

TEHNIČKI GLASNIK - TECHNICAL JOURNAL

Scientific-professional journal of University North

Volume 20
Varaždin, March 2026Issue 1
Pages 1-172**Editorial Office:**

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

Founder and Publisher:

Sveučilište Sjever / University North

Council of Journal:

Damir VUSIĆ, Chairman; Marin MILKOVIĆ, Member; Robert GEČEK, Member; Mario TOMIŠA, Member;
Marijana NEUBERG, Member; Danko MARKOVINOVIĆ, Member; Petar MIŠEVIĆ, Member

Editorial Board:**Domestic Members:**

Chairman Marin MILKOVIĆ (1), Milan KLJAJIN (1), Damir VUSIĆ (1), Krešimir BUNTAK (1), Anica HUNJET (1), Žiljka KONDIĆ (1), Goran KOZINA (1), Ljudevit KRPAN (1), Krunoslav HAJDEK (1), Marko STOJČIĆ (1), Božo SOLDIĆ (1), Mario TOMIŠA (1), Vlado TROPSA (1), Vinko VIŠNJIĆ (1), Sanja ŠOLIĆ (1), Dean VALDEC (1), Predrag PUTNIK (1), Petar MIŠEVIĆ (1), Duško PAVLETIĆ (5), Branimir PAVKOVIĆ (5), Mile MATIJEVIĆ (3), Damir MODRIĆ (3), Nikola MRVAC (3), Klaudio PAP (3), Ivana ŽILJAK STANIMIROVIĆ (3), Krešimir GRILEC (6), Biserka RUNJE (6), Sara HAVRLIŠAN (2), Dražan KOZAK (2), Roberto LUJIĆ (2), Leon MAGLIĆ (2), Ivan SAMARDŽIĆ (2), Antun STOIC (2), Katica ŠIMUNOVIĆ (2), Goran ŠIMUNOVIĆ (2), Ladislav LAZIĆ (7), Ante ČIKIĆ (1), Darko DUKIĆ (9), Gordana DUKIĆ (10), Srđan MEDIĆ (11), Sanja KALAMBURA (12), Marko DUNĐER (13), Zlata DOLAČEK-ALDUK (4), Dina STOBER (4)

International Members:

Boris TOVORNIK (14), Milan KUHTA (15), Nenad INJAC (16), Marin PETROVIĆ (18), Salim IBRAHIMEFENDIĆ (19), Zoran LOVREKOVIĆ (20), Igor BUDAK (21), Darko BAJIĆ (22), Tomáš HANÁK (23), Evgenij KLIMENKO (24), Oleg POPOV (24), Ivo ČOLAK (25), Katarina MONKOVA (26), Berenika HAUSNEROVA (8), Nenad GUBELJAK (27), Stefanija KLARIC (28), Bertrand MARESCHAL (29), Sachin R. SAKHARE (30), Suresh LIMKAR (31), Mandeep KAUR (32), Aleksandar SEDMAK (33), Han-Chieh CHAO (34), Sergej HLOCH (26), Grzegorz M. KRÓLCZYK (35), Djordje VUKELIC (21), Stanislaw LEGUTKO (17), Valentin POPOV (36), Dragan MARINKOVIC (36), Hamid M. SEDIGHI (37), Cristiano FRAGASSA (38), Dragan PAMUČAR (39), Imre FELDE (40), Levente KOVACS (40)

Editor-in-Chief:

Milan KLJAJIN

Technical Editor:

Milan KLJAJIN

Graphics Editor:

Snježana IVANČIĆ VALENKO

IT support:

Antonija MANDIĆ

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, Hrcak – Portal znanstvenih časopisa RH

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

Preparation ended:	Published (online):	Published (print):
February 6, 2026	February 23, 2026	March 13, 2026

Legend:

(1) University North, (2) University of Slavonki 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 Zlin, (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

CONTENT	I
G. K. Dayananda*, I. R. Manjunath, J. G. Sreerama Samartha, M. Vayusutha , D. N. Pratham, Sakshath, A. U. Shashank, Swasthi Prasad Shetty Hand Gesture Recognition System for Hearing and Speech Impaired People Using ML	1
T. Karuppusamy, K. Kathir Balaji, Senthil Kumar Madasamy, Beena Stanislaus Arputharaj, Parvathy Rajendran*, Arunkumar Karuppasamy, Vijayanandh Raja* Experimental Studies on Mechanical Characterizations of Silicon Nitride and Alumina added Al7075 Hybrid Composite for Various Engineering Applications	9
H. V. Srikanth, S. Vijay Kumar, R. Kousik Kumaar*, Senthil Kumar Madasamy, Beena Stanislaus Arputharaj, Parvathy Rajendran, Vijayanandh Raja* Effect of Rubber Seed Oil Biodiesel Additive on Compression Ignition Engines fuelled with Diesel-Ethanol Blends	16
Mostafa Abotaleb GLDM Algorithm for Big Data (SCADA) Wind Speed Modelling	23
Arash Shahin, Ehsan Masoudi*, Neda Rajabani The Influence of Green Technology Innovation and Supply Chain Learning on Sustainability Performance in SMEs	38
Darija Ćutić, Miroslav Mikota*, Maja Turčić Analysis of the Black and White Image Recording of the Converted RGB Photographic Image	47
Dalibor Viderščak*, Zdravko Schauerl, Sanja Šolić, Bojan Milovanović, Ivan Stanković, Mario Šerčer Determination of Thermal Characteristics of Fe-C Cast Iron Using Hot Disk Method	53
Dorijan Radočaj*, Daria Galić Subašić, Ivan Plaščak, Mladen Jurišić The Multicollinearity Effect on Machine Learning Accuracy for Leaf Chlorophyll Content Prediction of Indoor Plants	60
Konstantin P. Rukomojnikov*, Tatiana V. Sergeeva, Natalya N. Aleksagina Simulation Modeling of Labor Costs for Moving of Harvester between Working Positions in Cutting Area	66
Rrahim Sejdiu, Xhemajl Mehmeti*, Flamur Salihu, Labinot Topilla, Drin Krasniqi, Mohamad El Mehtedi, Agron Baraktari Impact of Infill Pattern Design on Stress-Strain behaviour of 3D Printed Parts	73
Sangchul Kim, Seunghyun Yeo, Seok-Heon Yun* Safety Management at Tower Crane in Construction Site	79
Ssang-Hee Seo Development of Drowsy Driving Detection System Using EEG	86
So-Yeon Jeon, Eui-Rim Jeong* Multi-Target Estimation in OFDM Radar using YOLOv8 for Integrated Sensing and Communication	94
Young Sook Park, Yun Kyoung Oh* Image Generation of Aesthetic Massage types Generative AI: Applying the Khizer Abbas Framework	101
Hyun Jung Kim, Sang Hyun Yoo* Real-Time Hybrid Query Transformation Method for Enhancing Search System Performance in Korean Language Applications	108
Guillaume Guerard*, Sonia Djebali, Maxime Hanus, Mark-Killian Zinenberg Research LLM: Using Large Language Models as a Research Companion	118
Abbas Fadhil Ibrahim*, Shakir M. Mousa, Semyan Omed Alyousifi, Diana Abed al Kareem Noori Experimental Investigation on Material Removal Rate and Tool Wear Rate for Machining Metal Matrix Composites in EDM Process	126
Angela Shvets*, Vladyslav Shaposhnyk, Leontii Muradian, Anzhelika Shvets, Alexei Reidemeister, Oleksandr Shykunov, Dmytro Molchanov Assessment of Forces Acting on the Gondola Car Body during Transportation of Rolled Wire Bundles	133
Attila Turi Fuelling Logistics from Refinery to Retail: Case Study of the Romanian Oil Supply Chain	141
Domagoj Zimmer*, Mladen Jurišić, Hrvoje Glavaš, Mirko Karakašić, Toni Beljo Application of Photogrammetry in Agriculture with an Emphasis on Close-Range	152
Amanda Wiana, Menul Teguh Riyanti, Virginia Suryani Setiadi, Bambang Tri Wardoyo, Muhammad Hafiz Hassan, Wegig Murwonugroho* The Crucial Role of Graphic Design in Shaping User Experience: A Study Case	158
Samuel O. Owwoeye*, Folasade O. Durodola, Babatunde M. Bisiriyu, Kehinde M. Adegbenro, Taofeek A. Adetoro Development of Solar-Powered Herbicide Spraying Robot Controlled Using a Mobile Application	166
INSTRUCTIONS FOR AUTHORS	III

AIR POLLUTION 2026

8-10 June 2026 | Bristol, UK

Organised by

Wessex Institute, UK
University of the West of England (UWE) Bristol, UK

Sponsored by

WIT Transactions on Ecology and the Environment



www.witconferences.com/air2026

Hand Gesture Recognition System for Hearing and Speech Impaired People Using ML

G. K. Dayananda*, I. R. Manjunath, J. G. Sreerama Samartha, M. Vayusutha, D. N. Pratham, Sakshath, A. U. Shashank, Swasthi Prasad Shetty

Abstract: Sign language has historically served as a primary form of communication for the deaf community. However, the lack of sign language proficiency among the general population and the limited availability of interpreters create barriers in communication between non-sign language speakers and deaf individuals. To address this issue, a proposed system aims to bridge this communication gap by assisting non-sign language speakers in recognizing and understanding American Sign Language (ASL) gestures. This system has the potential to enhance greatly communication between deaf individuals and the wider society. The process involves enhancing the quality and clarity of hand gestures by passing them through a filter. Subsequently, a classifier analyses the filtered hand gesture and predicts its corresponding meaning or category. This classification step enables non-sign language speakers to interpret accurately the intended message conveyed through these hand gestures. It is important to note that this method specifically focuses on recognizing the 26 letters of the ASL alphabet. Analysis of various datasets reveals that the enhanced network performs better than Faster R-CNN in terms of gesture categorization accuracy. Specifically, when using 175 photos for each move, the upgraded network achieves an 83% recognition accuracy, surpassing the performance of Faster R-CNN in gesture classification. Furthermore, with 600 images for each gesture, the improved network achieves an accuracy of 97. %. This finding suggests that employing a TOF camera to collect images based on depth information effectively reduces the influence from other elements during feature extraction. Overall, this system primarily focuses on recognizing fingerspelling gestures, which represent individual letters in ASL. It is commendable that efforts are being made to utilize neural networks and technology to enhance communication for individuals with disabilities.

Keywords: ASL; feature extraction; gesture recognition; hand gesture; image recognition; speech and hearing impaired

1 INTRODUCTION

Sign language has historically served as a primary form of communication within the deaf community. However, the limited knowledge of sign language among the general population, as well as the scarcity of interpreters, presents obstacles in bridging the communication gap between non-sign language speakers and deaf individuals. In an effort to address this issue, a proposed system aims to facilitate communication by enabling non-sign language speakers to recognize and comprehend American Sign Language (ASL) gestures [1]. This system has the potential to enhance greatly communication between deaf individuals and the wider society.

The methodology involves subjecting the hand gesture to a filtering process in order to improve its quality and clarity [2]. Subsequently, a classifier analyses the filtered hand gesture and predicts its corresponding meaning or category. This classification step allows non-sign language speakers to interpret accurately the intended message conveyed through the hand gestures. It is important to note that this methodology specifically focuses on recognizing the 26 letters of the ASL alphabet [3].

A visual language, sign language has three main components as shown in Fig. 1:

Manual Signs: Manual signs are the core elements of sign language. They involve the use of handshapes, movements, and orientations to convey meaning. Each sign represents a word, concept, or idea. Manual signs can be simple or complex, depending on the language and the specific sign being used.

Facial Expressions: Facial expressions play a crucial role in sign language as they provide grammatical and emotional information. They help convey nuances, such as questions, statements, negations, and emotions. Facial

expressions, including eye gaze, brow movement, and mouth patterns, add depth and clarity to the signs being produced.

Body Language and Postures: Body language and postures complement the manual signs and facial expressions in sign language. They contribute to the overall meaning and interpretation of the message. Body movements, such as the position of the torso, shoulders, and arms, along with head movements and body orientation, provide additional contextual information and convey the speaker's emphasis, location, and spatial relationships.

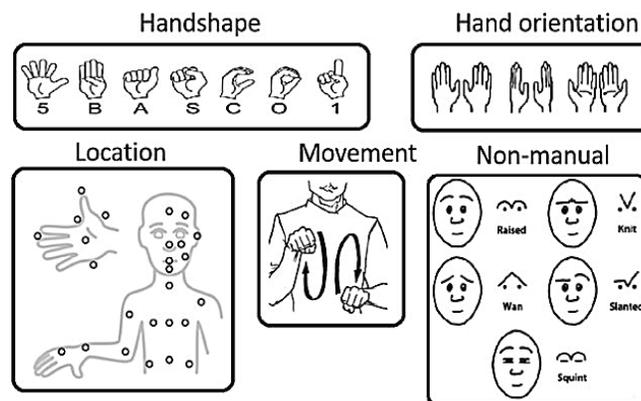


Figure 1 Sign Language major component [2]

Sign language is widely regarded as a vital means of communication and can be used to communicate with others who do not speak sign language, as well as people who have hearing or speech difficulties. Sign language, as opposed to spoken language, communicates meaning through physical gestures. It includes facial emotions, hand gestures, and movements of other body parts such as the eyes, legs, and so on. This study presents a design for recognising and interpreting ASL signs. Social contact, communication differences, education, behavioural issues, mental health

issues, and safety concerns are just a few of the challenges that speech and hearing-impaired people encounter while engaging with hearing-impaired people. A keyboard, mouse, or audio signal can be used to interface with a computer. The former demands constant physical contact, whilst the latter is sensitive to noise and distractions. A gesture is any physical movement done with the hand, eye, or other body part. The best motions for people to use and understand are hand gestures. Fig. 2 shows ASL Sign Language Gesture by hand.

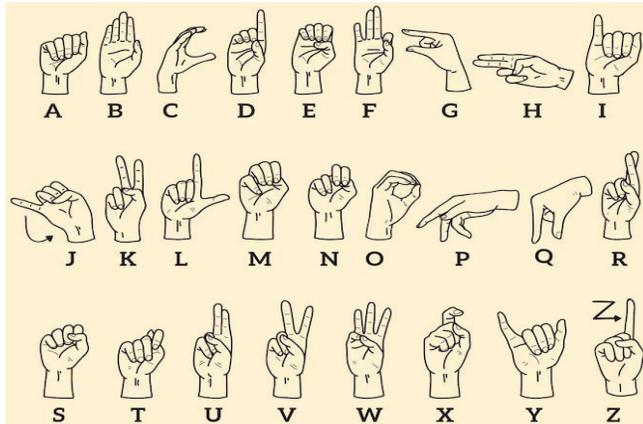


Figure 2 ASL Sign Language Gesture

2 LITERATURE SURVEY

In today's technological landscape, it is becoming increasingly common for electronic devices to be controlled without physical touch. This is primarily due to the rise of voice and gesture controls. Voice controls are capable of recognizing spoken words, enabling users to operate devices through verbal commands. On the other hand, gesture controls involve moving one's hand in specific patterns to execute desired tasks on the device [4]. Extensive research has been conducted to develop efficient and rapid hand gesture recognition systems [5]. Numerous experiments have been carried out to identify hand gestures, and these programs are often integrated into other applications, such as device control through hand gestures. Various methods exist for recognizing hand gestures from images.

Algorithms for recognizing hand gestures, such as multivariate fuzzy decision trees [6], Hidden Markov Models (HMMs) [7], dynamic time warping framework [8], latent regression forests [9], support vector machines, and surface electromyograms [10], are utilized to analyse and process these gestures. The captured movements are examined to identify occlusion and interactions involving close fingers, enabling the recognition of the correct gesture while disregarding intermittent gestures. Robust algorithms like HMM are necessary for real-time hand gesture detection that accurately identifies the intended gesture. The effectiveness of categorized gestures is evaluated using the Kth datasets and sign language alphabets as standard training and validation datasets. Hand gesture recognition plays a crucial role in various applications, including music orchestration, robotics, television control, rehabilitation, and sign language recognition.

This is the initial and most crucial step in the entire sign reformation process [11]. In order to capture photos using a webcam, camera interfacing is necessary. Nowadays, many laptops come with built-in cameras, which simplifies the process of capturing photos for later processing. Morphological operations involve shapes of images [12]. The two primary morphological processes are erosion and dilation, which are employed for noise reduction, element separation, element alignment, and even intensity fluctuation or gap detection in images. While dilation can fill in any gaps left by erosion or add pixels to boundary areas, erosion can reduce noise by widening gaps and reducing image boundaries. Dilation can also be used to link randomly placed pixels and add pixels to the boundaries. In order to construct a contour around the palm and identify finger positions, the convexity hull algorithm [13] is used in contour detection. Adaptive boosting is used in the convexity hull technique to detect hands, and the HAAR classifier algorithm [14] can be used to train classifiers. Segmenting the area of the image that contains the hand is the first step in the convexity hull algorithm. It is necessary to make some assumptions about the hand's characteristics, even if they could vary as the hand travels.

The gesture of the hand is classified based on the maximum score, which represents the minimum Euclidean distance between the input gesture and the dataset image of the corresponding character. The character associated with the input gesture is determined based on this minimum distance. It is important to note that the process described is a basic outline and does not include details of the actual segmentation and feature extraction techniques used. Depending on the complexity of the problem and the dataset, more advanced methods like those that deep learning-based approaches may be used for improved accuracy and efficiency. Additionally, the mention of "detecting human hands against white backgrounds" suggests a potential improvement in image pre-processing to enhance accuracy and speed by using a background subtraction technique to isolate the hand region. This could be beneficial when the images are consistently captured against a consistent background. Finally, the process involves two main steps: training and recognition [15]. During training, the model learns from a labelled dataset of hand gestures. During recognition, the learned model is used to classify new, unseen hand gestures based on their features and the calculated Euclidean distances.

To ensure accurate feature comparison and minimize errors during image processing, the system scales all the captured photographs to the same size. This step is essential when dealing with images captured using different camera resolutions, as it brings uniformity to the dataset and helps in achieving consistent results during recognition. Overall, the proposed system presents a comprehensive approach to assist disabled individuals in expressing themselves through sign language. By leveraging vision-based techniques [16] such as Histograms of Oriented Gradients (HOG) feature extraction and Euclidean distance-based classification, the system can recognize hand gestures and provide meaningful translations for communication.

3 SYSTEM REQUIREMENT

The identification of hand gestures has been the subject of extensive research in recent years. We discovered the following are the fundamental steps in hand gesture recognition with the aid of a literature review.

3.1 Relevance of the Work

Individuals who are deaf and unable to speak communicate using hand sign language. However, communicating with individuals who do not understand sign language can be challenging. To address this issue, we are currently developing a program that aims to facilitate communication between deaf individuals and those who do not understand sign language. Our program will detect and translate the hand signs made by the user into text, which will then be displayed on a screen for non-sign language users to read. By implementing this project, we aim to enhance communication capabilities for deaf individuals when interacting with non-sign language users.

3.1.1 Proposed Work

Different methods can be employed to gather data on hand gestures. One approach involves using electromechanical devices to capture accurately the configuration and position of the hand. Alternatively, glove-based methods can be utilized to extract information [17, 18], although these tend to be expensive and not user-friendly. Another option is to rely on vision-based techniques, where a computer camera serves as the input device to observe hand or finger movements. Vision-based methods offer a natural interface between individuals and computers, eliminating the need for additional hardware [19]. These systems, which can be implemented in software and/or hardware, simulate biological vision. By utilizing a vision-based strategy, the need for artificial devices in interaction is resolved, as solely hand movements represent all gestures. The design of the Gesture Recognition System is illustrated in Fig. 3.

To create own dataset using OpenCV [20], can follow these steps:

Install OpenCV: Install the OpenCV library on your system by following the installation instructions specific to your operating system.

Set up your camera: Connect a camera to computer or use the built-in camera of your device.

Capture frames: Use OpenCV to capture frames from the camera. Use the VideoCapture class to access the camera and retrieve frames.

Apply image processing: Apply the desired image processing techniques to the captured frames. In this case, Mentioned using a Gaussian blur filter and thresholding. OpenCV provides functions like GaussianBlur and threshold that you can use for these operations.

Extract features: After processing the image, extract relevant features from the processed image. The specific features you choose will depend on the problem trying to solve..

Pass processed image to CNN model: Extracted the features from the processed image, CNN (Convolutional Neural Network) model for prediction.

Letter detection and word formation: Keep track of the predicted letters over consecutive frames. If a letter is detected for more than 50 frames, consider it as a valid detection and store it. Whenever the letter changes or a blank symbol is detected, form a new word and store the previous word. Consider the space between words using the blank symbol.

Repeat the process: Continuously capture frames, apply image processing, pass the processed image to the CNN model, and detect and form words until collected sufficient dataset.

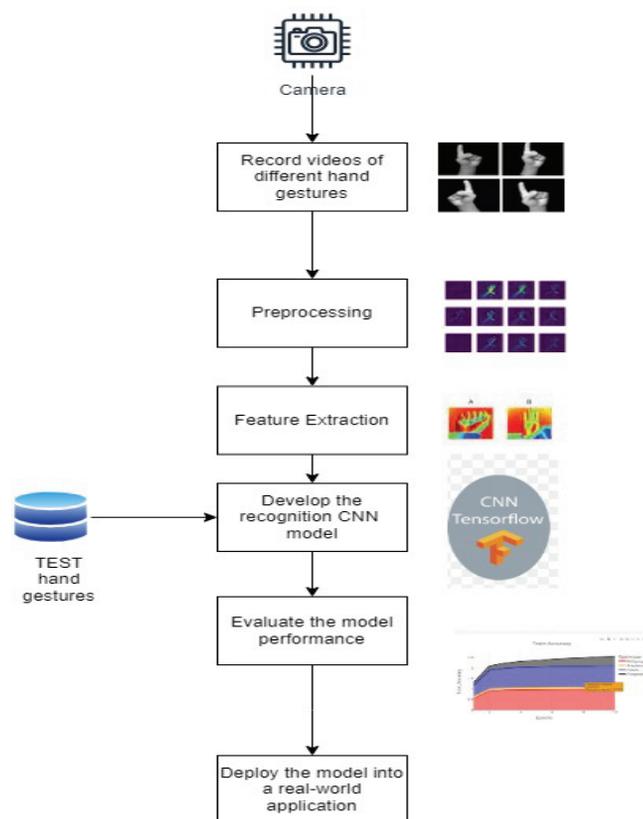


Figure 3 Flow diagram of Gesture Recognition System

3.2 Software Requirement

3.2.1 Artificial Neural Network

An Artificial Neural Network (ANN) is a computational model inspired by the structure and functioning of the human brain. It is a type of machine learning algorithm that can learn and adapt from examples, making it suitable for various tasks, including pattern recognition, classification, regression, and decision-making.

3.2.2 Convolutional Neural Networks (CNN) [22]

Convolution Layer: In this layer, small filters (also called kernels) with learnable parameters are applied to the input image. These filters move across the image to perform

convolutions (dot products) at different spatial positions. The output of this layer is a 2D activation map that represents the response of the filters at each position, capturing specific visual patterns or features like edges or colors.

Pooling Layer: After the convolution layer, the pooling layer is used to downsample the activation maps, reducing their spatial dimensions and the number of parameters. Two common types of pooling are max pooling and average pooling. In max pooling, a window (e.g., 2×2) slides over the activation map, and only the maximum value within each window is retained, effectively reducing the size of the map by half. In average pooling, the average of all values within the window is taken instead.

Fully Connected Layers: After several convolutional and pooling layers, the final stage of the CNN architecture usually involves fully connected layers (Fig. 4). These layers are similar to those in regular neural networks, where each neuron is connected to all neurons from the previous layer. The fully connected layers take the high-level abstract features learned by the earlier layers and combine them to make predictions about the input's class. The number of neurons in the output layer corresponds to the number of classes in the classification task, and the values of the output neurons represent the probabilities of the input belonging to each class.

CNNs have been very successful in computer vision tasks, such as image classification, object detection, and segmentation, because they can effectively learn hierarchical features from raw pixel data. The convolutional and pooling layers help capture local patterns and reduce the spatial dimensions of the data, while the fully connected layers make global decisions based on the extracted features.

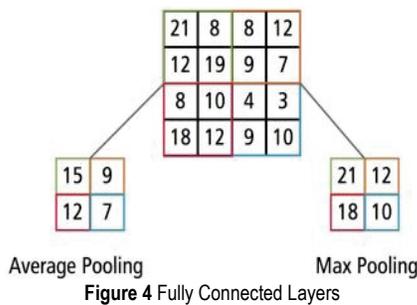


Figure 4 Fully Connected Layers

3.2.3 TensorFlow

TensorFlow is indeed an open-source software library that provides a framework for dataflow programming, especially well-suited for machine learning tasks involving neural networks. Tensor Flow's flexible design allows it to be deployed on different platforms seamlessly. This versatility makes it popular only not for research and development but for production-level machine learning applications. TensorFlow has been a cornerstone in the development and deployment of various machine-learning models, and it has played a significant role in advancing the field of deep learning [22]. Its ease of use, scalability, and compatibility with different hardware have made it a popular choice for researchers and developers alike. To extract different aspects

from an image, a Gaussian blur filter is applied. The resulting image after applying the Gaussian blur creation process is shown below.

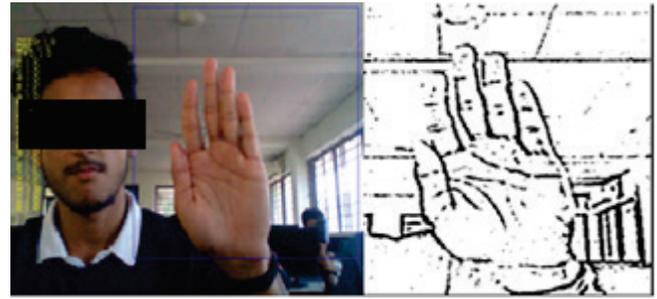


Figure 5 Display for speech to gesture output

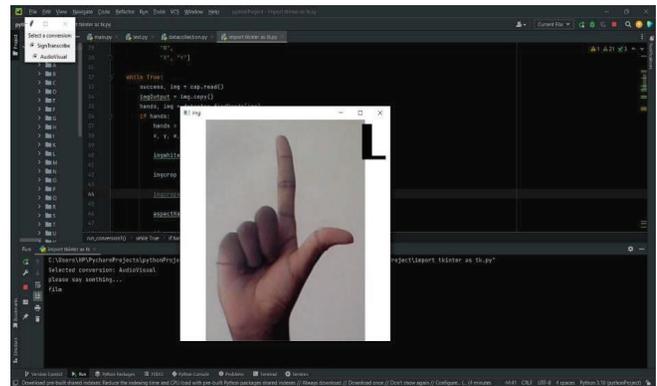


Figure 6 Speech to Gesture output Display

4 RESULTS AND DISCUSSION

4.1 Data Set Generation

The procedures you described involve using the Open Computer Vision (OpenCV) software to capture images of American Sign Language (ASL) symbols. Create own dataset to train and test a machine learning model for recognizing ASL symbols. This dataset will be used to teach the model to understand and classify different ASL signs. Captured approximately 800 photographs of each ASL symbol for training purposes. Collecting a sufficient number of images for each symbol ensures that the model has enough data to learn and generalize effectively. Additionally, collected about 200 images of each ASL symbol for testing purposes. The testing data is used to evaluate the model's performance and assess its ability to make accurate predictions on new, unseen data. To capture the images, used computer's webcam. This means acquired images in real-time by taking a picture of each frame presented to the webcam. Each frame captured by the webcam contained a specific ASL symbol. To focus the model's attention on the relevant area, defined a Region of Interest (ROI) around the ASL symbol. This ROI was visually indicated by a pink-bordered square. Utilized the OpenCV software for the image capture process. OpenCV is a popular open-source computer vision library that provides various tools and functions for image and video processing tasks. Fig. 7 likely depicts an example of the captured image with the pink-bordered square indicating the ROI around an ASL symbol.

The selection of a suitable filter is also a critical component of any ASL symbol recognition project. The filter that you select will determine the features that are extracted from your images, and the CNN model to identify the symbols will use these features. There are many different filters that you can select and it can be difficult to determine which filter will work best for your project. In your case, you experimented with a variety of filters before settling on the chosen filter.

The accuracy of the model is another important factor to consider. In this case, faced issues with the accuracy of the model in earlier rounds. This was likely because the dataset was not large enough or that the filter was not selecting the correct features. By increasing the input image size and improving the dataset, were able to improve the accuracy of the model.

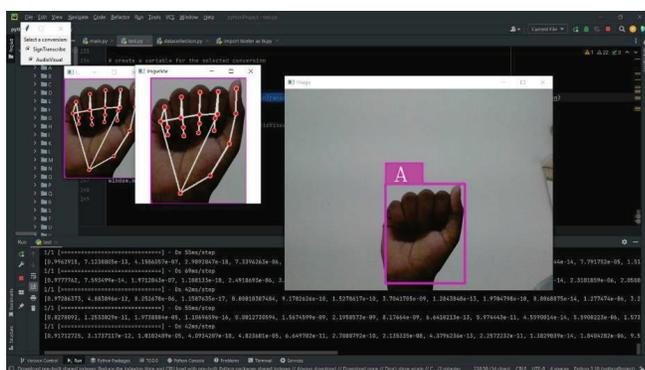


Figure 7 Region of interest (ROI) indicated by a pink-bounded square

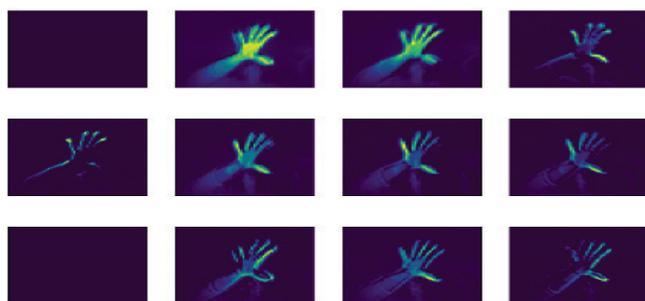


Figure 8 Gaussian blurred image

In order to extract different aspects from our image, we then apply our Gaussian blur filter to it. Fig. 8 displays the image after applying Gaussian blur.

4.2 CNN Configuration

The Stochastic Gradient Descent (SGD) algorithm is used to build the model, and a learning rate of 0.002 is used. The categorical cross-entropy function is used to assess the loss because the model is built to handle several classes. Additionally, measures like loss and accuracy are given to efficiently monitor the review process.

This particular arrangement was chosen after experimenting with various blends of nodes and layers. After experimenting with different combinations of nodes and

layers, this specific configuration was chosen. The CNN configuration details can be found in Tab. 1.

Table 1 Details of CNN configuration

Model Content	Details
First Convolution Layer	128 filters of size 3×3, ReLU, input size 72×72
First Max Pooling Layer	Pooling Size 3×3
Second Convolution Layer	128 filters of size 3×3, ReLU
Second Max Pooling layer	Pooling size 3×3
Dropout Layer	Excludes 30% neurons randomly
First Fully connected Layer	512 nodes, ReLU
Dropout Layer	Excludes 30% neurons randomly
Second Fully Connected Layer	512 nodes, ReLU
Dropout Layer	Excludes 30% neurons randomly
Output Layer	12 nodes for 12 classes, SoftMax
Optimization Function	Stochastic Gradient Descent (SGD)
Learning Rate	0.002
Metrics	Loss, Accuracy

4.3 Classification Result

This statement is supported by additional performance measures outlined in Tab. 2. As depicted in Fig. 9 and Fig. 10. The tight agreement between training and testing accuracy across each epoch suggests that the use of data augmentation strategies helped the model avoid overfitting.

To evaluate further the model's adaptability, the experiment was repeated using a train-test split of 75-45. Remarkably, the accuracy achieved for this split configuration was 95.2% and 97.5%, indicating the model's ability to perform well even on a larger test set.

Table 2 Details of performance measure

Performance measure	CNN without augmentation	CNN with augmentation
Precision	0.9176	0.9869
Recall	0.9432	0.9834
F-Measure	0.9365	0.9832
Accuracy, %	93.43	97.46

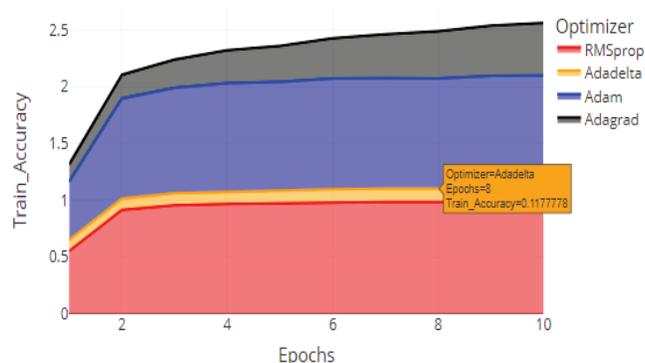


Figure 9 Train Accuracy for various optimizer (Adam and RMSprop shows best accuracies namely 95.2 and 97.5 %)

The training loss values you mentioned for RMSprop (0.071) and Adamgrad (0.0734) as shown in fig 10 indicate that both optimizers are performing similarly in terms of minimizing the loss during training. While there is a slight difference between the two values, it is not significant enough to distinguish easily their performance based solely

on the training loss. When comparing optimizers, it is important to consider multiple factors beyond just the training loss.

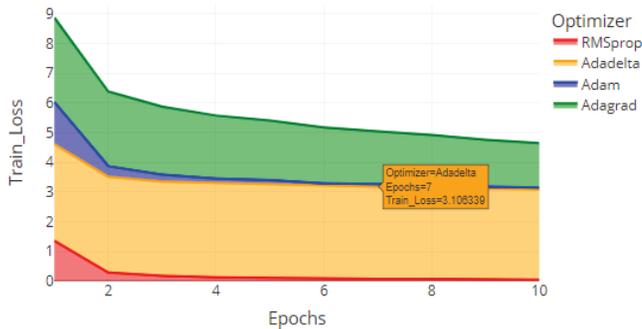


Figure 10 Train Loss for various optimizer (RMSprop 0.071 Adagrad 0.0734 again a close call to distinguish)

Based on the Tab. 3, it appears to contain results for different training epochs, optimizers, activation functions, and evaluation metrics for a specific model. Let us break down each column to understand the information it represents:

Epochs: The model underwent two epochs of training.

RMSprop was employed as the optimisation algorithm during the training process.

Activation_Function: The layers of the neural network made use of the ReLU activation function.

After two training epochs, the training loss was 0.28154. This shows the typical discrepancy between the training dataset's predicted and actual values. Better performance is indicated by lower values.

Train_Accuracy: After two epochs, the training accuracy was 0.91381. It displays the proportion of samples in the training set that were correctly categorised.

Train_Precision: After two epochs, the training precision was 0.23096. Precision is the percentage of all positively anticipated instances in the training set that were accurately predicted.

Train_Recall: After two epochs, the training recall was 0.99738. Recall quantifies the percentage of positive instances in the training set that were properly predicted out of all the actual positive instances.

Val_Loss: After two epochs, the validation loss was 0.26574. This shows the typical discrepancy between the validation dataset's predicted and actual values. Lower values imply more accurate generalisation.

Val_Accuracy: After two epochs, the validation accuracy was 0.98667. It shows the proportion of samples in the validation set that were correctly categorised.

Val_Precision: After two epochs, the validation precision was 0.99982. Precision is the percentage of all positively predicted cases in the validation set that were accurately predicted.

Val_Recall: After two epochs, the validation recall was 0.043908. Measures the percentage of correctly predicted favourable outcomes.

These values provide insights into the model's performance at the specified training stage. The table provides a snapshot of the model's performance metrics at different training epochs, displaying how the optimizer, activation function, and training progress influence the model's accuracy, loss, precision, and recall values.

Table 3 Detailed results

Epochs	Optimizer	Activation_Function	Train_Loss	Train_Accuracy	Train_Precision	Train_Recall	Val_Loss	Val_Accuracy	Val_Precision	Val_Recall
2	RMSprop	relu	0.28154	0.91381	0.23096	0.99738	0.043908	0.98667	0.26574	0.99982
7	RMSprop	softsign	0.07608	0.97992	0.25867	0.99937	0.130949	0.97315	0.31932	1
10	RMSprop	softsign	0.06242	0.98294	0.27277	0.99944	0.019199	0.99519	0.4014	0.99963
3	RMSprop	softsign	0.18316	0.94548	0.23709	0.99921	0.064611	0.97593	0.26	1
1	RMSprop	elu	1.41317	0.52857	0.18546	0.89111	0.289114	0.92093	0.23752	0.99926

In scenarios where there are limited datasets, particularly with only 175 images for each gesture, both networks achieve a recognition rate exceeding 96.8%. This observation suggests that using a TOF camera to collect images based on depth information effectively mitigates the interference caused by other factors during feature extraction.

As the number of datasets increases, the data recognition rate also improves. Augmenting the datasets enhances the network's resilience against complex environmental disturbances. A higher recognition rate establishes a solid foundation for subsequent classification tasks.

Based on the information presented in Fig. 9 and Fig. 10 it can be observed that the improved network outperforms

Faster R-CNN in terms of gesture classification accuracy across various datasets. Specifically, the improved network achieves a recognition accuracy of 83% when working with 175 images for each gesture. This accuracy surpasses the recognition performance of Faster R-CNN, indicating the superiority of the improved network in gesture classification.

The improved network demonstrates notable improvements in accuracy compared to the traditional network. When working with 300 images for each gesture, the improved network achieves a recognition accuracy of 94.8%. Remarkably, this accuracy is comparable to the recognition accuracy of the traditional network trained on 100 times the amount of data.

Furthermore, when there are 600 images for each gesture, the improved network achieves an accuracy of 97.5%, slightly surpassing the performance of Faster R-CNN. This demonstrates the effectiveness of the proposed network in achieving higher accuracy, regardless of whether the dataset is small or large.

5 CONCLUSION

In this project, a practical ASL (American Sign Language) recognition system has been developed specifically for individuals with hearing and speech impairments. The system utilizes real-time vision technology to identify accurately ASL alphabets. By implementing two layers of algorithms, the system is able to effectively predict and validate symbols that bear close resemblance to each other. Consequently, the system consistently recognizes the symbols as long as they are displayed clearly, without any background noise, and with sufficient illumination.

When there are few datasets available, especially when there are only 175 photos for each move, both networks are able to recognise more than 96.8% of the gestures. This finding implies that the influence from other elements during feature extraction can be efficiently reduced by employing a TOF camera to collect images based on depth information.

