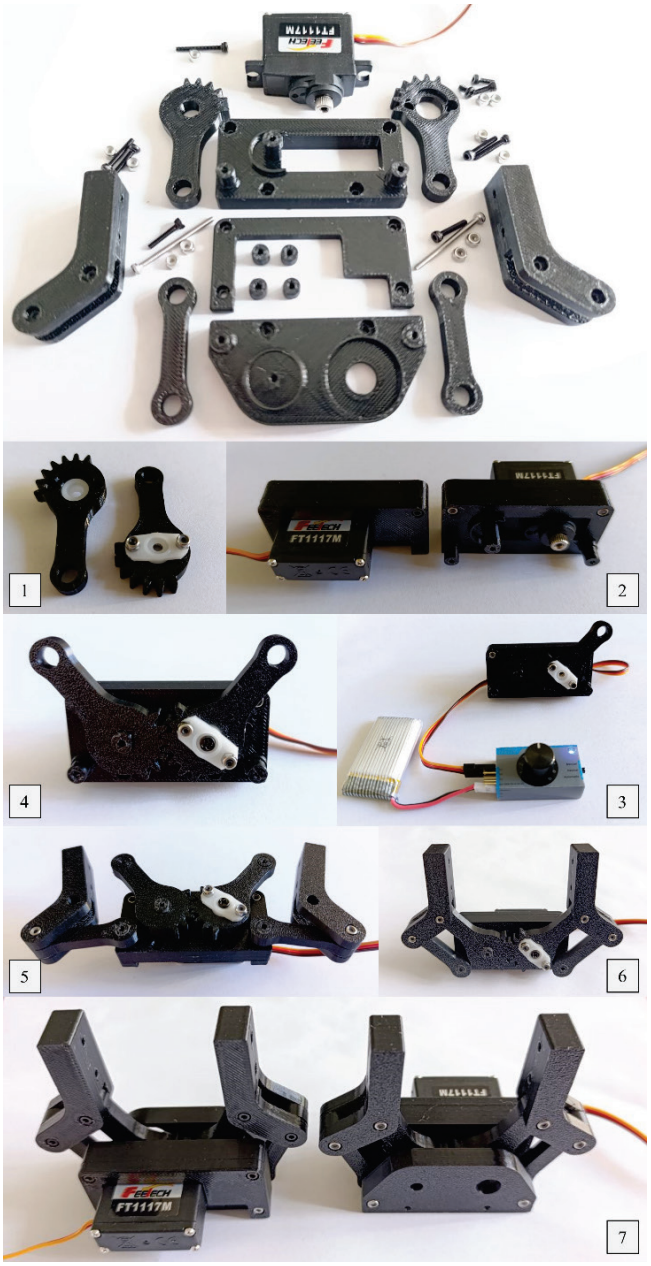


Figure 7 Assembly steps of the gripper components into a functional unit

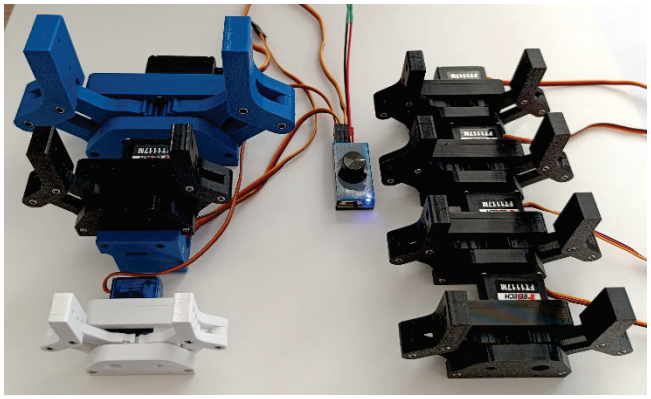


Figure 8 Display of three gripper sizes during preliminary testing

Once all the parts of the prototype or series of prototypes have been printed, along with connecting elements (bolts, nuts), the gripper assembly can begin, as shown in Fig. 7. The assembly process starts with mounting the servo adapter to the motor gear (Step 1) and continues with attaching the servo motor to the frame parts (Step 2). Step 3 involves testing the motor with the motor gear and adjusting the initial angle. Subsequently, in Steps 4, 5, and 6, the remaining mechanism parts are assembled. Finally, in Step 7, the cover is installed to secure all the elements into the assembly.

Fig. 8 illustrates the three prototyped robotic gripper sizes developed in this research during the preliminary testing phase of the full assembly. For this and previous testing stages, a module with four AA batteries and a servo tester, used to send the control PWM signal, were employed.

5 DISCUSSION

The design and prototyping of the robotic gripper demonstrate alignment with the learning outcomes of IUP11, IUP12, IUP14, IUP17, and IUP27 from the undergraduate professional mechatronics study. The mechanical gripper system can function both independently and as part of integrated robotic systems, whether industrial or mobile. For standalone use, the hardware gripper can be further enhanced by integrating sensors. An experiment to measure the gripping force using a sensor setup is under consideration, aligned with IUP21. To facilitate force measurement, 5 kg load cell sensors with HX711 amplifiers will be connected to the control system. From the perspective of the gripper, further work will focus on integrating sensors into the mechanism's jaws, as closed-loop systems require the use of sensors to perform tasks effectively. Given the gripper's intended low-cost design, the following sensors are being considered: a load cell sensor (similar to those used in force measurements) and a pressure sensor. As an extension of the experiment, a mechanical system with independent left and right motors for the parallel gripper mechanism can be incorporated. This setup allows for a comparison of the forces generated by a two-motor mechanism versus a single motor with a mechanical linkage. The integration of a computer and control system for data collection, measurement, and visualization is aligned with IUP25.

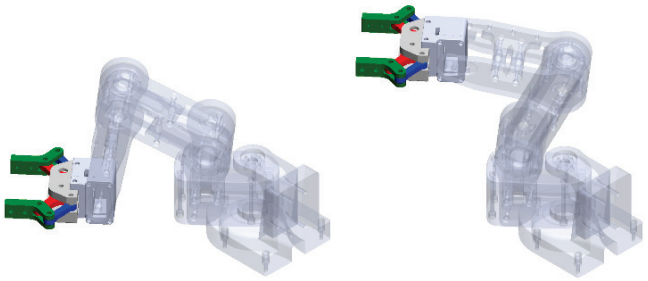


Figure 9 Integration of the robotic gripper and rotational degree of freedom with an RRR robotic arm configuration.

In further work, the integration of an inertial measurement unit (IMU) for rotational control, providing an additional degree of freedom, is under evaluation (Fig. 3b). Fig. 9 presents the CAD model of the proposed design for the robotic gripper integrated onto a robotic arm with three degrees of freedom (DoF). For this purpose, the design of robotic modules will be undertaken to integrate the proposed educational robotic gripper. These modules will include both the existing mobile platforms presented in previous research [22] and new robotic modules for industrial simulators with two, three, and four degrees of freedom.