Across several datasets, it can be seen that the enhanced network performs better than Faster R-CNN in terms of gesture categorization accuracy. Specifically, using 175 photos for each move, the upgraded network obtains an 83% recognition accuracy. This accuracy outperforms Faster R-CNN's performance in gesture classification, demonstrating the supremacy of the upgraded network.

6 REFERENCES

- [1] Corina, D. (1998). Chapter 21 - The Processing of Sign Language: Evidence from Aphasia, Editor(s): Stemmer, B. & Whitaker, H. A. *Handbook of Neurolinguistics*, Academic Press, 313-329. <https://doi.org/10.1016/B978-012666055-5/50025-3>
- [2] Prikhodko, A., Grif, M. & Bakaev, M. (2020). Sign Language Recognition Based on Notations and Neural Networks. https://doi.org/10.1007/978-3-030-65218-0_34
- [3] Ghorpade, S. R. & Waghmare, S. K. (2015). A Communication System for Deaf and Dumb People. *International Journal of Science and Research (IJSR)*, 4(9), 2319-7064.
- [4] Haleem, M. S. (2008). Voice controlled automation system. IEEE International Multitopic Conference, Karachi, Pakistan. 508-512. <https://doi.org/10.1109/INMIC.2008.4777791>
- [5] Kallio, S., Kela, J., Mäntyjärvi, J. & Plomp, J. (2006). Visualization of Hand Gestures for Pervasive Computing Environments. *Proceedings of the working conference on advanced visual interfaces*, ACM, Italy, 480-483. <https://doi.org/10.1145/1133265.1133363>
- [6] Zaman, K. R. & Noor, I. (2012). Hand Gesture Recognition: A Literature Review. *International Journal of Artificial Intelligence & Applications (IJAA)*, 3, 161-174. <https://doi.org/10.5121/ijaa.2012.3412>
- [7] Jeon, M.-J., Yang, S.-E. & Bien, Z. (2009). User adaptive hand gesture recognition using multivariate fuzzy decision tree and fuzzy garbage model. *IEEE International Conference on Fuzzy Systems*, Jeju, Korea (South), 474-479. <https://doi.org/10.1109/FUZZY.2009.5277156>
- [8] Wilson, A. D. & Bobick, A. F. (1999). Parametric hidden Markov models for gesture recognition. *IEEE transactions on pattern analysis and machine intelligence*, 21(9), 884-900. <https://doi.org/10.1109/34.790429>
- [9] Reyes, M., Dominguez, G. & Escalera, S. (2011, November). Featureweighting in dynamic timewarping for gesture recognition in depth data. In *2011 IEEE international conference on computer vision workshops (ICCV Workshops)*, 1182-1188. <https://doi.org/10.1109/ICCVW.2011.6130384>
- [10] Tang, D., Jin Chang, H., Tejani, A. & Kim, T. K. (2014). Latent regression forest: Structured estimation of 3D articulated hand posture. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 3786-3793. <https://doi.org/10.1109/CVPR.2014.490>
- [11] Rahman, M. H. & Afrin, J. (2013). Hand gesture recognition using multiclass support vector machine. *International Journal of Computer Applications*, 74(1), 39-43. <https://doi.org/10.5120/12852-9367>
- [12] Chen, X. & Wang, Z. J. (2013). Pattern recognition of number gestures based on a wireless surface EMG system. *Biomedical Signal Processing and Control*, 8(2), 184-192. <https://doi.org/10.1016/j.bspc.2012.08.005>
- [13] Chudasama, D., Patel, T., Joshi, S. & Prajapati, G. I. (2015). Image segmentation using morphological operations. *International Journal of Computer Applications*, 117(18). <https://doi.org/10.5120/20654-3197>
- [14] Madan, A. (2021). Face recognition using Haar cascade classifier. *Int. J. Mod. Trends Sci. Technol*, 7(01), 85-87.
- [15] Sonka, M., Hlavac, V., Boyle, R., Sonka, M., Hlavac, V. & Boyle, R. (1993). Image pre-processing. *Image processing, analysis and machine vision*, 56-111. https://doi.org/10.1007/978-1-4899-3216-7_4
- [16] Soni, N. S., Nagmode, M. S. & Komati, R. D. (2016). Online hand gesture recognition & classification for deaf & dumb. In *IEEE International Conference on Inventive Computation Technologies (ICICT2016)*, 3, 1-4. <https://doi.org/10.1109/INVENTIVE.2016.7830112>
- [17] Cabrera, M. E., Bogado, J. M., Fermin, L., Acuna, R. & Raley, D. (2012). Glove-based gesture recognition system. In *Adaptive Mobile Robotics* (pp. 747-753). https://doi.org/10.1142/9789814415958_0095
- [18] Garg, P., Aggarwal, N. & Sofat, S. (2009). Vision based hand gesture recognition. *International Journal of Computer and Information Engineering*, 3(1), 186-191.
- [19] Bradski, G. & Kaehler, A. (2008). *Learning OpenCV: Computer vision with the OpenCV library*. O'Reilly Media, Inc.
- [20] Wang, S. C. & Wang, S. C. (2003). Artificial neural network. *Interdisciplinary computing in java programming*, 81-100. https://doi.org/10.1007/978-1-4615-0377-4_5
- [21] Chua, L. O. & Roska, T. (1993). The CNN paradigm. *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, 40(3), 147-156. <https://doi.org/10.1109/81.222795>
- [22] Pang, B., Nijkamp, E. & Wu, Y. N. (2020). Deep learning with tensor flow: A review. *Journal of Educational and Behavioral Statistics*, 45(2), 227-248. <https://doi.org/10.3102/1076998619872761>

Authors' contacts:

G. K. Dayananda

(Corresponding author)

Faculty of Electronics & Communication Engineering,
Canara Engineering College,
Benjanapadavu, Bantwal Taluk, Mangalore, D. K. District,
574219 Karnataka, India
dayanand.gk@gmail.com

I. R. Manjunath

Faculty of Electronics & Communication Engineering,
Canara Engineering College,
Benjanapadavu, Bantwal Taluk, Mangalore, D. K. District,
574219 Karnataka, India

J. G. Sreerama Samartha

Faculty of Electronics & Communication Engineering,
Canara Engineering College,
Benjanapadavu, Bantwal Taluk, Mangalore, D. K. District,
574219 Karnataka, India

M. Vayusutha

Faculty of Electronics & Communication Engineering,
Canara Engineering College,
Benjanapadavu, Bantwal Taluk, Mangalore, D. K. District,
574219 Karnataka, India

D. N. Pratham

Electronics & Communication Engineering,
Canara Engineering College,
Benjanapadavu, Bantwal Taluk, Mangalore, D. K. District,
574219 Karnataka, India

Sakshath

Electronics & Communication Engineering,
Canara Engineering College,
Benjanapadavu, Bantwal Taluk, Mangalore, D. K. District,
574219 Karnataka, India

A. U. Shashank

Electronics & Communication Engineering,
Canara Engineering College,
Benjanapadavu, Bantwal Taluk, Mangalore, D. K. District,
574219 Karnataka, India

Swasthi Prasad Shetty

Electronics & Communication Engineering,
Canara Engineering College,
Benjanapadavu, Bantwal Taluk, Mangalore, D. K. District,
574219 Karnataka, India

Experimental Studies on Mechanical Characterizations of Silicon Nitride and Alumina added Al7075 Hybrid Composite for Various Engineering Applications

T. Karuppusamy, K. Kathir Balaji, Senthil Kumar Madasamy, Beena Stanislaus Arputharaj, Parvathy Rajendran*, Arunkumar Karuppasamy, Vijayanandh Raja*

Abstract: The present investigation is aimed at studying the influence of alumina and silicon nitride reinforcement particles in addition to the Al 7075 alloy on the hardness, wear, tensile, and impression creep behaviour. Three different composite specimens such as Al 7075 + 5 % alumina, Al 7075 + 5 % Silicon nitride, and Al 7075 + 2.5 % Silicon nitride + 2.5 % Alumina were fabricated using a stir casting process and the prepared composites were subjected to hardness, dry sliding wear test using pin on disc tribometer, tensile testing and impression creep tests. From the results, it was observed that the hardness of the Al 7075 alloy increases with the addition of both silicon nitride and alumina and Al 7075 + 2.5 % Silicon nitride + 2.5 % Alumina exhibited the maximum hardness of the composite specimens. Al 7075 + 2.5 % Silicon nitride + 2.5 % Alumina showed a 75% increase in wear resistance in comparison with the Al 7075 alloy and it also shows an improvement in creep and tensile properties.

Keywords: creep; friction; hardness; hybrid aluminium composite; metal matrix composite; tensile test

1 INTRODUCTION

Aluminium Matrix Composites (AMCs), a category of lightweight material systems. The unreinforced aluminium metal lacks the strength, rigidity, creep resistance, and wear resistance that the aluminium matrix composites do. Of all the commercial aluminium alloys, 7075 is an incredibly popular choice as a matrix material to prepare metal matrix composites. The addition of reinforcements like silicon nitride in the matrix of Al7075 alloy resulted in improvement in mechanical properties [1-3]. There are studies on the improvement in wear resistance with increase in weight percentage of combination of reinforcements such as TaC+Si₃N₄+Ti in the Al7075 matrix alloy [4-6]. A hybrid composite consists of addition of more than one reinforcement particle in the matrix. Al 7075 hybrid composites are widely used in aerospace and automobile industries owing to their higher specific strength. Some of the notable applications include aircraft fuselage, wings, tail section, satellite frames, battery enclosures etc [7-10]. This research work focusses on such Al7075 hybrid composites, its mechanical properties, etc. Al 7075 hybrid composites are manufactured by the addition of 5 % silicon carbide and 2.5 % alumina and were subjected to mechanical property evaluations. It was observed that the addition of secondary particles resulted in finer grain size microstructure and resulted in hardness value improvement [11-14]. Baradeswaran et al. fabricated Al7075 metal matrix composites by the addition of alumina particles in the matrix and studied their properties. From the results, they have concluded that the increase in weight percentage of alumina particles increases the tensile strength of the materials [15]. Moona et al. studied the Al7075 hybrid composite containing 1.5% of aluminium oxide, 1.5% of silicon carbide and 1.5 % eggshell particles and concluded that fatigue strength of such materials is greatly enhanced [16]. From the literature survey, it was found that there exists a significant gap in the research works in the area of silicon nitride and alumina added Al

7075 alloys and the characterization of such fabricated composites. There is minimal work reported in the area of impression creep behaviour of such composites. So, in the current study, silicon nitride and alumina are added to Al7075 alloy to create aluminium hybrid composites, and the mechanical characteristics such as hardness. Tensile, wear, impression creep of these composites are examined [17-20].

2 EXPERIMENTAL WORK

The chemical composition of the Al 7075 alloy is given in Tab. 1.

Table 1 Chemical composition of Al7075 alloy

Element	wt.%
Cr	0.18-0.28
Cu	1.2-2
Fe	0.5
Mg	2.1-2.9
Mn	0.3
Si	0.4
Ti	0.2
Zn	5.1-6.1
Al	Remaining

The silicon nitride and alumina particles are added to the base alloy to improve its strength, hardness, creep resistance, and wear resistance. The compositions of fabricated specimens are shown in Tab. 2.

Table 2 Fabricated specimen compositions

Sl. No	Compositions (wt.%)
1	Al 7075 Alloy
2	Al 7075 + 5% Silicon Nitride
3	Al 7075 + 5% Alumina
4	Al 7075 + 2.5% Silicon Nitride + 2.5% Alumina

2.1 Stir Casting Process

Stir casting process was used for the fabrication of nitride and alumina-reinforced Al7075 alloy hybrid composites. The

stir casting apparatus used for the manufacturing of composites are shown in Fig. 1. Al 7075 alloy is placed in the graphite crucible, and melting of the alloy is done at 720 °C and then mixing of reinforcement particles were carried out using a stirrer coated with graphite. Degassing of molten metal was carried out at 787 °C. Then the stirrer was inserted into the graphite crucible, and it is rotated at 600 rpm for maintaining proper vortex and achieving uniform mixing of the reinforcement particles. The reinforcement particles were preheated at a temperature of about 400 °C for a period of 2 hours and then the reinforcement particles were added to the molten aluminium alloy. Magnesium powders were added in small quantities during stirring to increase the wettability between matrix and reinforcement particles. Mechanical stirring was carried out for 5 minutes, and at the end of mixing, the molten materials were poured into a dried, preheated metallic mould of diameter 30 mm and length 150 mm. Bottom pouring was done to prevent the reaction between the atmosphere and molten composites during pouring. After solidification, the as-cast composites were removed from the die [9-12]. The solidified composite along with the die is shown in Fig. 2.



Figure 1 Stir Casting Apparatus



Figure 2 Solidified composite specimen

2.2 Hardness Studies

The hardness studies were carried out using a Rockwell Hardness testing machine at a load of 1 kg for a dwell time of 10 seconds. The hardness tests were performed according to ASTM E-92 standards. Specimens were polished before

taking hardness measurements to remove scratches and surface irregularities using the standard metallographic procedure. Ten readings were taken for each sample to ensure repeatability, and the average values are noted [13].

2.3 Wear Studies

A pin-on-disc wear testing apparatus (Model: TE-165-POD Magnum) connected to the data acquisition device was used to conduct dry sliding wear tests, as shown in Fig. 3. The experiments were carried out in accordance with ASTM G99-05 specifications. The experiments were carried out with a steady disc speed of 1 m/s at room temperature. The pin specimen is made to rub against the rotating steel disc for a sliding distance of 1000 m [14-16]. During the tests, the wear loss of the pin specimen is measured using the LVDT transducer with an accuracy of 1.0 µm. During the tests, the wear loss during the tests is continuously recorded in the system. The samples are weighed before and after wear tests manually using an electronic weighing balance to confirm the results obtained using a data acquisition system. The specimen dimensions used for wear testing are of 10 mm diameter and 25 mm length [17-18]. Fig. 4 shows the specimen which is used for wear test.



Figure 3 Pin-on-disc wear testing machine



Figure 4 Specimen used for Wear Test

2.4 Tensile Testing

Dark tensile testers are used to evaluate the tensile strength of newly developed composite materials (Model: UTB9052). The samples have been made in accordance with ASTM E-8 standards; Fig. 5 depicts the samples used for tensile testing. The tensile sample has a diameter of 6 mm and a length of 30 mm. The composition of specimens used for tensile studies is given in Tab. 1. The load-displacement data are obtained in the system attached with the tensile tester, and then a plot of stress versus strain is obtained for each specimen [19-20].

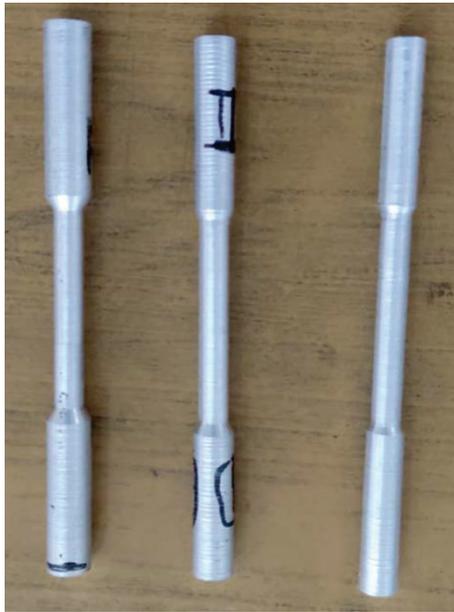


Figure 5 Specimens used for tensile testing



Figure 6 Impression creep testing machine

2.5 Creep Studies

The creep tests are carried out using impression creep testing machine (SPANKTRONICS). For creep tests, cylindrical specimens of 10 mm diameter and 10 mm height are used. The specimens are polished according to a standard metallographic procedure to remove the scratches present in the surface and the specimen is placed below the indenter. After sample preparation, the specimen is placed beneath the indenter as the first stage. The object is positioned beneath the 1.5 mm diameter tungsten carbide indenter after the lock screw is loosened to raise the furnace. The necessary temperature is established in the temperature controller after adding the small load. At this point, the thermocouple is introduced to measure the furnace's interior temperature. According to the literature study that was done, creep tests are conducted on all examples at a constant temperature of about 200 °C. Before adding significant weight, the sample is kept at this temperature for a 1 hour. Impression depth vs. time graph is collected by the data collection device and displayed on the monitor. The impression creep testing machine is shown in Fig. 6. The impression creep tests are carried out for Al7075 alloy and hybrid composites at a constant temperature of 200 °C and a constant load of 20 kg. The dwell time for carrying out creep tests was 2 hours.

3 RESULTS AND DISCUSSION

3.1 Hardness Results

Al7075+5% Silicon nitride specimen shows an increase in hardness of about 5% than Al7075 alloy due to the presence of Si₃N₄ particles in the matrix. Al7075+5% Al₂O₃ specimen shows an increase in hardness of about 3% than Al7075 alloy due to the presence of Al₂O₃ particles in the matrix. Al7075+2.5% Si₃N₄ +2.5% Al₂O₃ specimen shows an increase in hardness of about 14% than Al7075 alloy due to the presence of combination of Si₃N₄ particles and Al₂O₃ particles in the matrix. Al7075+2.5% Si₃N₄ +2.5% Al₂O₃ specimen exhibited higher hardness values of the investigated specimens. The increase in hardness is due to the addition of hard ceramic reinforcements in the matrix of Al7075 which acts as nucleation sites and refines the microstructure and thus resulting in increment in hardness values. The comprehensive hardness behaviour results of various combinations of this proposed aluminium metal matrix hybrid composites are revealed in Fig. 7.

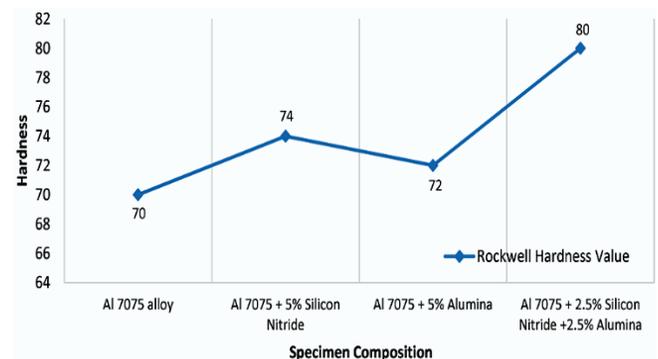


Figure 7 Rockwell Hardness values for Al7075 alloy and its composites

3.2 Wear Studies

The wear loss decreases by the addition of both silicon nitride and alumina reinforcements in the matrix. Fig. 8 shows the wear curves of Al7075, Al7075+5% Si₃N₄, Al7075 +5% Al₂O₃ and Al7075+2.5% Si₃N₄ +2.5% Al₂O₃. It is observed from Fig. 8, that the wear resistance of Al7075+5% silicon nitride specimen is improved of about 29% than Al7075 alloy. The wear resistance of Al7075+5% alumina specimen is improved of about 44% than Al7075 alloy. The wear resistance of Al7075+2.5%, silicon nitride+2.5% alumina specimen is improved of about 75% than Al7075 alloy. The Al7075 +2.5%silicon nitride+2.5% Alumina showed maximum wear resistance of the investigated specimens. The material loss due to wear for Al7075 + 2.5% silicon nitride + 2.5% Alumina particles at the end of wear test were nearly 50 microns whereas the weight loss due to wear for the Al7075 alloy was found to be 197 microns. The hard nature of silicon nitride particles and alumina particles lowers the wear loss of composite specimens. The higher wear resistance of hybrid composite can be attributed to the presence of hard reinforcement particles like silicon nitride and alumina, which bears the load at the surface and prevent the stress/load concentration on softer Al7075 matrix. The hardness of these ceramic particles is very high and acts as abrasion resistant phases leading to better wear resistance. The losses of material due to wear of all the investigated specimens are given in Fig. 9.

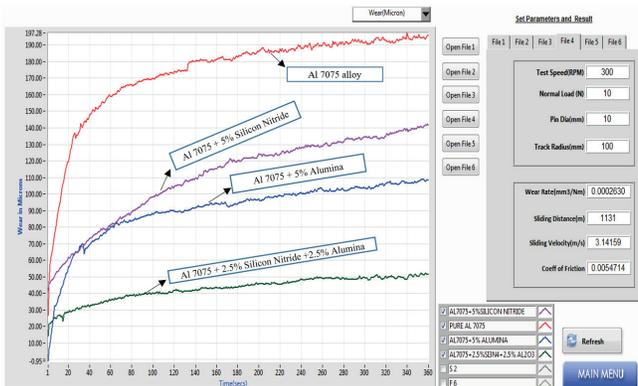


Figure 8 Wear curves of Al7075, Al7075+5% Si₃N₄, Al7075 +5% Al₂O₃ and Al7075+2.5% Si₃N₄ +2.5% Al₂O₃

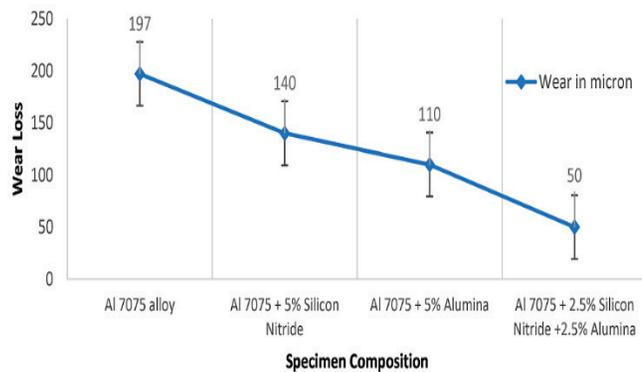


Figure 9 Wear loss (in microns) of Al7075 alloy and composites after wear testing

3.3 Frictional Force

Fig. 10 shows the variation of frictional force of Al7075, Al7075+5% Si₃N₄, Al7075 +5% Al₂O₃ and Al7075+2.5% Si₃N₄ +2.5% Al₂O₃. From the figure the variation of frictional force varies between 0 to 4.5 N for constant load. The composite specimens showed lower frictional force values due to the presence of reinforcement particles. The unreinforced alloy showed more fluctuation than the composite specimens because of the increasing tangential force between pin and disc surfaces. The lower frictional force values for hybrid composites indicate that the wear loss of the composites was less than the Al7075 alloy. Wear failure of the material will be more in the presence of higher frictional forces between the pin and the disc surface.

When alumina and silicon nitride mixes with Al7075 as reinforcements, the fluctuation of frictional force occurs initially. After it reaches 240 sec, friction reduces, and it is maintained constant.

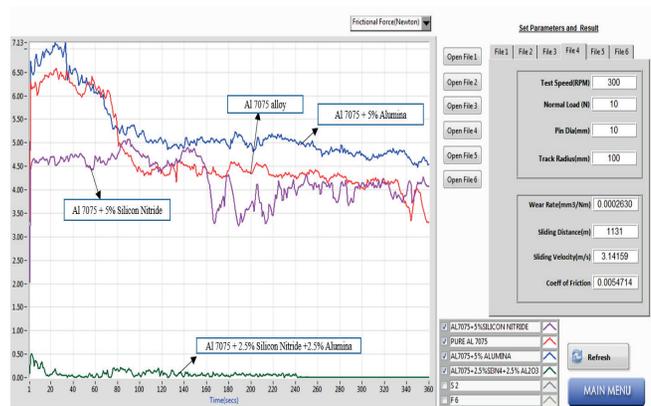


Figure 10 Variation of frictional force of Al7075, Al7075+5% Si₃N₄, Al7075 +5% Al₂O₃ and Al7075+2.5% Si₃N₄ +2.5% Al₂O₃

3.4 Coefficient of Friction

Fig. 11 depicts the variation in the coefficient of friction of aluminium hybrid composites. The coefficient of friction plays a significant role in many areas of engineering and physics, including the design of brakes, tires, and other frictional components.

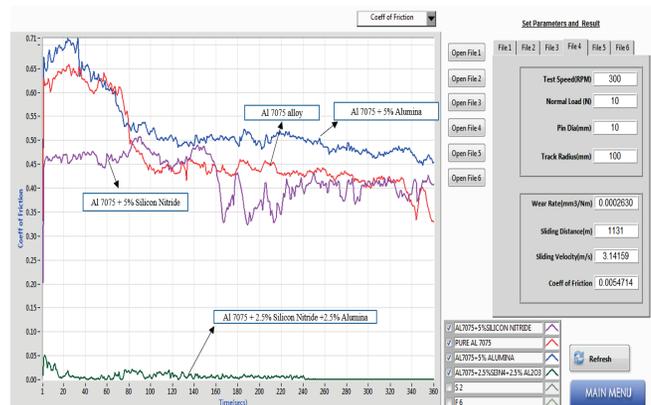


Figure 11 Variation in the coefficient of friction of Al7075 composites

From Fig. 11, the coefficient of friction of Al7075 composites varies from 0.03 to 0.5 than the Al7075 alloy. When alumina and silicon nitride mixes with Al7075 as reinforcements, the fluctuation of coefficient of friction occurs initially. After it reaches 240 sec, friction reduces and it is maintained constant.

3.5 Creep Studies

Al7075 hybrid composites are subjected to impression creep tests under cast conditions to study the effect of these composites' creep behaviour. Impression depth vs Time graph obtained from the creep tests for the Al7075 alloy is shown in Fig. 12.

The impression depth of Al7075 alloy is found to be 0.361 mm. The creep curves of Al7075+5% Si₃N₄ is shown in Fig. 13. The impression depth of Al7075 alloy+5% Si₃N₄ is found to be 0.195 mm. It is observed that the addition of reinforcement particles like silicon nitride can reduce creep deformation with increased creep resistance. The creep deformation of Al7075+5% Si₃N₄ composite is lower than the creep deformation of Al7075 alloy. The creep curves of Al7075+5% Al₂O₃ is shown in Fig. 14.

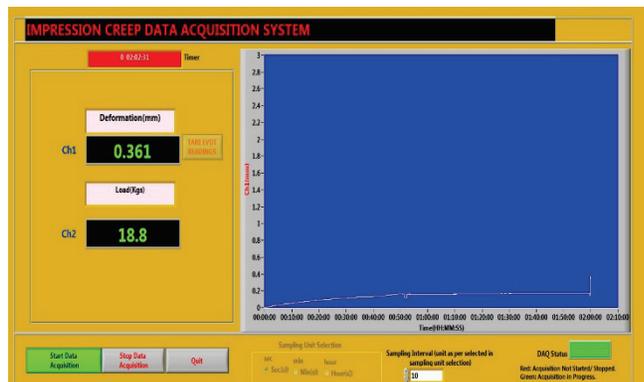


Figure 12 Impression Depth vs. Time graph for Al7075 alloy

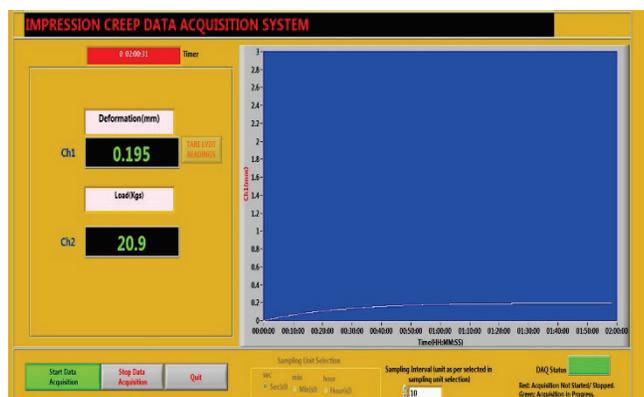


Figure 13 Impression Depth vs. Time graph for Al7075+5% silicon nitride

The impression depth of Al7075 alloy+5% Al₂O₃ is found to be 0.175 mm. It is observed that the addition of 5% alumina can reduce creep deformation. This is due to the hindrance of dislocation motion by precipitates, present in the matrix. This reduces dislocation motion of creep leading to enhanced creep resistance. The creep deformation of Al7075+5% Al₂O₃ composite is lower than the creep

deformation of Al7075 alloy. The creep curves of Al7075+2.5% Si₃N₄+2.5% Al₂O₃ is shown in Fig. 15.

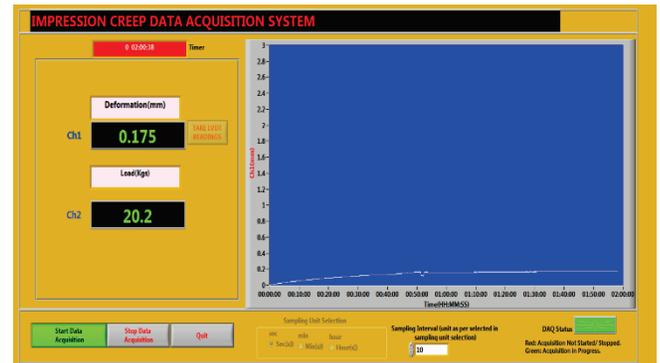


Figure 14 Impression Depth vs. Time graph for Al7075+5% alumina

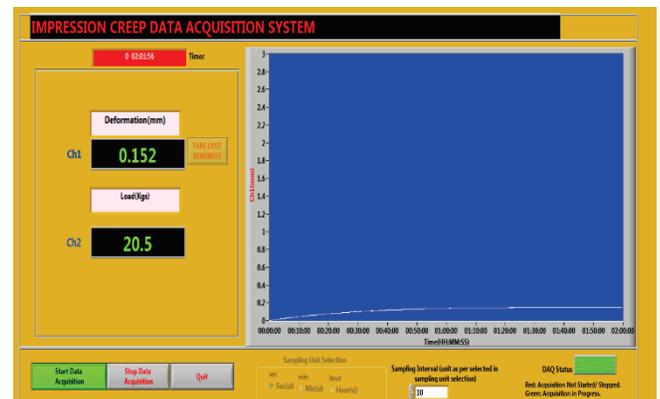


Figure 15 Impression Depth vs. Time graph for Al7075+2.5% Si₃N₄+2.5% Al₂O₃

The impression depth of Al7075 alloy+2.5% Si₃N₄+2.5% Al₂O₃ is found to be 0.152 mm. In comparison with the fabricated specimens Al7075 alloy+2.5% Si₃N₄+2.5% Al₂O₃ composites exhibit greater creep resistance with less creep deformation. From the Creep deformation graphs, it is evident that the creep resistance of the Al 7075 matrix composite increases. This is because silicon nitride and alumina particle acts as a barrier for dislocation movement. Hence dislocation pile-ups occur and therefore more amount of stress or higher temperature are required for the dislocation to overcome the barrier leading to enhanced creep strength. The comparison of creep curves for all the investigated specimens are shown in Fig. 16.

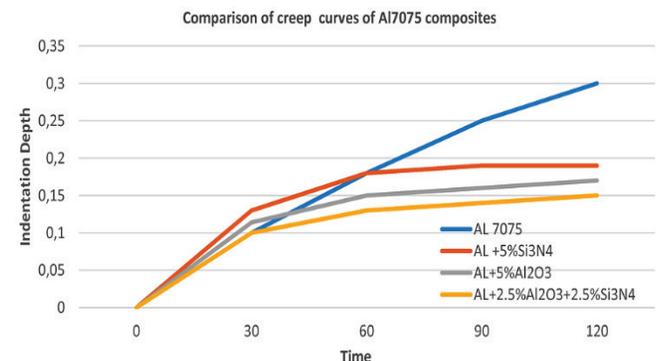


Figure 16 Comparison of creep curves (Indentation depth in mm & Time in minutes) for all the investigated specimens

3.6 Tensile Studies

The tensile tests have been performed for both Al7075 alloy and Al7075 hybrid composites. The hybrid composites exhibited higher tensile strength than that of the base 7075 alloy due to the fine grain structure in composite materials that have been obtained during solidification. The presence of reinforcement particles in the molten metal acts as nucleation sites for grain nucleation, and hence it resulted in finer grain size. According to the Hall-Petch relationship, the finer grain size will result in higher tensile strength. The presence of hard reinforcement particles like silicon nitride and alumina hinders the movement of dislocations and hence plastic deformation is restricted, which also contributes to an increase in tensile strength of the composites. The tensile stress values for the tested specimens are shown in Fig. 17.

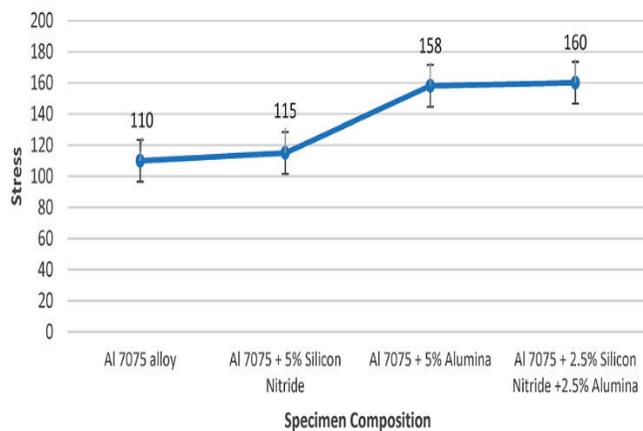


Figure 17 Stress values (MPa) of tested specimen

The tensile strength of Al 7075 alloy was 110 MPa, whereas Al7075 + 5% silicon nitride composite was 115 MPa, there is marginal improvement in tensile strength due to its addition. However, the 5% alumina added Al7075 alloy exhibited significant improvement in tensile strength, which is around 158 MPa. Al7075+2.5% Si₃N₄ +2.5% Al₂O₃ exhibits maximum tensile stress of 160 MPa when compared to another composite specimen, due to higher volume fraction of hard ceramic particles in the matrix contributing to Orowan strengthening, dislocation hindering mechanisms and fine grain structures.

4 CONCLUSIONS

The major conclusions drawn from the present study on experimental investigations on the effect of reinforcements on mechanical properties of hybrid Al7075 composites are summarized.

- 1) Al7075+2.5% Al₂O₃+2.5% Si₃N₄ exhibits high hardness when compared to other composite specimens.
- 2) The hybrid composite specimen exhibited lower wear loss (50 microns) than all the investigated composite samples and is more suitable for applications demanding higher wear resistance.
- 3) The composite specimens exhibited higher creep resistance than the Al 7075 alloy and Al7075+2.5% Al₂O₃+2.5% Si₃N₄ showed lower impression creep

deformation (0.152 mm) among the fabricated specimens.

Acknowledgements

Experimental test facilities are been provided by the authors' parent institution, which is Kumaraguru College of Technology, Coimbatore, Tamil Nadu, India. So, all the authors of this article would like to thank all the management of people and higher professionals.

5 REFERENCES

- [1] Ahmad, F., Lo, S. H. J., Aslam, M. & Haziq, A. (2013). Tribology behaviour of alumina particles reinforced aluminium matrix composites and brake disc materials. *Procedia Engineering*, 68(0), 674-680. <https://doi.org/10.1016/j.proeng.2013.12.238>.
- [2] Alhaji Ibrahim, M., Sahin, Y., Yusuf Gidado, A. & Said, M. (2019). Mechanical Properties of Aluminium Matrix Composite Including SiC/Al₂O₃ by Powder Metallurgy-A Review. *Global Scientific Journal*, 7(3), 23-38.
- [3] Natrayan, L. & Senthil Kumar, M. (2018). Study on Squeeze Casting of Aluminum Matrix Composites—A Review. In: Antony, K. & Davim, J. (eds) *Advanced Manufacturing and Materials Science. Lecture Notes on Multidisciplinary Industrial Engineering*. Springer, Cham. https://doi.org/10.1007/978-3-319-76276-0_8
- [4] Pradeep Kumar, J., Robinson Smart, D. S. & Selvan, C. P. (2018). Experimental evaluation of strength and wear rate of AA7075/TAC/Si₃N₄/Ti nano hybrid metal matrix composite. *International Journal of Mechanical Engineering and Technology*, 9(11), 1690-1698.
- [5] Surappa, M. K. (2003). Aluminium matrix composites: Challenges and opportunities. *Sadhana - Academy Proceedings in Engineering Sciences*, 28(1-2), 319-334. <https://doi.org/10.1007/BF02717141>
- [6] Senthil Kumar M. et al. (2018). Experimental investigations on mechanical and microstructural properties of Al₂O₃/SiC reinforced hybrid metal matrix composite. *IOP Conf. Ser.: Mater. Sci. Eng.* 402 012123 <https://doi.org/10.1088/1757-899X/402/1/012123>
- [7] Cui, Y., Wang, L. & Ren, J. (2008). Multi-functional SiC/Al composites for aerospace applications. *Chinese Journal of Aeronautics*, 21(6), 578-584. [https://doi.org/10.1016/S1000-9361\(08\)60177-6](https://doi.org/10.1016/S1000-9361(08)60177-6)
- [8] Pitchayapillai, G., Seenikannan, P., Raja, K. & Chandrasekaran, K. (2016). Al6061 Hybrid Metal Matrix Composite Reinforced with Alumina and Molybdenum Disulphide. *Advances in Materials Science and Engineering*, 2016. <https://doi.org/10.1155/2016/6127624>
- [9] Anand Babu, K., Venkataramaiah, P. & Dharma Reddy, K. (2018). Mechanical characterization of aluminium hybrid metal matrix composites synthesized by using stir casting process. *Materials Today: Proceedings*, 5(14), 28155-28163. <https://doi.org/10.1016/j.matpr.2018.10.058>
- [10] Arun Kumar, S., Hari Vignesh, J. & Paul Joshua, S. (2021). Investigating the effect of porosity on aluminium 7075 alloy reinforced with silicon nitride (Si₃N₄) metal matrix composites through STIR casting process. *Materials Today: Proceedings*, 39(1), 414-419. <https://doi.org/10.1016/j.matpr.2020.07.690>
- [11] Pratap Singh, A., Senthil Kumar, M., Deshpande, A., Jain, G., Khamesra, J., Mhetre, S., Awasthi, A. & Natrayan, L. (2021). Processing and characterization mechanical properties of

- AA2024/Al₂O₃/ZrO₂/Gr reinforced hybrid composite using stir casting technique. *Materials Today: Proceedings*, 37(2), 1562-1566. <https://doi.org/10.1016/j.matpr.2020.07.156>
- [12] Mistry, J. M. & Gohil, P. P. (2019). Experimental investigations on wear and friction behaviour of Si₃N₄p reinforced heat-treated aluminium matrix composites produced using electromagnetic stir casting process. *Composites Part B: Engineering*, 161, 190-204. <https://doi.org/10.1016/j.compositesb.2018.10.074>
- [13] Devegowda, T. M. & Gurupavan, H. R. (2022). Evaluation of Microstructure and Hardness of Al-SiC Metal Matrix. *Composite Material*, 10(10), 14-16.
- [14] Liu, S., Wang, Y., Muthuramalingam, T. & Anbuechziyan, G. (2019). Effect of B₄C and MOS₂ reinforcement on microstructure and wear properties of aluminum hybrid composite for automotive applications. *Composites Part B: Engineering*, 176, 107329. <https://doi.org/10.1016/j.compositesb.2019.107329>
- [15] Baradeswaran, A. & Elaya Perumal, A. (2014). Study on mechanical and wear properties of Al7075/Al₂O₃/graphite hybrid composites. *Composites Part B: Engineering*, 56, 464-471. <https://doi.org/10.1016/j.compositesb.2013.08.013>
- [16] Moona, G., Rastogi, V., Walia, R. S. & Sharma, R. (2020). Microstructure and Wear Study of Al 7075-T6/Eggshell/SiC/Al₂O₃ Hybrid Composites. *Lecture Notes in Mechanical Engineering*, 471-481. https://doi.org/10.1007/978-981-15-1071-7_39
- [17] Paranthaman, V., Sundaram, K. S. & Natrayan, L. (2022). Influence of SiC Particles on Mechanical and Microstructural Properties of Modified Interlock Friction Stir Weld Lap Joint for Automotive Grade Aluminium Alloy. *Silicon*, 14, 1617-1627. <https://doi.org/10.1007/s12633-021-00944-5>
- [18] Suryakumari, T. S. A. & Ranganathan, S. (2018). Preparation and Study the Wear Behaviour of Aluminium Hybrid Composite. *Materials Today: Proceedings*, 5(2), 8104-8111. <https://doi.org/10.1016/j.matpr.2017.11.497>
- [19] Kousik Kumar, R., Somasundara Vinoth, K., Senthil Kumar, M., Beena, S. A., Parvathy, R., Ahmed, H. D. & Vijayanandh, R. (2024). Experimental Studies on Microstructural, Physical, and Mechanical Characterizations of Hybrid Aluminium Metal Matrix Composites Incorporated with Cenosphere and Molybdenum Disulphide. *Journal of Engineering*, 5964256, 15 pages. <https://doi.org/10.1155/2024/5964256>
- [20] David Raja Selvam, J. & Dinaharan, I. (2017). In situ formation of ZrB₂ particulates and their influence on microstructure and tensile behavior of AA7075 aluminum matrix composites. *Engineering Science and Technology, an International Journal*, 20(1), 187-196. <https://doi.org/10.1016/j.jestch.2016.09.006>

Authors' contacts:

T. Karuppusamy

Department of Mechanical Engineering, Kumaraguru College of Technology, Coimbatore - 641049, Tamil Nadu, India
E-mail: karuppusamy.t.mec@kct.ac.in

K. Kathir Balaji, Senior Engineer

Titan Company Limited,
Bangalore-560100, Karnataka, India
E-mail: kathirbalaji99@gmail.com

Senthil Kumar Madasamy

Department of Aeronautical Engineering, Kumaraguru College of Technology, Coimbatore - 641049, Tamil Nadu, India
E-mail: senthilkumar.m.aeu@kct.ac.in

Beena Stanislaus Arputharaj

Department of Research and Innovation, Saveetha School of Engineering, SIMATS, Chennai-602105, Tamil Nadu, India
E-mail: beena2192@gmail.com

Parvathy Rajendran

(Corresponding author)
Department of Mechanical and Aerospace Engineering,
United Arab Emirates University,
Al Ain, UAE
E-mail: aeparvathy@uaeu.ac.ae

Arunkumar Karuppasamy

Department of Aeronautical Engineering,
MLR Institute of Technology, Hyderabad, Telangana, India
E-mail: arunssksamy@gmail.com

Vijayanandh Raja

(Corresponding author)
Department of Aeronautical Engineering, Kumaraguru College of Technology,
Coimbatore - 641049, Tamil Nadu, India
E-mail: vijayanandh.raja@gmail.com

Effect of Rubber Seed Oil Biodiesel Additive on Compression Ignition Engines fuelled with Diesel-Ethanol Blends

H. V. Srikanth, S. Vijay Kumar, R. Kousik Kumar*, Senthil Kumar Madasamy, Beena Stanislaus Arputharaj, Parvathy Rajendran, Vijayanandh Raja*

Abstract: In this study, the primary focus revolved around the utilization of rubber seed oil (RSO) biodiesel as a supplement in blends of diesel-ethanol (DE) for a diesel engine. The DE blends were formulated by combining ethanol with diesel and emulsifying them with RSO biodiesel in a 10% (v/v) proportion. The ethanol concentrations in the blends varied between 5% and 15% (v/v). Under conditions of maximum loading, it was observed that the blend labelled DE15B10, comprising 15% ethanol and 10% RSO biodiesel, demonstrated the highest brake thermal efficiency (BTE). Although all the examined fuels exhibited an elevated Brake Specific Fuel Consumption (BSFC) compared to conventional diesel, DE15B10 displayed a 4.30% increase in BSFC over fossil diesel. Nevertheless, the exhaust emission characteristics of DE15B10 were found to be superior to those of conventional diesel. These results indicate that DE-biodiesel blends, especially DE15B10, show potential as a viable alternative fuel option without requiring any modifications to the engine hardware.

Keywords: biodiesel; diesel-ethanol; emission; performance; rubber seed oil; transesterification

1 INTRODUCTION

The economy of a nation is significantly influenced by energy, a pivotal determinant. According to statistical information on global energy provision presented by British Petroleum (BP 2020), the primary share is attributed to fossil fuels, with crude oil at 33.1%, coal at 27.0%, and natural gas at 24.2%. Subsequent contributors include nuclear at 4.3%, hydro at 6.4%, and other renewable sources at 5%. The overreliance on fossil fuels has resulted in environmental contamination and global warming, prompting researchers worldwide to explore sustainable and environmentally friendly fuel alternatives such as biodiesel, ethanol, and butanol. These substitutes play a vital role in addressing transportation energy requirements while adhering to stringent pollution regulations [1, 2, and 3].

Biodiesel derived from first and second-generation oils is acknowledged as a promising alternative renewable fuel source for both transportation and gas turbine applications [4-6]. In the Indian context, opting for non-edible oil feedstocks in biodiesel production is considered a more secure approach to address the energy-food conflict [7, 8]. Nevertheless, the accessibility of non-edible oils exhibits significant regional disparities [9, 10]. To ensure continuous biodiesel production, numerous studies have explored the utilization of non-edible to produce biodiesel through various sources [11-14].

Several researchers have explored alternative fuels for internal combustion (IC) engines, including bio-ethanol derived from the fermentation of agricultural waste, in addition to biodiesel. The utilization of these fuels, known as Diesel-Ethanol (DE) blends, has shown promise in enhancing engine efficiency and reducing emissions due to their higher oxygen content [15, 16]. However, employing DE directly in IC engines presents notable technical challenges such as a lower cetane number, flash point, and ethanol solubility in diesel.

To address these issues, researchers have investigated the impact of additives, specifically emulsifiers, on the solubility of ethanol in diesel fuel blends like DE. The incorporation of emulsifiers prevents phase separation in Diesel-Ethanol. Emulsifiers function by strengthening the attraction between the two liquid phases, thereby reducing interfacial tension forces and ensuring stable emulsions [17-20].

Some studies suggest that the addition of biodiesel, a fatty acid methyl ester (FAME), as an emulsifier in DE blends can mitigate phase separation issues. FAME acts as an amphiphilic substance, contributing to the stabilization of ethanol and diesel mixtures. Despite these findings, further research is needed to fully establish the viability of biodiesel as an effective emulsifier in Diesel-Ethanol blends, as existing studies on this subject are limited [21-24].

The main objective of this research is to substitute fossil diesel with ethanol in DE blends, employing RSO biodiesel as an emulsifying agent. RSO exhibits considerable potential as a sustainable alternative fuel derived from non-edible oil sources. The biodiesel was synthesized from RSO using the transesterification process. The study focuses on evaluating the operational and emission features of a diesel engine running on Diesel-Ethanol blends with the inclusion of RSO methyl esters as the emulsifying agent.

2 MATERIALS AND METHODOLOGY

2.1 Synthesis of RSO Biodiesel

Rubber seeds sourced from the coastal region of Karnataka, India, underwent processing utilizing a hydraulic pressing machine for oil extraction. Following extraction, the oil underwent heating at 100 °C and subsequent filtration to eliminate suspended particles and moisture. Chemicals required for biodiesel production, namely methanol, sodium hydroxide, and sulfuric acid, were obtained from Sigma-Aldrich, India. The acid value of the oil was assessed through

the standard titrimetric method, initially measuring at 20 mg KOH/g oil. The acid value reduction was achieved through the implementation of saponification during the transesterification process, employing a two-step transesterification method, acid catalyzed esterification followed by base catalyzed transesterification process. The process is explained in the authors previously published papers [25, 26]. The properties of the biodiesel and crude oil are given in Tab. 1.

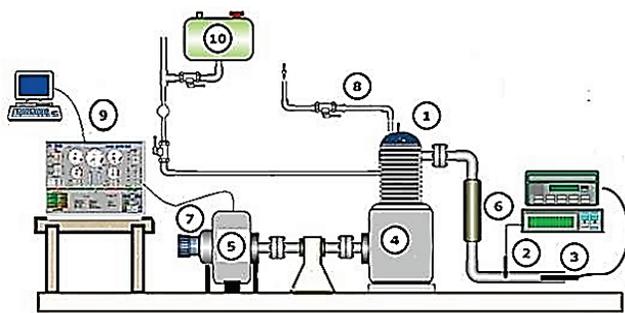
Table 1 Properties of RSO biodiesel

Fuel Properties	RSO	*RSO Biodiesel
Specific gravity	0.92	0.886
Kinematic viscosity 40 °C (mm ² /s)	7.8	3.5
Flashpoint (°C)	198	125
Calorific value (kJ/kg)	38545	40100
Acid value (mg KOH/gm)	10	0.38
Cloud point (°C)	6	16
Pour point (°C)	2	10
Sulphur content (%w/w)	-	0.01
Cetane number	65.50	63.5

*Analysis Results

3 TEST FUELS AND EXPERIMENTAL SETUP

In this present investigation, an analysis is conducted on diesel engines powered by conventional fossil fuel, as well as blends such as DE5B10, DE10B10, DE15B10, and pure B100, assessing their performance and emissions. Different ethanol-diesel mixtures were formulated using RSO biodiesel at a concentration of 10% (v/v) to serve as an emulsifier. To guarantee proper homogenization, the fuel blends underwent agitation for duration of 15 minutes and were then stored for 24 hours to detect any potential phase separation.



- 1-Pressure sensor
- 2-Smoke meter
- 3-Exhaust gas analyser
- 4-Dual cylinder CI engine
- 5-Eddy current dynamometer
- 6-Exhaust gas calorimeter
- 7-Crank angle encoder
- 8-Water inlet to engine
- 9-Control panel & computer
- 10-Fuel tank

Figure 1 Diesel engine test rig

3.1 Engine Experimental Setup

For the investigation, we employed a Kirloskar TV 1 single-cylinder, water-cooled 4-stroke diesel engine featuring a compression ratio of 17.5:1, a rated speed of 1500 rpm, and a rated brake power (BP) of 5.5 kW. The engine's rotational speed remained constant at 1500 rpm throughout

the entire experimental duration. To gauge the emissions of CO, CO₂, NO_x, and HC, an AVL 444 di-gas analyzer was utilized, while smoke opacity was assessed using an AVL 437C smoke meter. The experimental setup, illustrated in Fig. 1, was employed for the study.

4 TEST FUELS AND EXPERIMENTAL SETUP

Within this section, the focus is on the outcomes of the test engine. Various performance metrics, including Brake Specific Energy Consumption (BSEC) and Brake Thermal Efficiency (BTE), were pinpointed. Subsequent sections utilize graphical representations to analyze and elucidate emission parameters such as carbon dioxide (CO₂), hydrocarbon (HC), nitrogen oxide (NO_x), carbon monoxide (CO), and smoke opacity.

4.1 Performance Characteristics

4.1.1 Variation in BTE

The outcomes of the investigation into the brake thermal efficiency (BTE) of diesel, RSO biodiesel (B100), and ethanol-biodiesel blends (DE5B10, DE10B10, and DE15B10) are illustrated in Fig. 2. The results indicate that, across the entire load spectrum, all tested blends exhibited superior BTE compared to both diesel and B100. Under maximum load conditions, DE5B10, DE10B10, and DE15B10 blends achieved BTE increases of 7.29%, 8.86%, and 10.19%, respectively, over diesel. Furthermore, these blends exhibited BTE enhancements of 10.69%, 12.12%, and 13.51% over B100 under the same conditions. The superior performance of DE15B10 can be attributed to the synergistic effects of viscosity and density reduction, leading to improved combustion characteristics. These findings underscore the potential of ethanol-biodiesel blends as efficient and cleaner alternatives to conventional diesel fuel [27-29].

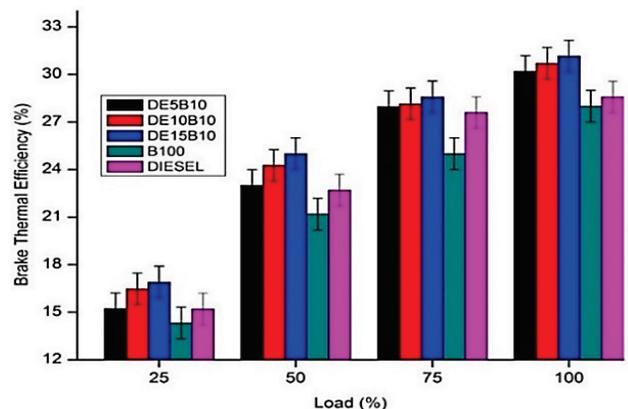


Figure 2 BTE with load

4.1.2 Variation in BSFC

Fig. 3 illustrates the variations in brake-specific fuel consumption (BSFC) across different fuel blends under varying load conditions. As the load increased, a

corresponding decrease in BSFC was observed for all tested fuel blends. At full load, B100 exhibited a 12.3% higher BSFC compared to diesel fuel, indicating increased fuel consumption. Similarly, DE5B10 recorded a 7.39% rise in BSFC, while DE10B10 and DE15B10 displayed increases of 10.69% and 4.30%, respectively, relative to diesel. The higher BSFC observed with increasing ethanol content can be primarily attributed to leaner combustion characteristics and the lower calorific value of ethanol, which results in reduced energy output per unit of fuel. These results align with previous studies, reinforcing the influence of ethanol's oxygenation on combustion efficiency and fuel consumption trends [30, 31].

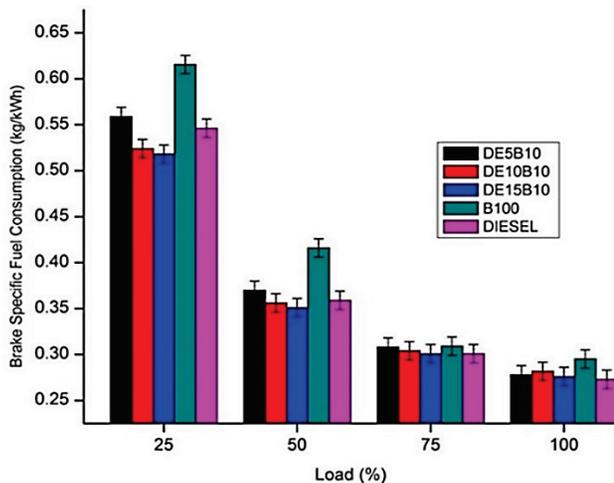


Figure 3 BTE with load

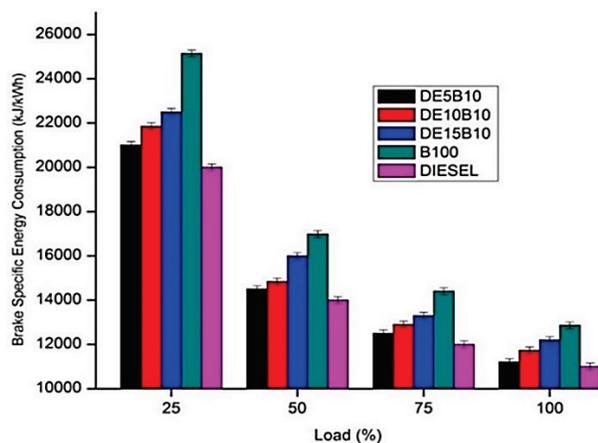


Figure 4 BSEC with load

4.1.3 Variation in BSEC

Fig. 4 illustrates the variations in brake-specific energy consumption (BSEC) across different fuel blends under varying load conditions. The results indicate a consistent decline in BSEC as load increases for all tested blends. This reduction can be attributed to the fact that the percentage increase in brake power (BP) is lower than the corresponding rise in fuel consumption, leading to improved energy utilization at higher loads. Additionally, reduced heat losses at elevated loads further contribute to this trend. Among the

tested fuels, DE5B10 exhibited a 1.82% higher BSEC than conventional diesel, while DE10B10 and DE15B10 recorded increases of 6.39% and 9.95%, respectively. B100 demonstrated the highest BSEC, registering a 14.9% increase over diesel due to its higher calorific value and lower energy efficiency. These results align with previous studies, reaffirming the impact of biodiesel and ethanol blends on fuel energy consumption, combustion efficiency, and overall engine performance [32, 25]. Future research should further explore optimization strategies to enhance fuel efficiency in alternative fuel applications.

4.2 EMISSION CHARACTERISTICS

4.2.1 CO₂ Emission

Fig. 5 illustrates the variations in carbon dioxide (CO₂) emissions across different load conditions for various fuel blends. The results indicate a consistent upward trend in CO₂ emissions with increasing load, regardless of the fuel blend. This increase can be attributed to enhanced combustion efficiency at higher loads, which promotes the complete oxidation of fuel carbon content. Compared to conventional diesel, all tested blends exhibited slightly elevated CO₂ emissions. Specifically, DE5B10 recorded a 1.93% increase, DE10B10 showed a 3.75% rise, and DE15B10 demonstrated the highest increase at 7.2%. This trend is primarily due to the enhanced conversion of carbon monoxide (CO) to CO₂, facilitated by the higher oxygen content introduced through ethanol blending. The greater oxygen availability in ethanol-rich blends promotes more complete combustion, leading to increased CO₂ formation. These findings are consistent with previous research, underscoring the trade-off between reduced CO emissions and slightly higher CO₂ emissions in ethanol-biodiesel-diesel fuel blends [33].

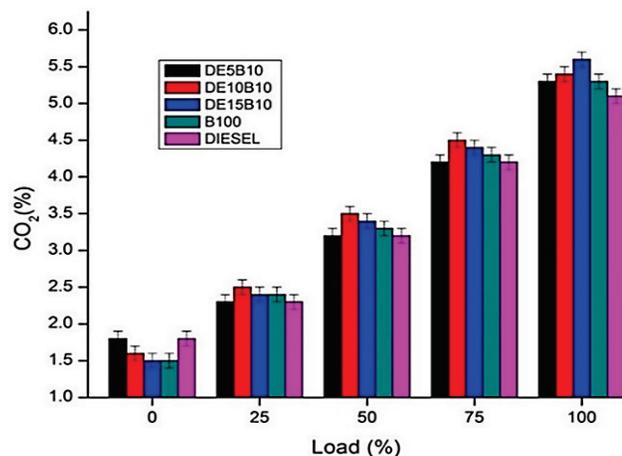


Figure 5 CO₂ emissions with load

4.2.2 HC Emissions

Fig. 6 illustrates the hydrocarbon (HC) emissions for various fuel blends under different load conditions. The results show that HC emissions are highest at full load for all tested fuel mixtures. The increase in HC emissions with higher ethanol content is attributed to unburned ethanol in the

fuel blends, which may lead to incomplete combustion. Compared to diesel, DE5B10 exhibited a 21.3% reduction in HC emissions, while DE10B10 and DE15B10 recorded reductions of 26.52% and 13.92%, respectively. Notably, B100 demonstrated the lowest HC emissions, showing a 31.65% reduction compared to diesel fuel. The improved combustion efficiency of biodiesel and ethanol blends, due to their oxygen content, plays a crucial role in lowering HC emissions. These findings indicate that biodiesel and ethanol-enriched blends contribute to cleaner combustion, making them a promising alternative to conventional diesel fuel in reducing unburned hydrocarbon pollutants [27].

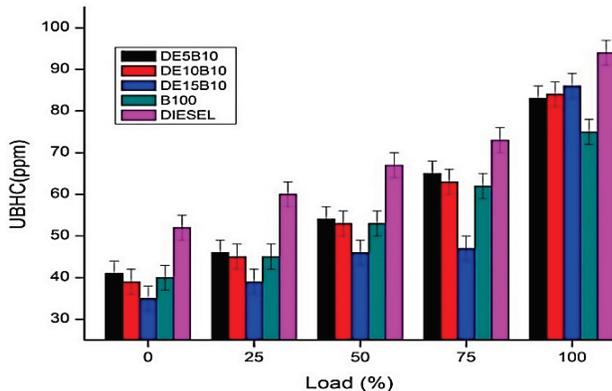


Figure 6 HC emissions with load

4.2.3 NO_x Emissions

Fig. 7 presents a detailed analysis of nitrogen oxide (NO_x) emissions for various fuel blends under different load conditions. The results indicate that, unlike conventional diesel fuel, NO_x emissions increase with higher loads across all tested blends. This trend is primarily attributed to elevated combustion temperatures and increased oxygen availability in ethanol-biodiesel blends, which enhance the formation of NO_x during combustion.

At full load, the NO_x emissions of DE5B10 were 2.15% higher than those of diesel, while DE10B10 exhibited a 3.3% increase. Interestingly, the DE15B10 blend showed NO_x emissions comparable to those of diesel, suggesting a potential balance between combustion efficiency and emissions control. Meanwhile, B100 recorded the highest NO_x emissions, largely due to its high oxygen content, which facilitates more complete combustion but also intensifies NO_x formation.

The increased proportion of ethanol in the blends also plays a significant role in NO_x emissions. Ethanol addition reduces fuel viscosity, improves atomization, and enhances air-fuel mixing, leading to more efficient combustion. However, this also results in higher in-cylinder temperatures and a prolonged premixed combustion phase, which further contribute to NO_x formation.

Despite the slight increase in NO_x emissions, the overall benefits of ethanol-biodiesel blends, such as reduced particulate matter and hydrocarbon emissions, highlight their potential as cleaner alternatives to fossil diesel. Future research should focus on optimizing combustion strategies

and exhaust after-treatment technologies to mitigate NO_x emissions while maximizing efficiency and sustainability [27, 34].

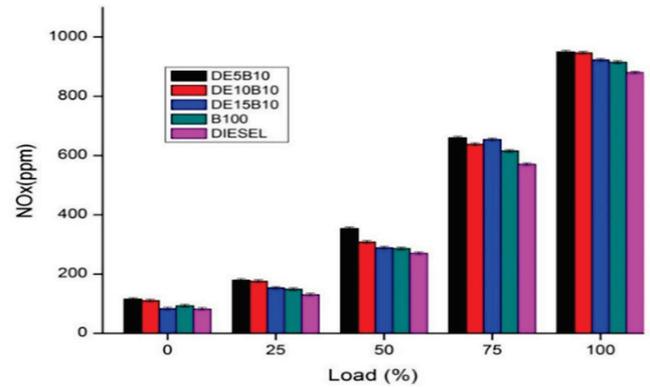


Figure 7 NO_x emissions with load

4.2.4 CO Emission

Fig. 8 illustrates the carbon monoxide (CO) emissions for various fuel blends under different load conditions. The results indicate that at lower loads, CO emissions remain relatively constant, whereas at higher loads, there is a notable increase in CO emissions. However, biodiesel blends consistently exhibit lower CO emissions compared to conventional diesel fuel. This reduction is primarily attributed to the higher oxygen content in biodiesel, which enhances oxidation and promotes the conversion of unburned CO into carbon dioxide (CO₂).

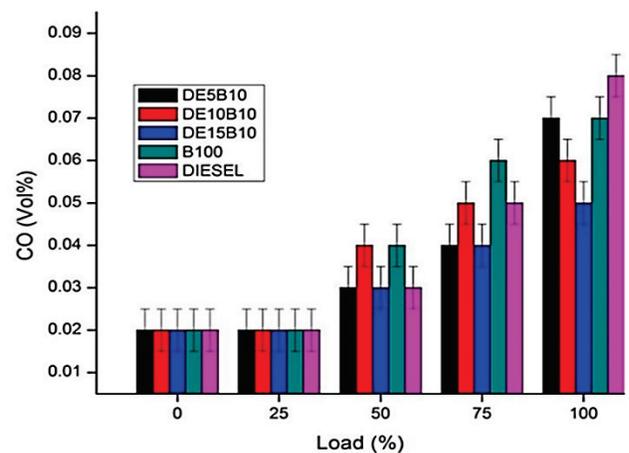


Figure 8 NO_x emissions with load

Under maximum load conditions, DE5B10 recorded a 25.5% reduction in CO emissions, while DE10B10 and DE15B10 demonstrated reductions of 38.3% and 50.5%, respectively, compared to diesel fuel. These significant reductions highlight the potential of ethanol-biodiesel blends in improving combustion efficiency and reducing harmful emissions. The findings confirm that biodiesel-enriched blends serve as an effective alternative to diesel, offering cleaner combustion and contributing to lower environmental pollution levels [34, 35].

4.2.5 Smoke Opacity

Fig. 9 illustrates the smoke opacity levels for various fuel blends under different loading conditions. The results indicate that as the load increases, smoke opacity rises for all tested fuel combinations. Notably, the B100 biodiesel blend exhibits 14.1% higher smoke opacity than diesel fuel across the entire load spectrum, primarily due to its higher viscosity and incomplete combustion at certain operating conditions.

Conversely, ethanol-biodiesel blends (DE5B10, DE10B10, and DE15B10) demonstrate reduced smoke opacity compared to diesel fuel. This reduction is attributed to the presence of free oxygen in ethanol, which enhances combustion efficiency and minimizes particulate matter formation. Specifically, DE5B10 recorded a 3.69% decrease in smoke opacity, while DE10B10 and DE15B10 exhibited reductions of 6.96% and 11.36%, respectively, relative to diesel fuel. These findings highlight the effectiveness of ethanol-biodiesel blends in reducing soot emissions, making them a cleaner alternative to conventional diesel for compression ignition engines [36-39].

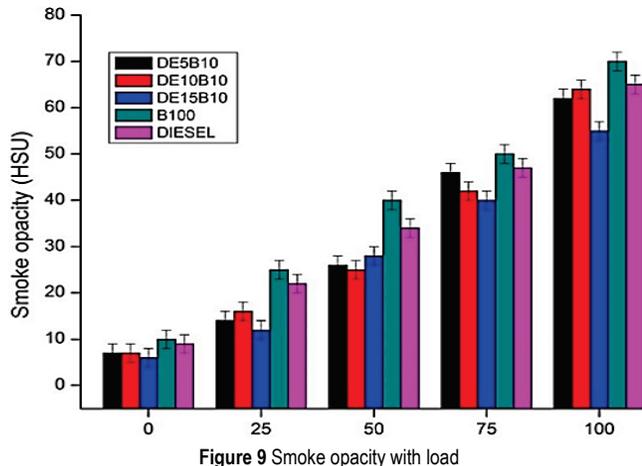


Figure 9 Smoke opacity with load

5 CONCLUSIONS

The primary objective of this study was to synthesize biodiesel from rubber seed oil (RSO) via transesterification and evaluate its feasibility as an additive in diesel-ethanol (DE) blends for compression ignition (CI) engines. Under maximum load conditions, the DE5B10, DE10B10, and DE15B10 blends exhibited increases in brake thermal efficiency (BTE) by 7.29%, 8.86%, and 10.19%, respectively, compared to diesel blends. While all RSO biodiesel blends demonstrated higher brake-specific fuel consumption (BSFC) than diesel, the DE15B10 blend achieved an optimal balance between efficiency and emissions. In terms of emission characteristics, the addition of RSO biodiesel significantly reduced carbon monoxide (CO) emissions. The CO emissions for the DE5B10, DE10B10, and DE15B10 blends were lower than those of traditional diesel fuel, with reductions of 25.5%, 38.3%, and 50.5%, respectively. Hydrocarbon (HC) emissions also decreased, with reductions of 21.3% for the DE5B10 blend, 26.52% for the DE10B10 blend, and 13.92% for the DE15B10 blend. Furthermore, nitrogen oxide (NO_x)

emissions were notably lower under low-load conditions for all biodiesel blends compared to diesel fuel, a critical factor for regulatory compliance with emission standards. Additionally, the integration of RSO biodiesel and ethanol contributed to a significant reduction in smoke opacity, enhancing combustion efficiency and decreasing particulate matter emissions. This study underscores the viability of using RSO biodiesel as a sustainable and renewable emulsifier in DE blends, ensuring phase stability and improved fuel performance. The findings indicate that the DE15B10 blend, with its optimal combination of performance enhancement and emission reduction, is a promising alternative fuel for CI engines without requiring modifications to existing engine hardware. Future research should focus on long-term durability tests and economic feasibility assessments to further validate the commercial applicability of RSO biodiesel blends in the automotive industry [40-42].

6 REFERENCES

- [1] Theogene, N., & Damasenkwa, P. (2019). The economic feasibility of Jatropha cultivation for biodiesel production in Rwanda: A case study of Kirehe district. *International Journal of Sustainable Energy and Development*, 50, 27-37. <https://doi.org/10.1016/j.esd.2019.03.001>
- [2] Thakur, A. K., Kaviti, A. K., Singh, R., & Gehlot A. (2020). An overview of butanol as compression ignition engine fuel. *International Journal of Energy for Clean Environment*, 21(4), 333-354. <https://doi.org/10.1615/InterJEnrCleanEnv.2020033667>
- [3] Govindaraj, E., & Duraisamy, K. (2022). Investigation of Performance and Emission Characteristics of an SI Engine Fitted With Zeolite-Coated Catalytic Converter for Various Ethanol-Gasoline Blends. *International Journal of Energy for Clean Environment*, 23(2), 1-17. <https://doi.org/10.1615/InterJEnrCleanEnv.2021036582>
- [4] Ayoob, A. K., & Fadhil, A. (2019). Biodiesel production through transesterification of a mixture of non-edible oils over lithium supported on activated carbon derived from scrap tires. *Energy Conversion and Management*, 201, 112-149. <https://doi.org/10.1016/j.enconman.2019.112149>
- [5] Ayoob, A. K., & Fadhil, A. (2019). Valorization of waste tires in the synthesis of an effective carbon-based catalyst for biodiesel production from a mixture of non-edible oils. *Fuel*, 264, 116754. <https://doi.org/10.1016/j.fuel.2019.116754>
- [6] Chiong, M.-C., Valera-Medina, A., Chong, W. W. F., Chong, C. T., Mong, G. R., & Mohd Jaafar, M. N. (2020). Effects of swirler vane angle on palm biodiesel/natural gas combustion in swirl-stabilized gas turbine combustor. *Fuel*, 277, 118213. <https://doi.org/10.1016/j.fuel.2020.118213>
- [7] Demirbas, F. (2009). Bio refineries for biofuel upgrading: a critical review. *Applied Energy*, 86, 151-161. <https://doi.org/10.1016/j.apenergy.2009.04.043>
- [8] Kumar, A., & Sharma, S. (2010). Potential Non-Edible Oil Resources as Biodiesel Feedstock: An Indian Perspective, Renewable and Sustainable. *Energy reviews*, 15(4), 1791-1800. <https://doi.org/10.1016/j.rser.2010.11.020>
- [9] Sharma, Y. C., & Singh, B. (2010). A hybrid feedstock for a very efficient preparation of biodiesel. *Fuel Processing Technology*, 91, 1267-73. <https://doi.org/10.1016/j.fuproc.2010.04.008>
- [10] Girish, A. C., Rangnath S. C., & Gowda B. (2015). A Field Survey on Abundance of Biofuel Plant Species in Alur Taluk of Hassan District, Karnataka, India. *Journal of Applied and Natural Science*, 7(1) 26-29.

- <https://doi.org/10.31018/jans.v7i1.558>
- [11] Jena, P. C., Raheman, H., Prasanna Kumar, G. V., & Machavaram, R. (2010). Bio-diesel production from mixture of Mahua and Simarouba oils with high free fatty acids. *Biomass and Bioenergy*, 34(8) 1108-1116. <https://doi.org/10.1016/j.biombioe.2010.02.019>
- [12] Balaji, S., Saravanan, R., & Kapilan, N. (2018). Biodiesel production from azadirachta indica oil and its process optimization. *International Journal of Mechanical and Production Engineering Research and Development*, 8(6), 952-957.
- [13] Miraculas, A. G., Bose, N., & Raj, E. R. (2018). Process Parameter Optimization for Biodiesel Production from Mixed Feedstock Using Empirical Model. *Sustainable Energy Technologies and Assessments*, 28, 54-58. <https://doi.org/10.1016/j.seta.2018.06.004>
- [14] Hong, K. I., Park, W. J., & Lee, B. S. (2012). Optimization of fish oil based biodiesel synthesis. *Journal of Industrial and Engineering Chemistry*, 19, 764-768. <https://doi.org/10.1016/j.jiec.2012.10.011>
- [15] He, B.-Q., Shuai, S. J., Wang, J. X., & He, H. (2003). The effect of ethanol blended diesel fuels on emissions from a diesel engine. *Atmospheric Environment*, 37, 4965-4971. <https://doi.org/10.1016/j.atmosenv.2003.08.029>
- [16] Zhang, R.-D., He, H., Shi, X.-Y., Zhang, C.-B., He, B.-Q., & Wang, J.-X. (2004). Preparation and emission characteristics of ethanol-diesel fuel blends. *Journal of Environmental Science*, 16, 793-796.
- [17] Sandun, F. D., & Milford, H. (2005). Phase behavior of the ethanol-biodiesel-diesel micro emulsion system. *Transactions of ASAE*, 48, 903-908. <https://doi.org/10.13031/2013.18494>
- [18] Kwanchareon, P., Luengnaruemitchai, A., & Jai-In, S. (2007). Solubility of a diesel-biodiesel-ethanol blend, its fuel properties, and its emission characteristics from diesel engine. *Fuel*, 86(7-8), 1053-1061. <https://doi.org/10.1016/j.fuel.2006.09.034>
- [19] Shi, X., Pang, X., Mu, Y., He, H., Shuai, S., Wang, J., Hu, C., & Li, R. (2006). Emission reduction potential of using ethanol-biodiesel-diesel fuel blend on a heavy-duty diesel engine. *Atmospheric Environment*, 40(14), 2567-2574. <https://doi.org/10.1016/j.atmosenv.2005.12.026>
- [20] Zhu, L., Cheung, C. S., Zhang, W. G. & Huang, Z. (2011). Combustion, performance and emission characteristics of a DI diesel engine fuelled with ethanol biodiesel blends. *Fuel*, 90(5), 1743-1750. <https://doi.org/10.1016/j.fuel.2011.01.024>
- [21] Srikanth, H. V., Sharanappa, G. Manne, B., & Kumar, S. B. (2021). Niger seed oil bio-diesel as an emulsifier in diesel-ethanol blends for compression ignition engine. *Renewable Energy*, 163, 1-12. <https://doi.org/10.1016/j.renene.2020.07.010>
- [22] Hansen, A. C., Zhang, Q., & Lyne, P. W. L. (2005). Ethanol-diesel fuel blends – a review. *Bioresource Technology*, 96, 277-285. <https://doi.org/10.1016/j.biortech.2004.04.007>
- [23] Shi, X., Yu, Y., He, H., Shuai, S., Wang, J., & Li, R. (2005). Emission characteristics using methyl soyate-ethanol-diesel fuel blends on a diesel engine. *Fuel*, 84(12-13), 1543-1549. <https://doi.org/10.1016/j.fuel.2005.03.001>
- [24] Balaji, S., Saravanan, R., & Kapilan, N. (2021). Influence of Propyl gallate antioxidant on performance and emissions of a compression ignition engine fueled with MadhucaIndica B20 ester blends. *Energy Sources, Part A: Recovery, Utilization, and Environmental Effects*, 43(18), 2197-2209. <https://doi.org/10.1080/15567036.2019.1644396>
- [25] Srikanth, H. V., Venkatesh, J., Godiganur, S., Manne, B., Bharath Kumar, S., & Spurthy, S. (2019). Combustion, performance, and emission characteristics of dairy-washed milk scum biodiesel in a dual cylinder compression ignition engine. *Energy Sources, Part A: Recovery, Utilization, and Environmental Effects*, 42(23), 2873-2890. <https://doi.org/10.1080/15567036.2019.1618996>
- [26] Srikanth, H. V., Venkatesh, J., & Godiganur, S. (2018). Box-Behnken Response Surface Methodology for Optimization of Process Parameters for Dairy Washed Milk Scum Biodiesel Production. *Biofuels*, 12(1), 113-123. <https://doi.org/10.1080/17597269.2018.1461511>
- [27] Venkata Subbaiah, G., Raja Gopal, K., Hussain, S. A., Durga Prasad, B., & Tirupathi Reddy, K. (2010). Rice bran oil biodiesel as an additive in diesel-ethanol blends for diesel engines. *International Journal of Research and Reviews in Applied Sciences*, 3(3), 334-342.
- [28] Klajn, F. F., Gurgacz, F., Lenz, A. M., Iacono, G. E. P., Souza, S. N. M. de, & Ferruzzi, Y. (2018). Comparison of the emissions and performance of ethanol-added diesel-biodiesel blends in a compression ignition engine with those of pure diesel. *Environmental Technology*, 41(4), 511-520. <https://doi.org/10.1080/09593330.2018.1504122>
- [29] Racopoulos, C. D., Antonopoulos, K. A., & Racopoulos, D. C. (2007). Experimental heat release analysis and emissions of a HSDI diesel engine fueled with ethanol-diesel fuel blends. *Energy*, 32(10), 1791-1808. <https://doi.org/10.1016/j.energy.2007.03.005>
- [30] Wie, L., Cheung, C. S., & Ning, Z. (2018). Effects of biodiesel-ethanol and biodiesel-butanol blends on the combustion, performance and emissions of a diesel engine. *Energy*, 155, 957-970. <https://doi.org/10.1016/j.energy.2018.05.049>
- [31] Tongroon, M., Saisirirat, P., Suebwong, A., Aunchaisri, J., Kananont, M., & Chollacoop N. (2019). Combustion and emission characteristics investigation of diesel-ethanol biodiesel blended fuels in a compression-ignition engine and benefit analysis. *Fuel*, 255, 115728. <https://doi.org/10.1016/j.fuel.2019.115728>
- [32] Abhishek, P., Rajsekhar, P., & Durbadal, D. (2017). An experimental study of combustion, performance, exergy and emission characteristics of a CI engine fueled by diesel-ethanol-biodiesel-blends. *Energy*, 141, 839-852. <https://doi.org/10.1016/j.energy.2017.09.137>
- [33] Krishna, S. M., Abdul Salam, P., Tongroon, M., & Chollacoop, N. (2019). Performance and emission assessment of optimally blended biodiesel-diesel-ethanol in diesel engine generator. *Applied Thermal-Engineering*, 155, 525-533. <https://doi.org/10.1016/j.applthermaleng.2019.04.012>
- [34] Shahir, S. A., Masjuki, H. H., Kalam, M. A., Imran, A., & Ashraful, A. M. (2015). Performance and emission assessment of diesel-biodiesel-ethanol/bioethanol blend as a fuel in diesel engines: A review. *Renewable and Sustainable Energy Reviews*, 48, 62-78. <https://doi.org/10.1016/j.rser.2015.03.049>
- [35] Jamrozik, A., Tutak, W., Pyrc, M., & Sobiepański, M. (2017). Effect of diesel-biodiesel-ethanol blend on combustion, performance, and emissions characteristics on a direct injection diesel engine. *Thermal Science*, 21(1), 591-604. <https://doi.org/10.2298/TSC160913275J>
- [36] Alptekin, E., Canakci, M., Ozsezen, A. N., Turkcan, A., & Sanli, H. (2015). Using waste animal fat-based biodiesels bioethanol-diesel fuel blends in a DI diesel engine. *Fuel*, 157, 245-254. <https://doi.org/10.1016/j.fuel.2015.04.067>
- [37] Açikkalp, E., & Yerel Kandemir, S. (2019). A method for determining optimum insulation thickness: Combined economic and environmental method. *Thermal Science and Engineering Progress*, 11, 249-253. <https://doi.org/10.1016/j.tsep.2019.04.004>
- [38] Açikkalp, E., Yerel Kandemir, S., & Ahmadi, M. H. (2020). Performance Evaluation of the Thermophotovoltaic-Driven Thermoionic Refrigerator. *J. Energy Resour. Technol.*, 142(3), 032001. <https://doi.org/10.1115/1.4044558>