6 CONCLUSION

This paper presents the design, prototyping, and testing of a scalable and low-cost robotic gripper. The aligned learning outcomes have been demonstrated within the undergraduate professional mechatronics study program through various aspects of the robotic gripper. A robotic gripper with a parallel mechanism was described, and its system parameters were defined. A low-cost, scalable, and easily upgradeable design for the gripper mechanism was proposed. FDM technology was selected for the production of the mechanism parts, and the additive manufacturing process was illustrated. The assembled robotic grippers were tested and are now ready for integration into the educational process. An additional advantage of the proposed design and manufacturing technology is its applicability across various levels of education. In future work, the proposed gripper design will be integrated with robotic modules to create complete robotic systems.

Acknowledgement

This research was funded by Karlovac University of Applied Sciences through an internal project, Development of innovative educational robot modules (iEduBot).

7 REFERENCES

[1] Hernandez-de-Menendez, M., Escobar Díaz, C. A., & Morales-Menendez, R. (2020). Engineering education for smart 4.0 technology: a review. *International Journal on Interactive Design and Manufacturing, 14*, 789–803. https://doi.org/10.1007/s12008-020-00672-x

[2] Long, Z., Jiang, Q., Shuai, T., Wen, F., & Liang, C. (2020). A Systematic Review and Meta-analysis of Robotic Gripper. *IOP Conference Series: Materials Science and Engineering, Resource Exploration and Mining Engineering, 782*, 042055,

https://doi.org/10.1088/1757-899X/782/4/042055
[3] Tai, A., Chun, M., Gan, Y., Selamet, M., & Lipson, H. (2021). PARA: A one-meter reach, two-kg payload, three-DoF open source robotic arm with customizable end effector. *HardwareX*, 10. https://doi.org/10.1016/j.ohx.2021.e00209
[4] Wang, C., Zhou, Y., Xie, B., Xie, J., & Zhang, J. (2022). An Underactuated Universal Gripper: Design, Analysis, and Experiment. *Energies, 15*, 9151. https://doi.org/10.3390/en15239151
[5] Fiaz, U. A., Abdelkader, M., & Shamma, J. S. (2018). An Intelligent Gripper Design for Autonomous Aerial Transport with Passive Magnetic Grasping and Dual-Impulsive Release. *IEEE/ASME International Conference on Advanced Intelligent Mechatronics (AIM)*, Auckland, New Zealand, 1027-1032. https://doi.org/10.1109/AIM.2018.8452383
[6] Dai, Y., Xiang, C., Qu, W., & Zhang, Q. (2022). A Review of End-Effector Research Based on Compliance Control. *Machines, 10*, 100. https://doi.org/10.3390/machines10020100
[7] Romano, J.M., Hsiao, K., Niemeyer, G., Chitta, S., & Kuchenbecker, K. J. (2011). Human-Inspired Robotic Grasp Control With Tactile Sensing. *IEEE Transactions on Robotics, 27*(6), 1067-1079. https://doi.org/10.1109/TRO.2011.2162271
[8] Hassan, A., & Abomoharam, M. (2014). Design of a Single DOF Gripper based on Four-bar and Slider-crank Mechanism for Educational Purposes. *Procedia CIRP, 21*, 379-384. https://doi.org/10.1016/j.procir.2014.02.062
[9] Kang, L., Seo, J.-T., Kim, S.-H., Kim, W.-J., & Yi, B.-J. (2019). Design and Implementation of a Multi-Function Gripper for Grasping General Objects. *Applied Sciences, 9*, 5266. https://doi.org/10.3390/app9245266
[10] Kang, L., Kim, S.-H., & Yi, B.-J. (2021). Modeling, Design, and Implementation of an Underactuated Gripper with Capability of Grasping Thin Objects. *Machines, 9*, 347. https://doi.org/10.3390/machines9120347
[11] Cui, Y., An, X., Lin, Z., Guo, Z., Liu, X-J., & Zhao, H. (2024). Design and implementation of an underactuated gripper with enhanced shape adaptability and lateral stiffness through semi-active multi-degree-of-freedom endoskeletons. *The International Journal of Robotics Research, 43*(6), 873-896. https://doi.org/10.1177/02783649231220674
[12] Piljek, P., Kotarski, D., Šćuric, A., & Petanjek, T. (2023). Prototyping and Integration of Educational Low-Cost Mobile Robot Platform. *Tehnicki glasnik, 17*(2), 179-184. https://doi.org/10.31803/tg-20220714131724
[13] 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*. IntechOpen, Ch. 4. https://doi.org/10.5772/intechopen.102781
[14] Ma, R. R., Odhner L. U., & Dollar, A. M., (2013). A modular, open-source 3D printed underactuated hand. *IEEE International Conference on Robotics and Automation*, Karlsruhe, Germany, 2737-2743. https://doi.org/10.1109/ICRA.2013.6630954
[15] Tlegenov, Y., Telegenov, K., & Shintemirov, A., (2014). An open-source 3D printed underactuated robotic gripper. *IEEE/ASME 10th International Conference on Mechatronic and Embedded Systems and Applications (MESA)*, Senigallia, Italy, 1-6. https://doi.org/10.1109/MESA.2014.6935605
[16] Telegenov, K., Tlegenov, Y., & Shintemirov, A., (2015). A Low-Cost Open-Source 3-D-Printed Three-Finger Gripper Platform for Research and Educational Purposes. *IEEE Access*, 3, 638-647. https://doi.org/10.1109/ACCESS.2015.2433937
[17] Białek, M., & Rybarczyk, D. (2023). A Comparative Study of Different Fingertips on the Object Pulling Forces in Robotic Gripper Jaws. *Applied Sciences*, *13*, 1247. https://doi.org/10.3390/app13031247
[18] Paduraru, E., Dumitras, C.-G., Chitariu, D.-F., Horodinca, M., & Chifan, F. (2023). Research on Additive Technique Parameter Optimization for Robotic Gripper Construction. *Machines, 11*, 621. https://doi.org/10.3390/machines11060621
[19] Curkovic, P. (2024). Comparative Analysis of Topology Optimization Platforms for Additive Manufacturing of Robot Arms. *Designs, 8,* 98. https://doi.org/10.3390/designs8050098.
[20] Program preddiplomskog stručnog studija Mehatronika usklađen s HKO registrom i kvalifikacijom stručni/a prvostupnik/prvostupnica (baccalaureus/baccalaurea) inženjer/inženjerka mehatronike. Available online: https://vub.hr/hko-mehatronika/doc/e1/Program%20preddiplomskog%20stru%C4%8Dnog%20studija%20Mehatronika.pdf (accessed on 23 June 2023). (in Croatian)
[21] Šančić, T.; Brčić, M.; Kotarski, D., & Łukaszewicz, A. (2023). Experimental Characterization of Composite-Printed Materials for the Production of Multirotor UAV Airframe Parts. *Materials, 16,* 5060. https://doi.org/10.3390/ma16145060
[22] Kotarski, D., Piljek, P., & Šančić, T. (2025). Design and Development of Educational Modular Mobile Robot Platform. *Tehnicki glasnik, 19*(1), 1-8. https://doi.org/10.31803/tg-20221010113555

Authors' contacts:

Denis Kotarski, PhD
(Corresponding author)
Karlovac University of Applied Sciences,
Josipa Jurja Strossmayera 9, 47000, Karlovac, Croatia
denis.kotarski@vuka.hr