- [39] Açikkalp, E., & Yerel Kandemir, S. (2018). Optimum insulation thickness of the piping system with combined economic and environmental method. *Energy Sources, Part A: Recovery, Utilization, and Environmental Effects*, 40(23), 2876-2885. <https://doi.org/10.1080/15567036.2018.1512683>
- [40] Bibin, C., Sheeja, R., Devarajan, Y., Madhu, S., Venkatesan, S. P., & Natrayan, L. (2023). Environment impact assessment of agricultural diesel engines utilizing biodiesel derived from phoenix sylvestris oil. *Environmental Research*, 224, 115432. <https://doi.org/10.1016/j.envres.2023.115432>
- [41] Vijayaragavan, M., Subramanian, B., Sudhakar, S., & Natrayan, L. (2022). Effect of induction on exhaust gas recirculation and hydrogen gas in compression ignition engine with simarouba oil in dual fuel mode. *International Journal of Hydrogen Energy*, 47(88), 37635-37647. <https://doi.org/10.1016/j.ijhydene.2021.11.201>
- [42] Mahesha, C. R., Jamuna Rani, G., Dattu, V. S. N. C. H., Rao, Y. K. S. S., Madhusudhanan, J., Natrayan L., Sekhar, S. C., & Sathyamurthy, R. (2022). Optimization of transesterification production of biodiesel from Pithecellobium dulce seed oil. *Energy Reports*, 8(16), 489-497. <https://doi.org/10.1016/j.egyr.2022.10.228>

Authors' contacts:

H. V. Srikanth

Department of Aeronautical Engineering,
Nitte Meenakshi Institute of Technology,
Nitte (Deemed to be University)
Yelahanka, Bangalore – 560064, Karnataka, India

S. Vijay Kumar

Department of Mechanical Engineering,
Nitte Meenakshi Institute of Technology,
Nitte (Deemed to be University)
Yelahanka, Bangalore – 560064, Karnataka, India

Kousik Kumaar Rajagopal

(Corresponding author)
Department of Aeronautical Engineering,
Nitte Meenakshi Institute of Technology,
Nitte (Deemed to be University)
Yelahanka, Bangalore – 560064, Karnataka, India
kousikkumaarphd@gmail.com

Senthil Kumar Madasamy

Department of Aeronautical Engineering,
Kumaraguru College of Technology,
Coimbatore - 641049, Tamil Nadu, India
senthilkumar.m.aeu@kct.ac.in

Beena Stanislaus Arputharaj

Department of Research and Innovation,
Saveetha School of Engineering, SIMATS,
Chennai-602105, Tamil Nadu, India
beena2192@gmail.com

Parvathy Rajendran

Department of Mechanical and Aerospace Engineering,
United Arab Emirates University,
Al Ain, UAE
aeparvathy@uaeu.ac.ae

Vijayanandh Raja

(Corresponding author)
Department of Aeronautical Engineering,
Kumaraguru College of Technology,
Coimbatore - 641049, Tamil Nadu, India
vijayanandh.raja@gmail.com

GLDM Algorithm for Big Data (SCADA) Wind Speed Modelling

Mostafa Abotaleb

Abstract: This study enhances wind speed forecasting by implementing the second-order Generalized Least Deviation Method (GLDM), focusing on wind turbines in Turkey. The research aims to improve predictive accuracy and operational efficiency in renewable energy systems through advanced mathematical modeling in meteorology. The GLDM, utilizing a quasilinear recurrence equation, addresses the inherent non-linearity and variability of wind speed data. By applying the method to extensive SCADA data, this study minimizes residuals in nonlinear big data environments, integrating both linear and nonlinear components to refine predictions. A critical aspect of this research is the comparison between the second-order GLDM and traditional forecasting models, including statistical methods and machine learning approaches. The results demonstrate the superior performance of GLDM, as indicated by lower prediction errors and greater accuracy across key metrics. The study also underscores the importance of GLDM coefficients, a_i , in improving predictive capabilities. The findings advocate for the adoption of GLDM in wind speed forecasting, highlighting its potential to significantly enhance wind energy management through increased accuracy. This study also sets a precedent for broader applications of advanced mathematical models in environmental science, illustrating the effectiveness of GLDM in optimizing renewable energy resources.

Keywords: atmospheric dynamics; Generalized Least Deviation Method (GLDM); renewable energy optimization; SCADA Data Analysis; statistical model validation; wind speed forecasting; wind turbine efficiency

1 INTRODUCTION

The advancement of renewable energy technologies has become a global imperative in the face of the pressing environmental challenges posed by conventional energy sources. Among the various renewable options, wind power has emerged as a leading contender, harnessing the abundant and sustainable resource of wind to generate electricity. However, the inherent variability and intermittency of wind power pose significant challenges for its seamless integration into the energy grid. Accurate forecasting of wind speeds is thus a crucial prerequisite for optimizing the operations and efficiency of wind energy systems. Wind speed forecasting has become a pivotal area of research, drawing the attention of scientists, engineers, and policymakers alike. The development of sophisticated statistical models and data-driven approaches has enabled significant strides in enhancing the precision and reliability of wind speed predictions. These advancements not only improve the management of wind farms and energy distribution but also contribute to broader environmental sustainability goals, such as reducing greenhouse gas emissions and mitigating the impacts of climate change. As the global commitment to renewable energy continues to strengthen, the field of wind speed forecasting remains a dynamic and interdisciplinary domain, poised to unlock new frontiers in renewable energy optimization and environmental stewardship. The continuous release of greenhouse gases by conventional energy resources derived from fossil fuels contributes significantly to global warming and its detrimental impacts on the Earth's atmosphere. Consequently, there is an urgent need to expand the capacity of power plants based on renewable energy sources to satisfy the steadily increasing global energy demand. Notably, onshore wind and photovoltaic solar power plants, which are among the most technologically advanced renewable resources, offer the benefit of having the lowest levelized cost of energy [12]. As such, it is anticipated that these resources will constitute a larger proportion of the

total renewable energy mix in the future. It has been observed that the power generation from these two resources often exhibits an inverse relationship, particularly on a monthly scale [13]. Additionally, wind energy is recognized for its significant variability on an hourly basis, making it an intermittent source of power. Furthermore, as wind power's integration into the electrical grid system expands, not only do operational costs escalate, but the reliability of the system also diminishes [14]. Thus, precise wind forecasting is essential to ensure energy supply security and to address the intermittency associated with wind energy.

Wind speed forecasting can be categorized into three distinct types: physical, statistical, and hybrid models. Physical models, often requiring extensive computational resources, are predominantly applied for long-term forecasting. In contrast, statistical models are more commonly utilized for short-term predictions of wind speed. Although numerical weather prediction (NWP) models serve as a direct basis for physical wind speed forecasting, the outputs from these NWP models can also function as initial and boundary condition data in subsequent NWP models for the purpose of downscaling [15].

Wind energy stands as a cornerstone of the global transition towards renewable power sources, driven by its sustainability and the growing imperative to reduce greenhouse gas emissions. Accurate wind speed forecasting emerges as a pivotal challenge within this context, underpinning the efficiency and reliability of wind turbines. The ability to predict wind speeds with high precision is instrumental for operational planning, energy yield optimization, and minimizing the gap between generated and demanded power. This study introduces a novel approach to wind speed forecasting, leveraging the rich datasets provided by Supervisory Control and Data Acquisition (SCADA) systems from operational wind turbines in Turkey. Big Data has emerged as a transformative force across various industries, revolutionizing how we collect, analyse, and utilize vast amounts of information. In the realm of

renewable energy, particularly wind power, Big Data plays a pivotal role in optimizing operations, enhancing efficiency, and ensuring reliability. One crucial aspect where Big Data is indispensable is in the monitoring and analysis of wind speed through Supervisory Control and Data Acquisition (SCADA) systems. Wind energy generation heavily relies on the availability and consistency of wind speeds. Understanding wind patterns, variations, and potential disruptions is paramount for maximizing energy output and maintaining the integrity of wind turbines. SCADA systems serve as the backbone of this process, continuously collecting real-time data from various sensors installed on wind turbines and meteorological stations. These systems capture an extensive range of parameters, including wind speed, direction, and temperature, humidity, and turbine performance metrics. The volume of data generated by SCADA systems can be staggering, often reaching terabytes or even petabytes of information. Managing and processing this vast dataset requires sophisticated Big Data analytics capabilities. Advanced algorithms and machine learning techniques are employed to analyze historical data, identify trends, and predict future wind patterns with a high degree of accuracy. This predictive capability is invaluable for optimizing turbine operation, scheduling maintenance activities, and mitigating potential downtime.

Moreover, Big Data analytics enables wind farm operators to enhance resource allocation and decision-making processes. By integrating SCADA data with geographical information systems (GIS) and other external datasets, operators can gain deeper insights into local weather patterns, topographical features, and environmental factors that influence wind behaviour. This holistic understanding allows for more informed site selection, layout optimization, and strategic planning, ultimately improving the overall performance and profitability of wind farms. Furthermore, the application of Big Data in wind speed analysis extends beyond operational optimization to include grid integration and energy forecasting. By leveraging historical SCADA data along with meteorological forecasts and market trends, energy providers can accurately predict future energy production levels and adjust grid operations accordingly. This proactive approach enhances grid stability, facilitates renewable energy integration, and supports the transition towards a more sustainable energy infrastructure.

Modelling univariate time series data for wind speed is essential for understanding and predicting the behaviour of this critical variable in renewable energy production. Time series modelling involves analysing the sequential nature of data points collected over time to uncover patterns, trends, and dependencies. In the context of wind speed, univariate time series models focus solely on the historical variations of wind speed without considering other variables. A frequently utilized method for simulating wind speed time series is the Autoregressive Integrated Moving Average (ARIMA) model. ARIMA models are exceptionally appropriate for data that is stationary, characterized by constant mean and variance throughout the time series. This makes them highly effective in analysing and predicting patterns where data points tend to return to a long-term average, providing a

robust framework for understanding wind speed fluctuations over time. By harnessing the autocorrelation and seasonality inherent in wind speed data, ARIMA models can offer critical insights into both short-term fluctuations and long-term trends. This capability allows for a detailed understanding of wind speed dynamics, facilitating more accurate forecasting and strategic planning in fields that rely on wind speed data, such as renewable energy management and meteorological research. Additionally, extensions such as seasonal ARIMA (SARIMA) can account for periodic patterns that are characteristic of wind behaviour. Another popular method for modelling wind speed time series is the use of machine learning algorithms, such as recurrent neural networks (RNNs) and long short-term memory (LSTM) networks. These deep learning models excel at capturing complex temporal dependencies and nonlinear relationships in the data. By training on historical wind speed observations, RNNs and LSTMs can learn to make accurate predictions of future wind speeds, enabling better resource planning and operational decision-making for wind farms. Moreover, hybrid approaches that combine traditional statistical methods with machine learning Techniques provide a complete approach for simulating time series for wind speed. For example, integrating ARIMA models with neural networks can leverage the strengths of both approaches, resulting in improved forecasting accuracy and robustness. Additionally, ensemble methods such as bagging and boosting can further enhance prediction performance by aggregating multiple models' outputs.

Through the application of the second-order Generalized Least Deviation Method (GLDM), our research not only seeks to enhance forecasting accuracy but also to contribute to the broader understanding of wind dynamics. The focus on SCADA data, reflecting actual turbine performance over time, provides a robust foundation for our analytical models, offering insights into the complex interplay of meteorological and operational factors affecting wind speed. This introductory exploration sets the stage for a detailed investigation into the potential of advanced mathematical modelling techniques to revolutionize wind speed forecasting and, by extension, the efficiency of wind energy production. The prevailing methodologies for forecasting wind speed encompass the persistence technique, which adopts a physical stance through numerical time predictors, alongside various statistical approaches. These include the Autoregressive with Exogenous Input (ARX), Autoregressive Moving Average (ARMA), and Autoregressive Integrated Moving Average (ARIMA) models, as well as several neural network architectures such as the Artificial Neural Network (ANN), Adaptive Linear Element Network (ADALINE), Feed-forward Neural Network (FNN), and Multilayer Perceptron (MLP), Recurrent Neural Network (RNN), and Radial Basis Function (RBF) networks [16]. Recent advancements have seen machine learning algorithms (MLAs) like spatial correlation, Fuzzy Logic (FL), and Ensemble prediction being integrated with Artificial Neural Networks (ANNs) during the forecasting process. This integration has given rise to hybrid models, offering an innovative alternative that

enhances the accuracy of wind speed predictions [17]. Conversely, methodologies like the Autoregressive Moving Average (AR), ARMA, ARIMA, ARX-Type, and Fractional-Autoregressive Integrated Moving Average (f-ARIMA) [18]. Necessitate a foundational understanding of mathematical models and are predominantly applied to long-term forecasting. These models require linearity, offering straightforward implementation and the establishment of dependable intervals. Consequently, analysing the nature of the time series is crucial for selecting a suitable model. Conversely, models utilizing Artificial Neural Networks (ANN) are adept at handling non-linear systems without necessitating preliminary mathematical modelling knowledge. Furthermore, certain hybrid models leverage time series data within artificial intelligence frameworks to forecast wind speeds. Such models prove invaluable by offering critical insights into harnessing a locale's wind potential for prospective wind energy installations, through the projection of future wind speeds [19]. Furthermore, these models exhibit flexibility in handling inline measurements and possess fault tolerance capabilities. Hence, it is advisable to utilize extensive time series data for network training to enhance the accuracy of wind speed forecasting outcomes.

Among various forecasting methodologies, the Generalized Least Deviation Method (GLDM) stands out as an innovative technique for analysing complex environmental systems. Initially conceptualized within the realm of mechanical engineering, GLDM has proven to be an effective framework for understanding the intricate dynamics of wind speed patterns. This method adeptly captures both linear and non-linear interactions among crucial atmospheric variables, offering a comprehensive approach to wind speed forecasting. In our investigation, we explore the utility of GLDM in the context of wind energy generation, specifically focusing on its application to wind speed data collected from turbines. By analysing historical wind speed records, our objective is to evaluate the capacity of GLDM to accurately predict wind speeds, thereby facilitating more efficient wind turbine operation. Through an in-depth examination of the model's coefficients, error assessments, and the statistical validation of its predictions, we aim to discern the model's performance across varying complexities and its general applicability to environmental forecasting. The outcomes of this analysis are anticipated to provide valuable insights not only for the enhancement of wind energy production but also for the broader field of environmental modelling. By demonstrating the efficacy and versatility of GLDM in capturing wind speed fluctuations, this research contributes to the advanced methodologies available for environmental prediction and management. Our ultimate aim is to offer findings that aid energy specialists, meteorologists, and environmental policymakers in harnessing wind resources more effectively, promoting sustainable energy solutions amidst changing climatic conditions. In our research [1, 20], we elaborate on the methodologies used to determine the parameters of a unique difference equation with quasilinear characteristics, which has been identified as addressing regression analysis challenges pertaining to interdependent observed variables. This method facilitates the application of

the Generalized Least Deviations Method (GLDM) to wind speed data. In this study, we conduct computational experiments using both a short and a long wind speed dataset to demonstrate the statistical relevance of the model's coefficients. Exploring this model is particularly critical as, in contrast to neural network-based approaches, it provides a transparent mechanism for deriving high-quality quasilinear difference equations. These equations offer a more precise description of wind speed dynamics, underscoring the significance of our method in enhancing the accuracy and reliability of wind speed forecasting [21-30].

At the heart of the Generalized Least Deviation Method (GLDM)'s success in wind speed forecasting lies its robust mathematical formulation. This method optimizes an objective function $F(\mathbf{a}) = \sum_{i=1}^n |y_i - f(x_i; \mathbf{a})|$, where y_i denotes the actual wind speeds observed, $f(x_i; \mathbf{a})$ represents the predicted wind speeds by the model, and \mathbf{a} symbolizes the coefficients vector. The optimization process seeks to find the best set of coefficients $\mathbf{a} = \{a_1, a_2, \dots, a_m\}$, with m indicating the model's complexity level. The determination of m is crucial, directly affecting the model's ability to accurately capture the dynamics of wind speed fluctuations. Our research employs computational experiments alongside statistical assessments to clarify how different model complexities can be optimally chosen for various wind speed datasets. This endeavour not only aims to validate the GLDM's applicability to wind speed forecasting but also seeks to enhance the synergy between the model's theoretical framework and its practical execution. This study's insights aim to further the understanding of GLDM's application scope, demonstrating its utility in delivering precise and reliable wind speed forecasts in the context of renewable energy. Time series analysis is a fundamental tool for understanding the dynamics of sequential data, ranging from economic indicators to physiological signals. Quasilinear recurrence equations offer a versatile framework for modeling the evolution of univariate time series, allowing for the incorporation of both linear and nonlinear dependencies. In this paper, we explore the concept of quasilinear recurrence equations and discuss their adaptation for modeling univariate time series data.

A quasilinear recurrence equation represents the evolution of a variable x_t at time t as a function of its lagged values $x_{t-1}, x_{t-2}, \dots, x_{t-k}$, along with an error term ϵ_t . Mathematically, it can be expressed as: $x_t = f(x_{t-1}, x_{t-2}, \dots, x_{t-k}) + \epsilon_t$, where f represents a possibly nonlinear function capturing the dependencies between the lagged values, and ϵ_t denotes the error term at time t . The parameter k indicates the lag order, determining how many previous observations influence the current value.

In the context of univariate time series analysis, quasilinear recurrence equations are often used to model the dynamics of a single variable over time. The function f in Equation (1) can take various forms, depending on the characteristics of the data and the underlying dynamics. For example:

- In autoregressive models, f might represent a linear combination of lagged values, expressed as

$f(x_{t-1}, x_{t-2}, \dots, x_{t-k}) = \sum_{i=1}^k \phi_i x_{t-i}$, where ϕ_i are the autoregressive coefficients.

- In nonlinear models, f could involve nonlinear transformations or interactions between lagged values, such as $f(x_{t-1}, x_{t-2}) = \sin(x_{t-1}) + \cos(x_{t-2})$.

The choice of f depends on the complexity of the underlying dynamics and the assumptions about the data-generating process. Estimating the parameters of the function f is a crucial step in modeling univariate time series data. Various estimation techniques can be employed, including:

- **Least Squares Estimation:** This method minimizes the sum of squared residuals between the observed and predicted values of the time series.

- **Maximum Likelihood Estimation:** ML estimation maximizes the likelihood function of the observed data, assuming a specific distribution for the error term ϵ_t .

- **Bayesian Inference:** Bayesian methods incorporate prior beliefs about the parameters and update them using Bayes' theorem to obtain posterior distributions.

These techniques allow us to identify the functional form of f and estimate its parameters from observed data, providing insights into the underlying dynamics of the time series. Quasilinear recurrence equations can be extended to incorporate additional factors such as seasonality, trend, and exogenous variables. For instance, in seasonal time series analysis, f might incorporate seasonal dummies or seasonal autoregressive terms to capture recurring patterns. Furthermore, the error term ϵ_t can be assumed to follow certain distributions such as Gaussian, Student's t, or GARCH processes, depending on the characteristics of the data and the desired properties of the model.

Quasilinear recurrence equations provide a flexible framework for modelling univariate time series data, allowing for the incorporation of both linear and nonlinear dependencies. By adapting these equations to the specific characteristics of the data and employing appropriate estimation techniques, we can develop accurate and informative models for analysing and forecasting time series. Future research could explore further extensions of quasilinear recurrence equations and their applications in diverse domains.

The utilization of quasilinear recurrence equations offers a versatile approach to modelling the dynamics of univariate time series data, accommodating both linear and nonlinear dependencies. In this context, these equations represent the evolution of a variable over time as a function of its lagged values, along with an error term to account for stochastic fluctuations. The formulation allows for diverse representations of the underlying dynamics, ranging from autoregressive models with linear combinations of lagged values to more complex nonlinear transformations and interactions. Estimating the parameters of these equations involves various techniques such as least squares estimation, maximum likelihood estimation, and Bayesian inference, which provide insights into the structure and behaviour of the time series. Additionally, quasilinear recurrence equations can be adapted to include other elements such as seasonality,

trends, and external variables, improving their relevance for real-world data. Therefore, these equations are a valuable resource for examining and predicting time series data in various fields. Future research endeavours could explore further extensions and applications of quasilinear recurrence equations, thereby advancing our understanding of complex temporal dynamics and facilitating more accurate predictions in diverse fields. The method utilized in this study, while conceptually analogous to the traditional Least Absolute Deviation (LAD) method, is referred to as the Generalized Least Deviation Method (GLDM) to emphasize the specific enhancements introduced for wind speed forecasting. Although both methods aim to minimize the sum of absolute deviations between observed and predicted values, GLDM has been extended through the incorporation of the arctan function in the optimization process. This modification enhances the method's robustness against outliers and improves its capacity to address the non-linear characteristics inherent in wind speed data. The coefficients within the GLDM framework are derived by solving a constrained optimization problem, where the objective function $F(a) = \sum_{i=1}^n \arctan(|y_i - \hat{y}_i(a)|)$ is minimized. The optimization process is approached through both primal and dual formulations. The primal problem focuses on minimizing the deviation function under specific constraints, while the dual problem involves maximizing the associated Lagrangian, thus providing a complementary perspective on the solution space. Through the use of both primal and dual optimization techniques, a thorough exploration of the solution space is achieved, leading to a stable and reliable estimation of the model parameters. These methodological advancements, which include enhanced handling of non-linearities and outliers, distinguish GLDM from the conventional LAD method. The adoption of the GLDM nomenclature reflects these innovations, underscoring its appropriateness for the complex and variable conditions encountered in wind speed forecasting.

2 DATASET DESCRIPTION

The dataset pivotal to our analysis is sourced from a comprehensive SCADA (Supervisory Control and Data Acquisition) system dataset of a wind turbine, made publicly available on Kaggle at <https://www.kaggle.com/datasets/berkerisen/wind-turbine-scada-dataset>. This dataset represents an invaluable asset for wind energy research, encapsulating high-resolution operational data from a wind turbine actively generating power in Turkey.

Stored in CSV format and approximately 2 megabytes in size, the dataset encompasses 50,531 records spread across five columns. The attributes include 'Date/Time', signifying the timestamp of data recording; 'LV Active Power (kW)', detailing the power output; 'Wind Speed (m/s)', which is the primary variable of interest for our forecasting model; 'Theoretical Power Curve (kWh)', representing the expected power generation based on wind speed; and 'Wind Direction (°)', indicating the wind's direction at the time of measurement.

For the purposes of this study, our analysis is focused on the ‘Wind Speed (m/s)’ column. This choice is driven by our objective to refine wind speed forecasting methods, thereby enhancing the predictive accuracy and operational efficiency of wind turbines. The dataset’s granularity, with recordings at ten-minute intervals, provides a rich temporal resolution that is instrumental in capturing the dynamic and fluctuating nature of wind speeds. This high level of detail supports the development and validation of our second-order Generalized Least Deviation Method (GLDM) forecasting model, offering new insights into wind behaviour and its implications for wind energy production.

The dataset’s wind speed variable, crucial for forecasting turbine energy output, is summarized in Tab. 1. With a total count of 50,530 observations, the wind speed showcases a mean value of 7.56 m/s, indicative of the site’s moderate wind conditions. The standard deviation of 4.23 m/s reflects significant variability, underscoring the challenging nature of accurate wind speed forecasting. The minimum recorded wind speed is 0.00 m/s, highlighting periods of calm, while the maximum speed reaches 25.21 m/s, pointing to instances of very high wind conditions. The distribution’s quartiles, with the 25th percentile at 4.20 m/s, the median at 7.10 m/s, and the 75th percentile at 10.30 m/s, further describe the dataset’s spread, illustrating the common wind speeds that turbines are likely to encounter. This statistical overview lays the groundwork for our analysis, emphasizing the importance of developing robust forecasting models capable of accommodating the broad range of wind speeds observed.

Table 1 Statistical summary of wind speed measurements

Statistic	Value (m/s)
Count	50,530
Mean	7.56
Standard Deviation	4.23
Minimum	0.00
25 th Percentile	4.20
Median (50 th Percentile)	7.10
75 th Percentile	10.30
Maximum	25.21

3 METHOD

Wind speed forecasting is crucial for efficient energy generation and distribution, especially in the context of renewable energy sources such as wind farms. With the proliferation of sensor technology and the advent of big data analytics, there has been a growing interest in developing accurate forecasting models for wind speed based on historical data. Univariate time series forecasting, focusing solely on the wind speed variable, offers a practical approach to predicting future wind conditions, enabling better resource allocation and grid management. In this study, we delve into the complexities and approaches associated with forecasting univariate time series data, specifically focusing on large-scale wind speed datasets. Prior to the development of predictive models, it is imperative to conduct a thorough pre-processing of the wind speed data. This step is crucial to enhancing the data’s quality and ascertaining its appropriateness for subsequent analytical procedures.

Through this paper, we aim to outline effective strategies and challenges in the predictive modelling of wind speed, emphasizing the importance of rigorous data preparation to ensure accurate and reliable forecasting outcomes. This involves steps such as data cleaning to remove outliers and missing values, data normalization to scale the values within a certain range, and feature engineering to extract relevant information from the raw data. In the context of wind speed forecasting, additional considerations may include dealing with seasonality, trend, and periodic fluctuations caused by diurnal and weather patterns. Despite the advancements in forecasting techniques, several challenges remain in implementing univariate time series forecasting for wind speed big data. These challenges include dealing with data heterogeneity and spatiotemporal dependencies, handling non stationarity and seasonality, incorporating external factors such as weather patterns and topography, and addressing computational scalability issues for analysing large-scale datasets. Addressing these challenges requires interdisciplinary collaboration between domain experts, data scientists, and engineers, leveraging advanced analytics, machine learning, and high-performance computing techniques. univariate time series forecasting offers a practical approach to predicting wind speed based on historical data, enabling better resource allocation and grid management in renewable energy systems. By preprocessing the data, selecting appropriate forecasting models, evaluating their performance, and addressing implementation challenges, stakeholders can develop accurate and reliable forecasting solutions for wind speed big data. Future research could explore hybrid approaches integrating multiple forecasting techniques and incorporating real-time data streams for enhanced prediction capabilities.

Before delving into our methodological advancements in wind speed forecasting, it is essential to grasp the mathematical and computational principles that underpin the Generalized Least Deviation Method (GLDM). At the core of GLDM lies the objective of minimizing the differences between actual and predicted wind speeds, encapsulated in the optimization challenge: minimize $L(\mathbf{a}) = \sum_{i=1}^n |y_i - \hat{y}_i(\mathbf{a})|$, where y_i represents the observed wind speed measurements, $\hat{y}_i(\mathbf{a})$ denotes the model’s estimated wind speeds, and $\mathbf{a} = \{a_1, a_2, \dots, a_k\}$ denotes the array of model coefficients. The robustness of the Generalized Least Deviation Method (GLDM) in wind speed forecasting is largely due to its unique approach to minimizing deviations through the use of the arctan function in the optimization process. Unlike conventional methods that focus on minimizing squared errors, GLDM minimizes $\arctan(|y_i - \hat{y}_i|)$, where y_i represents the observed values and \hat{y}_i denotes the predicted values. This approach inherently reduces sensitivity to outliers, as the arctan function tempers the influence of extreme values, ensuring that the model remains stable and reliable even in the presence of anomalies.

Moreover, GLDM’s ability to effectively model non-linear dynamics is enhanced by its integration of quasilinear recurrence equations, which allow the method to capture both linear and non-linear patterns within wind speed data. This

dual capability to manage outliers and complex relationships makes GLDM particularly well-suited for forecasting in challenging environments where traditional linear models may fall short. Through the application of GLDM across varying orders and diverse datasets, our investigation seeks to identify the optimal model configuration that balances simplicity with predictive precision. This comprehensive examination forms the basis for assessing the model's effectiveness and its suitability for enhancing the accuracy of wind speed forecasts.

The initial stage of the forecasting procedure involves a Time Series dataset, denoted as $\{y_t\} \in \mathbb{R}_{t=1-m}^T$, where each y_t signifies a datum at time t , encapsulated within a period from 1 to T , with the initiation at an earlier point indexed by m .

Subsequent to the collection of time series data, the process incorporates a GLDM Estimator algorithm. GLDM, postulated as an acronym for Generalized Least Deviation Method, is postulated to calibrate the data, deducing a set of pivotal factors $\{a_1, a_2, \dots, a_m\} \in \mathbb{R}$. These factors, intrinsic real numbers, epitomize the inferred parameters obtained from the time series data.

These extracted factors are then harnessed by a Predictor mechanism to prognosticate future values. This predictor is designed to generate outputs encapsulating the Forecasting Horizon (FH) and prospective forward-looking values, indicative of the temporal scope and expected data points for this horizon respectively.

The considered algorithm operates as follows (see Fig. 1).

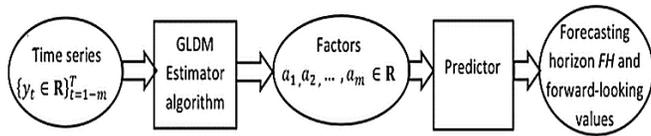


Figure 1 The approach taken for implementing the model

3.1 GLDM

Consider analysing a single time series from a chosen tile. The same logic applies to other tiles, with adjustments based on specific parameters. Linear autoregressive models often have limited forecast horizons. Building appropriate nonlinear models or neural networks might be impractical due to technical constraints. However, quasi-linear models can extend the forecasting horizon. Let's proceed with implementing the approach we have discussed. Ref. [1] to determine the coefficients $a_1, a_2, a_3 \dots, a_m \in \mathbb{R}$ of a m^{th} order quasilinear autoregressive model

$$y_t = \sum_{j=1}^{n(m)} a_j g_j(\{y_{t-k}\}_{k=1}^m) + \varepsilon_t, \quad t = 1, 2, \dots, T \quad (1)$$

by up-to-date information about of values of state variables $\{y_t \in \mathbb{R}\}_{t=1-m}^T$ at time instants t ; here $g_j: (\{y_{t-k}\}_{k=1}^m) \rightarrow \mathbb{R}$, $j = 1, 2, \dots, n(m)$ are given $n(m)$ functions, and $\{\varepsilon_t \in \mathbb{R}\}_{t=1}^T$ The methodology includes establishing the parameters for the recurrence equation, represented as Equation (1). This process is critical as it addresses potential unknown errors

that could arise during the parameterization phase. Ensuring precision in this step is essential for the accuracy of the forecasting models, as inaccuracies in parameter settings could significantly affect the outcomes. This paper discusses the techniques and considerations necessary to mitigate such errors, thereby enhancing the reliability of the model's predictions. The GLDM estimation algorithm [1] gets a time series $\{y_t \in \mathbb{R}\}_{t=-1-m}^T$ of length $T + m \geq (1 + 3m + m^2)$ as an input data and determines the factors $a_1, a_2, a_3 \dots, a_m \in \mathbb{R}$ by solving the optimization task

$$\sum_{t=1}^T \arctan \left| \sum_{j=1}^{n(m)} a_j g_j(\{y_{t-k}\}_{k=1}^m) - y_t \right| \rightarrow \min_{\{a_j\}_{j=1}^{n(m)} \subset \mathbb{R}} \quad (2)$$

The Cauchy distribution $F(\xi) = \frac{1}{\pi} \arctan(\xi) + \frac{1}{2}$.

It exhibits the highest entropy among distributions of random variables devoid of a mathematical expectation and variance. Hence, the $\arctan(*)$ function is adopted as the loss function.

In the context of wind speed forecasting, we assume that the absolute differences between observed and predicted values follow a standard Cauchy distribution. The standard Cauchy distribution, with its probability density function $f(x) = \frac{1}{\pi(1+x^2)}$, is characterized by its heavy tails, making it particularly well-suited for modeling data that exhibit significant variability and are susceptible to outliers. Unlike the normal distribution, the Cauchy distribution has infinite variance and an undefined mean, which allows it to accommodate extreme values without disproportionately influencing the overall model.

This assumption is crucial in wind speed forecasting, where the data often include extreme deviations that can impact the accuracy of predictive models. By assuming that the absolute differences $|y_i - \hat{y}_i|$ are distributed according to the standard Cauchy distribution, the Generalized Least Deviation Method (GLDM) gains enhanced robustness against outliers, thereby improving the model's reliability and stability. This approach ensures that the predictive model remains effective even in the presence of anomalies, which is essential for accurate wind speed forecasting in complex and variable environments.

Consideration is given to an m^{th} order model that is characterized by quadratic nonlinearity. The foundational set, denoted as $g_j(*)$, is designed to potentially encompass a range of subsequent functions. In this approach, a comprehensive analysis of the nonlinear dynamics within the dataset is facilitated. By accommodating quadratic elements within the model's structure, the ability to capture more complex patterns and interactions within the data is enhanced, thereby improving the accuracy and robustness of the predictions. Further elaboration on the specific functions included within this set and their roles in refining the model's predictive capabilities will be provided in this paper.

$$\begin{aligned} g_{(k)}(\{y_{t-k}\}_{k=1}^m) &= y_{t-k}, \\ g_{(kl)}(\{y_{t-k}\}_{k=1}^m) &= y_{t-k} \cdot y_{t-l}, \\ k &= 1, 2, \dots, m; \quad l = k, k + 1, \dots, m. \end{aligned} \quad (3)$$

Obviously, in this case $n(m) = 2m + C_m^2 = m(m + 3)/2$, and the numbering of $g_{(*)}$ functions can be arbitrary. Specifically, for $m=2$, the functions $g_{(*)}$ are as follows:

$$g_1 = y_1, \quad g_2 = y_2, \quad g_3 = y_1^2, \quad g_4 = y_2^2, \quad g_5 = y_1 \cdot y_2.$$

In this scenario, the model is structured as follows:

$$y_t = (a_1 y_{t-1} + a_2 y_{t-2}) + (a_3 y_{t-1}^2 + a_4 y_{t-2}^2 + a_5 y_{t-1} y_{t-2}). \quad (4)$$

Predictor forms the indexed by $t = 1, 2, \dots, T-1, T$ family of the m^{th} order difference equations

$$\overline{y[t]_\tau} = \sum_{j=1}^{n(m)} a_j^* g_j(\{\overline{y[t]_{\tau-k}}\}_{k=1}^m),$$

$$\tau = t, t+1, t+2, t+3, \dots, T-1, T, T+1, \quad (5)$$

For lattice functions $\overline{y[t]}$ with values $\overline{y[t]_\tau}$ which interpreted as constructed at time moment t the forecasts for y_τ . Let us use the solution of the Cauchy problem for its difference Eq. (5) under the initial conditions

$$\overline{y[t]_{t-1}} = y_{t-1}, \quad \overline{y[t]_{t-2}} = y_{t-2}, \dots, \quad \overline{y[t]_{t-m}} = y_{t-m}$$

$$t = 1, 2, \dots, T-1, T \quad (6)$$

To find the values of the function $\overline{y[t]}$.

So we have the set $\overline{Y}_\tau = \{\overline{y[t]_\tau}\}_{t=1}^T$ of possible prediction values of y_τ . Further we use this set to estimate the probabilistic characteristics of the y_τ value.