Alen Šćuric
University of Zagreb, Faculty of Mechanical Engineering and Naval Architecture,
Ivana Lučića 5, 10002 Zagreb, Croatia,
Trivium STEM Edu association
Ciglenica Zagorska 10, 49223 Sveti Križ Začretje, Croatia
alenscuric9@gmail.com

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-20250418163132

Professional paper
Received: 2025-04-18, Accepted: 2025-05-02

Analysis of the Possibilities of Applying Artificial Intelligence to Modeling and 3D Printing of Drone Elements from Polymeric Materials

Grzegorz Budzik*, Łukasz Przeszłowski, Andrzej Paszkiewicz, Tomasz Dziubek, Tomasz Lis, Marek Magniszewski

Abstract: The subject of the article is to present the possibilities of using artificial intelligence for modeling and 3D printing of drone structural elements from polymer materials. The traditional three-dimensional modeling process based on 3D-CAD systems and the modeling method using artificial intelligence algorithms were analyzed. The project implementation time, design assumptions and accuracy of models obtained using both methods of creating numerical models were compared. Test models produced by 3D printing from polymer materials based on previously prepared numerical data were also presented.

Keywords: Artificial Intelligence (AI); drone; polymer materials; 3D modelling; 3D printing

1 INTRODUCTION

The term "drone" is a commonly accepted term for an unmanned aerial vehicle (UAV) or an unmanned aircraft system (UAS). The above names refer to an aircraft that does not have a pilot, crew or passengers on board [1]. Objects defined in this way can be remotely controlled or constitute autonomous systems. The term drone appeared in the 1920s and 1930s in military technology in solutions for which the presence of a human on board was unnecessary or dangerous. The first solutions concerned target aircraft intended for shooting practice. Currently, with the development of technology, the use of drones has expanded to many non-military areas. Drones are widely used in photography, broadly understood monitoring (agriculture, forest areas, power lines, etc.), inspection and control, and even product delivery. The commonness and availability of modern solutions means that they have even become elements of entertainment [2]. Drone races can serve as an example. Depending on the type of mission performed, drones can be fixed-wing or rotary-wing aircraft [5]. The wide use of unmanned systems has become possible thanks to the miniaturization of electronics and the introduction of modern polymer construction materials. Composites consisting of a polymer matrix and reinforcement in the form of glass, carbon or aramid fibers enable the serial production of elements with favorable stiffness and strength parameters at low weight [14]. Another direction of development concerns the practically unlimited possibilities of shaping structural elements using additive manufacturing methods [11].

Unmanned aerial vehicles are currently an important element of transport and logistics in many sectors of the economy, and they play a particularly important role in the area of offensive and defensive military technologies [15]. The production of drones can be based on the use of serial components, but prototype solutions manufactured using 3D printing can also be successfully used [3]. In such a case, special-purpose drones with a structure dedicated to specific tasks can be designed and built [4]. Additive technologies allow for a fairly large degree of freedom in shaping the drone's structure, and various methods of optimizing this structure can also be used in terms of its strength and mass [6]. The classic approach to designing a structure takes into account its strength properties, which are the basis for developing a model and prototype [7]. Additive technologies allow for the production of complex geometric shapes and structures, which is why topological optimization can be used in the design process [13]. A separate design issue is the possibility of using programs using artificial intelligence algorithms for three-dimensional modeling of objects. Such programs have recently experienced dynamic development, especially in the area of modeling objects for which maintaining specific dimensions or manufacturing tolerances is not important, such as: building models, human and animal figures, etc. [8].

When analyzing the issue of potential application of AI in the process of 3D modeling and printing, various areas can be distinguished. One of them is the selection of materials for printing [12, 13], in order to optimize the matching of material properties to the needs of the application and working conditions of a given drone. Such an approach can take into account various structural, environmental or mechanical requirements and at the same time correlate them with the mechanical properties of a wide range of polymers. Considering the reduction of the time of the entire design and manufacturing process, and at the same time reducing the costs associated with defective components, prediction and control of print quality play a very important role [9]. Especially in this area, AI can control the course of the 3D printing process in real time. For this purpose, both AI models identifying anomalies and deformations in the structure of the printed object using advanced image analysis, as well as AI tools processing sensory data from the environment and the printed object (e.g. temperature, humidity, growth rate of the printed object, etc.) can be used [10]. Another area of AI application is simulations. Trained models based on theoretical, catalog, experimental and simulation data can improve the process of optimizing strength and aerodynamic properties. In such a case, Big Data can support the prediction of the behavior of the developed structures in various operating conditions. As can be seen, the presented issue is very broad [11]. Therefore, this work

focuses on the use of software based on artificial intelligence algorithms, which allows for modeling objects based on entered data in the form of descriptive (text) or raster images (e.g. photos or screenshots). Entering data in text form is limited by the number of characters, which in turn are responsible for the detailing of the data. This detail does not always translate into the mapping of these details in the geometry of the developed model and geometric accuracy. A very detailed description of the object can introduce interference in the model generation process in the form of additional elements that may not be geometrically and programmatically consistent with the assumptions of the object's geometry. Taking the above into account, the issue of entering input data as an object description may constitute a separate research issue.

2 DESIGNING DRONE ELEMENTS USING 3D-CAD MODELLING

Modeling objects using 3D-CAD software allows for the creation of almost any three-dimensional objects, which can then be manufactured using production systems based on numerically controlled machining. In the case of structural elements manufactured using 3D printing used to build drones using multi-rotor drives, the design of the structure must take into account the structure and dimensions of individual elements such as the engine, propeller, control system or batteries. An example of a drone shown in Figure 1 manufactured using 3D printing can be an unmanned ship made as part of drone construction research at the Department of Machine Design of the Rzeszów University of Technology and the implementation of diploma theses [5].

Figure 1 Drone prototype made using 3D printing [5]

The dimensions and geometry of the drone must be adapted to the capabilities of the 3D printer. For 3D printers with a large workspace, drone bodies can be manufactured in their entirety, without the need to divide them into component parts.

The drone prototype shown in Fig. 1 was made of components, the housing of the control system and battery, and the arms mounting the engines. All the elements were connected to each other using screws. The use of a structure consisting of several elements was dictated by the use of a 3D printer for production with a workspace that does not allow for printing the entire body. The use of printers with a larger workspace allows for modeling the body elements as a whole. An example of such an approach is the model of the drone body made using 3D-CAD CATIA software (Fig. 2).

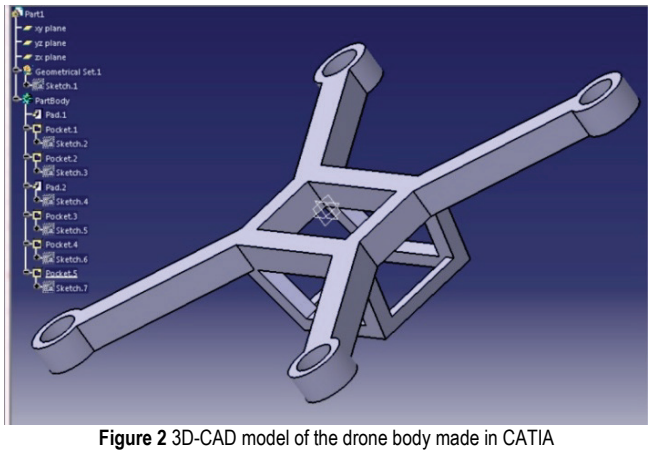


Figure 2 3D-CAD model of the drone body made in CATIA

Rotor drones do not have to meet special aerodynamic requirements, because the lift force and flight trajectory control are implemented using the engine control system. Taking this into account, the geometry of such a drone can be relatively simple, it can be made in the process of solid modeling and does not require the use of special methods of surface modeling, which in most cases takes more time than solid modeling. To make a drone body using the 3D printing method, data must be prepared in a form readable by the printer software, e.g. in STL format (Fig. 3). Geometric data in this form is generated based on the solid model.

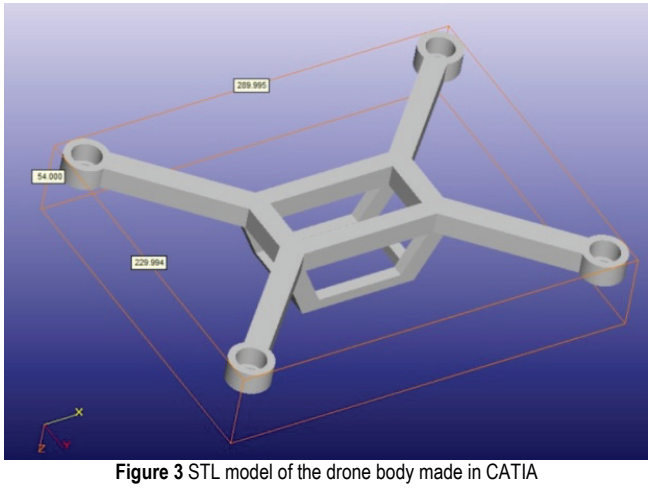


Figure 3 STL model of the drone body made in CATIA

Analyzing the complexity of the structure and the applied 3D-CAD modeling procedures, the time of modeling and data preparation for the drone body shown in Figs. 2 and 3 was recorded, which amounted to 44 minutes in total. The presented numerical model was used to make a 3D print of

the body from PLA polymer material using MEX process, FFF technology (Fig. 4).

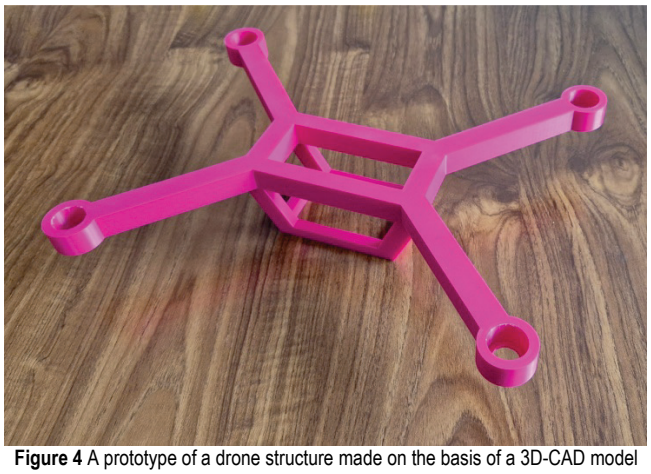


Figure 4 A prototype of a drone structure made on the basis of a 3D-CAD model using the 3D printing method

3D-CAD modeling allows for taking into account a number of design assumptions, while the accuracy of the numerical model is high, in most engineering programs it is within the value of 0.001 mm, which should be considered sufficient both in the use of 3D printing technology and machining on CNC machines. An important feature of 3D-CAD modeling is the ability to quickly modify the drone's structure in the output file area and adapt the structure to individual applications and operational activities, e.g. transporting unusual loads.

3 DESIGNING DRONE ELEMENTS USING ARTIFICIAL INTELLIGENCE-BASED PROCEDURES

As part of the conducted research, commercially available applications for modeling three-dimensional objects based on artificial intelligence algorithms Meshy AI 3D Model Generator were used. Two methods of entering data as information about the designed model were selected, i.e. the descriptive (text) method and the method of generating a model based on a photograph of an existing object. In the case of the descriptive method, several variants of the description were introduced with varying degrees of detail and type of language. For modeling based on the analysis of photographs, a photo of a four-engine drone previously developed at the Rzeszow University of Technology, shown in Fig. 1, was used. The photo in JPG format was loaded into the software based on AI algorithms and the process of generating a 3D model was started. After about 2 minutes, a three-dimensional numerical model was generated, shown in Figure 5. This drawing, which is a screenshot from the program, shows some similarities to the model saved on a digital photograph. The software used allows for the generation of many versions of models, however, due to the volume of the article, the one closest to the original was selected. The developed numerical model can be saved in the form of files in various formats, e.g. STL, which allows for 3D printing (Fig. 6).

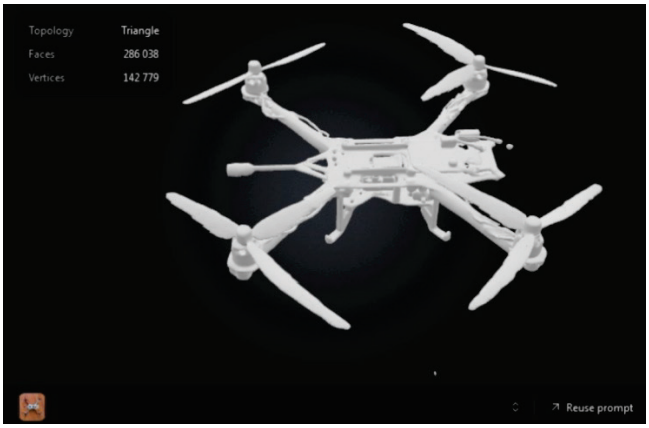


Figure 5 3D drone model generated using AI based on photography

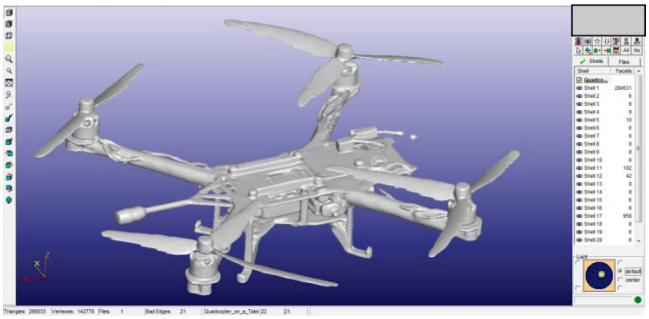


Figure 6 STL model generated using AI based on photography

Analyzing the STL model shown in Figure 6, it can be seen that it consists of several dozen objects generated as open and closed triangle meshes. From the point of view of the manufacturing process based on 3D printing, making a physical prototype requires the use of closed mesh models. Fragments of meshes are most often read by the software controlling the 3D printer as file defects that prevent the 3D printing process from starting. For this reason, the physical prototype shown in Figure 6 of the STL model was not made. In order to develop the drone construction model, the descriptive method was also used, for which the following text was entered into the program window: “supporting structure of a quadrocopter drone with the attachment of four electric motors, battery and control system attachment”. The program generated a model of a drone equipped with a drive and control system, although the description content presented the task of making a model of the supporting structure itself (Fig. 7). Additionally, one can see irregularities in the model geometry or model defects in the form of propeller blades with random geometry. Considering the disruptions of the developed geometry, in this case, it was decided not to make a 3D print for this model. The algorithm, based on its own assessment, determined the importance of individual words that make up the model description, and the presented model shows that it used the key word quadrocopter to describe a four-engine drone. The generation time of the model shown in the figure did not exceed 2 minutes.