Task (2), involving GLDM estimation, constitutes a concave optimization problem. The introduction of supplementary variables streamlines it into the subsequent linear programming task:

$$\sum_{t=1}^T p_t z_t \rightarrow \min_{\substack{(a_1, a_2, \dots, a_{n(m)}) \in \mathbb{R}^m, \\ (z_1, z_2, \dots, z_T) \in \mathbb{R}^T}} \quad (7)$$

$$-z_t \leq \sum_{j=1}^{n(m)} [a_j g_j(\{y_{t-k}\}_{k=1}^m)] - y_t \leq z_t, \quad (8)$$

$$t = 1, 2, \dots, T,$$

$$z_t \geq 0, \quad t = 1, 2, \dots, T. \quad (9)$$

The task outlined in Eqs. (7)–(9) is identified as a canonical form, incorporating variables $n(m)+T$ and subject to $3n$ inequality constraints. These constraints are critical as they include conditions that guarantee the non-negativity of $z_j, j = 1, 2, \dots, T$. Correspondingly, the dual task associated with Eq. (7) is presented, providing an alternative perspective on the problem. This dual formulation is essential for exploring different solution strategies and for understanding the underlying structure and limitations of the model. The relationship between the primal and dual tasks enriches the analysis, offering insights into the feasibility and optimization of the model within the defined constraints.

$$\sum_{t=1}^T (u_t - v_t) y_t \rightarrow \max_{u, v \in \mathbb{R}^T} \quad (10)$$

$$\sum_{t=1}^T a_j g_j(\{y_{t-k}\}_{k=1}^m) (u_t - v_t) = 0, \quad j = 1, 2, \dots, n(m), \quad (11)$$

$$u_t + v_t = p_t, \quad u_t, v_t \geq 0, \quad t = 1, 2, \dots, T. \quad (12)$$

Let's introduce the following variables: $w_t = u_t - v_t$, $t = 1, 2, \dots, T$. Conditions (12) suggest that:

$$u_t = \frac{p_t + w_t}{2}, \quad v_t = \frac{p_t - w_t}{2}, \quad -p_t \leq w_t \leq p_t,$$

$$t = 1, 2, \dots, T.$$

Thus, the optimal solution for task (10)–(12) is equivalent to the optimal solution for the corresponding task. This equivalence signifies that solving one effectively resolves the other, underscoring a fundamental symmetry in their mathematical structure.

$$\sum_{t=1}^T w_t \cdot y_t \rightarrow \max_{w \in \mathbb{R}^T} \quad (13)$$

$$\sum_{t=1}^T g_j(\{y_{t-k}\}_{k=1}^m) \cdot w_t = 0, \quad j = 1, 2, \dots, n(m), \quad (14)$$

$$-p_t \leq w_t \leq p_t, \quad t = 1, 2, \dots, T. \quad (15)$$

Constraints (14) establish a linear variety \mathcal{L} that is $(T - n(m))$ -dimensional, characterized by a matrix of dimensions $(n(m) \times T)$.

$$S = \begin{bmatrix} g_1(\{y_{1-k}\}_{k=1}^m) & g_1(\{y_{2-k}\}_{k=1}^m) & \dots & g_1(\{y_{T+1-k}\}_{k=1}^m) \\ g_2(\{y_{1-k}\}_{k=1}^m) & g_2(\{y_{2-k}\}_{k=1}^m) & \dots & g_2(\{y_{T+1-k}\}_{k=1}^m) \\ \vdots & \vdots & \ddots & \vdots \\ g_{n(m)}(\{y_{1-k}\}_{k=1}^m) & g_{n(m)}(\{y_{2-k}\}_{k=1}^m) & \dots & g_{n(m)}(\{y_{1-k}\}_{k=1}^m) \end{bmatrix}$$

Constraints (15) establish a T -dimensional parallelepiped, denoted as \mathcal{T} . This delineation specifies the straightforward structure of the feasible set for task (13)–(15). The geometric configuration of the parallelepiped facilitates an easier visualization and understanding of the feasible solution space, thus enhancing the approachability and manageability of the task. By clearly defining the bounds and dimensions of this set, the analysis becomes more focused and structured, allowing for more efficient exploration of potential solutions within these defined parameters, which is the intersection of the $(T - n(m))$ -dimensional linear variety \mathcal{L} defined in (14) and the T -dimensional parallelepiped \mathcal{T} described in (15), facilitates its resolution through an algorithm. This algorithm utilizes the gradient projection of the objective function detailed in (13), represented by the vector $(\nabla = \{y_t\}_{t=1}^T)$, onto the permissible region $\mathcal{L} \cap \mathcal{T}$. The region is delineated by the constraints (14)–(15). The projection matrix for LL is given by: $S_{\mathcal{L}} = E - S^T \cdot (S \cdot S^T)^{-1} \cdot S$, and the gradient projection onto LL is calculated as $\nabla_{\mathcal{L}} = S_{\mathcal{L}} \cdot \nabla$. Additionally, if the external normal of any face of the parallelepiped forms an acute angle with the gradient projection $\nabla_{\mathcal{L}}$, then movement along this face will be zero. This ensures that the gradient projection method effectively navigates the feasible set, optimizing the objective function within the defined constraints.

GLDM estimates demonstrate robustness against correlations among values in $\{y_t \in \mathbb{R}\}_{t=1-m}^T$, and, with proper adjustments, provide superior results for error probability distributions that exhibit tails heavier than those of a normal distribution (refer to [2]). This supports the practicality of employing an algorithm based on the Weighted Least Deviation Method (WLDM) for solving the identification problem. According to the findings published in [3], the task of computing GLDM estimates can be converted into an iterative process utilizing WLDM estimates [1].

The operational procedure of the algorithm is depicted in Fig. 2. The initial data required includes:

- $S = \{S_t \in \mathbb{R}^N\}_{t \in T}$, matrix representing the linear variety
- $\nabla_{\mathcal{L}}$, the gradient projection of the objective function on \mathcal{L} ;
- Weight factors $\{p_t \in \mathbb{R}^+\}_{t=1}^T$;
- The values of the specified state variables $\{y_t \in \mathbb{R}^+\}_{t=1-m}^T$.

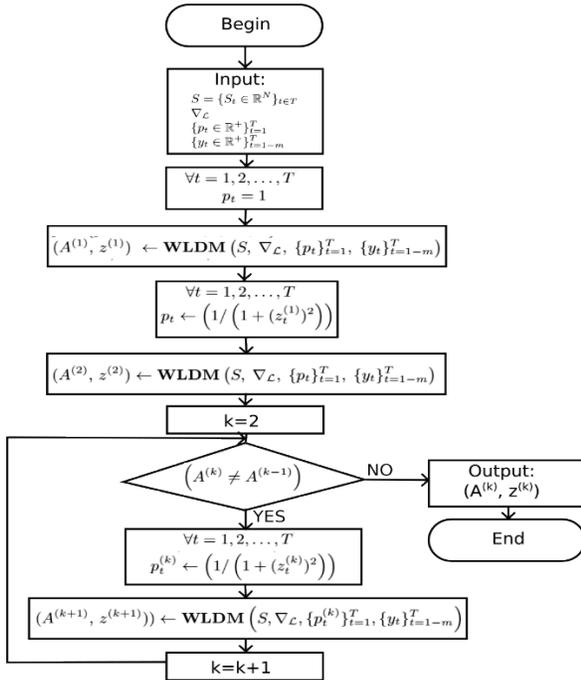


Figure 2 The strategy employed by the GLDM estimation algorithm

The algorithm operates through an iterative process aimed at obtaining the optimal GLDM solution $A \in \mathbb{R}^{n(m)}$ and the vector of residuals $z \in \mathbb{R}^T$. This process terminates when the solution for the current iteration $(A^{(k)})$ equals the solution from the previous iteration $(A^{(k-1)})$. To derive A and z , the Weighted Least Deviation Method (WLDM) estimation algorithm is employed [4]. This algorithm uses the same input data as the GLDM algorithm and calculates the factors necessary for convergence.

$$a_1, a_2, a_3, \dots, a_{n(m)} \in \mathbb{R}$$

By solving the optimization task

$$\sum_{t=1}^T p_t \cdot \left| \sum_{j=1}^{n(m)} a_j g_j (\{y_{t-k}\}_{k=1}^m) - y_t \right| \rightarrow \min_{\{a_j\}_{j=1}^{n(m)} \in \mathbb{R}^{n(m)}} \quad (16)$$

The scheme of this algorithm is shown in Fig. 3.

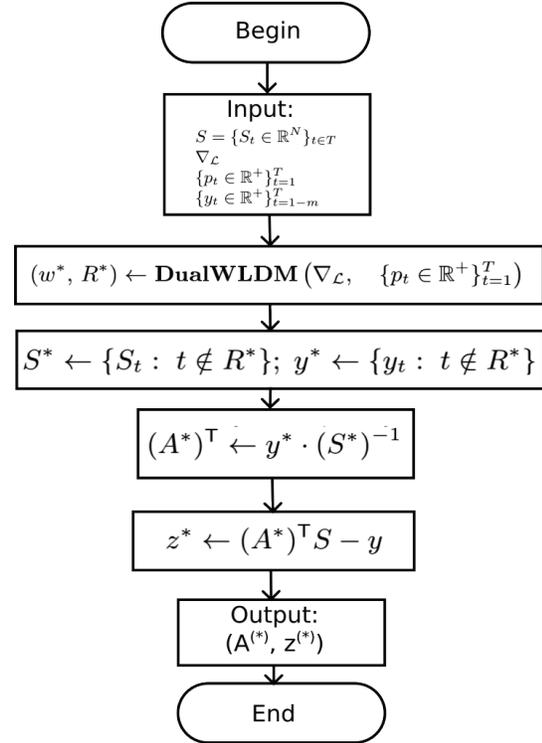


Figure 3 The framework of the WLDM estimation algorithm

The computational complexity of this algorithm is capped at $O(T^2)$ because of the straightforward structure of the permissible set, which consists of the intersection between a T -dimensional cuboid, as described in (15), and a $(T - n(m))$ -dimensional linear variety, as mentioned in (14).

The algorithm designed for the dual task outlined in equations (13)–(15) commences the quest for the optimal solution at the zero point, methodically advancing along a specified vector $\nabla_{\mathcal{L}}$. This structured progression is designed to methodically traverse the solution space, adhering closely to the defined constraints and systematically evaluating each potential solution along the way. If the current point lands on the face of the cuboid \mathcal{T} , then the corresponding coordinate in the movement direction is set to zero, effectively halting movement in that direction. This method ensures that the search remains within the confines of the defined feasible region.

If (w^*, R^*) is the outcome of implementing the gradient projection algorithm [1], then w^* represents the optimal solution for the task defined in equations (13)–(15). Consequently, the optimal solution for the task outlined in equations (10)–(12) corresponds to R^* . This alignment ensures that both solutions are effectively optimized through the same procedural framework.

$$u_t^* = \frac{p_t + w_t^*}{2}, \quad v_t^* = \frac{p_t - w_t^*}{2}, \quad t = 1, 2, \dots, T.$$

It follows from the complementarity condition for a pair of mutually dual tasks (7)–(9) and (10)–(12) that:

$$y_t = \sum_{j=1}^{n(m)} [a_j g_j(\{y_{t-k}\}_{k=1}^m)] \quad \forall t \notin R^*, \quad (17)$$

$$y_t = \sum_{j=1}^{n(m)} [a_j g_j(\{y_{t-k}\}_{k=1}^m)] + z_t^*, \quad \forall t \in R^*: w_t^* = p_t, \quad (18)$$

$$y_t = \sum_{j=1}^{n(m)} [a_j g_j(\{y_{t-k}\}_{k=1}^m)] - z_t^*, \quad \forall t \in R^*: w_t^* = -p_t. \quad (19)$$

Indeed, the solution $(\{a_j^*\}_{j=1}^{n(m)}, z^*)$ of the system of linear algebraic equations (17)–(19) functions as the dual optimal solution for the task (13)–(15) and simultaneously as the optimal solution for the task (16). This dual role confirms the validity of the theorem cited in reference [5].

Theorem 3.1 *Let*

- w^* be the optimal solution of the task (13)–(15),
- $(\{a_j^*\}_{j=1}^{n(m)}, z^*)$ be solution of a system of linear algebraic equations (17)–(19).

Therefore, $\{a_j^*\}_{j=1}^{n(m)}$ constitutes the optimal solution for the task (16).

The primary challenge associated with the utilization of the WLDM estimator lies in the lack of universal formal guidelines for selecting weight coefficients. As a result, this approach necessitates further investigation and research to establish effective methodologies.

Theorem 3.2 [4]: The sequence $\{(A^{(k)}, z^{(k)})\}_{k=1}^{\infty}$ generated by the GLDM estimator Algorithm, converges to the global minimum (a^*, z^*) of the task (2).

The computational complexity of the GLDM estimation algorithm appears to be directly correlated with that of the algorithm employed for solving the primal and/or dual WLDM tasks. Various computational experiments indicate that the average iteration count necessary for the GLDM estimation algorithm aligns with the number of coefficients in the identified equation. Should this hypothesis prove valid, the computational complexity associated with solving practical problems would not exceed these parameters.

$$O((n(m))^3 T + n(m) \cdot T^2).$$

It should be noted that the search for and discovery of high-order autoregression equations are subject to specific conditions. Among these conditions, one noteworthy factor is the algorithm's high sensitivity to rounding errors. To mitigate the risk of calculation errors, it is imperative to execute basic arithmetic operations meticulously within the realm of rational numbers, supplemented by parallelization [6–11].

4 ERROR METRICS FOR WIND SPEED FORECASTING

The process of forecasting wind speed through time series analysis plays a vital role in optimizing wind energy production. This task, which centers on predicting future wind speeds from past and present observations, requires a high degree of accuracy to ensure effective operational planning for wind farms. To ascertain the performance of our wind speed forecasting models, specifically those devised utilizing first and second order Generalized Least Deviation Method (GLDM) algorithms, we employ a set of critical error

metrics. These metrics are essential for measuring the forecasts' accuracy, reliability, and the potential bias within them. In this section, we delve into key error metrics: Root Mean Square Error (*RMSE*), R-Squared (R^2), Mean Absolute Percentage Error (*MAPE*), Mean Absolute Error (*MAE*), and Mean Squared Error (*MSE*). Each of these metrics sheds light on different facets of our models' performance, offering a well-rounded evaluation. Through the analysis of *RMSE*, R^2 , *MAPE*, *MAE*, and *MSE*, we identify the strengths and limitations of our forecasting methodologies, pinpointing opportunities for enhancement. Such meticulous scrutiny is pivotal for refining our wind speed forecasting methods, thereby pushing forward the capabilities of wind energy management practices.

The selection of diverse error metrics, including *RMSE*, *MAE*, *MSE*, *MAPE*, and R^2 , is essential for a comprehensive evaluation of the Generalized Least Deviation Method (GLDM) in wind speed forecasting. Each metric captures different aspects of model performance: *RMSE* emphasizes large errors, *MAE* provides a balanced view of average error, *MSE* highlights overall error magnitude, *MAPE* offers insights into relative performance, and R^2 assesses the explanatory power of the model. Together, these metrics enable a thorough assessment, guiding the refinement of our forecasting methods. This iterative improvement enhances predictive precision, thereby advancing the capabilities of wind energy management practices.

4.1 Root Mean Square Error (*RMSE*) for Wind Speed Forecasting

RMSE is a crucial metric for assessing the accuracy of wind speed forecasts, quantifying the standard deviation of the residuals or prediction errors. It effectively captures the magnitude of error between the forecasted and actual wind speeds, offering insights into the overall precision of the forecasting model. Specifically for wind speed forecasting, *RMSE* evaluates the average magnitude of errors across all predictions, providing a clear measure of model performance in predicting wind speeds. The formula for calculating *RMSE* in the context of wind speed forecasting is presented as:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (20)$$

Here, y_i represents the observed wind speeds, \hat{y}_i denotes the predicted wind speeds by the model, and n is the total number of observations. A lower *RMSE* value indicates a model with higher accuracy in forecasting wind speeds, underscoring the model's efficacy in energy production planning and operational optimization for wind farms.

4.2 R-Squared (R^2) for Wind Speed Forecasting

The R^2 metric offers insights into the accuracy of wind speed forecasts by indicating the goodness of fit between the predicted and actual wind speed values. It quantifies the proportion of variance in observed wind speed data that is predictable from the forecasting model. Specifically, R^2

assesses the extent to which variations in actual wind speed can be explained by the model's predictions, thereby serving as a measure of the model's explanatory power. The formula for calculating R^2 in the context of wind speed forecasting is expressed as:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (21)$$

where y_i denotes the observed wind speed values, \hat{y}_i represents the predicted wind speed values based on the model, \bar{y} is the mean of observed wind speeds, and n is the total number of observations. A higher R^2 value indicates a model that more accurately reflects the observed wind speed variations, making it a crucial metric for evaluating the performance of wind speed forecasting models.

4.3 Mean Absolute Percentage Error (MAPE) for Wind Speed Forecasting

MAPE is a critical metric in wind speed forecasting as it quantifies the average magnitude of prediction errors as a percentage of actual wind speeds. This measure provides a clear indicator of the model's accuracy in percentage terms, making it particularly useful for understanding the relative size of forecast errors in predicting wind speeds. The *MAPE* calculation for wind speed forecasts is given by:

$$MAPE = \frac{100\%}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad (22)$$

where y_i represents the observed wind speed values, \hat{y}_i denotes the predicted wind speed values from the model, and n is the number of observations. A lower *MAPE* value indicates a higher accuracy of the wind speed forecasting model, highlighting its effectiveness in closely matching the actual wind speed measurements.

4.4 Mean Absolute Error (MAE) for Wind Speed Forecasting

MAE plays a pivotal role in assessing the accuracy of wind speed forecasts by measuring the average magnitude of errors across predictions without accounting for their direction. This metric is particularly useful in wind energy studies as it provides a straightforward indication of the model's performance in predicting wind speeds. By calculating the average of the absolute differences between the predicted and the actual wind speed observations, *MAE* offers a clear, interpretable measure of forecast accuracy. The *MAE* for wind speed forecasting is defined as:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (23)$$

where y_i denotes the actual observed wind speed values, \hat{y}_i represents the predicted wind speed values from the forecasting model, and n is the total number of wind speed observations in the dataset. A lower *MAE* value indicates a higher precision of the wind speed forecasting model,

underscoring its reliability in predicting wind speed with minimal error.

4.5 Mean Squared Error (MSE) for Wind Speed Forecasting

MSE is a crucial metric in evaluating the accuracy of wind speed forecasting models, as it quantifies the average squared difference between the estimated values and what is actually observed. This squared difference penalizes larger errors more severely than smaller ones, making *MSE* particularly informative for understanding the performance of wind speed predictions. The formula for calculating *MSE* in the realm of wind speed forecasting is expressed as:

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (24)$$

where y_i represents the actual observed wind speeds, \hat{y}_i denotes the wind speed values predicted by the model, and n is the total count of observations. A lower *MSE* value indicates a model that more accurately forecasts wind speeds, highlighting its effectiveness in closely mirroring actual wind speed variations. This measure provides insights into the model's precision, with the aim of minimizing forecasting errors and enhancing the reliability of wind energy production forecasts.

In this thorough investigation, we have explored various error metrics pivotal for evaluating the precision and efficacy of univariate time series models in forecasting wind speed. Our extensive examination encompasses *RMSE*, R^2 , *MAPE*, *MAE*, *MSE*, each providing distinct perspectives on the forecasting model's accuracy. This methodical assessment constructs a solid foundation for evaluating wind speed predictions, allowing for an in-depth analysis of the model's strengths and areas needing enhancement. As we continually refine our forecasting approaches, the insights derived from these metrics prove indispensable. The ultimate objective is to elevate the accuracy of wind speed forecasts, contributing significantly to more efficient wind farm management and energy production strategies. The pursuit of high-fidelity predictions is a progressive endeavor, with each cycle of evaluation edging us closer to delivering precise and actionable forecasts for wind energy optimization.

5 RESULTS

In the development of predictive models for wind speed, the GLDM approach employs a set of coefficients to encapsulate the system dynamics at various orders of complexity. Tab. 2 delineates the coefficients for the first and second orders of the model, highlighting the foundational parameters that govern model behavior. Specifically, the first order model utilizes a straightforward formulation with coefficients a_1 and a_2 , aiming for a balance between simplicity and predictive capability. Conversely, the second order model expands this basis with a total of five coefficients (a_1 through a_5), thereby enhancing the model's ability to capture more intricate patterns in wind speed data. This gradation in model complexity is instrumental in

refining our understanding and forecasting of wind speed variations.

Table 2 Coefficients from First to Second Order for Wind Speed

GLDM Order	Coefficients
First	$a_1 = 1.0092, a_2 = -0.0011$
Second	$a_1 = 0.9300, a_2 = 0.0764, a_3 = 0.0248, a_4 = 0.0241, a_5 = -0.0499$

Tab. 3 showcases an error matrix comparing the performance metrics of the first and second-order Generalized Least Deviation Method (GLDM) models for wind speed prediction. Metrics include Root Mean Square Error (*RMSE*), R^2 , Mean Absolute Percentage Error (*MAPE*), Mean Absolute Error (*MAE*), and Mean Squared Error (*MSE*).

Table 3 Error Matrix for Wind Speed Prediction using GLDM model

GLDM Order	<i>RMSE</i>	R^2	<i>MAPE</i>	<i>MAE</i>	<i>MSE</i>
First	0.75	0.97	9.98	0.52	0.56
Second	0.74	0.97	9.50	0.52	0.55

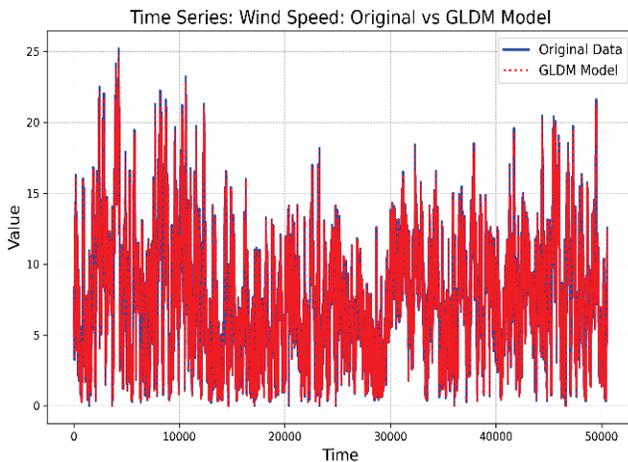


Figure 4 GLDM from the first order for Wind speed

The second-order model exhibits a marginal improvement over the first-order model across several metrics. Specifically, it achieves a slightly lower *RMSE* (0.74 compared to 0.75) and *MSE* (0.55 compared to 0.56), indicating a more accurate fit to the observed wind speed data. Similarly, the *MAPE* is reduced from 9.98 in the first order to 9.50 in the second order, further evidencing enhanced predictive accuracy. The consistent R^2 value of 0.97 for both models suggests a strong explanatory power, yet the improvements in other metrics for the second-order model underscore its superior capability in modeling wind speed with greater precision and reliability.

Fig. 6 illustrates the performance comparison between first and second-order Generalized Least Deviation Method (GLDM) models for predicting wind speed. Fig. 4 shows the first-order GLDM model, while Fig. 5 depicts the second-order GLDM model. The graphical representation indicates that the second-order GLDM model captures the wind speed dynamics with greater accuracy, as evidenced by the closer alignment of its predictions (red dotted line) with the original data (blue solid line). The enhanced precision of the second-

order model makes it a preferable choice for wind speed analysis and forecasting.

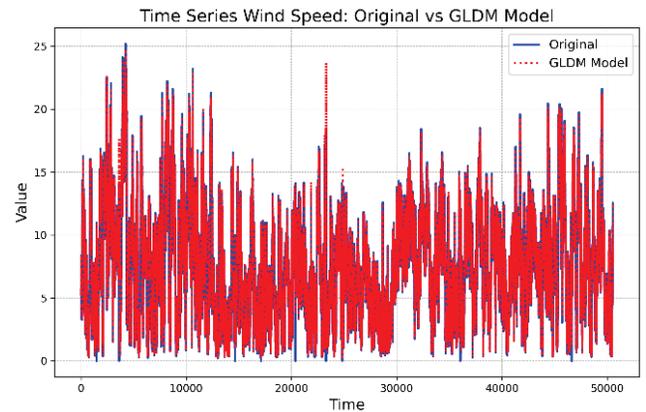


Figure 5 GLDM from the second order for Wind speed

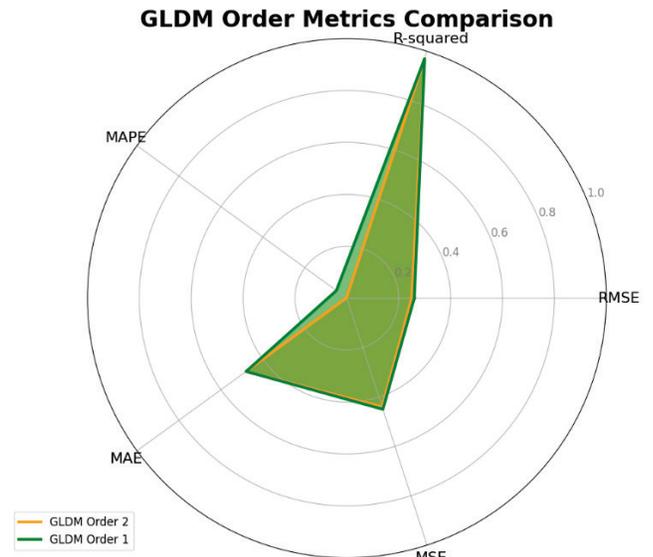


Figure 6 GLDM models from the first and second order for Wind Speed

Fig. 6 presents a radar chart to compare the performance of the first and second-order Generalized Least Deviation Method (GLDM) models for wind speed prediction, detailed in Tab. 3. This chart encapsulates several error metrics: Root Mean Square Error (*RMSE*), R^2 , Mean Absolute Percentage Error (*MAPE*), Mean Absolute Error (*MAE*), and Mean Squared Error (*MSE*). In this visual assessment, the second-order GLDM model is represented by the orange area, which shows a marked improvement over the first-order model, indicated in green, in all measured metrics. Notably, the second-order model achieves a lower *RMSE* and *MSE*—0.74 and 0.55 respectively—compared to the first-order model’s 0.75 and 0.56. Moreover, there is a decrease in *MAPE* from 9.98 for the first order to 9.50 for the second order. Despite both models having an R^2 value of 0.97, the reduction in the error metrics for the second-order model demonstrates its superior predictive accuracy and better fit with the observed wind speed data.

The application of the second-order Generalized Least Deviation Method (GLDM) in our wind speed forecasting

methodology represents a significant innovation. This model is crafted to predict future wind speeds by analyzing past wind speed data exclusively, eliminating the dependency on additional meteorological variables. The strength of the model lies in its ability to identify and decode the patterns present within the wind speed time series data.

Table 4 Performance of Wind Speed Models

Model	RMSE	MSE	MAE	R^2	MAPE
MLP	5.070	25.70	4.05	0.04969	102.61%
SVM	0.80	0.64	0.638	0.96	12.13%
Autoarima	0.7453	0.5555	0.5209	0.9689	10.00%
Exponential Smoothing	0.7493	0.5614	0.5226	0.9686	9.99%
BATS Model	2.4	5.76	1.915	0.30	19.15%
TBATS Model	4.6	21.16	3.671	0.45	24.36%
Prophet Model	3.8472	14.8009	3.0759	0.1717	78.03%
Hybrid autoarima-ES	0.90	0.81	0.718	0.19	71.80%
Hybrid autoarima-Polynomial	0.88	0.7744	0.702	0.2254	70.20%
GLDM Second Order	0.74	0.55	0.52	0.97	9.50%

Tab. 4 provides a comprehensive comparison of the performance metrics for various models employed in wind speed prediction. The table presents key indicators such as Root Mean Square Error (*RMSE*), Mean Squared Error (*MSE*), Mean Absolute Error (*MAE*), R^2 , and Mean Absolute Percentage Error (*MAPE*) for each model. Notably, the GLDM Second Order model demonstrates superior performance, achieving the lowest *RMSE* of 0.74, *MSE* of 0.55, and *MAE* of 0.52. Additionally, it attains the highest R^2 value of 0.97, indicating that it explains 97% of the variance in the wind speed data. The model also exhibits the lowest *MAPE* of 9.50%, reflecting its exceptional predictive accuracy.

The AutoARIMA and Exponential Smoothing models also perform robustly, with *RMSE* values of 0.7453 and 0.7493, respectively, and R-squared values closely aligned with the GLDM model at 0.9689 and 0.9686. Their *MAPE* values of 10.00% and 9.99%, respectively, further underscore their strong predictive capabilities. The SVM model, while exhibiting an *RMSE* of 0.80 and an R^2 value of 0.96, shows a slightly higher *MAPE* of 12.13%, indicating marginally reduced accuracy relative to the GLDM and AutoARIMA models.

In contrast, models such as MLP and TBATS show significantly less accurate predictions, as evidenced by their higher *RMSE* and *MAPE* values. The MLP model reports an *RMSE* of 5.070 and an exceedingly high *MAPE* of 102.61%, while the TBATS model has an *RMSE* of 4.6 and a *MAPE* of 24.36%. Similarly, the Prophet model, Hybrid AutoARIMA-ES, and Hybrid AutoARIMA-Polynomial models present higher *RMSE* and *MAPE* values, reflecting their lower predictive performance relative to the top-performing models. In summary, the GLDM Second Order model emerges as the most effective method for wind speed prediction, delivering the highest levels of accuracy and reliability among the evaluated models.

The mathematical foundation of our second-order GLDM, tailored specifically for wind speed time series forecasting, is articulated by the following equation:

$$\hat{y}_t = a_1 y_{t-1} + a_2 y_{t-2} + a_3 y_{t-1}^2 + a_4 y_{t-1} y_{t-2} + a_5 y_{t-2}^2, \quad (25)$$

where \hat{y}_t represents the forecasted wind speed at time t , and y_{t-1} and y_{t-2} denote the wind speeds at the one and two preceding time steps, respectively. The coefficients a_1 through a_5 are optimized to best align with the historical data, reflecting the influence of the past wind speeds' linear and non-linear relationships.

This model's prowess in predicting future wind speeds stems from a composite approach that captures both the direct effects of the recent past wind speeds as well as their squared values, embodying non-linear interactions. The incorporation of squared and product terms allows the model to account for more complex dynamical patterns that may influence future wind speed values, thereby improving forecast accuracy and reliability.

Given the crucial role of wind speed prediction in various applications such as renewable energy management and weather forecasting, our study illustrates the importance of advanced statistical models in environmental science. The second-order GLDM, with its firm mathematical basis and empirical support, offers a substantial step forward in understanding and predicting the nuanced dynamics of wind speeds. Future research should focus on enhancing these models, extending their meteorological applications, and integrating them into holistic weather forecasting systems. The intersection of mathematical insights and practical utility remains a promising domain for future advancements, likely to yield sophisticated tools for environmental predictions and analyses.

6 DISCUSSION

This study's investigation into the second-order Generalized Least Deviation Method (GLDM) for wind speed forecasting signifies a notable progression in meteorological data analysis. Implementing the second-order GLDM model, based on the mathematical formulation

$$\hat{y}_t = a_1 y_{t-1} + a_2 y_{t-2} + a_3 y_{t-1}^2 + a_4 y_{t-1} y_{t-2} + a_5 y_{t-2}^2,$$

our research not only narrows the divide between theoretical models and practical forecasting applications but also reveals the intricate dynamics influencing wind speed variability.

The coefficients a_1 to a_5 , derived through the GLDM framework, quantitatively illustrate the complex interplay between the linear and non-linear aspects of historical wind speed data. This refined comprehension is crucial for the enhanced prediction of future wind speeds, an important factor in numerous sectors such as renewable energy, weather forecasting, and aviation safety.

Our results indicate that incorporating both the direct and squared impacts of past wind speeds, as well as their interactions, into the forecasting model offers unprecedented detail and accuracy. This is especially vital in the current climate, where unpredictable weather patterns necessitate reliable forecasting methods.

In conclusion, our findings emphasize the pivotal role of advanced mathematical models in environmental sciences, particularly meteorology. The second-order GLDM model, with its empirical validity and mathematical robustness, marks a considerable leap in our predictive capabilities for wind speed. Future studies should strive to enhance these models, broaden their meteorological applicability, and incorporate them into holistic weather forecasting systems. The confluence of mathematical theory and practical implementation continues to be a promising avenue for innovation, likely to produce more advanced tools for environmental forecasting and analysis.

The adapted second-order GLDM equation for forecasting wind speed, utilizing all relevant coefficients a_1, a_2, a_3, a_4 , and a_5 , is given by:

$$\hat{y}_t = 0.9300 \cdot y_{t-1} + 0.0764 \cdot y_{t-2} + 0.0248 \cdot y_{t-1}^2 + 0.0241 \cdot y_{t-1} \cdot y_{t-2} - 0.0499 \cdot y_{t-2}^2 \quad (26)$$

This formulation effectively replaces the G functions with their corresponding operations based on the historical wind speed data at time steps $t - 1$ and $t - 2$. The equation succinctly captures both linear and non-linear influences of past wind speeds on the forecasted wind speed \hat{y}_t , providing a robust model for wind speed prediction.

The equation represents the second-order Generalized Least Deviation Method (GLDM) for wind speed forecasting. Let's break down each coefficient and its significance:

$$\hat{y}_t = 0.9300 \cdot y_{t-1} + 0.0764 \cdot y_{t-2} + 0.0248 \cdot y_{t-1}^2 + 0.0241 \cdot y_{t-1} \cdot y_{t-2} - 0.0499 \cdot y_{t-2}^2 \quad (27)$$

- $0.9300 \cdot y_{t-1}$: This term represents the linear influence of the wind speed at time $t - 1$ on the forecasted wind speed \hat{y}_t . It indicates how much the previous wind speed directly affects the current forecast.
- $0.0764 \cdot y_{t-2}$: Similar to the first term, this coefficient denotes the linear impact of the wind speed at time $t - 2$ on the forecasted value. It captures the delayed effect of the wind speed from two time steps ago on the current forecast.
- $0.0248 \cdot y_{t-1}^2$: This term accounts for the non-linear influence of the wind speed at time $t - 1$ on the forecast. It represents the squared value of the wind speed at the previous time step, indicating the presence of non-linear relationships in the data.
- $0.0241 \cdot y_{t-1} \cdot y_{t-2}$: Here, we have the cross-product term between the wind speeds at $t - 1$ and $t - 2$. This captures the interaction between the wind speeds at adjacent time steps and provides insight into how their combined effect influences the forecast.
- $-0.0499 \cdot y_{t-2}^2$: Similar to the third term, this coefficient accounts for the non-linear influence of the wind speed at time $t - 2$, but squared. It captures any non-linear patterns specific to the wind speed at the time step two periods ago.

Each coefficient reflects a different aspect of the historical wind speed data and its impact on the forecasted value. By combining linear and non-linear terms, the equation captures the complex dynamics of wind behavior, leading to more accurate forecasts.

7 CONCLUSION

In this investigation, we have meticulously explored the application of the second-order Generalized Least Deviation Method (GLDM) to the field of wind speed forecasting. Through a comprehensive analysis, we demonstrated that the model, characterized by its incorporation of both linear and non-linear historical data through the mathematical formula

$$Y(t) = \sum_{i=1}^5 a_i G_i(Y_{t-1}, Y_{t-2}),$$

profoundly enhances the accuracy of wind speed predictions. The G functions, designed to capture various dynamic aspects of wind behavior, including direct impacts, squared influences, and interactions between different time steps, represent a significant leap in our methodological toolkit for understanding and predicting meteorological phenomena.

This study's findings not only underscore the importance of sophisticated mathematical models in meteorology but also highlight the intricate relationship between wind speed's past values and its future trajectory. By leveraging the nuanced capabilities of the second-order GLDM, we have laid a solid foundation for more reliable and precise wind speed forecasting models. Such advancements are crucial for various practical applications, ranging from renewable energy management to disaster preparedness and climate research.

This investigation into the efficacy of the second-order Generalized Least Deviation Method (GLDM) for forecasting wind speed has yielded insightful and promising results. Through the application of a sophisticated mathematical model that integrates both linear and non-linear dynamics of historical wind speed data, we have demonstrated a significant improvement in forecast accuracy over traditional first-order models. The mathematical formulation of our model,

$$Y(t) = \sum_{i=1}^5 a_i G_i(Y_{t-1}, Y_{t-2}),$$

effectively captures the complexity of wind speed variations by incorporating past data points and their interactions. This study not only validates the theoretical merits of the GLDM approach but also underscores its practical utility in meteorological forecasting and related applications.

Our findings highlight the critical role of advanced statistical models in understanding and predicting environmental phenomena. By embracing the non-linear nature of weather patterns and the intrinsic value of historical data, the second-order GLDM model offers a robust tool for meteorologists, environmental scientists, and industries reliant on accurate weather forecasting.

8 FURTHER RESEARCH

The promising results of this study open several avenues for further research. Future investigations could explore the following areas:

- **Model Extension:** Extending the GLDM model to incorporate third-order or higher interactions could uncover deeper insights into the wind speed dynamics and potentially offer even greater forecasting precision.
- **Variable Integration:** Including additional meteorological variables such as temperature, humidity, and atmospheric pressure might enhance the model's comprehensive understanding of weather patterns, leading to improved predictive capabilities.
- **Cross-Disciplinary Applications:** Applying the GLDM framework to other fields, such as hydrology or oceanography, could test the model's versatility and contribute to a broader spectrum of environmental science.
- **Real-time Forecasting Systems:** Developing real-time forecasting systems based on the GLDM model could have significant implications for disaster preparedness, renewable energy management, and aviation safety, among other sectors.
- **Machine Learning Integration:** Investigating the integration of GLDM with machine learning algorithms could yield advanced models that learn from and adapt to changing weather patterns dynamically.

In conclusion, the second-order GLDM model represents a significant step forward in the predictive modeling of wind speed. Its success paves the way for further research aimed at enhancing our understanding of meteorological phenomena and our ability to forecast them accurately. As we continue to refine and expand upon this model, the potential for impactful discoveries and practical applications in weather prediction and beyond remains vast and compelling.