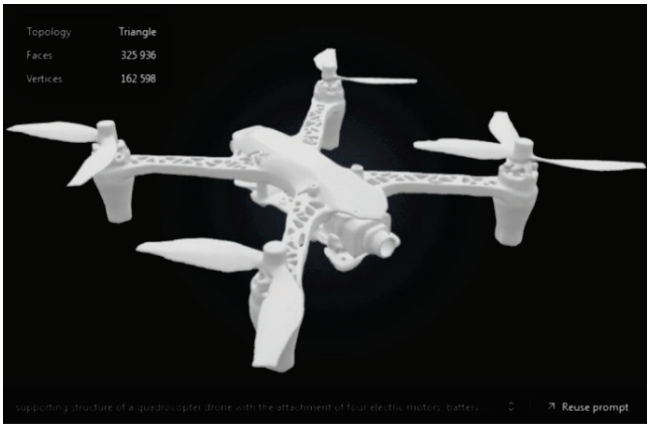


Figure 7 3D model of a drone generated using AI based on text description

An additional attempt was made to modify the data in terms of shortening the text description in the form of: "supporting structure of a quadrocopter drone". The view of this variant of the model generated using artificial intelligence is presented in Fig. 8.

Figure 8 3D model of the drone generated using AI based on a shortened text description

This model was then saved in STL format (Fig. 9) in order to test the production of a physical prototype using 3D printing.

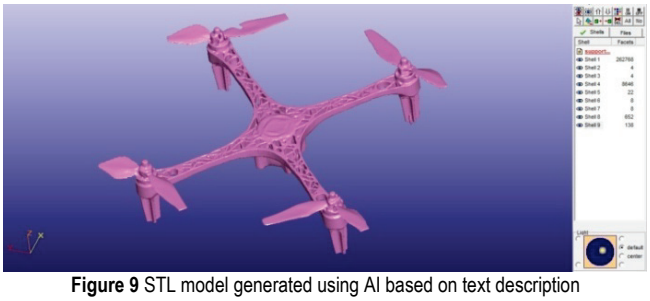


Figure 9 STL model generated using AI based on text description

On this basis, a 3D print was made from PLA polymer material using the MEX process and FFF technology. The prototype view is shown in Fig. 10.

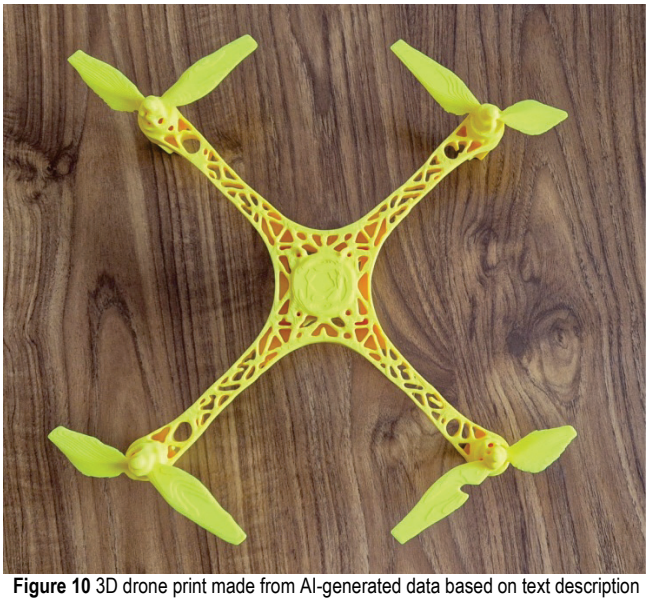


Figure 10 3D drone print made from AI-generated data based on text description

The presented model has an irregular shape, which is unacceptable for technical objects. The automatic modeling process introduced design assumptions aimed at reducing the model's mass by using a skeleton structure. However, discontinuities in this structure are visible, which will negatively affect its mechanical, static and fatigue strength. Additionally, the possibility of changing the language from English to Polish was analyzed and the description of the object was formulated as an unmanned aerial vehicle in Polish: "bezzałogowy statek powietrzny". Based on such a description, the four-engine drone shown in Fig. 11 was generated. It should be noted here that the text did not assume the number of engines used, but the program assumed that the model would have four engines, perhaps it was a kind of continuation of previous activities, which assumed the number of engines in the description.

Figure 11 STL model generated using AI based on text description in Polish

The generated 3D model was saved as an STL file for 3D printing purposes and, as can be seen in Fig. 12, four geometrically flawless mesh objects appeared in the dialog box on the right (shell 1 – Shell 4), which could be used directly to make a prototype using the 3D printing method.

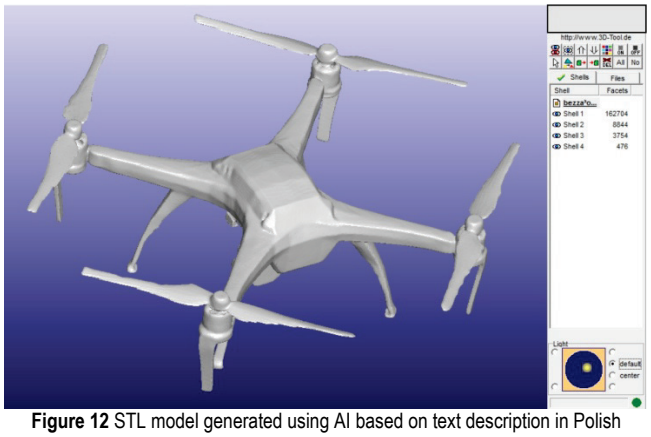


Figure 12 STL model generated using AI based on text description in Polish

Fig. 13 shows a prototype made as before in the MEX additive process, using FFF technology from PLA material.

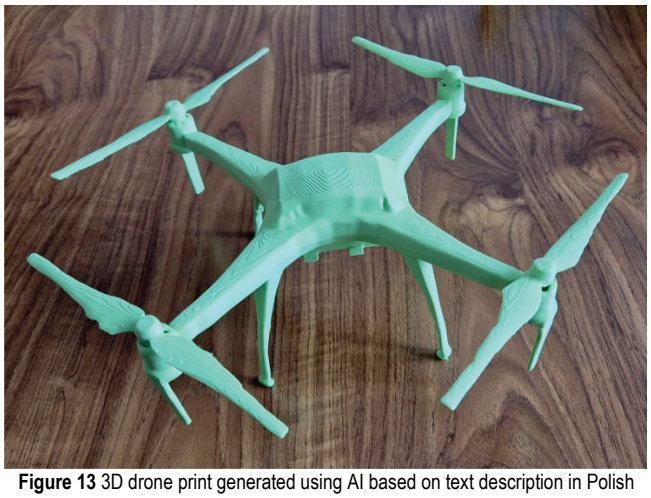


Figure 13 3D drone print generated using AI based on text description in Polish

4 CONCLUSIONS

As part of the work carried out, a total of several models were generated using software based on artificial intelligence algorithms Meshy AI 3D Model Generator. The time of generating subsequent models based on the AI algorithm was similar and did not exceed two minutes on average. On the other hand, the time of making the model using 3D-CAD software in the traditional way took about 45 minutes. Comparing only the time values, it can be seen that the AI algorithm generates a three-dimensional model very quickly. However, attention should be paid to the quality of the obtained geometric data from the engineering point of view. In general, the models generated by the AI algorithm meet the assumptions entered into the program as descriptive or raster input data. Objects were generated that geometrically meet the assumptions of its description of a four-engine drone. Models generated using AI are burdened with a large amount of geometry interference. Models generated using AI did not meet the technical assumptions of the engineering project, both in terms of the accuracy of the model execution and any definition of dimensions, especially in relation to the assembly of the drive system or control. Additionally, the software using AI allows the geometric data of models to be saved in formats that are difficult to edit and introduce design changes. Ultimately, it may translate into a significant extension of the production time of a 3D model, often exceeding the modeling time using 3D-CAD software.

An engineering project is implemented based on an assumed methodology defining the purpose of the object, strength, materials used, geometric accuracy and method of execution. Each of these stages is controlled by the designer. In the case of using an AI algorithm for design, the software accepts the description as input data without the possibility of using feedback. It is also difficult to determine in the case of using a descriptive algorithm which words the algorithm will consider important from the point of view of the project and which will be treated as secondary. This is an important issue in the case of using AI for design, but it goes beyond the scope of this study.

An important criterion for using the method of modeling three-dimensional objects is their purpose. If we intend to develop a preliminary conceptual model, it is reasonable to use software that generates geometry using AI algorithms. In the case of work on the design of an engineering structure and a functional prototype, classic 3D-CAD modeling methods should be used, which allow the creation of a precise numerical model intended for production using the 3D printing method or other technologies based on numerical data.

5 REFERENCES

[1] Altschuld, J. (2021). *Drone technology in architecture, engineering and construction*. John Wiley & Sons Inc.

[2] Alyassi, R., Khonji, M., Karapetyan, A., Chau, S. C. K., Elbassioni, K., & Tseng, C. M. (2023). Autonomous recharging and flight mission planning for battery-operated autonomous drones. *IEEE Transactions on Automation Science and Engineering, 20*(2), 1034–1046. https://doi.org/10.1109/TASE.2022.3175565

[3] Budzik, G., Tomaszewski, K., & Soboń, A. (2022). Opportunities for the application of 3D printing in the critical infrastructure system. *Energies, 15*(5), 1656. https://doi.org/10.3390/en15051656

[4] Cui, J., Ren, L., Mai, J., Zheng, P., & Zhang, L. (2022). 3D printing in the context of cloud manufacturing. *Robotics and Computer-Integrated Manufacturing, 74*, 102256. https://doi.org/10.1016/j.rcim.2021.102256

[5] Dziedzic, K. (2024). *Influence of propeller geometry on the flight properties of an unmanned aircraft* (Diploma thesis, supervised by Łukasz Przeszłowski). Rzeszow University of Technology.

[6] Erps, T., et al. (2021). Accelerated discovery of 3D printing materials using data-driven multiobjective optimization. *Sci. Adv., 7*(42). https://doi.org/10.1126/sciadv.abf7435

[7] Gibson, I., Rosen, D., Stucker, B., & Khorasani, M. (2021). Materials for additive manufacturing. In *Additive manufacturing technologies* (pp. 379–428). Springer. https://doi.org/10.1007/978-3-030-56127-7_14

[8] Goh, G. D., Hamzah, N. M. B., & Yeong, W. Y. (2023). Anomaly detection in fused filament fabrication using machine learning. *3D Printing and Additive Manufacturing, 10*(3), 428–437. https://doi.org/10.1089/3dp.2021.0231

[9] ISO/ASTM. (2021). *ISO/ASTM 52900: Additive manufacturing — General principles — Fundamentals and vocabulary*.
[10] Jadhav, Y., Pak, P., & Farimani, A. B. (2024). LLM-3D Print: Large language models to monitor and control 3D printing. *arXiv*. https://doi.org/10.48550/arXiv.2408.14307
[11] Niu, M. C. (2006). *Composite airframe structures*. Conmilit Press Ltd.
[12] Wohlers Associates. (2024). *Wohlers report 2024: Analysis. Trends. Forecasts. 3D printing and additive manufacturing state of the industry*.
[13] Woźniak, J., Budzik, G., Przeszłowski, Ł., Fudali, P., Dziubek, T., & Paszkiewicz, A. (2022). Analysis of the quality of products manufactured with the application of additive manufacturing technologies with the possibility of applying the Industry 4.0 conception. *International Journal for Quality Research, 16*(3), 831–850. https://doi.org/10.24874/IJQR16.03-12
[14] Zhang, L., Luo, X., Ren, L., Mai, J., Pan, F., Zhao, Z., & Li, B. (2020). Cloud based 3D printing service platform for personalized manufacturing. *Science China Information Sciences, 63*, 124201. https://doi.org/10.1007/s11432-018-9942-y
[15] Shuaibu, A. S., Mahmoud, A. S., & Sheltam, T. R. (2025). A review of last-mile delivery optimization: Strategies, technologies, drone integration, and future trends. *Drones, 9*(3), 158. https://doi.org/10.3390/drones9030158

Authors' contacts:

Grzegorz Budzik, Prof. PhD, DSc, Eng.
(Corresponding author)
Rzeszów University of Technology,
Al. Powstańców Warszawy 12, 35-959 Rzeszów, Poland
gbudzik@prz.edu.pl

Łukasz Przeszłowski, PhD., DSc. Eng. Assistant Prof.
Rzeszów University of Technology,
Al. Powstańców Warszawy 12, 35-959 Rzeszów, Poland
lprzesz@prz.edu.pl

Andrzej Paszkiewicz, PhD., DSc. Eng. Assistant Prof.
Rzeszów University of Technology
Al. Powstańców Warszawy 12, 35-959 Rzeszów, Poland
andrzejp@prz.edu.pl

Tomasz Dziubek, PhD., DSc. Eng. Associate Prof.
Rzeszów University of Technology
Al. Powstańców Warszawy 12, 35-959 Rzeszów, Poland
tdziubek@prz.edu.pl

Tomasz Lis, PhD., DSc. Eng. Assistant Prof.
Rzeszów University of Technology,
Al. Powstańców Warszawy 12, 35-959 Rzeszów, Poland
list@prz.edu.pl

Merek Magniszewski, PhD., DSc. Eng. Associate Prof.
Rzeszów University of Technology,
Al. Powstańców Warszawy 12, 35-959 Rzeszów, Poland
magniszewski@prz.edu.pl