9 REFERENCES

- [1] Panyukov, A., Makarovskikh, T., & Abotaleb, M. (2022). Forecasting with using quasilinear recurrence equation. In *International Conference on Optimization and Applications* (pp. 183–195). Springer. https://doi.org/10.1007/978-3-031-22990-9_13
- [2] Makarovskikh, T., Panyukov, A., & Abotaleb, M. (2023). Using general least deviations method for forecasting of crops yields. In *International Conference on Mathematical Optimization Theory and Operations Research* (pp. 376–390). Springer. https://doi.org/10.1007/978-3-031-43257-6_28
- [3] Panyukov, A. V., & Mezaal, Y. A. (2018). Stable estimation of autoregressive model parameters with exogenous variables on the basis of the generalized least absolute deviation method. *IFAC-PapersOnLine*, 51(11), 1666–1669. <https://doi.org/10.1016/j.ifacol.2018.08.217>
- [4] Panyukov, A. V., & Mezaal, Y. A. (2020). Improving of the identification algorithm for a quasilinear recurrence equation. In *Advances in Optimization and Applications: 11th International Conference, OPTIMA 2020, Moscow, Russia, September 28–October 2, 2020, Revised Selected Papers 11* (pp. 15–26). Springer. https://doi.org/10.1007/978-3-030-65739-0_2
- [5] Panyukov, A., Makarovskikh, T., & Abotaleb, M. (2022). Forecasting with using quasilinear recurrence equation. In *International Conference on Optimization and Applications* (pp. 183–195). Springer. https://doi.org/10.1007/978-3-031-22990-9_13
- [6] Gwet, K. L. (2016). Testing the difference of correlated agreement coefficients for statistical significance. *Educational and Psychological Measurement*, 76(4), 609–637. <https://doi.org/10.1177/0013164415596420>
- [7] LaHuis, D. M., & Ferguson, M. W. (2009). The accuracy of significance tests for slope variance components in multilevel random coefficient models. *Organizational Research Methods*, 12(3), 418–435. <https://doi.org/10.1177/1094428107308984>
- [8] Anderson, M. J., & Legendre, P. (1999). An empirical comparison of permutation methods for tests of partial regression coefficients in a linear model. *Journal of Statistical Computation and Simulation*, 62(3), 271–303. <https://doi.org/10.1080/00949659908811936>
- [9] Ståhl, L., Wold, S., et al. (1989). Analysis of variance (ANOVA). *Chemometrics and Intelligent Laboratory Systems*, 6(4), 259–272. [https://doi.org/10.1016/0169-7439\(89\)80095-4](https://doi.org/10.1016/0169-7439(89)80095-4)
- [10] Rouder, J. N., Engelhardt, C. R., McCabe, S., et al. (2016). Model comparison in ANOVA. *Psychonomic Bulletin & Review*, 23, 1779–1786. <https://doi.org/10.3758/s13423-016-1026-5>
- [11] Graham, M. H., & Edwards, M. S. (2001). Statistical significance versus fit: Estimating the importance of individual factors in ecological analysis of variance. *Oikos*, 93(3), 505–513. <https://doi.org/10.1034/j.1600-0706.2001.930317.x>
- [12] Shen, W., Chen, X., Qiu, J., et al. (2020). A comprehensive review of variable renewable energy leveled cost of electricity. *Renewable and Sustainable Energy Reviews*, 133, 110301. <https://doi.org/10.1016/j.rser.2020.110301>
- [13] Widén, J. (2011). Correlations between large-scale solar and wind power in a future scenario for Sweden. *IEEE Transactions on Sustainable Energy*, 2(2), 177–184. <https://doi.org/10.1109/TSTE.2010.2101620>
- [14] Ren, G., Liu, J., Wan, J., et al. (2017). Overview of wind power intermittency: Impacts, measurements, and mitigation solutions. *Applied Energy*, 204, 47–65. <https://doi.org/10.1016/j.apenergy.2017.06.098>
- [15] Ruiz, S. A. G., Barriga, J. E. C., Martínez, J. A. M., et al. (2021). Wind power assessment in the Caribbean region of Colombia, using ten-minute wind observations and ERA5 data. *Renewable Energy*, 172, 158–176. <https://doi.org/10.1016/j.renene.2021.03.033>
- [16] Augustyn, A., & Kamiński, J. (2018). A review of methods applied for wind power generation forecasting. *Polityka Energetyczna*, 21(2), 139–150. <https://doi.org/10.33223/epj/96214>
- [17] Soman, S. S., Zareipour, H., Malik, O., & Mandal, P. (2010). A review of wind power and wind speed forecasting methods with different time horizons. In *IEEE North American Power Symposium 2010* (pp. 1–8). <https://doi.org/10.1109/NAPS.2010.5619586>
- [18] Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time series analysis: Forecasting and control*. John Wiley & Sons.
- [19] do Nascimento Camelo, H., Sérgio Lucio, P., Verçosa Leal Junior, J. B., et al. (2018). Innovative hybrid modeling of wind speed prediction involving time-series models and artificial neural networks. *Atmosphere*, 9(2), 77. <https://doi.org/10.3390/atmos9020077>
- [20] Makarovskikh, T., Panyukov, A., & Abotaleb, M. (2023). Using general least deviations method for forecasting of crops yields. In *International Conference on Mathematical Optimization Theory and Operations Research* (pp. 376–390). Springer. https://doi.org/10.1007/978-3-031-43257-6_28

- [21] Ochoa, G. V., Alvarez, J. N., & Vanegas Chamorro, M. N. (2019). Data set on wind speed, wind direction and wind probability distributions in Puerto Bolivar – Colombia. *Data in Brief*, 27, 104753. <https://doi.org/10.1016/j.dib.2019.104753>
- [22] Bebi, E., Stermasi, A., Alcani, M., & Cenameri, M. (2023). Assessment of wind potential: The case of Puka region in Albania. *International Journal of Innovative Technology and Interdisciplinary Sciences*, 6(1), 1112–1120. <https://doi.org/10.35741/issn.0258-2724.57.6.97>
- [23] Baballëku, M., Verzivolli, A., Luka, R., & Zgjanolli, R. (2023). Fundamental basic wind speed in Albania: An adoption in accordance with Eurocodes. *Journal of Transactions in Systems Engineering*, 1(2), 56–72. <https://doi.org/10.15157/JTSE.2023.1.2.56-72>
- [24] Qosja, S., Rolle, R., & Gebremedhin, A. (2022). Solving the bottleneck issue of energy supply: Case study of a wind power plant. *International Journal of Innovative Technology and Interdisciplinary Sciences*, 5(2), 874–891. <https://doi.org/10.15157/IJITIS.2022.5.2.874-891>
- [25] Dhoska, K., Bebi, E., Markja, I., et al. (2024). Modelling the wind potential energy for metallurgical sector in Albania. *Scientific Reports*, 14, 1302. <https://doi.org/10.1038/s41598-024-51841-x>
- [26] Jabbar, R. I. (2021). Statistical analysis of wind speed data and assessment of wind power density using Weibull distribution function (Case study: Four regions in Iraq). *Journal of Physics: Conference Series*, 1804, 012010. <https://doi.org/10.1088/1742-6596/1804/1/012010>
- [27] Manusov, V., Matrenin, P., Nazarov, M., Beryozkina, S., Safaraliev, M., & Zicmane, I., Ghulomzoda, A. (2023). Short-term prediction of wind speed based on a learning process control algorithm in isolated power systems. *Sustainability*, 15, 1730. <https://doi.org/10.3390/su15021730>
- [28] Simankov, V., Buchatskiy, P., Teploukhov, S., Onishchenko, S., Kazak, A., & Chetyrbok, P. (2023). Review of estimating and predicting models of the wind energy amount. *Energies*, 16(16), 5926. <https://doi.org/10.3390/en16165926>
- [29] Kadhem, A. A., Abdul Wahab, N. I., Aris, I., Jasni, J., & Abdalla, A. N. (2017). Advanced wind speed prediction model based on a combination of Weibull distribution and an artificial neural network. *Energies*, 10, 1744. <https://doi.org/10.3390/en10111744>
- [30] Hocaoglu, F. O., Gerek, O. N., & Kurban, M. (2010). A novel wind speed modeling approach using atmospheric pressure observations and hidden Markov models. *Journal of Wind Engineering and Industrial Aerodynamics*, 98, 472–481. <https://doi.org/10.1016/j.jweia.2010.02.003>

Author's contacts:

Mostafa Abotaleb, Assoc. Prof.
Engineering School of Digital Technologies,
Yugra State University,
Khanty-Mansiysk, 628012, Russia
E-mail: abotalebmostafa@bk.ru

The Influence of Green Technology Innovation and Supply Chain Learning on Sustainability Performance in SMEs

Arash Shahin, Ehsan Masoudi*, Neda Rajabani

Abstract: Increasing attention to issues such as energy consumption, waste reduction, and resource efficiency has made the importance of sustainable performance more visible. The current research was conducted to investigate the impact of green technology innovation (GTI) and supply chain learning (SCL) on sustainability performance (SP). The study's statistical population consisted of managers of Small and medium-sized enterprises (SMEs) in Mobarakeh industrial town in Iran. The sampling method was a simple random method and 130 companies were determined by G*Power software to determine the sample size. The validity of the questionnaires was confirmed using content validity and construct validity and their reliability was obtained by Cronbach's alpha. The data was analyzed by structural equation modelling in SMART_PLS software. The research findings showed that GTI and SCL directly and significantly affect SP. According to the obtained results, GTI and SCL can be effective in improving SP in SMEs.

Keywords: environment; green technology innovation; SMEs; supply chain learning; supply chain management; sustainability performance

1 INTRODUCTION

In a world where most businesses are focused on generating wealth through economic performance, in today's dynamic business environment, companies face many challenges, including environmental issues [1]. These challenges, which are the result of inefficient methods of production and consumption, have become worrying for many countries in the world in recent years. To deal with these challenges, countries have introduced several strategic measures such as "Sustainable Development Goals" [2]. Some sources have said that small and medium-sized enterprises (SMEs) can cause more than 50% of industrial pollution that leads to environmental damage [3]. Therefore, organizations must focus their progress on measuring sustainability performance (SP) instead of focusing only on operational and economic performance to improve their sustainability indicators [4].

In providing new services and products, organizations adopt different methods to improve SP, which can be mentioned in green technology innovation (GTI) and supply chain learning (SCL). With the rapid development of the economy, innovation has become a key element for sustainable and healthy competitive development [5], and in this regard, to develop sustainable practices and pay attention to issues such as greenhouse gas emissions, energy efficiency and waste reduction, companies should focus on technological innovations [6]. The innovative activities of technology in the traditional form have indeed produced many material improvements, but also caused a lot of environmental damage, which is why even with the high costs of implementation and training of employees, green technology innovation replaces the traditional approach to improvement. It has become the SP of companies [7]. For shareholders of manufacturing companies, it is attractive in financing that GTI can improve the image, credit and reputation of the companies. In addition to creating a competitive advantage for companies, innovation in green technology can also improve social welfare and

environmental sustainability in companies. In addition, GTI reduces the emission of greenhouse gases, which leads to the optimal use of renewable energy sources [8].

In addition, SMEs as forward-looking organizations are of great importance due to their poor knowledge of their supply chain and sustainable production processes, and to compete at the national and global level, they need effective supply chain management and must cooperate with all members of the chain. Provide cooperation [9]. Today, companies focus on achieving sustainable performance, apart from competitive advantage, by collaborating with various supply chain partners. Companies mainly work together to improve individual and supply chain performance against turbulent environments. Finally, the focus of the supply chain performance has been extended to environmental and social aspects in addition to the conventional economic aspect [10]. SCL shows the knowledge related to sustainability in the learning process with the aim of social, environmental and economic issues. This is essential for the successful implementation of sustainability and performance among supply chain members [11].

Masoudi and Shahin (2025) [8] and Masoudi and Emamiyan (2025) [12], in their research concluded that GTI has an impact on SP. Also, according to the research done by Yang et al. (2023) [13] and Sun et al. (2023) [14], SCL affects sustainability performance.

According to the review of the available literature, few studies have specifically addressed the hypotheses of the present research. It is also the first time that the impact of innovation and learning on performance is studied simultaneously. In addition, the current research has been studied in a developing country, and environmental issues in developing countries are more important than in developed countries. Today, paying attention to environmental issues, the quality of products and services, and sustainable performance in organizations such as small and medium-sized companies is considered one of the main characteristics of any developing organization, which plays an important role in the economic development of countries. Accordingly,

the current research discusses the effect of GTI and SCL on SP among SMEs.

In the subsequent sections of this study, we begin by outlining the theoretical underpinnings that form the basis of our research. Following this introduction, we will detail the proposed model and the associated hypotheses guiding our investigation. The methodology employed in this research will then be thoroughly described. To validate the proposed model, we conduct an empirical study, and based on the findings derived from this analysis, we will engage in a comprehensive discussion and draw conclusions.

2 LITERATURE REVIEW

In this section, we will review the research literature. First, we will explain the concepts of sustainability performance, GTI and SCL. Next, we will discuss the relationship between GTI and SP and the relationship between SCL and SP.

2.1 Sustainability Performance

The most well-known view about sustainability performance refers to three economic (financial), social and environmental aspects. Economic performance involves a plan of production that meets current and future needs. Economic performance shows the continued profitability and financial survival of an organization whose measurement depends on long-term operational value, cost-effectiveness, output, return on investment and market value. Social performance is concerned with the relationship between social conditions such as deprivation and ecological destruction, and the concept of social performance suggests that the reduction of deprivation should entail neither unjustified environmental destruction nor economic instability. Its purpose is to reduce deprivation in the dominant economic and environmentally compatible asset base of humanity. Environmental performance requires maintaining important environmental functions and therefore, maintaining the capacity of capital stock to provide those functions, in other words, the result of managing an organization is defined from its environmental aspects [15].

The coordination of environmental, social and economic goals to maximize values and growth in the long term in providing business activities is called the function of sustainability. For this reason, the efficient use of environmental, social and economic resources for current use and at the same time protection of current resources for future generations is called sustainable performance. Environmental sustainability performance is a vital centre among the dimensions of SP, which indicates the ability of a company to reduce its environmental impact. The evaluation of environmental performance includes a detailed examination of operational effects on the environment to reduce environmentally destructive activities. Economic SP is a valid sustainable aspect among the sustainability characteristics to increase the market values of companies. It includes financial benefits and serves as an indicator of financial stability and sustainable survival of a company.

Finally, social SP refers to diverse social factors such as improving employee performance, health, safety and social well-being, equal opportunities for advancement, and support for humanitarian initiatives [8].

2.2 Green Technology Innovation

Innovation, as a fundamental source of competitive advantage, has been considered in a wide range of studies, and for this reason, agents also attach great importance to innovation processes and consider it the basis for having a good competitive position [16]. Braun and Wield (1994) were the first to introduce the concept of GTI, which is defined as technologies and processes that improve the quality of the environment and reduce the use of raw materials [5]. GTI, which aims to promote technological innovation in green processes or products, has also been used in pollution prevention, waste reduction, resource recycling and other environmental management practices [17].

Green technology innovation consists of three dimensions: product innovation, process innovation and end-of-line management innovation. Green product innovation refers to the creation of environmentally friendly products using less, non-toxic resources and degradability. With this type of innovation, the sustainability, durability and recyclability of products can be increased. Green process innovation improves the performance of manufacturing and recycling processes. This type of innovation increases the design and implementation of production processes that are more suitable and sustainable for the natural environment. End-of-line management innovation refers to the integration of technological innovations that improve the efficiency of the treatment of three wastes: wastewater, waste gas, and waste residue. This type of technology contributes to the development of sustainable products by reducing input materials and energy requirements and increasing the life cycle of products [12].

2.3 Supply Chain Learning

The concept of SCL originates from the theory of inter-organizational learning. Bessant and Tsekouras in 2001 were the first to use the supply chain as a knowledge-sharing network [18]. SCL is an important process that is widely accepted by manufacturing and service organizations as an approach to business improvement through the collection and exploitation of information and knowledge by suppliers and customers [19]. To achieve sustainability, companies are increasingly establishing collaborative relationships with their supply chain partners. Supply chain literature shows that over time, the focus of companies' performance in the supply chain has progressed from economic aspects to environmental and social aspects [11].

Supply chain learning consists of three dimensions: supplier learning, customer learning and internal learning. Supplier learning refers to the exchange of knowledge between the company and suppliers. The company uses the experiences of suppliers to improve products and adopt new technologies, and monitoring systems are in place to

effectively utilize this knowledge. Customer learning implies that primary customers play a vital role in the SCL process. Using the information and knowledge provided by customers, the company improves product innovation, develops new technologies, and creates effective monitoring systems for optimal use of this knowledge. Internal learning in the supply chain refers to the exchange of information and knowledge between the internal units of the organization. This process includes effective communication of product information and innovation, learning from each other, promoting the adoption of new technologies and evaluating the application of product knowledge in different units, which helps to improve the overall performance of the organization [20].

2.4 The Influence of GTI on Sustainability Performance

Green technologies that affect energy become a leading factor in reducing environmental problems caused by the industrial revolution and effectively reduce carbon emissions by more than 60%. Pollution monitoring, waste management and clean technology are examples of GTI [21]. GTI refers to the development of new technologies or the improvement of existing technologies to minimize carbon production, conserve existing resources, and reduce pollution for the health of employees and the public. GTI refers to technological advances aimed at reducing harmful environmental impacts. For these reasons, researchers pay much attention to GTI and its role in sustainable development and competitive advantage [12].

Green technology innovation leads to rapid progress in new systems and equipment to protect the environment and natural resources, and as a result, improves the understanding of stakeholders and managers and helps companies improve economic output. Also, adopting this innovation reduces production, operation and processing costs and ultimately increases revenues and the overall economic performance of the company. Green technology innovation as a transfer mechanism enhances environmental performance through the successful deployment of digital finance helps reduce carbon emissions and improves renewable energy sources [8].

2.5 The Influence of SCL on Sustainability Performance

SCL has been proposed as an important way to manage inter-firm resources to achieve sustainability. Case findings show that SCL by considering environmental, social and economic aspects leads to better sustainability performance. Pursuing sustainability requires acquiring and sharing new knowledge about sustainability-related products, processes, and initiatives [11].

Learning in the supply chain helps improve environmental performance. Sustainable supply chain management theory emphasizes the importance of learning in responsible supply chain management, because this process facilitates the exchange of environmental knowledge between companies and suppliers and customers, and leads to achieving sustainable supply chain configurations and

improving environmental performance. In particular, information on environmentally friendly raw materials provided by suppliers is essential for companies to make responsible production decisions. Also, sharing information such as inventory levels and sales data among supply chain members can improve logistics efficiency and visibility and contribute to the effective use of resources, which ultimately leads to improved sustainability performance [14].

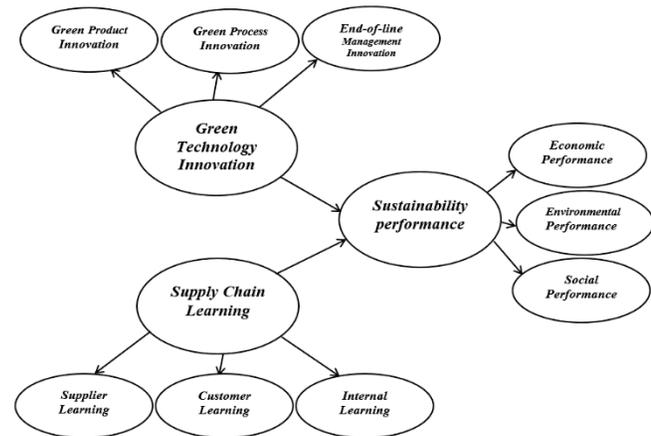


Figure 1 Research model

Fig. 1 presents the conceptual framework of the study. In this research, the independent variables identified are GTI and SCL, while the dependent variable is sustainability performance. Based on this, the hypotheses of this research are as follows:

H1: GTI impact sustainability performance.

H2: SCL impact sustainability performance.

3 METHODOLOGY

In this section, we will discuss the tools used to measure research variables and confirm the validity of its content. In the following, we will discuss the statistical population of the research, the method of data collection the sampling method and the reason for its use.

3.1 Measurement Instrument

A survey instrument was employed to assess the research variables. This questionnaire was divided into two sections. The initial section gathered demographic details about the participants, comprising three questions (Respondent's age, Respondent's level of education and Duration of the company). The subsequent section contained items pertinent to the research variables, i.e. GTI, SCL and SP. For this part of the questionnaire, 44 items were used. Also, the items were measured on a five-point Likert scale from 1 (completely disagree) to 5 (completely agree). The following studies have been used to measure the research variables.

Green technology innovation: For the GTI variable, we used the Wang et al. (2022) [22] research. This research measures GTI with 15 items and includes 3 dimensions: green product innovation (5 items), green process innovation (5 items), and end-of-line management innovation (5 items).

Supply chain learning: For the SCL variable, we used the Huo et al. (2021) [20] research. This research measures SCL with 15 items and includes 3 dimensions: supplier learning (5 items), customer learning (5 items) and internal learning (5 items).

Sustainability performance: For the SP variable from the research of Yang et al. (2023) [13] we used. This research measures SP with 14 items and includes 3 dimensions of economic performance (5 items), environmental performance (5 items) and social performance (4 items).

3.2 Sample Size and Data Analysis Approach

The statistical population of this research consists of quality managers of small and medium companies in Mobarakeh Industrial City. Data collection was conducted using a straightforward random sampling technique during the winter season of 2024. For this purpose, the list of companies was used for simple random sampling. This industrial townhouses more than 500 active small and medium companies. To establish the necessary sample size for PLS-SEM, we utilized G*Power software to conduct a power analysis based on the model configurations [23]. In this research, 130 data (quality managers of small and medium-sized enterprises) to achieve an 80% power level and identify R^2 values of no less than 0.1 while maintaining a 5% error margin, G*Power software was utilized to determine the required sample size. For this purpose, 15% more, i.e. 150 questionnaires were distributed among the companies, after collecting them and removing the questionnaires with incomplete information, the rest of the questionnaires were used for analysis. Also, the questionnaires were distributed to the respondents in person and reminders were made by telephone to increase the response rate. In addition, two expert academicians and two senior experts from the companies reviewed the survey items and the content validity of the questionnaire was confirmed.

The Kolmogorov-Smirnov (K-S) test indicated that the distribution of the constructs was not normal at a significance level of $p = 0.05$, which was attributed to the limited sample size ($n = 130$). Consequently, PLS-SEM was employed for data analysis [24]. In this paper, SEM was used because, according to Byrne (2013) [25], confirmatory factor analysis (CFA) is used to examine the theoretical and conceptual dimensions of visible and hidden scales in complex models. Causal inference is due to the nature of accessing and correcting the measurement errors observed in the detailed analysis, and for this reason, in the present study, PLS-SEM was used in the hypothesis analysis.

4 RESULTS

These results were obtained in the analysis of demographic variables. In terms of the duration of company activity, 46.93% of companies with more than 15 years of activity, equivalent to 61 companies, had the highest frequency, and 10.77% of companies, equivalent to 14 companies with less than 5 years of activity, had the lowest frequency. In terms of age, most of the respondents were

between 41 and 50 years old with a frequency of 44.61%, equivalent to 58 people, and managers less than 30 years old with 13.08%, equivalent to 17 people, had the lowest frequency. In terms of educational level, most of the respondents with a bachelor's degree with a frequency of 48.46%, equivalent to 63 people, and those with a higher degree than a bachelor's degree with 20%, equivalent to 20 people, had the lowest frequency in the statistical sample.

Table 1 Measurement Model

Construct	Label	Factor	C- α	CR	AVE
Green product innovation	GPt11	0.851	0.904	0.929	0.722
	GPt12	0.845			
	GPt13	0.884			
	GPt14	0.849			
	GPt15	0.818			
Green process innovation	GPs11	0.838	0.871	0.906	0.661
	GPs12	0.866			
	GPs13	0.876			
	GPs14	0.724			
	GPs15	0.749			
End-of-line management innovation	EMI1	0.786	0.879	0.911	0.673
	EMI2	0.765			
	EMI3	0.775			
	EMI4	0.788			
	EMI5	0.865			
Supplier learning	SL1	0.825	0.856	0.897	0.635
	SL2	0.842			
	SL3	0.845			
	SL4	0.795			
	SL5	0.793			
Customer learning	CL1	0.829	0.854	0.896	0.633
	CL2	0.846			
	CL3	0.768			
	CL4	0.770			
	CL5	0.761			
Internal learning	IL1	0.786	0.808	0.867	0.566
	IL2	0.719			
	IL3	0.731			
	IL4	0.783			
	IL5	0.741			
Economic performance	EcP1	0.830	0.853	0.895	0.631
	EcP2	0.738			
	EcP3	0.771			
	EcP4	0.769			
	EcP5	0.857			
Environmental performance	EnP1	0.753	0.840	0.887	0.611
	EnP2	0.737			
	EnP3	0.793			
	EnP4	0.787			
	EnP5	0.832			
Social performance	SP1	0.827	0.817	0.879	0.645
	SP2	0.785			
	SP3	0.836			
	SP4	0.762			

Regarding the research variables, Chin (2010) outlines that the analysis of studies using Partial Least Squares (PLS) involves two main stages: first, the assessment of the measurement model (external) and the evaluation of the structural model (internal); second, conducting tests on the measurement model. These tests are essential for verifying both the reliability and validity of the constructs involved [26]. As demonstrated in Tab. 1, all factor loadings for the constructs exceed the threshold of 0.7, thereby confirming the validity of the proposed model. This suggests that the

model is robust and adequately represents the underlying theoretical framework [27].

To assess the reliability of the external model, we utilized Cronbach's alpha ($C-\alpha$), composite reliability (CR), and average variance extracted (AVE). As illustrated in Table 1, both $C-\alpha$ and CR exceed the threshold of 0.7, indicating that the reliability of the variables is established. Additionally, the AVE for all constructs was found to be greater than 0.5, further confirming their reliability [24]. Furthermore, the composite reliability is confirmed as the value of CR exceeds that of AVE [28]. The results pertaining to discriminant validity are presented in Tables 2 and 3. According to Table 2, the square root of the AVE for each variable surpasses its correlations with other variables, indicating compliance with the standards set forth by Fornell and Larcker (1981) [29]. Additionally, Table 3 reveals that the HTMT ratios are below 0.9, thereby confirming the establishment of discriminant validity [30].

Table 2 Fornell and Larcker coefficients

	<i>SP</i>	<i>EcP</i>	<i>EnP</i>	<i>GP<i>s</i>i</i>	<i>GP<i>t</i>i</i>	<i>EMI</i>	<i>SL</i>	<i>IL</i>	<i>CL</i>
<i>SP</i>	0.803								
<i>EcP</i>	0.356	0.794							
<i>EnP</i>	0.310	0.709	0.781						
<i>GP<i>s</i>i</i>	0.300	0.337	0.365	0.813					
<i>GP<i>t</i>i</i>	0.559	0.394	0.359	0.292	0.850				
<i>EMI</i>	0.599	0.498	0.471	0.310	0.466	0.820			
<i>SL</i>	0.349	0.726	0.513	0.271	0.346	0.409	0.797		
<i>IL</i>	0.347	0.598	0.652	0.237	0.277	0.454	0.599	0.752	
<i>CL</i>	0.297	0.605	0.519	0.291	0.443	0.403	0.684	0.735	0.795

Table 3 HTMT criterion

	<i>SP</i>	<i>EcP</i>	<i>EnP</i>	<i>GP<i>s</i>i</i>	<i>GP<i>t</i>i</i>	<i>EMI</i>	<i>SL</i>	<i>IL</i>	<i>CL</i>
<i>SP</i>									
<i>EcP</i>	0.418								
<i>EnP</i>	0.367	0.835							
<i>GP<i>s</i>i</i>	0.360	0.390	0.426						
<i>GP<i>t</i>i</i>	0.640	0.446	0.409	0.316					
<i>EMI</i>	0.705	0.575	0.544	0.340	0.504				
<i>SL</i>	0.411	0.848	0.604	0.318	0.397	0.471			
<i>IL</i>	0.430	0.718	0.797	0.292	0.328	0.538	0.713		
<i>CL</i>	0.350	0.704	0.614	0.332	0.389	0.460	0.793	0.819	

In analyzing the structural framework of the research, a detailed three-phase approach was utilized. This approach involved measuring the R^2 value, assessing the quality of the Q^2 model, and evaluating the significance of the path coefficients present in the structural framework [31]. The findings from this analysis are presented in Tabs. 4 and 5, as well as Figs. 2 and 3, providing a clear overview of the model's performance and the relationships between variables. This systematic approach ensures a thorough understanding of how well the model fits the data and its predictive capabilities.

Table 4 R^2 , cross validity redundancy and communality

Variables	R^2	Redundancy index	Communality index
Sustainability performance	0.679	0.252	0.315

$$GOF = \sqrt{AVE \cdot R^2} = 0.661 \quad (1)$$

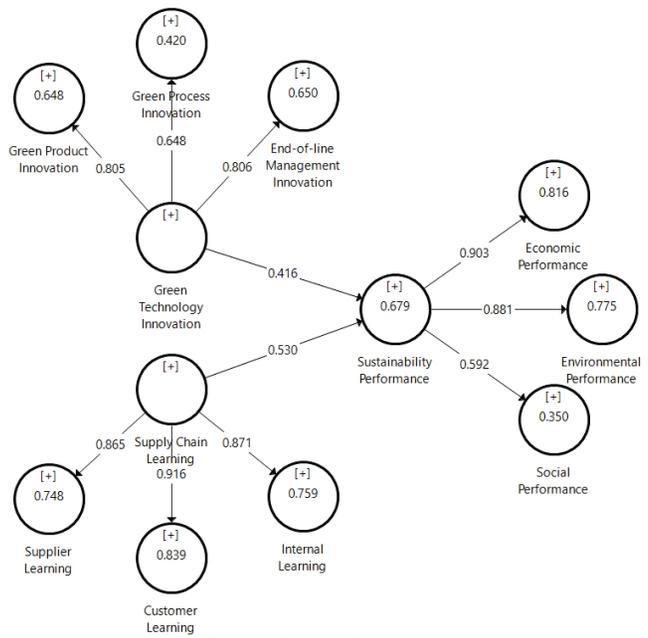


Figure 2 Path coefficient result

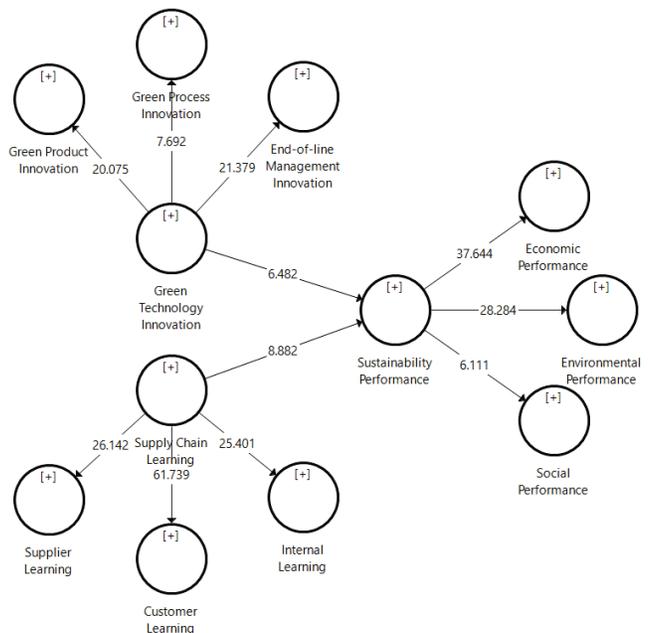


Figure 3 Z-value significant coefficients

Based on the thresholds of 0.25, 0.50, and 0.75, which represent low, medium, and high levels, the R^2 value indicating sustainability performance approached a high level [32]. Similarly, using the thresholds of 0.02, 0.15, and 0.35 for the redundancy index in the structural model of sustainability performance, it was found to be medium to high [23]. Furthermore, the communality index for sustainability performance also fell within the medium to high range. Lastly, employing the goodness of fit (GOF) test as outlined by Tenenhaus et al. (2005) and following Eq. (1), the overall research model was assessed. The resulting value of 0.661 indicated that the overall quality of the research model was high [33].

Also, In the evaluation of the structural model, the results indicate a good fit of the model with the data. The RMSEA is 0.06, which suggests an acceptable model fit. Additionally, the CFI is reported at 0.93 and the TLI is also 0.93, both of which are above the accepted thresholds for a good fit. These results demonstrate that the proposed model effectively explains the relationships among the variables and is statistically valid [34].

In conclusion, the analysis of the structural model involved utilizing path coefficient testing alongside the bootstrap method to evaluate the strength of the connections among the research hypotheses. As illustrated in Fig. 3 and Tab. 5, it is evident that both GTI and SCL influence sustainability performance.

Table 5 Hypotheses testing results

Hypothesis	Path coefficient	SE	t-value	p-value	Decision
GTI → SP	0.416	0.064	6.482	0.000	Supported
SCL → SP	0.530	0.060	8.882	0.000	Supported

As shown in Tab. 5, GTI has an effect of 0.416 (t -value = 6.482) on sustainability performance. Also, SCL has an effect of 0.530 (t -value = 8.882) on sustainability performance. The findings of this research hold significant practical implications for SMEs, as they indicate that investing in GTI and enhancing SCL can contribute to the improvement of their SP. In light of increasing social and environmental pressures, these companies can not only mitigate their negative environmental impacts by adopting these approaches but also gain a competitive advantage. These results can guide managers in designing effective strategies to achieve their sustainability objectives.

5 DISCUSSION

This part offers the examination and explanation of the research results, using a systematic method [34]. In recent years, the concepts of GTI and SCL have been linked with economic ideology, and a key result of GTI and SCL is sustainable performance. Utilizing GTI and SCL within both manufacturing and service sectors enables organizations to concentrate on their intended market, optimize their material and human resources, and enhance their competitive standing in the industry. However, due to the small sample size of the current research, which includes 130 small and medium-sized companies and is limited to only one industrial town in Iran, caution should be exercised in generalizing the findings to other fields. For example, it can be examined from different points of view: 1) Cultural and economic characteristics: Every country or region has its own cultural, economic and social characteristics that can affect the acceptance and implementation of GTI and SCL concepts. For example, small and medium enterprises (SMEs) in developed countries may have better infrastructure and financial resources to implement these concepts. 2) Diversity in industry: Different industries may have different needs. For example, companies in the IT sector may use GTI and SCL more than companies in the manufacturing or service

industries. 3) Resource limitations: Small and medium enterprises in developing countries may face more limitations in human and financial resources, which can affect their ability to implement the desired concepts. 4) Local laws and regulations: Local laws and regulations can affect how companies operate. For example, environmental requirements in one country may be much stricter or easier than in another.

The present research was conducted on SMEs in Mobarakeh Industrial Town in Iran. For this purpose, a questionnaire was used in data collection. The quality managers of these companies were responsible. According to Tab. 5, in the first hypothesis of the research regarding the effect of GTI on SP, considering that its t -value is 6.058, which is outside the range $(-1.96, 1.96)$, the p -value is less than 0.05, and the intensity of the effect 0.416 shows the positive and significant effect of GTI on SP. In the second hypothesis of the research about the effect of SCL on SP, considering that its t -value is 9.021, which is outside the range $(-1.96, 1.96)$ the p -value is less than 0.05, and the intensity of the effect is 0.530. It shows the positive and significant effect of SCL on sustainability performance.

5.1 Theoretical Implications

About the importance of SMEs in environmental issues and the impact of GTI and SCL on sustainable performance in these companies, the results of this research have important theoretical implications. The first research hypothesis, the impact of GTI on sustainability performance, was accepted with an impact rate of 0.416. The result of the research with the research results of Masoudi and Shahin (2025) [8] and Masoudi and Emamiyan (2025) [12] is similar. The confirmation of this hypothesis shows that the performance and survival of companies strongly depend on their innovation, and one of the ways to achieve business sustainability is to adapt to technological advances. The goal of innovation is to create new technologies and features, optimize the current process by integrating existing technologies, and increase knowledge about the interrelationship of processes and tasks, as well as about adding values to current structures and processes. Growth and development in green technologies increase environmental innovations and consequently increase environmental productivity. Specifically, GTI helps companies find creative solutions to environmental, economic, and social issues. From another perspective, sustainable performance can be created when companies develop innovative strategies. The company's innovation acts as an internal driver in creating green capabilities. Eco-innovation enhances brand reputation, attracts environmentally conscious consumers, and creates new market opportunities, resulting in long-term financial benefits. In addition, the impact of clean technology on sustainable performance cannot be ignored.

The second hypothesis of the research, the effect of SCL on sustainability performance, was accepted with an effect rate of 0.530. The results of the research with the results of Yang et al. (2023) and Sun et al. (2023) is similar. The

confirmation of this hypothesis shows that Learning enables organizations to interact with the internal and external environment. From this interaction, knowledge about the changing environment, market and stakeholder demands is created, acquired, shared, integrated and used to increase organizational value. Learning provides an opportunity for the organization to actively generate information relevant to understanding and addressing the ever-changing economic, social and environmental demands of stakeholders simultaneously. The information obtained through learning is used in the pursuit of strategies and policies that address the demands of sustainability performance. SCL represents the sustainability-related knowledge of a learning process aimed at solving social, environmental and economic issues, which is necessary for the successful implementation of sustainability and performance among supply chain members. SCL creates competitive advantages and leads to improved sustainability performance of companies. Learning the supply chain in the manufacturing industry also helps to increase the company's long-term values and ultimately sustainability performance. Effective supply chain management encourages companies to adopt environmentally friendly approaches along with economic and social performance in sourcing, production and distribution processes. By incorporating sustainable practices into SCL, such as eco-design, green packaging, and renewable energy sources, organizations can achieve cost savings through reduced energy consumption, reduced waste, and improved resource efficiency. Finally, it increases the overall sustainability of their operations.

5.2 Managerial Implications

Enhancing Sustainability Performance in SMEs through GTI and SCL. To strengthen sustainability performance in SMEs through GTI and SCL, managers must implement a comprehensive and strategic approach. This approach should encompass four key pillars: 1) Development of a Green Innovation Culture, 2) Strengthening Collaboration with Suppliers, 3) Establishing Sustainable Performance Evaluation Systems, and 4) Utilizing Digital Technologies. 1) Development of a Green Innovation Culture: Managers should pay special attention to fostering an organizational culture that promotes green innovations. This involves encouraging employees to contribute creative ideas aimed at reducing environmental impacts. To facilitate this, organizing workshops and training sessions related to green technologies can enhance employees' awareness and skills. Additionally, establishing a reward system for innovative ideas can provide further motivation for employee participation. 2) Strengthening Collaboration with Suppliers: Managers should focus on building close and effective relationships with suppliers and supply chain partners. Such collaboration may involve forming joint working groups to exchange best practices and green technologies. For instance, regular meetings can be held to share successful experiences and discuss challenges and solutions. This type of collaboration not only aids in improving environmental performance but can also lead to cost reductions and

increased efficiency. 3) Establishing Sustainable Performance Evaluation Systems: Managers need to develop sustainable performance evaluation criteria that encompass environmental, social, and economic indicators. These criteria should be designed to enable the measurement of progress and facilitate data-driven decision-making. For example, utilizing performance management software can assist in data collection and analysis, helping managers identify strengths and weaknesses. 4) Utilizing Digital Technologies: The adoption of digital technologies such as the Internet of Things (IoT) and data analytics can contribute to process optimization and waste reduction. Managers should take the investment in these technologies seriously and design implementation plans accordingly. For instance, deploying smart sensors on production lines can help monitor energy consumption and raw material usage, leading to cost reductions and improved environmental performance.

6 CONCLUSION

The purpose of this article was to analyze the impact of GTI and SCL on sustainability performance in SMEs. For this purpose, first by collecting information on the theoretical foundations and literature of the subject, library sources, articles and scientific databases were used, and finally, the results were discussed by testing the proposed model in SMEs in Mobarakeh Industrial town in Iran. The findings showed the effect of GTI on sustainability performance. The effect of SCL on stability performance was also confirmed. The results showed that SCL has a greater impact on stability performance than GTI. In addition, the results showed that green product innovation from GTI and customer learning from SCL had the greatest impact on sustainability performance. Green process innovation from GTI and supplier learning from SCL had the least impact on sustainability performance.

Among the advantages of the proposed model of the current research, we can mention the comprehensiveness of the examined variables and the novelty of the research topic in the sense that the effect of GTI and SCL simultaneously affects SP. Overall, it can be stated that the implementation of GTI and SCL is essential for enhancing sustainability outcomes in small and medium-sized enterprises (SMEs). The sustainability of companies is critical to realizing an organization's vision without compromising its market advantage while meeting the needs of economic growth, and environmental and social responsibilities. The transition of an organization to a sustainable future should be realized through the efforts and responsibility of executive directors, stakeholders and employees.

While our initial discussion emphasized the benefits of GTI and SCL, such as enhanced sustainability, improved operational efficiency, and competitive advantage, we acknowledge that there are inherent challenges and costs that organizations may encounter during implementation. The primary concern related to GTI is the increase in operational costs, which can impose a significant financial burden on SMEs. The initial investment in green technologies involves purchasing advanced equipment, training personnel, and

integrating these technologies into existing processes. Although the long-term benefits and positive environmental impacts of GTI are considerable, the short-term costs may hinder organizations from adopting these initiatives. Additionally, implementing SCL practices comes with complex challenges. Transitioning to SCL may require significant adjustments in supplier relationships, transportation methods, and inventory management. For instance, sourcing materials from environmentally friendly suppliers can limit options and increase procurement costs. Moreover, the need for transparency in the supply chain can elevate administrative burdens and complexities associated with ensuring supplier compliance with sustainability standards. Resistance to change within organizations is another challenge; employees and stakeholders accustomed to traditional practices may be reluctant to embrace new technologies or processes. This resistance can lead to a lack of engagement or even active opposition, disrupting the successful implementation of these initiatives. Therefore, addressing this human factor and developing effective strategies for change management, such as promoting a sustainability culture and providing appropriate training, is essential. Finally, organizations must pay attention to the economic and regulatory context in which these innovations are implemented, as market fluctuations and political changes can significantly impact the success of their sustainability efforts.

A significant constraint of this study is the limited number of participants, which could hinder the ability to apply the results more broadly. A limited number of respondents can lead to statistical insignificance and may not adequately represent the broader population of SMEs. Additionally, the geographical focus on SMEs in Iran introduces further constraints; the unique economic, cultural, and regulatory contexts of this region may not reflect the experiences of SMEs in other countries. Consequently, caution should be exercised when extrapolating these findings to different settings. Another significant limitation is the reliance on self-reported data, which can introduce bias due to subjective perceptions and potential inaccuracies in responses. Self-reported measures are often influenced by social desirability and personal biases, potentially skewing the results. Furthermore, the timing of data collection during the winter of 2024 may also limit the applicability of findings to other periods, as external factors such as economic conditions and seasonal variations can impact business operations and attitudes toward innovation. Overall, these limitations highlight the need for further research with larger, more diverse samples and objective measures to validate and expand upon the current findings. Furthermore, the subsequent recommendations may be considered for upcoming studies:

1) Future investigations could explore the influence of various elements such as total quality management, hard and soft quality management, green supply chain management, green entrepreneurship and supply chain agility on sustainability performance in manufacturing and service organizations.

- 2) In future research, the effect of each dimension of GTI, i.e. green product innovation, green process innovation and end-of-line management innovation, on sustainability performance can be studied in three hypotheses.
- 3) In future research, the impact of GTI on the dimensions of sustainability performance, i.e. economic performance, social performance and environmental performance, can be studied in three hypotheses.
- 4) The current research can be studied among companies that are more related to green products and processes, such as food manufacturing industries.

7 REFERENCES

- [1] Fok, L., Morgan, Y. C., Zee, S., & Mock, V. E. (2023). The impact of organizational culture and total quality management on the relationship between green practices and sustainability performance. *International Journal of Quality & Reliability Management*, 40(6), 1564-1586. <https://doi.org/10.1108/IJQRM-12-2021-0450>
- [2] Fatimah, Y. A., Kannan, D., Govindan, K., & Hasibuan, Z. A. (2023). Circular economy e-business model portfolio development for e-business applications: Impacts on ESG and sustainability performance. *Journal of Cleaner Production*, 415, 137528. <https://doi.org/10.1016/j.jclepro.2023.137528>
- [3] Jum'a, L., Alkalha, Z., Al Mandil, K., & Alaraj, M. (2023). Exploring the influence of lean manufacturing and total quality management practices on environmental sustainability: the moderating role of quality culture. *International Journal of Lean Six Sigma*, 14(7), 1626-1654. <https://doi.org/10.1108/IJLSS-11-2021-0203>
- [4] Ali, S. S., Kaur, R., & Khan, S. (2023). Evaluating sustainability initiatives in warehouse for measuring sustainability performance: an emerging economy perspective. *Annals of Operations Research*, 324(1), 461-500. <https://doi.org/10.1007/s10479-021-04454-w>
- [5] Zhou, L., & Dai, Y. (2023). Green Production Management and Innovation Nexus: Evidence from Technology-Based SMEs of China. *Sustainability*, 15(6), 4710. <https://doi.org/10.3390/su15064710>
- [6] Khan, A. N., Mehmood, K., & Kwan, H. K. (2024). Green knowledge management: A key driver of green technology innovation and sustainable performance in the construction organizations. *Journal of Innovation & Knowledge*, 9(1), 100455. <https://doi.org/10.1016/j.jik.2023.100455>
- [7] Zhou, N., Pan, L., Tian, Y., Zhu, N., Cai, X., & Gao, J. (2023). How sustainable business model innovation and green technology innovation interact to affect sustainable corporate performance. *Frontiers in Environmental Science*, 11, 1049295. <https://doi.org/10.3389/fenvs.2023.1049295>
- [8] Masoudi, E., & Shahin, A. (2025). The influence of TQM on sustainability performance: the mediating role of green technology innovation in manufacturing firms. *Technological Sustainability*, 4(4), 353-379. <https://doi.org/10.1108/TECHS-04-2025-0082>
- [9] Kot, S. (2018). Sustainable supply chain management in small and medium enterprises. *Sustainability*, 10(4), 1143. <https://doi.org/10.3390/su10041143>
- [10] Sudusinghe, J. I., & Seuring, S. (2022). Supply chain collaboration and sustainability performance in circular economy: A systematic literature review. *International Journal of Production Economics*, 245, 108402.

- <https://doi.org/10.1016/j.jipe.2021.108402>
- [11] Yang, Y., Jia, F., & Xu, Z. (2019). Towards an integrated conceptual model of supply chain learning: an extended resource-based view. *Supply Chain Management: An International Journal*, 24(2), 189-214. <https://doi.org/10.1108/SCM-11-2017-0359>
- [12] Masoudi, E., & Emamiyan, S. A. (2025). The influence of green product innovation, green process innovation, and end-of-line management innovation on sustainability performance dimensions. *SEISENSE Journal of Management*, 8(1), 59-75. <https://doi.org/10.33215/mxs5kw74>
- [13] Yang, Y., Jiang, Y., Chen, H., & Xu, Z. (2023). Linking relation-specific investments and sustainability performance: the mediating role of supply chain learning. *International Journal of Operations & Production Management*, 43(8), 1329-1356. <https://doi.org/10.1108/IJOPM-07-2022-0463>
- [14] Sun, J., Xie, X., Shi, Y., Zhou, M., Yan, L., & Venkatesh, M. (2023). The effect of supply chain learning on corporate sustainability performance: the perspective of ambidextrous innovation capability. *International Journal of Logistics Research and Applications*, 1-17. <https://doi.org/10.1080/13675567.2023.2226058>
- [15] Antwi-Boateng, C., Mensah, H. K., & Asumah, S. (2023). Eco-intellectual capital and sustainability performance of SMEs: The moderating effect of eco-dynamic capability. *Cogent Business & Management*, 10(3), 2258614. <https://doi.org/10.1080/23311975.2023.2258614>
- [16] Masoudi, E. (2021). The impact of total quality management on innovation in small and medium-sized enterprises with the mediating role of organizational learning. *Quarterly Journal of Industrial Technology Development*, 19(43), 77-92. <https://doi.org/10.22034/jtd.2021.244744>
- [17] Feng, Y., Wang, X., & Liang, Z. (2021). How does environmental information disclosure affect economic development and haze pollution in Chinese cities? The mediating role of green technology innovation. *Science of the Total Environment*, 775, 145811. <https://doi.org/10.1016/j.scitotenv.2021.145811>
- [18] Masoudi, E., & Shahin, A. (2021). Structural Relation between Green Entrepreneurial Orientation and Green Innovation by role of Supply Chain Learning as a Mediator in SMEs. *Journal of Entrepreneurship Development*, 14(3), 521-539. <https://doi.org/10.22059/jed.2021.316511.653554>
- [19] Haq, M. Z. U., Gu, M., & Huo, B. (2021). Enhancing supply chain learning and innovation performance through human resource management. *Journal of Business & Industrial Marketing*, 36(3), 552-568. <https://doi.org/10.1108/JBIM-12-2019-0534>
- [20] Huo, B., Haq, M. Z. U., & Gu, M. (2021). The impact of information sharing on supply chain learning and flexibility performance. *International Journal of Production Research*, 59(5), 1411-1434. <https://doi.org/10.1080/00207543.2020.1824082>
- [21] Ye, M., Hao, F., Shahzad, M. & Kamran, H. W. (2022). How green organizational strategy and environmental CSR affect organizational sustainable performance through green technology innovation amid COVID-19. *Frontiers in Environmental Science*, 10, 959260. <https://doi.org/10.3390/su16010277>
- [22] Wang, M., He, Y., Zhou, J., & Ren, K. (2022). Evaluating the effect of Chinese environmental regulation on corporate sustainability performance: the mediating role of green technology innovation. *International Journal of Environmental Research and Public Health*, 19(11), 6882. <https://doi.org/10.3390/ijerph19116882>
- [23] Hair, J. F., Hult, G. T. M., Ringle, C. M., & Sarstedt, M. (2017). *A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM)*, Sage, Thousand Oaks, CA.
- [24] Hair, J. F., Sarstedt, M., Pieper, T. M., & Ringle, C. M. (2012). The use of partial least squares structural equation modeling in strategic management research: a review of past practices and recommendations for future applications. *Long range planning*, 45(5-6), 320-340. <https://doi.org/10.1016/j.lrp.2012.09.008>
- [25] Byrne, B. M. (2013). *Structural equation modeling with Mplus: Basic concepts, applications, and programming*. Routledge.
- [26] Chin, W. W. (2010). How to write up and report PLS analyses. In *Handbook of Partial Least Squares*. Springer, Berlin, Heidelberg, 655-690.
- [27] Hair, J. F., Anderson, R. E., Tatham, R. L. & Black, W. C. (2010). *Multivariate data analysis (7th ed)*. New Jersey: Prentice Hall.
- [28] Bagozzi, R. P., & Yi, Y. (1988). On the evaluation of structural equation models. *Journal of the Academy of Marketing Science*, 16, 74-94. <https://doi.org/10.1007/BF02723327>
- [29] Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research*, 18(1), 39-50. <https://doi.org/10.1177/002224378101800104>
- [30] Henseler, J., Ringle, C. M., & Sarstedt, M. (2015). A new criterion for assessing discriminant validity in variance-based structural equation modeling. *Journal of the Academy of Marketing Science*, 43, 115-135. <https://doi.org/10.1007/s11747-014-0403-8>
- [31] Moradi, M., and Miralmasi, A. (2020), Pragmatic research method. School of Quantitative and Qualitative Research. Retrieved from. <https://analysisacademy.com>
- [32] Hair, J. F., Ringle, C. M., & Sarstedt, M. (2011). PLS-SEM: Indeed a silver bullet. *Journal of Marketing Theory and Practice*, 19(2), 139-152. <https://doi.org/10.2753/MTP1069-6679190202>
- [33] Tenenhaus, M., Vinzi, V. E., Chatelin, Y. M., & Lauro, C. (2005). PLS path modeling. *Computational Statistics & Data Analysis*, 48(1)159-205. <https://doi.org/10.1016/j.csda.2004.03.005>
- [34] Moradi, M., & Kheiri, B. (2025). Crafting cohesive narratives: a step-by-step innovative pattern for writing structured discussions in research. *Quality & Quantity*, 1-35. <https://doi.org/10.1007/s11135-025-02185-9>

Authors' contacts:

Arash Shahin, Prof.
Department of Management, University of Isfahan,
Azadi Square, Isfahan, Iran
E-mail: shahin@ase.ui.ac.ir

Ehsan Masoudi, MA
(Corresponding author)
Quality Management Research Group, University of Isfahan,
Azadi Square, Isfahan, Iran
E-mail: Ehsanma74@yahoo.com

Neda Rajabani, PhD
PhD in Industrial Management, Operations Research, University of Tehran,
16th Azar St, Enghelab Square, Tehran, Iran
E-mail: neda_rajabani@ut.ac.ir

Analysis of the Black and White Image Recording of the Converted RGB Photographic Image

Darija Ćutić, Miroslav Mikota*, Maja Turčić

Abstract: A digital camera is generally used for digital recording of photographic images in the RGB system. When such a multicoloured photographic image is converted to black and white using the standard Grayscale conversion in Adobe Photoshop, the colour information, expressed as R, G and B channel values, is lost. This paper analyses the darkening caused by different conversions of the RGB record of a photographic image to a black and white image record. The conversions were performed on three photos, with 10 random colours separated for each photo before conversion. After conversion, the L parameter, which indicates the brightness value, was analysed. The changes in the L parameter due to different methods of photo conversion from multicolour to black and white, such as desaturation, extraction of the R channel, extraction of the G channel and extraction of the B channel, were also analysed.

Keywords: black and white image; Grayscale system; L parameter; photo conversion; RGB system

1 INTRODUCTION

The digital multicolour records of photographic images are generally recorded in RGB colour spaces such as sRGB or Adobe RGB. The main purpose of the RGB colour system is to reproduce the photograph on the screen [1-3]. The photograph is then converted to black and white as this allows the manipulation of the photograph to achieve a sufficient range of tones to control blackness and emphasise the contrast between light and dark tones [4, 5]. When converting a photo from a multicolour to a black and white system, certain details are lost and the display information is reduced. This loss is caused by the reduction of the colour range [6-8]. In the RGB system, the colours contain parameters that determine the amount of red, green and blue components for each RGB channel [9, 10]. When converting an image from a multicolour to a black and white display with a standard Grayscale conversion, the parameter L from the $L^*a^*b^*$ system, which defines the brightness, remains the same [11, 12].

To perform a more detailed analysis, the L parameter was observed in other methods of conversion from a multicolour to a black and white display (desaturation, extraction of the R channel, extraction of the G channel and extraction of the B channel) [13, 14]. Colours that have the same L parameter in the RGB system are displayed in the same grey colour when we convert the photograph to the standard Grayscale system. The display information is lost because conversion to a black and white system using the standard Grayscale conversion causes the reduction of the RGB colour system. In this paper with three different multicoloured photos, 10 colours were separated for each of these photos. The colours were observed in different types of conversion to a black and white display and the change in the parameter L after conversion was analysed. Thus, a new parameter ΔL was determined to monitor the difference and change in the parameter L .

1.1 Conversion of a Photographic Image from a Multicolour RGB System to a Black and White Image Record

There are several ways to convert a photographic image from multicolour to black and white. The basic method of converting the original multicolour image to a converted black and white image is the standard Grayscale conversion, in which the photo is transferred from the RGB system to the Grayscale system. In this conversion mode, the brightness of the grey colour is determined by the parameter L , which in this case remains the same for the original and the converted image [15].

Other conversions of photographic images from RGB multicolour to black and white leave the images in the RGB system. One of these is desaturation, i.e. reducing the colour saturation to a minimum, which converts the image to black and white. There are other possibilities such as separating the red, green or blue channels of the image records, resulting in a separate black and white representation.

1.2 Monitoring the L Parameter

The parameter L determines the darkening or brightness of the grey colour in the black and white display. The L parameter only remains the same for the standard Grayscale conversion of the original multicolour image into a black and white image. With the other methods mentioned for converting multicolour images to black and white, the L parameter is also changed for some, but not all colours. By monitoring the parameter L , the optimal conversion method for a photo can be determined depending on the subject of the photo and the proportion of a particular colour in the RGB system.

If we compare the L parameter of the original RGB image record with the L parameter of the converted black and white image records, the increase or decrease of the L parameter can be expressed by ΔL .

Accordingly, ΔL can be expressed by the equation:

$$\Delta L = L_{RGB} - L_{BW} \quad (1)$$

Where L_{RGB} is a parameter that specifies the brightness of an RGB multicolour image, and L_{BW} is a parameter that specifies the brightness of a black and white image that has been converted using one of the various conversions (1). According to specific conversion L_{BW} can be shown as L_{GS} , for standard Grayscale conversion to black and white, L_{DES} for conversion to black and white with desaturate tool, L_{R-C} for conversion to black and white via extracting the red channel, L_{B-C} for conversion to black and white via extracting the blue channel or L_{G-C} for conversion to black and white via extracting the green channel.

L_{RGB} and L_{GS} which is parameter of Grayscale standard conversions have the same L parameter value as it is shown in next equation:

$$L_{RGB} = L_{GS} \quad (2)$$

If the ΔL parameter is greater than 10 in different conversions compared to L_{RGB} or L_{GS} , this difference is noticeable and affects the contrast in different conversions.

2 EXPERIMENT PREPARATION

2.1 Colour Separation

For the experiment, three photographs were taken in which the separation of randomly selected colours from the recorded data set was performed, and the specified separated colours were observed after conversion to a black and white display. Figs. 1, 2 and 3 show randomly separated colours from photographic images. Fig. 1 shows the first photo, in which the main object is a portrait. The skin contains a handful of red tones, which is why this subject was chosen.



Figure 1 First photography with ten separated colours

In Fig. 2 you can see how the sea was chosen as the main object of the picture because it is rich in shades of grey and blue. In Fig. 3, the main object in the photo is a human figure in the greenery of the park, and this scene was chosen to observe the photo, which is rich in blue tones.



Figure 2 Second photography with ten separated colours



Figure 3 Third photography with ten separated colours

2.2 Image Conversion to Black and White

All photos taken were converted to black and white using five different conversion methods: "Desaturate", "Grayscale", "R-channel", "B-channel" and "G-channel". The "Desaturate" method converts a multicoloured photo to black and white by reducing the saturation of the entire photo to zero, but leaving the photo in the RGB system. The "Grayscale" method is a standard method for converting a photo to black and white, in which the photo is converted from the RGB system to the Grayscale display system. In the "R-channel" method, the red channel is extracted while the green and blue channels are switched off so that they are not visible in the photo. The "G-channel" conversion method extracts the green channel, which is the only one visible, while switching off the display of the red and blue channels. The last method for converting a multicolour image to black and white is the "B-channel" method, in which we separate the blue channel for the photographic capture of the image and set it as the only visible channel, while switching off the red and green channels in this capture.

Fig. 4 shows the conversion of the first photo to black and white using five different conversion methods, while Fig. 5 shows the conversion of the second photo and Fig. 6 the conversion of the third photo.



Figure 4 First photograph converted to black and white with five different methods



Figure 5 Second photograph converted to black and white with five different methods



Figure 6 Third photograph converted to black and white with five different methods

3 RESULTS VALUES OF THE SEPARATED COLOURS FROM CAPTURED PHOTOGRAPHS

The following tables show the values corresponding to the separated colours. The values were measured using Adobe Photoshop software. The red, green and blue components in the RGB system and the L value are measured for each colour.

After different types of conversion to black and white, the L values were measured for each converted display in order to compare them and analyse the changes.

Tab. 1 shows the measured values for the first five separated colours of the first photo, while Tab. 2 shows the other five separated colours of the first photo. Tab. 3 and Tab. 4 show the measured values for the ten separated colours from the second photo. Tab. 5 and Tab. 6 show the measured values for the separated colours of the third photo.

Table 1 Table values of first 5 separated colours from first photography

Photography 1 – first 5 separated colours					
RGB colour code	#49271d	#b4866f	#333d34	#3d4b28	#4a887d
R	73	180	51	61	74
G	39	134	61	75	136
B	29	111	52	40	125
L_{RGB} (RGB)	20	60	25	30	52
L_{GS} (Grayscale)	20	60	25	30	52
L_{DES} (Desaturate)	21	60	24	24	44
L_{R-C} (R channel)	40	79	30	35	41
L_{G-C} (G channel)	23	64	34	41	65
L_{B-C} (B channel)	17	56	30	24	61

Table 2 Table values of second 5 separated colours from first photography

Photography 1 – second 5 separated colours					
RGB colour code	#d4aa94	#914f51	#9d9789	#6a625f	#d2624a
R	212	145	157	106	210
G	170	79	151	98	98
B	148	81	137	95	74
L_{RGB} (RGB)	73	42	63	42	56
L_{GS} (Grayscale)	73	42	63	42	56
L_{DES} (Desaturate)	73	47	61	42	59
L_{R-C} (R channel)	88	68	71	54	88
L_{G-C} (G channel)	76	43	70	51	50
L_{B-C} (B channel)	68	44	65	50	40

Table 3 Table values of first 5 separated colours from second photography

Photography 2 – first 5 separated colours					
RGB colour code	#73abe4	#3c2e2b	#525e4a	#c0a9a3	#3a5280
R	115	60	82	192	58
G	171	46	94	169	82
B	228	43	74	163	128
L_{RGB} (RGB)	68	20	38	71	35
L_{GS} (Grayscale)	68	20	38	71	35
L_{DES} (Desaturate)	70	21	36	72	39
L_{R-C} (R channel)	57	34	44	82	33
L_{G-C} (G channel)	76	27	49	75	44
L_{B-C} (B channel)	93	25	41	74	62

Table 4 Table values of second 5 separated colours from second photography

Photography 2 – second 5 separated colours					
RGB colour code	#9ec5e2	#857671	#1e2e0a	#5a4b44	#272935
R	158	133	30	90	39
G	197	118	46	75	41
B	226	113	10	68	53
L_{RGB} (RGB)	77	51	17	33	17
L_{GS} (Grayscale)	77	51	17	33	17
L_{DES} (Desaturate)	78	52	10	34	19
L_{R-C} (R channel)	72	64	18	47	23
L_{G-C} (G channel)	84	58	27	41	24
L_{B-C} (B channel)	92	56	3	38	31



Figure 7 Results of RGB conversion to black and white with the emphasized value of the L Grayscale parameter (L_{GS})

Table 5 Table values of first 5 separated colours from third photography

Photography 3 – first 5 separated colours					
RGB colour code	#352c25	#283112	#8c917d	#252a16	#4e632b
R	53	40	140	37	78
G	44	49	145	42	99
B	37	18	125	22	43
L_{RGB} (RGB)	19	19	59	16	39
L_{GS} (Grayscale)	19	19	59	16	39
L_{DES} (Desaturate)	18	13	56	12	30
L_{R-C} (R channel)	31	24	66	22	42
L_{G-C} (G channel)	26	29	68	25	51
L_{B-C} (B channel)	22	9	61	12	25

Table 6 Table values of second 5 separated colours from third photography

Photography 3 – first 5 separated colours					
RGB colour code	#70724d	#b1a28f	#301814	#355f5d	#d8a995
R	112	177	48	53	216
G	114	162	24	95	169
B	77	143	20	93	149
L_{RGB} (RGB)	47	68	12	37	73
L_{GS} (Grayscale)	47	68	12	37	73
L_{DES} (Desaturate)	40	66	13	31	74
L_{R-C} (R channel)	56	78	28	31	89
L_{G-C} (G channel)	57	73	14	49	75
L_{B-C} (B channel)	42	67	11	49	69

4 DISCUSSION

In all tables, the values where the ΔL parameter exceeds the value of 10 are marked with orange text.

Tab. 1 and Tab. 2 refer to the first photograph which contains many red tones due to the portrait photography. The deviation of the L parameter becomes visible exactly when the image is converted to black and white by extracting the R channel (red channel). In Fig. 1 we can see the colours that were separated, namely the first (#49271d), second (#b4866f), sixth (#d4aa94), seventh (#914f51), ninth (#6a625f) and tenth (#d2624a) colour. If we look at the separate colours listed in Fig. 1, we can see that they all contain a certain amount of red colour, i.e. if we look at the R value in the listed tables, it is higher than the G and B value in the table for all colours.

Also, the increased values of the L parameter in Tab. 1 for the fourth (#3d4b28) and fifth (#4a887d) separated colours containing green show that the value of the L parameter increases exactly when the image is converted to black and white using the G-channel (green channel) extraction method. If we look at the G value in Tab. 1 for these colours, we see that it is higher compared to their R and B values in the original photographic record, which means that they contain mainly green colours in their composition.

In the case of the third (#333d34) and eighth colours (#9d9789), there was no significant change in the L value in the different conversions of the image to black and white, and when the R, G and B values of these colours are examined in Tab. 1 and Tab. 2, it can be concluded that neither colour dominates over the others, so it is assumed that there was therefore no significant change in the L parameter.

In Tab. 3 and Tab. 4, we can analyse the increase in the L value for the separated colours of the second photography, which contains many shades of blue. We can see the increase in the L parameter in Tab. 3 for the first (#73abe4) and fifth

colour (#3a5280) and in Tab. 4 for the sixth (#9ec5e2) and tenth colour (#272935). All the colours listed have a more dominant B value compared to the R and G values of the original photographic record, as can be seen in Tab. 3 and Tab. 4, confirming that they have mainly blue colours in their composition. Similarly, the L value is significantly higher when the image is converted to black and white using the B channel (blue channel) extraction method.

In Tab. 3, the second (#3c2e2b) and fourth colours (#c0a9a3) and in Tab. 4, the seventh (#857671) and ninth colours (#5a4b44) show an increase in the L parameter when they are converted from a multicolour to a black and white image using the R-channel extraction method. If we check the amount of R colour in Tab. 3 and Tab. 4, we can see that the R value of these colours is higher than the G and B values of the original photographic record of the image.

In Tab. 3, the third colour (#525e4a) has an increased L -parameter when the image is converted from multicolour to black and white using the G-channel extraction method. Looking at the original record of the image of this colour in Tab. 3, we can see that the G value is higher compared to the R and B values of the colour.

Tab. 4 shows that the eighth colour (#1e2e0a), when converted from a multicolour to a black and white image using the B-channel extraction conversion method, has an L -value up to five times lower than the L -value of the original photographic image record.

If we look at the R, G and B values of the original record in Tab. 4, we see that the B value is three times smaller than the R value and four times smaller than the G value.

In Tab. 5 and Tab. 6 we can analyse the behaviour of the L values for the colours separated from the third photo. The separated colours can be seen in Fig. 3 and are mainly dominated by a green tone. It is interesting to see that only the fifth colour (#4e632b) in Tab. 5 shows an increase in the L parameter when a multicolour image is converted to black and white using the G-channel extraction method. If we examine the R, G and B values of the original image record for the given colour, we can see that the G value is the highest.

In Tab. 6, for the ninth colour (#355f5d), we see an increase in the L parameter when converting a multicolour image to black and white using the G-channel extraction method and the B-channel extraction method. If we examine the R, G and B values of the original image record for the given colour, we can see that the G and B values are very similar, and almost double compared to the R value.

In Tab. 5, the first colour (#352c25) and in Tab. 6, the eighth colour (#301814) and the tenth colour (#d8a995) show a visible increase in the L parameter when they are converted from a multicolour image to black and white using the R-channel extraction method. When we examine the R, G and B values in Tab. 5 and Tab. 6 for the original image record for the specified colours, we can see that the R value is higher compared to the G and B values.

In Tab. 5, when converting an image from a multicolour to a black and white record using the B-channel extraction method, the second colour (#283112) and the fifth colour (#4e632b) have an L -parameter twice as large as the original

multicolour image record. If we examine the R, G and B values of the colours listed in Tab. 5 in the original image record, we can see that the B value is up to two times smaller compared to the R and G values.

It is interesting to examine the behaviour of the L parameter in the standard Grayscale conversion of a multicolour image to black and white, as Fig. 7 shows that in these 30 separate colours, 6 colour pairs were converted to the same grey colour. The colour space that is showed on Fig. 7 before conversion is RGB colour space and on the right-side Fig. 7 shows converted Grayscale colour space and the brightness parameter L_{GS} . We follow the L_{RGB} and L_{GS} parameters, which are the same, so that certain colours that are visually different in the RGB system become the same grey colour in the standard Grayscale conversion to black and white due to the same L parameter.

The RGB colours #1e2e0a and #272935 have the same shade of grey with the L_{GS} value 17 when converted to black and white using the standard Grayscale conversion. The RGB colours #352c25 and #283112 become the same grey colour with the L_{GS} value 19 when converted to black and white with the standard Grayscale conversion.

The RGB colours #49271d and #3c2e2b become the same grey colour with the L_{GS} value 20 when converted to black and white with the standard Grayscale conversion. The RGB colours #914f51 and #6q625f become the same grey colour with an L_{GS} value of 42 when converted to black and white using the standard Grayscale conversion.

The RGB colours #73abe4 and #b1a28f become the same grey colour with an L_{GS} value of 68 when converted to black and white using the standard Grayscale conversion. The RGB colours #d4aa94 and #d8a995 become the same grey colour with an L_{GS} value of 73 when converted to black and white using the standard Grayscale conversion.

For the colours indicated in Fig. 7, the tables show at which conversion to black and white ΔL exceeds the value of 10. Using these conversions would reduce the loss of contrast when converting to black and white.

Tab. 1 shows for the colour in RGB #49271d that the conversion to black and white by extracting the red channel is the one with ΔL greater than 10, since the L_{RGB} value is 20 and the L_{R-C} value is 40.

In Tab. 2, it can be seen for the colour RGB #914f51 that the conversion to black and white by extracting the red channel is the one with ΔL greater than 10, since the L_{RGB} value is 42 and the L_{R-C} value is 68. In the same table, it can be seen that for the colour in RGB #6a625f, the conversion to black and white by extracting the red channel is the one with ΔL greater than 10 because the L_{RGB} value is 42 and the L_{R-C} value is 54. Tab. 2 also shows that for the colour RGB #d4aa94, the conversion to black and white by extracting the red channel is the one with ΔL greater than 10 because the L_{RGB} value is 73 and the L_{R-C} value is 88.

In Tab. 3, for the colour in RGB #3c2e2b, it can be seen that the conversion to black and white by extracting the red channel is the one with ΔL greater than 10, since the L_{RGB} value is 20 and the L_{R-C} value is 34. In the same table, it can be seen for the colour in RGB #73abe4 that the conversion to black and white by extracting the blue channel is the one with

ΔL greater than 10 because the L_{RGB} value is 68 and the L_{B-C} value is 93.

Tab. 4 shows for the colour in RGB #1e2e0a that the conversion to black and white by extracting the blue channel is the one with ΔL greater than 10 because the L_{RGB} value is 17 and the L_{B-C} value is 3. In the same table, it can be seen for the colour in RGB #272935 that the conversion to black and white by extracting the blue channel is that with ΔL greater than 10 because the L_{RGB} value is 17 and the L_{B-C} value is 31.

Tab. 5 shows for the colour RGB #352c25 that the conversion to black and white by extracting the red channel is the one with ΔL greater than 10, since the L_{RGB} value is 19 and the L_{R-C} value is 31. It can also be seen in Tab. 5 for the colour in RGB #283112 that the conversion to black and white by extracting the blue channel is the one with ΔL equal to 10 because the L_{RGB} value is 19 and the L_{B-C} value is 9, which is closest to the desired ΔL for this colour deviation.

In Tab. 6, for the colour in RGB #b1a28f, it can be seen that the conversion to black and white by extracting the red channel is the one where ΔL is equal to 10, since the L_{RGB} value is 68 and the L_{R-C} value is 78. In the same table, it can be seen for the colour RGB #d8a995 that the conversion to black and white by extracting the red channel is the one where ΔL is greater than 10 because the L_{RGB} value is 73 and the L_{R-C} value is 89.

Considering the above, it would be optimal for photographs such as the first captured photograph shown in Fig. 1, which contain many red tones, to be converted to black and white using the red channel extraction method in order to increase the contrast and obtain the widest possible tonal range in the converted black and white display.

Photographs such as the second captured photograph shown in Fig. 2, which contain many blue tones, should be converted to black and white using blue channel extraction.

Photographs that contain many green tones, such as the third captured photograph shown in Fig. 3, can be optimally converted to black and white by also extracting the red channel in order to obtain a contrast that corresponds to that of the original captured record.

5 CONCLUSION

Records of digital photography captured in the RGB system can be converted to black and white in various ways. One of the most commonly used methods for converting to black and white is the standard Grayscale conversion in Adobe Photoshop. As can be seen from the research conducted, the L_{RGB} value, which represents the L parameter for brightness in the RGB colour system, is equal to the L_{GS} value, which represents the L parameter of the converted standard Grayscale record. This confirms that in the standard Grayscale image conversion, the colour system is greatly reduced because some colours that are visually different in the RGB system become the same shade of grey in the converted standard Grayscale display because the L_{RGB} and L_{GS} parameters are the same.

Monitoring the L parameter in other conversions of multicolour digital photos to black and white, this research

has shown that other methods of converting multicolour photos to black and white are more suitable than the standard Grayscale conversion if there are certain subjects in the photo that are shown by a predominant amount of one of the primary colours of the RGB system. The ΔL value, which must be greater than 10, can be used to determine which conversion is optimal for certain photos.

For photos that contain a lot of red tones in the RGB system, such as photos of human portraits or nudes, it is optimal to convert them from a multicolour system to black and white using the red channel extraction method. The L_{R-C} parameter fulfils the condition that ΔL is greater than 10 compared to the L_{RGB} parameter.

For photos whose subject is the sea, for example, and which contain many blue tones, the optimum conversion from the multicolour RGB system to black and white would be the method of extracting the blue channel, as the L_{B-C} parameter fulfils the condition that ΔL is greater than 10 compared to the L_{RGB} parameter.

Photos of nature taken in the RGB colour system contain a large amount of green tones. It is optimal to convert them to black and white by extracting the red channel, because ΔL is greater than 10 when we compare the L_{R-C} and L_{RGB} parameters.

By using the optimal conversion method of different subjects with different predominant primary colours in the RGB system, the loss of contrast and colour space in the converted black and white system would be reduced, leaving the converted black and white record more similar to the original multicolour photographic record.

6 REFERENCES

[1] Togban, E, Kerouh, F., & Ziou, D. (2023). Identifying color space for improved image display. *Twelfth International Conference on Image Processing Theory, Tools and Applications (IPTA)*, Paris, France, 1-6. <https://doi.org/10.1109/IPTA59101.2023.10320079>

[2] Kao, W. C., Hong, K. D., & Hsu, C. Y. (2023). Color Reproduction of Multi-Pigment Color Electronic Papers. *The IEEE International Conference on Consumer Electronics (ICCE2023)*, Las Vegas, NV, USA, 1-2. <https://doi.org/10.1109/ICCE56470.2023.10043485>.

[3] Vezina, M., Ziou, D., & Kerouh, F. (2015). Color Space Identification for Image Display. In: Kamel, M., Campilho, A. (eds) *Image Analysis and Recognition. ICIAR 2015. Lecture Notes in Computer Science, vol 9164*. Springer, Cham. https://doi.org/10.1007/978-3-319-20801-5_51

[4] Saravanan, C. (2010). Color Image to Grayscale Image Conversion, *Second IEEE International Conference on Computer Engineering and Applications*. <https://doi.org/10.1109/ICCEA.2010.192>

[5] Vishali, G., Meghana, H., Kumar M. A., & Rajesha, N. (2019). An Investigation on Image Conversions and Edge Detection Patterns in Image Processing. *International Conference on Emerging Trends in Science and Engineering (ICESE)*, Hyderabad, India, 1-8. <https://doi.org/10.1109/ICESE46178.2019.9194692>

[6] Morović, J., & Luo, M. R. (2000). Calculating medium and image gamut boundaries for gamut mapping. *Color Research & Application*, 25(6), 394-401.

[https://doi.org/10.1002/1520-6378\(200012\)25:6%3C394::AID-COL3%3E3.0.CO;2-Y](https://doi.org/10.1002/1520-6378(200012)25:6%3C394::AID-COL3%3E3.0.CO;2-Y)

[7] L. de Queiroz R., & Braun K. (2006). Color to Gray and Back: Color Embedding into Textured Gray Images, *IEEE Transactions on Image Processing*, 15(6), 1464-1470. <https://doi.org/10.1109/tip.2006.871181>

[8] Bala, R., & Eschbach, R. (2004). Spatial color-to-grayscale transform preserving chrominance edge information. In *Proc. IS&T 12th Color and Imaging Conf.*, 82-86. <https://doi.org/10.2352/CIC.2004.12.1.art00016>

[9] Mikota, M., Pavlović, I., & Matijević, M. (2015). Effect of the changes in the RGB digital image channel on the perception of fashion photography while retaining its iconicity. *Tekstil*, 64(1-2), 13-18.

[10] Čutić, D., & Mikota, M. (2019) Comparison of separate channels of RGB system converted to grayscale in nude photography. *Proceedings of 23rd international conference on printing, design and graphic communications Blaž Baromić 2019*, 128-136

[11] Ganchovska, V., & Krasteva, I. (2022). Converting color to grayscale image using LabVIEW. *International Conference Automatics and Informatics (ICAI2022)*, Varna, Bulgaria, 320-323. <https://doi.org/10.1109/ICAI55857.2022.9960062>

[12] Erhu, Z., Jia, Y., Yajun, C., & Yang, Y., (2010). A study of image color quality evaluation based on S-CIELAB. *The 3rd International Congress on Image and Signal Processing*. <https://doi.org/10.1109/CISP.2010.5646866>

[13] Kumar, E. B., & Thiagarasu, V. (2017). Color channel extraction in RGB images for segmentation. *The 2nd International Conference on Communication and Electronics Systems (ICCES2017)*, Coimbatore, India, 234-239. <https://doi.org/10.1109/CESYS.2017.8321272>.