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

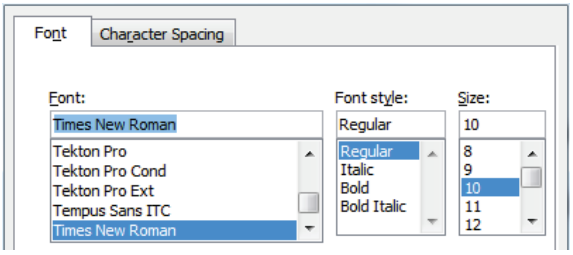


Figure 1 Text under the figure [1]
(Style: Arial Narraw, 8pt, Align Centre)

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

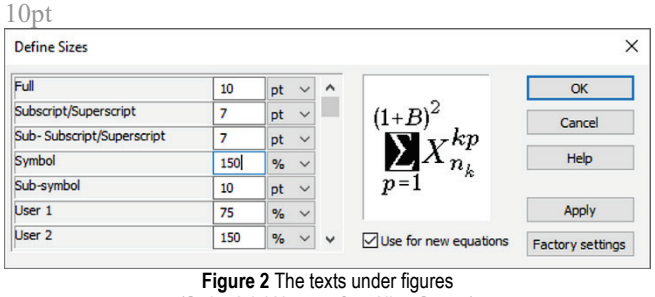


Figure 2 The texts under figures
(Style: Arial Narraw, 8pt, Align Centre)

10pt

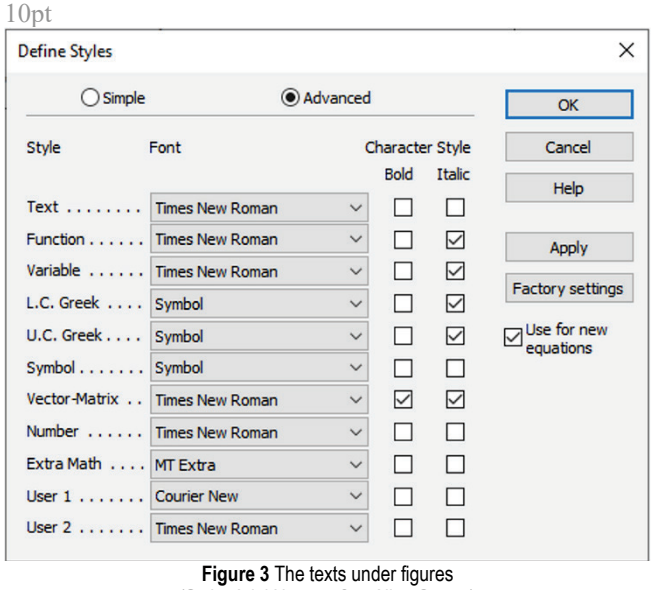


Figure 3 The texts under figures
(Style: Arial Narraw, 8pt, Align Centre)

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.

Oberwolfach Meetings 2027

Mathematisches Forschungsinstitut Oberwolfach

Member of Leibniz Association

Workshops

Participants of the Workshops at Oberwolfach are invited personally by the Director of the Institute. Participation is subject to such an invitation. Interested researchers, in particular early career scientists, can contact the administration of the Institute. Since the number of participants is restricted, not all inquiries can be considered.

03.01. – 08.01.2027 ID: 2701

Moduli Theory in Differential Geometry and Analysis
B. Hanke (Augsburg), I. Alonso Lorenzo (Piscataway), C. Mantoulidis (Houston), M. Upmeier (Aberdeen)

10.01. – 15.01.2027 ID: 2702

Geometry of Motives
S. Gille (Edmonton), C. Ravi (Mumbai), P. Sechin (Regensburg), A. Vishik (Nottingham)

17.01. – 22.01.2027 ID: 2703

Wave-Kinetic Interfaces: Dispersive Dynamics Meets Kinetic Theory
M. Iacobelli (Zürich), N. Pavlovic (Austin), G. Staffilani (Cambridge MA), N. Visciglia (Pisa)

24.01. – 29.01.2027 ID: 2704

Geometric Methods in Data Science
M. Bauer (Tallahassee), X. Pennec (Sophia Antipolis), S. Sommer (København), G. Steidl (Berlin)

31.01. – 05.02.2027 ID: 2705

Non-commutative Function Theory: A Link between Operator Theory and Random Matrices
M. Banna (Abu Dhabi), G. Cébron (Toulouse), M. Hartz (Saarbrücken), J. E. McCarthy (St Louis)

07.02. – 12.02.2027 ID: 2706a

Mahler Measures and L-Functions
M. Lalín (Montréal), D. van Straten (Mainz), W. Zudilin (Nijmegen)

07.02. – 12.02.2027 ID: 2706b

Lagrangian Torus Fibrations
M. Abouzaid (Stanford), E. Mazzon (Paris), H. Ruddat (Stavanger)

14.02. – 19.02.2027 ID: 2707

Scientific Machine Learning: Mathematical Foundations and Algorithmic Advances
M. Han Veiga (Columbus), A. Heinlein (Delft), C. B. Schönlieb (Cambridge UK), J. Zech (Heidelberg)

21.02. – 26.02.2027 ID: 2708

Hypoelliptic Operators in Geometry
W. Bauer (Hannover), X. Ma (Tianjin/Paris), N. Savale (Dublin), Y. Wang (Baltimore)

28.02. – 05.03.2027 ID: 2709

Statistical Analysis of Black Box Models
P. Bartlett (Berkeley), S. Langer (Bochum), J. Rousseau (Paris), S. Wang (Berlin)

Proposals invited

Mini-Workshops

01.08. – 06.08.2027 (2731a/b)
deadline for appl.: 31.01.2027

08.08. – 13.08.2027 (2732a/b)
deadline for appl.: 31.01.2027

15.08. – 20.08.2027 (2733a/b)
deadline for appl.: 31.01.2027

22.08. – 27.08.2027 (2734a/b)
deadline for appl.: 31.01.2027

Time slots for four further Mini-Workshops in winter 2027 will be announced on our website in 2026.

Tandem- or Mini-Workshops

05.09. – 10.09.2027 (2736a/b)
deadline for appl.: 01.09.26/31.01.27

Oberwolfach Seminars

16.05. – 21.05.2027 (2720a/b)
to be announced in winter 2026

24.10. – 29.10.2027 (2743a/b)
to be announced in winter 2026

21.11. – 26.11.2027 (2747a/b)
to be announced in winter 2026

Arbeitsgemeinschaft (study group)

14.03. – 19.03.2027 (2711)
to be announced in spring 2026

10.10. – 15.10.2027 (2741)
to be announced in fall 2026

17.10. – 22.10.2027 (2742),
to be announced in fall 2026

07.03. – 12.03.2027 ID: 2710

Representations of Finite Groups
O. Dudas (Marseilles), R. Kessar (Manchester), B. Späth (Wuppertal), P. H. Tiep (Rutgers)

21.03. – 26.03.2027 ID: 2712

Homotopy Theory
J. Grodal (København), M. Hill (Minneapolis), B. Richter (Hamburg)

28.03. – 02.04.2027 ID: 2713

Generalized Computability Theory
J. P. Aguilera (Wien), N. Greenberg (Wellington), B. Löwe (Hamburg/Cambridge UK), M. I. Soskova (Madison)

04.04. – 09.04.2027 ID: 2714

Aperiodic Order: Spectral Theory, Harmonic Analysis and Dynamical Systems in Interaction
D. Damanik (Houston), D. Lenz (Jena), N. P. Frank (Poughkeepsie), R. Yassawi (London)

Check online!

11.04. – 16.04.2027 ID: 2715

Geometric, Algebraic and Topological Combinatorics
M. Juhnke (Osnabrück), E. Nevo (Jerusalem/Valladolid), F. Santos (Santander), V. Welker (Marburg)

18.04. – 23.04.2027 ID: 2716

Structure Preserving Methods for Computational Geometric Mechanics
E. Celledoni (Trondheim), S. Ober-Blöbaum (Paderborn), A. Palha (Delft), A. Stern (St Louis)

25.04. – 30.04.2027 ID: 2717

Mathematical Analysis of Fluid Dynamics Models
I. Kukavica (Los Angeles), A. Mazzucato (University Park), E. S. Titi (Cambridge UK), E. Wiedemann (Erlangen)

02.05. – 07.05.2027 ID: 2718

Statistical Physics and Random Surfaces
C. Garban (Lyon), J. Miller (Cambridge UK), S. Sheffield (Cambridge MA)

09.05. – 14.05.2027 ID: 2719

Complexity Theory
M. Bläser (Saarbrücken), I. Dinur (Princeton), R. O'Donnell (Pittsburgh), S. Vadhan (Harvard)

23.05. – 28.05.2027 ID: 2721

Interactions between Algebraic Geometry and Non-commutative Algebra
W. Crawley-Boevey (Bielefeld), O. Schiffmann (Orsay), S. Spenko (Bruxelles), M. Wemyss (Glasgow)

30.05. – 04.06.2027 ID: 2722a

Nonlinear Optics: Physics, Analysis, and Numerics
K. Busch (Berlin), R. El-Ganainy (St Louis), M. Hochbruck (Karlsruhe), C. Lasser (München)

30.05. – 04.06.2027 ID: 2722b

Methods of Mathematization around 1900
H. Junker (Bonn), D. Kent (St Andrews), J. Lorenat (Claremont)

06.06. – 11.06.2027 ID: 2723

Arithmetic of Shimura Varieties
K. Madapusi (Chestnut Hill), T. Richarz (Darmstadt), E. Viehmann (Münster), M. Zhang (Princeton)

13.06. – 18.06.2027 ID: 2724

Homological Commutative Algebra and Applications to Geometry
H. Brenner (Osnabrück), D. Eisenbud (Berkeley), E. Faber (Graz), C. Miller (Syracuse)

20.06. – 25.06.2027 ID: 2725

Recent Trends in Algebraic Geometry
G. Farkas (Berlin), M. Mustata (Ann Arbor), R. Pandharipande (Zürich), C. Voisin (Paris)

27.06. – 02.07.2027 ID: 2726

Mathematical Horizons in Artificial Intelligence
C. Boyer (Paris), M. Hein (Tübingen), R. Nowak (Madison), F. Yang (Zürich)

04.07. – 09.07.2027 ID: 2727

Differentialgeometrie im Grossen
O. Chodosh (Stanford), J. Lotay (Oxford), A. Lytchak (Karlsruhe), L. Wang (New Haven)

11.07. – 16.07.2027 ID: 2728

Dynamische Systeme
S. Crovisier (Gif-sur-Yvette), U. L. Hryniewicz (Aachen), T. M-Seara (Barcelona)

18.07. – 23.07.2027 ID: 2729

Discrete Nonlinear Optimization: Mathematics and Algorithms
J. Lee (Ann Arbor), R. Sotirov (Tilburg), A. Wiegele (Klagenfurt)

25.07. – 30.07.2027 ID: 2730

Partial Differential Equations
A. Fraser (Vancouver), T. Lamm (Karlsruhe), C. Mantoulidis (Houston), X. Ros-Oton (Barcelona)

29.08. – 03.09.2027 ID: 2735

Mathematical Aspects of General Relativity
C. Cederbaum (Tübingen), G. Holzegel (Münster), J. Luk (Stanford), T. Oliynyk (Clayton)

12.09. – 17.09.2027 ID: 2737

Homotopical Algebra and Higher Structures
T. Dyckerhoff (Hamburg), J. Holstein (Hamburg), M. Livernet (Paris), M. Markl (Prague)

19.09. – 24.09.2027 ID: 2738

Proof Complexity and Beyond
P. Beame (Seattle), M. Mahajan (Chennai/Mumbai), J. Nordström (Copenhagen/Lund), P. Pudlák (Praha)

26.09. – 01.10.2027 ID: 2739

Quantum Groups – Algebra, Analysis and Category Theory
S. Neshveyev (Oslo), D. Nikshych (Durham NH), J. Plavnik (Bloomington), A. Skalski (Warsaw)

03.10. – 08.10.2027 ID: 2740

Polytopal Finite Element Methods for PDEs: Theory, Algorithms, and Machine Learning Enhancement
P. Antonietti (Milano), L. Beirão da Veiga (Milano), J. Droniou (Montpellier), P. Wriggers (Hannover)

31.10. – 05.11.2027 ID: 2744

New Directions in Tomography: Theory and Applications
G. Ambartsoumian (Arlington), S. Arridge (London), B. Hahn-Rigaud (Stuttgart), A. Wald (Göttingen)

07.11. – 12.11.2027 ID: 2745

Applications of Optimal Transportation
G. Carlier (Paris), E. Radici (L'Aquila), F. Santambrogio (Lyon), B. Schmitzer (Göttingen)

14.11. – 19.11.2027 ID: 2746

Combinatorial Optimization
J. de Loera (Davis), V. Traub (Zürich), L. Végh (Bonn)

28.11. – 03.12.2027 ID: 2748

Applied Harmonic Analysis and Data Science
R. Alaifari (Aachen), A. Bandeira (Zürich), F. Krahmer (Garching), R. Ward (Austin)

05.12. – 10.12.2027 ID: 2749

Machine Learning in Infinite Dimensions
M. de Hoop (Houston), Y. Korolev (Bath), N. Mücke (Braunschweig), M. Thorpe (Coventry)

12.12. – 17.12.2027 ID: 2750

Convex Geometry and its Applications
S. Artstein-Avidan (Tel Aviv), D. Hug (Karlsruhe), E. Werner (Cleveland)

TEHNIČKI GLASNIK / TECHNICAL JOURNAL – GODIŠTE / VOLUME 20 – BROJ / ISSUE 3

RUJAN 2026 / SEPTEMBER 2026 – STRANICA / PAGES 343-510

Sveučilište Sjever

SVEUČILIŠTE SJEVER / UNIVERSITY NORTH – CROATIA – EUROPE

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

TEHNICKIGLASNIK@UNIN.HR – HTTP://TEHNICKIGLASNIK.UNIN.HR