[14] Chang Y., & Mukai, N. (2022). Color Feature Based Dominant Color Extraction. *IEEE Access*, 10, 93055-93061. <https://doi.org/10.1109/ACCESS.2022.3202632>

[15] Čutić, D., Mikota, M., & Žeželj, T. (2019). Zacrnljenja grayscale zapisa konvertiranog rgb zapisa fotografske slike. *PRINTING&DESIGN2019*, Zagreb: FotoSoft (FS), 128-135. (in Croatian)

Authors' contacts:

Darija Čutić, senior lecturer
University of Applied Sciences Zagreb,
Vrbik 8, 10000 Zagreb, Croatia
dcutic@tvz.hr

Miroslav Mikota, PhD, associate professor
(Corresponding author)
University of Zagreb Faculty of Graphic Arts,
Getaldićeva ulica 2, 10000 Zagreb, Croatia
miroslav.mikota@grf.unizg.hr

Maja Turčić, PhD, higher lecturer
University of Applied Sciences Zagreb,
Vrbik 8, 10000 Zagreb, Croatia
mturcic@tvz.hr

Determination of Thermal Characteristics of Fe-C Cast Iron Using Hot Disk Method

Dalibor Viderščak*, Zdravko Schauperl, Sanja Šolić, Bojan Milovanović, Ivan Stanković, Mario Šercer

Abstract: This paper explores the influence of graphite shape, type, and dimensions on the thermal characteristics of Fe-C cast iron. Test samples were prepared from blocks of grey and nodular cast iron, specifically examining castings with D and A types of graphite as well as nodular graphite in a ferrite matrix. To analyze the effect of surface texture, sample surfaces were prepared with varying parameters of surface roughness, achieving values of 0.5 μm for polished and 12.6 μm for unpolished surfaces. Thermal characteristics, including thermal conductivity and effusivity, were evaluated using unsteady hot disk methods, with water as the contact agent, to simulate practical thermal conditions. Additionally, the microstructure of each sample was analyzed using optical microscopy. Results indicate that graphite content, type, shape, and surface roughness collectively have the most significant impact on the thermal properties of cast iron. This study's findings provide valuable insights into optimizing cast iron for applications requiring efficient thermal management and highlight the importance of graphite morphology and surface finish in enhancing thermal performance.

Keywords: graphite; grey cast iron; hot disk method; nodular cast iron; thermal conductivity; thermal effusivity

1 INTRODUCTION

According to recent data, the global production of grey cast iron reached approximately 31 million tons in 2023, and it is projected to increase to 41 million tons by 2034, with an average annual growth rate (CAGR) of 2.5%. This growth is attributed to its extensive use in the automotive, construction, and industrial consumer goods sectors [1].

Regarding nodular (ductile) cast iron, global production continues to expand due to its widespread application in industries such as automotive and heavy machinery [1, 2]. In Europe, for example, in 2021, nodular cast iron accounted for approximately 43.5% of the total production of ferrous castings, while grey cast iron represented 49.3% (CAEF - The European Foundry Association and Future Market Insights). These figures indicate a continuous increase in the production of both types of cast iron, underscoring their critical roles in various industrial sectors and infrastructure projects [3, 4].

Thermal and mechanical properties of Fe-C cast iron primarily depend on its structure, concretely amount and the form of graphite [5]. The quantity of extracted graphite within the matrix of Fe-C cast iron directly depends on the equivalent of graphite (CE): $CE = \%C + 1/3 \times (\%Si + \%P)$ [4]. This indicates that the higher value of the coefficient of CE generally results in a higher amount of extracted graphite. At the same time, there will be an increase of values of thermal characteristics, while a decrease of values of mechanical characteristics [6, 7]. Besides chemical composition, type of extracted graphite mostly depends on the method of cast iron production [8]. Based on the reasons above ASTM suggests a standard for the classification of graphite according to ASTM A247-47 and 67 and upgrades it to the International Standards Organization ISO Recommendation 9, according to which graphite differs according to its shape, size and layout of graphite structure [9–11]. Because there are many kinds of grey and nodular casts iron which differ chemically, by phases in the matrix (ferrite, perlite, bainite, cementite etc.) and their quantity content, according to the shape and dimensions of graphite,

and likewise, it is important to note that there are no standard samples of Fe-C cast iron [4]. For this reason, this paper shows the results of the characterization according to the structure, chemical composition and thermal characteristics of these types of cast iron, as well as a comparison of the thermal characteristics of the sample, depending on the roughness of the surface tested by the hot disk method [4, 8].

Recent studies have emphasized the importance of these thermal and mechanical properties in various high-stress applications, such as brake components, which are subject to significant thermal loads. Dubale's (2019) doctoral research, for instance, analyzed the thermo-mechanical behavior of disc brake rotors, highlighting the critical role of thermal conductivity and structural resilience in components exposed to friction-induced heating. Similarly [12] Venugopal et al. (2023) examined grey cast iron under electrical discharge machining (EDM) processes, emphasizing the relationship between thermal management and wear resistance in high-temperature environments. These findings underscore the need to accurately characterize cast iron's thermal properties to predict its behavior in thermally demanding applications.

Moreover, the tribological properties of Fe-C cast iron also play a significant role in component durability, as highlighted by [13] Skulić & Bukvić (2016), who studied the impact of surface roughness on piston-cylinder sets in internal combustion engines. Their findings demonstrate how surface roughness and graphite morphology influence both thermal conductivity and wear resistance in mechanical assemblies. This underscores the necessity of assessing thermal behavior in cast iron with varying graphite shapes and roughness levels to optimize material performance in practical applications.

Pretested samples of grey cast iron containing A and D type flake (lamellar) graphite in a ferritic matrix showed that they give cast iron a fragile structure of low ductility, relatively low tensile strength and high compressive strength and relatively high thermal conductivity. In pretested samples of nodular cast iron containing graphite in the form of nodules in a ferrite matrix it is shown that they give cast iron a relatively tough structure with good mechanical

characteristics and lower thermal conductivity compared to the grey cast iron with type A and D graphite [11, 14-16]. It is important to note that preliminary tests show that reducing the amount of graphite in grey cast iron increases the tensile strength and fatigue strength, respectively the values of mechanical properties [5, 15], while the values of thermal characteristics decrease [18].

Grey and nodular graphite iron are primarily used in at manufacture of brake discs, heating plates on ovens, molds for casting glass, gears, motor shafts, cover sliding bearings, pump impellers, pistons and gear motors, etc. [2, 12]. As most of these elements in practice are extremely thermally loaded, it is important to know the thermal conductivity of grey and nodular cast iron castings to be able to predict the thermodynamic processes in the machine elements [10, 17].

Given the diverse types of grey and nodular cast iron, each with unique chemical compositions, matrix phases (e.g., ferrite, pearlite, bainite), and graphite configurations, this study aims to provide a comprehensive characterization of these materials. Specifically, it examines the thermal properties of cast iron samples with A and D type graphite in a ferritic matrix, as well as nodular cast iron with graphite nodules. By investigating surface roughness through hot disk testing, this study evaluates how surface texture and graphite morphology impact thermal conductivity. The findings from these analyses are essential for designing cast iron components in high-temperature applications, such as brake discs, heating elements, and engine parts, where understanding thermodynamic processes is crucial.

2 MATERIALS AND METHODS

In the experimental part of the study three samples of Fe-C cast iron of different surface roughness with different forms of graphite were tested, all three in the ferrite matrix, Fig. 1.

Sample 9D is a type of grey cast iron with a type D flake graphite in a ferritic matrix, and an average volume fraction of 8.90% of graphite in the alloy in the solid state. Sample 1A is grey cast iron with a type A flake graphite in ferritic matrix and an average volume fraction of 10.50% graphite within the alloy in the solid state. Type A graphite, with a uniform appearance and shape within a grey cast iron, occurs because of moderate cooling, while type D graphite occurs because of deeper subcooling during the production process [8, 14, 18]. In laboratory research, it was obtained that these two samples of grey cast iron with different types of graphite have the same chemical composition and hardness but have different appearance and shape of graphite sheets which is achieved by controlled cooling of samples during the casting process. Sample 13 is a nodular cast iron with graphite nodular type ferrite in the matrix, and the average volume fraction of 6.70% in the alloy in the solid state. Nodular graphite inside cast iron forms its nodular form in melt during the manufacturing process and is characterized by that it is of proper spherical shape [1, 17, 19]. A quantity content of graphite or a volume percentage of graphite within the three Fe-C casts is determined according to ASTM E 562-11 [22].

Brinell hardness value of 202 HBW10/3000 was obtained for samples 9D and 1A, and for sample 13 the value of obtained hardness was 152 HBW10/3000, both according to HRN EN ISO 6506-1:2014 [23].

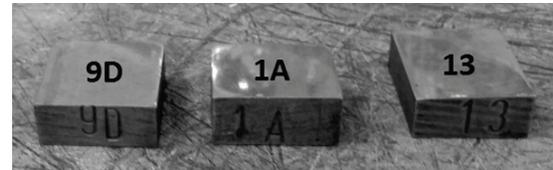


Figure 1 Samples for testing – grey cast iron with D (sample 9D) and A (sample 1A) type flake (lamellar) graphite, and nodular cast iron with graphite in a form of nodules (sample 13)

Table 1 Percentage of chemical elements in three Fe-C test samples

Grey cast iron with D type flake graphite – sample 9D (%)										
C	Si	Mn	P	S	Cr	Ni	Mo	V	Ti	Fe
3.21	2.41	0.59	0.162	0.044	0.00	0.10	0.02	0.06	0.01	Rest.
Grey cast iron with A type flake graphite – sample 1A (%)										
C	Si	Mn	P	S	Cr	Ni	Mo	V	Ti	Fe
3.21	2.41	0.59	0.162	0.044	0.00	0.10	0.02	0.06	0.01	Rest.
Nodular cast iron with graphite in the form of nodules – sample 13 (%)										
C	Si	Mn	P	S	Cr	Ni	Mo	V	Ti	Fe
3.55	2.50	0.66	0.051	0.007	0.003	0.003	0.008	0.03	0.126	Rest.

Since there have been two test surfaces with different profile roughness parameters produced in each group of samples, it was necessary to determine their different profile roughness parameters. The measurement has been carried out by contact method in electronic - mechanical device stylus Perthometer C5D. The Perthometer C5D specializes in measuring surface roughness parameters, such as R_a (average roughness) and R_z (mean peak-to-valley height), with high precision, typically $\pm 0.1 \mu\text{m}$. However, it requires stable conditions as environmental vibrations and reflective surfaces can affect accuracy. Different profile roughness parameters for the non-polished side of samples is $R_a - 12.6$

$\pm 0.1 \mu\text{m}$, while for the polished side of the sample is as follows: $R_a - 0.5 \pm 0.1 \mu\text{m}$. Roughness levels were consistent across all tested samples. This consistency in roughness values was essential for accurately assessing the impact of surface finish on thermal conductivity and effusivity across different graphite-containing cast iron samples. Broad range of surface roughness values was chosen ($0.5 \mu\text{m}$ for polished and $12.6 \mu\text{m}$ for unpolished) to comprehensively investigate the impact of surface finish on thermal characteristics. This range allows a detailed analysis of impact surface roughness affect thermal conductivity and effusivity, offering insights

applicable to real-world applications where surface roughness varies.

The chemical composition of all three samples is determined by using a Leco GDS500 A that is based on the method of Glow Discharge Atomic Emission Spectrometry (AES) [24], and the percentage of resulting chemical elements of all three casts is shown in Tab. 1.

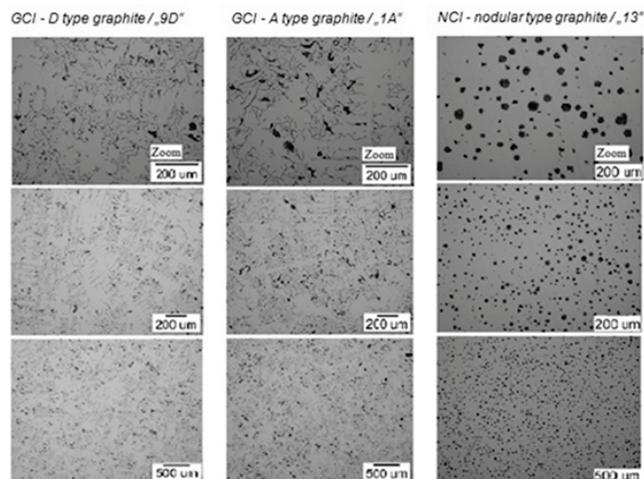


Figure 2 Microstructures of test samples; light microscope

2.1 Microstructure Analysis

A microstructure of the samples was analyzed. The samples were prepared in such a way that they were ground, polished and pitted. Grinding was conducted with abrasive grit paper P 4000, polishing was conducted by diamond paste, while the etching was performed in Nital, a 3% solution of nitric acid in ethyl alcohol. Analysis of the microstructure was conducted on a light microscope Olympus GX51 Analyzer. The Olympus GX51 Analyzer is a metallurgical microscope offering magnification up to 1000 \times , designed for detailed examination of material microstructures, including grain size and phase distribution. Although highly capable, it requires precise calibration and smooth sample surfaces for accurate analysis, with limitations on rough or highly reflective samples. The resulting microstructures of the three test samples are shown in Fig. 2.

2.2 Testing Method

Testing was carried out by a device Mathis TCi. The Mathis TCi instrument measures thermal properties, specifically thermal conductivity, diffusivity, and effusivity, using the modified transient plane source (MTPS) technique. It operates with $\pm 5\%$ accuracy across a range of 0.03 to 30 W/mK, suitable for diverse materials from insulators to conductors. Accurate results with the Mathis TCi depend on smooth sample contact, as inconsistencies in surface quality may impact measurements. Together, these devices provide a robust profile of surface, structural, and thermal properties essential for in-depth material analysis. This kind of device is based on a non-stationary method of testing thermal performance of a material which is referred to as a hot disk

method. The device consists of a sensor, a control unit and a computer program that manages and collects data measurements [24]. This type of device directly measures thermal conductivity and thermal effusivity as well as thermal diffusivity and heat capacity [25]. The method is based on passing a one-dimensional heat flow from the sensor to the sample so that the sensor is located within several concentrically wrapped rings, spirals, or 10-micron thick wires made of platinum, which are a constant source of heat transferred to the upper glass surface of the sensor, and through which a one-dimensional heat flow travels towards the sample [26].

The spiral is made of platinum because platinum is a material that has the feature of excellent linear stretching and contraction at a wide range of temperatures. The main components of the device are shown in Fig. 3.

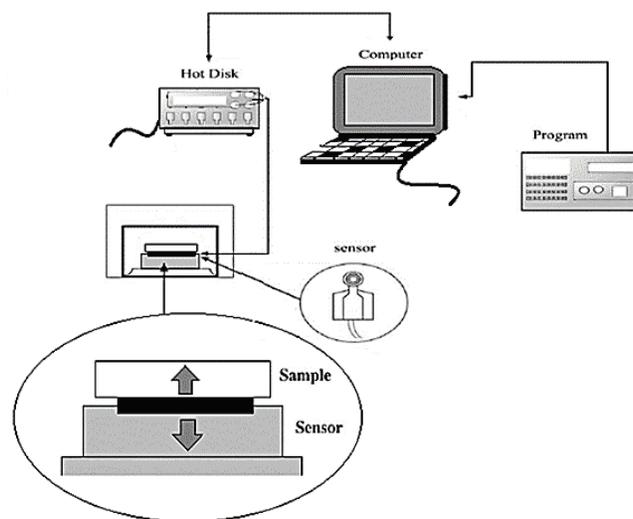


Figure 3 Components of the device and a diagram of the device for testing thermal characteristics based on the hot disk method

The device assumes a one-dimensional heat flow in the direction of the sensor towards the sample, while at the same time ignoring the thermal resistance of the touch sensor and the sample. With that in mind, a change of contact surface temperature sensor ($x = 0$) is given by the Eq. (1):

$$\Delta T(x = 0, t) = \frac{1.1284 \cdot G \cdot \sqrt{t}}{e_1 + e_2}, \quad (1)$$

where ΔT ($^{\circ}\text{C}$) is a temperature change of contact temperature, G (W/m^2) a density of heat flow that comes into the sensor, t (sec) time measured from the beginning of the process, e_1 ($\text{W}\sqrt{\text{s}/\text{m}^2\text{K}}$) is equivalent effusivity of the sensor and e_2 ($\text{W}\sqrt{\text{s}/\text{m}^2\text{K}}$) is effusivity of the tested sample [26, 25]. Value the change of the contact temperature to equivalent voltage change is converted by the temperature coefficient of resistance TCR. The relation between the sensor resistance and its temperature is determined by the Eq. (2):

$$R(T) = R_0(1 + \alpha \cdot T) = R_0 + \Delta T, \quad (2)$$

where $R(T)$ of the sensor resistance at the temperature T , the sensor resistance R_0 at 0°C , the sensor temperature T ($^\circ\text{C}$), α the temperature coefficient of resistance (TCR) and A slope of correlation [25, 26]. Applying Ohm's Law to connect Eq. (1) and Eq. (2) yields Eq. (3).

$$\Delta V(t) = \frac{1.1284 \cdot I \cdot A \cdot G \cdot \sqrt{t}}{e_1 + e_2} \quad (3)$$

Then, the Eq. (3) can be written as Eq. (4):

$$\Delta V(t) = m \cdot \sqrt{t}, \quad (4)$$

where m is the slope of the line that describes the change in voltage with respect to the root of time, according to the value of the Eq. (5):

$$m = \frac{1.1284 \cdot I \cdot A \cdot G}{e_1 + e_2} \quad (5)$$

The assumption is that in the short period, all parameters on the right side of the Eq. (5) are constant, then the heat effusivity e_2 would be calculated according to the Eq. (6):

$$\frac{1}{m} = M_2 \cdot e_2^2 + M_1 \cdot e_2 + C, \quad (6)$$

where M_1 and M_2 are sensor constants that depend on a type of testing sample and C is a characteristic of the sensor dependent on the temperature T during the testing, the Eq. (7).

$$C = C_0 + C_1 \cdot T + C_2 \cdot T^2, \quad (7)$$

where C_0 , C_1 and C_2 are constants of a sensor [25, 26].

Measurement of thermal conductivity λ is based on data collected for obtaining heat effusivity, through the algorithm m^* for calibration and calculation of thermal conductivity. The device uses an iterative process to calibrate the sensor, based on known values of coefficients of thermal conductivity. Thermal conductivity is calculated according to the Eq. (8):

$$\frac{1}{m - m^*} = \alpha \cdot \lambda + z, \quad (8)$$

where α is the slope, and z cutting values on the calibration curve [25, 26]. The formulas used in this paper are derived from a combination of established sources and original derivations tailored to the specific experimental setup. Formulas related to thermal conductivity and effusivity are based on the Mathis TCi Thermal Conductivity Instrument User Manual by Mathis Instruments, Inc., which details the MTPS (Modified Transient Plane Source) method for thermal property measurements. Equations involving

thermal resistance and voltage-based calculations are grounded in the principles found in Introduction to Heat Transfer by Incropera and DeWitt, covering fundamental concepts of heat transfer and thermal resistance. Additionally, general equations for thermal conductivity in alloys, particularly Fe-C alloys, are referenced from Materials Science and Engineering: An Introduction by Callister and Rethwisch. Any formulas unique to this study have been specifically derived by the authors to meet the experimental conditions and are indicated as original contributions in the paper.

2.3 Research

In addition to the influence of graphite, the paper also explored the influence of surface roughness on the sample's thermal characteristics.

Table 2 Characteristics of test Fe-C samples

Characteristics	Grey cast iron		Nodular cast iron
	9D	1A	13
Type of graphite	D graphite	A graphite	Nodules
Percentage of graphite (%)	8.90	10.50	6.70
Dimensions of sample (mm)	10.98/27.87/26.05	12.32/27.83/23.41	10.72/30.89/25.87
Weight of sample (kg)	0.055	0.056	0.059
Density of sample (kg/m ³)	6838.14	6959.12	6.892.52
Ambient temperature ($^\circ\text{C}$)	23 \pm 2		
Coupling agent	Water		

Three measurements was carried out for each surface sample, each measurement had 10 readings, with each reading lasting 60 seconds. It takes from 0,8 to 5 seconds for the reading and the rest of the 60 seconds during the interval readings serve for a complete cool down of the sample [25]. The test was carried out with water as a coupling agent at a room temperature of 23 \pm 2 $^\circ\text{C}$ - water is used for testing in temperature ranges from 5 $^\circ\text{C}$ to 70 $^\circ\text{C}$ [24, 26]. The main characteristics of Fe-C test samples are shown in Tab. 2.

3 RESULTS AND DISCUSSION

Tab. 3 shows the results of mean values of specific heat capacity and thermal diffusivity of test Fe-C samples depending on the surface roughness and the type of graphite.

Table 3 Results of mean values of specific heat capacity and thermal diffusivity of test Fe-C samples

Fe-C samples	Specific heat capacity (J/kg/K)	Thermal diffusivity (m ² /s)
Sample 9D - polished side ($Ra = 0.5 \mu\text{m}$)	166.40	1.25×10^{-4}
Sample 9D - non polished side ($Ra = 12.6 \mu\text{m}$)	354.96	1.43×10^{-5}
Sample 1A - polished side ($Ra = 0.4 \mu\text{m}$)	121.72	2.54×10^{-4}
Sample 1A - non polished side ($Ra = 12.4 \mu\text{m}$)	168.80	1.17×10^{-4}
Sample 13 - polished side ($Ra = 0.6 \mu\text{m}$)	327.68	1.89×10^{-5}
Sample 13 - non polished side ($Ra = 12.5 \mu\text{m}$)	378.22	9.94×10^{-6}

Fig. 4 and Fig. 5 show graphs of the results of the thermal conductivity and heat effusivity of the Fe-C samples depending on the surface roughness and type of graphite. Graphs show the minimum, maximum and mean value for each test sample depending on the surface roughness, and its comparison.

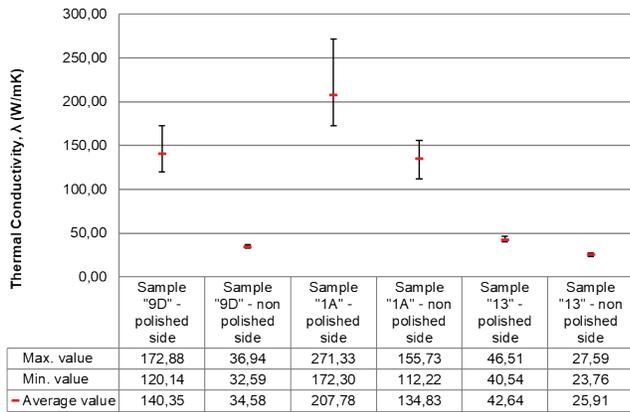


Figure 4 Results of mean values of thermal conductivity of test Fe-C samples with their minimum and maximum values

Since pure graphite is an anisotropic element having different thermal conductivities depending on the orientation of the crystal lattice, a large deviation of the results is expected [27]. Therefore, it is necessary to analyze the value of standard deviations and coefficients of variation of test samples obtained based on the results of thermal conductivities and thermal effusivities, shown in Tab. 4.

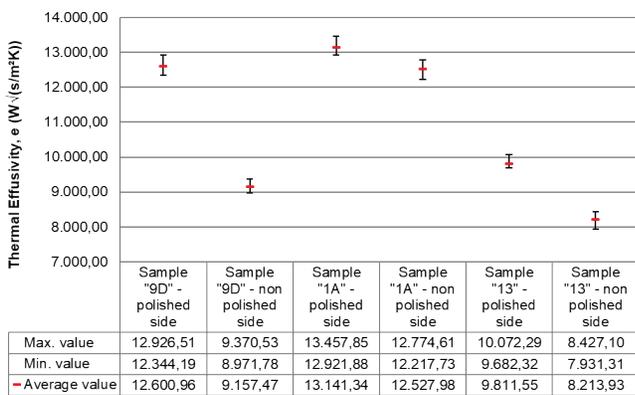


Figure 5 Results of mean values of thermal effusivity of test Fe-C samples with their minimum and maximum values

Obtained values of standard deviation and coefficient of variation concerning the tested side of the sample are relatively small, which indicates that the average deviation of individual values from the mean value is relatively small, and it shows relatively high uniformity of the results. Tab. 4 shows that sample 13 obtained at least an average deviation of the individual values of the arithmetic mean, indicating a relatively high uniformity of the results. A bit larger average deviation of the individual values from the mean was obtained at sample 9D, but also shows a relatively high uniformity of the obtained results. The largest average deviation of the individual values of arithmetic mean was

obtained for sample 1A which tells us that there is a relatively small uniformity of the results obtained, which of course is not desirable. Also, in Tab. 4, a smaller average deviation of the individual values from the mean was obtained with non-polished samples compared to polished samples.

Table 4 Measures of dispersion mean values of thermal conductivity and thermal effusivity of test Fe-C samples

Measures of dispersion mean values of thermal conductivity of Fe-C test samples			
Fe-C samples		Standard deviation, σ	Coefficient of variation, V (%)
Sample 9D	Polished side ($Ra = 0.5 \mu\text{m}$)	12.88	9.19
	Non polished side ($Ra = 12.6 \mu\text{m}$)	0.77	2.19
Sample 1A	Polished side ($Ra = 0.4 \mu\text{m}$)	27.33	12.77
	Non polished side ($Ra = 12.4 \mu\text{m}$)	10.87	7.88
Sample 13	Polished side ($Ra = 0.6 \mu\text{m}$)	1.18	2.75
	Non polished side ($Ra = 12.5 \mu\text{m}$)	0.44	1.68

Measures of dispersion mean values of thermal effusivity of Fe-C test samples			
Fe-C samples		Standard deviation, σ	Coefficient of variation, V (%)
Sample 9D	Polished side ($Ra = 0.5 \mu\text{m}$)	145.80	1.16
	Non polished side ($Ra = 12.6 \mu\text{m}$)	69.95	0.76
Sample 1A	Polished side ($Ra = 0.4 \mu\text{m}$)	149.72	1.14
	Non polished side ($Ra = 12.4 \mu\text{m}$)	128.49	1.02
Sample 13	Polished side ($Ra = 0.6 \mu\text{m}$)	82.78	0.84
	Non polished side ($Ra = 12.5 \mu\text{m}$)	55.95	0.68

The results show a significant difference in thermal conductivity and effusivity depending on the amount, type and distribution of graphite particles and the surface roughness. The greater thermal conductivity of polished samples can be explained by the fact that at non polished side of the sample there is a better chance for at presence of residual air in microcavities at the contact of the sensors and the sample. Because of that, it has an impact on a decrease of read value of thermal conductivity due to the low value of thermal conductivity of air at room temperature, which is $\lambda_{\text{AIR}} = 0,026 \text{ W/mK}$, in relation to the resulting thermal conductivities of casts [1, 28].

From the obtained results of measurements, it has been shown that thermal conductivity and effusivity of materials depend on the type and quantity of graphite inside the cast at the micro level. It can be seen in Figs. 4-5 that the values of thermal conductivity and effusivity are lowest in nodular cast iron with 6.70% graphite in the form of nodules (sample 13), while in grey cast iron with 8.90% D type flake (lamellar) graphite (sample 9D) they are somewhat higher. The greatest values of thermal conductivity and effusivity were obtained by testing a sample of grey cast iron with 10.50% A type flake (lamellar) graphite (sample 1A). The above shows the fact that the thermal conductivity of tested iron cast samples

increases with the amount of graphite in test samples. The same is true for the values of thermal effusivities, specific heat capacities and thermal diffusivities. The thermal conductivity of iron is $\lambda_{\text{IRON}} \approx 70$ to 80 W/mK, while the thermal conductivity of graphite differs whether it is measured parallelly ($\lambda_{\text{GRAPHITE}, \parallel} \approx 25$ to 274 W/mK) or perpendicularly to the basal plane ($\lambda_{\text{GRAPHITE}, \perp} \approx 10$ W/mK) [26, 27]. Hence the conclusion is that thermal conductivity of test casts depends on the amount of graphite and the orientation of the basal plane of graphite relative to the position one-dimensional heat flow sensor device for testing. It tells us that a higher value of thermal conductivity was obtained in samples with a higher amount of graphite where one-dimensional heat flow traveled parallelly to the basal graphite plane, and vice versa. Because of that, in anisotropic materials it is necessary, by hot disk method, to measure thermal conductivity in several orientations relative to the test sample to obtain a more realistic result.

4 CONCLUSION

In this paper, the results of testing thermal conductivity and effusivity are shown on samples of grey cast iron with A and D type flake graphite in a ferritic matrix and nodular cast iron with nodular graphite in a ferritic matrix using the non-stationary method of the hot disk. The method of hot disk is a fast and simple, but sensitive method that is most often used in quality control in production and development of materials.

The test concluded that the value of the thermal conductivity and effusivity of tested samples mostly depends on the roughness of the surface on which the testing is being conducted and, on the form, type, layout, orientation and the amount of graphite in cast iron. The results show a higher value of thermal conductivity and effusivity tested on the polished side of the sample. Also, it has been shown that for the samples with a higher quantity content of graphite and denser layout of graphite granules at the microstructural level higher values of thermal conductivity and effusivity are obtained.

In the end, by statistical measures of dispersion, it is proved that a relatively low deviation of the individual values is obtained from the arithmetic mean of the same, which shows a relatively high consistency of the results obtained by this method of testing.

The results of this study indicate a significant influence of graphite type, shape, and distribution, along with surface roughness, on the thermal conductivity and effusivity of Fe-C cast iron. This lays a strong foundation for future research into optimizing thermal and mechanical characteristics for industrial applications. Expanding this work, further studies could investigate Fe-C cast iron with varied carbon percentages and explore additional graphite morphologies, as these factors could yield new insights into enhancing both thermal management and structural resilience of cast iron components.

Future experiments could also involve testing under diverse environmental conditions, such as higher or variable temperatures, to simulate real-world applications more

accurately. Exploring alternative alloying elements or incorporating nanoparticles within the Fe-C matrix could further advance the material's thermal properties. These additional conditions would not only support broader industrial uses, such as in brake discs and engine components but might also reduce operational and maintenance costs by extending the material's lifespan in thermal applications. Ultimately, these studies would help refine Fe-C cast iron's applicability in mechanical constructs requiring specific thermal characteristics, such as cooling systems or high-wear components, while promoting cost efficiency through improved material performance.

Acknowledgement

This research was supported by the project KK.01.1.1.07.0013 - Innovative Ti-Mg dental implants for the world market.

5 REFERENCES

- [1] Zhou, X., Li, Y., Huang, X., & Wu, W. (2023). Recent Advances in Thermal Conductivity of Cast Irons: A Review. *Materials Science and Engineering: A*, 870, 144664. <https://doi.org/10.1016/j.msea.2023.144664>
- [2] Xiao, Y., Liu, J., Wang, Z., & Sun, Q. (2020). Study on the Thermal Conductivity of Compacted Graphite Cast Iron and Its Application in Automotive Brake Discs. *Automotive Engineering*, 42(5), 495-502. <https://doi.org/10.3969/j.issn.1671-7277.2020.05.005>
- [3] Staff report MC. (2014). 48th Census of World Casting Production. Steady Growth in Global Output. *Modern Casting*, 17-21.
- [4] Zhang, Y., Wei, X., Sun, J., & Liu, Y. (2021). Thermal Conductivity and Microstructure of Cast Irons Under Various Heat Treatment Conditions. *Materials Characterization*, 180, 111354. <https://doi.org/10.1016/j.matchar.2021.111354>
- [5] Wang, J., Zhang, L., & Liu, X. (2022). Effects of Graphite Morphology on the Mechanical and Thermal Properties of Gray Cast Iron. *Journal of Materials Science*, 57(10), 5123-5136. <https://doi.org/10.1007/s10853-022-07067-9>
- [6] Hecht, R. L., Dinwiddie, R., Porter, W. D., & Wang H. (1996). Thermal Transport Properties of Grey Cast Irons. *SAE transaction*, 105, 1699-1705. <https://doi.org/10.4271/962126>
- [7] Bockus, S. (2006). A study of the microstructure and mechanical properties of continuously cast iron products. *METABK*, 45(4), 287-290.
- [8] Gomez, M., Dominguez, F., & Ortega, A. (2021). Evaluation of Graphite Nodularity and Its Effects on the Thermal Properties of Nodular Cast Irons. *Journal of Engineering Materials and Technology*, 143(3), 31012. <https://doi.org/10.1115/1.4049351>
- [9] Sjögren, T. (2007). Influences of the Graphite Phase on Elastic and Plastic Deformation Behaviour of Cast Irons. <https://doi.org/10.1007/s11661-007-9115-8>
- [10] Angus, H. T. (1976). *Cast Iron: Physical and Engineering Properties*. Second. 1-33. <https://doi.org/10.1016/B978-0-408-70933-0.50006-2>
- [11] Prakash, P., Mytri, V. D., & Hiremath, P. S. (2011). Digital Microstructure Analysis System for Testing and Quantifying the Ductile Cast Iron. *International Journal of Computer Applications*, 19(3).
- [12] Venugopal, P., Saravanan, K. G., & Thanigaivelan, R. (2023).

- Performance Analysis of Edm on Grey Cast Iron Using Rsm and Topsis Method. *Applied Engineering Letters*, 8(1), 10-16. <https://doi.org/10.18485/aeletters.2023.8.1.2>
- [13] Skulić, A., & Bukvić, M. (2016). Tribological properties of piston-cylinder set in internal combustion engines. *Applied Engineering Letters*, 1(1), 29-33.
- [14] Shifani Madtha, L., & Narendra Babu, B. R. (2013). Experimental Behavioural Study of Ductil Cast Iron Microstructure and Its Mechanical Properties. *International Journal of Engineering Research and Applications (IJERA)*, 3(3), 1470-1475.
- [15] Samuelsson, D. P. B. (2011). Analysis of microstructural strain-fields in grey cast iron. Diploma work No. 47/2011. <https://odr.chalmers.se/server/api/core/bitstreams/b8adf42c-9daf-4045-a88f-5dfb832a47ce/content>
- [16] Martinez, F., Gonzalez, A., & Sanchez, L. (2021). Thermal Conductivity Measurement of Graphitic Cast Irons Using Hot Disk Method. *Journal of Thermal Analysis and Calorimetry*, 144(3), 1747-1758. <https://doi.org/10.1007/s10973-020-09619-3>
- [17] Collini, L., Nicoletto, G., & Konečná, R. (2008). Microstructure and mechanical properties of pearlitic gray cast iron. *Materials Science and Engineering: A*, 488, 529-539. <https://doi.org/10.1016/j.msea.2007.11.070>
- [18] Lee, J., Kim, S., & Park, H. (2023). Analysis of Microstructural Effects on Thermal Conductivity in High-Alloy Cast Irons. *Materials Today: Proceedings*, 74, 208-216. <https://doi.org/10.1016/j.matpr.2023.07.017>
- [19] Rodriguez, E., Fernandez, J., & Ortega, R. (2020). Influence of Heat Treatment on the Thermal Conductivity of Gray Cast Iron. *Materials Research Express*, 7(11), 115501. <https://doi.org/10.1088/2053-1591/abc544>
- [20] Ohser, J., Sandau, K., Stets, W., & Gerber, W. (2003). Image Analytical Characterization of Graphite in Grey Cast Iron and Classification on Lamellar Arrangement. *Practical Metallography*, 40(9), 454-473. <https://doi.org/10.1515/pm-2003-400906>
- [21] Moumeni, E. (2013). *Solidification of cast iron - A study on the effect of microalloy elements on cast iron*.
- [22] ASTM E 562 - 01. (2011). *Standard Test Method for Determining Volume Fraction by Systematic Manual Point Count*. West Conshohocken, PA.
- [23] EN ISO 6506-1:2014. *Metallic materials -- Brinell hardness test -- Part 1: Test method*.
- [24] Maul, C. L. (2008). *Glow Discharge Atomic Emission Spectrometry - The Methodology, Calibration and Analytical Performance for Bulk and Quantitative Depth Profile Analysis*.
- [25] Mikulić, D. & Milovanović, B. (2010). TCi system for non-destructive determination of thermal properties of materials. In: *10th European Conference on Non Destructive Testing*. Publishing house Spektr, 256-266.
- [26] Instruments M. Mathis TCi User Manual. c-Therm Technologies Ltd., 21 Alison Blvd., Fredericton, Canada; 2008.
- [27] Velichko, A. (2008). Quantitative 3D Characterization of Graphite Morphologies in Cast Iron using FIB Microstructure Tomography. *Chemie, Pharmazie, Bio- und Werkstoffwissenschaften der Universität des Saarlandes, Saarbrücken, Germany*. <https://doi.org/10.22028/D291-22506>
- [28] McQuillan, F. J., Culham, J. R. & Yovanovich, M. M. (1984). *Properties of Dry Air at One Atmosphere*. https://www.mhtlab.uwaterloo.ca/pdf_reports/mhlt_G01.pdf
- [29] Peet, M. J., Hasan H. S., & Bhadeshia, H. K. D. H. (2011). Prediction of thermal conductivity of steel. *International Journal of Heat and Mass Transfer*, 54(11-12), 2602-2608. <https://doi.org/10.1016/j.ijheatmasstransfer.2011.01.025>

Authors' contacts:

Dalibor Viderščak, PhD, assistant
(Corresponding author)
University of Zagreb, Faculty of Mechanical Engineering and Naval Architecture,
Department of Materials,
Ivana Lučića 5, 10000 Zagreb, Croatia
dalibor.viderscak@fsb.unizg.hr

Zdravko Schauerl, PhD, Prof.
University of Zagreb, Faculty of Mechanical Engineering and Naval Architecture
Department of Materials,
Ivana Lučića 5, 10000 Zagreb, Croatia
zdravko.schauerl@fsb.unizg.hr

Sanja Šolić, PhD, Assoc. Prof.
University North, Department of Mechanical Engineering,
Jurja Križanića 31b, 42000 Varaždin, Croatia
ssolic@unin.hr

Bojan Milovanović, PhD, Assoc. Prof.
University of Zagreb, Faculty of Civil Engineering, Department of Materials,
Fra Andrije Kačića-Miošića 26, 10000 Zagreb, Croatia
bojan.milovanovic@grad.unizg.hr

Ivan Stanković, Project Coordinator
Rotarex SA,
24 rue de Diekirch, L-7440 Lintgen, Luxembourg
stankovic.ivan@rotarex.com

Mario Šercer, PhD, director
Development and Training Centre for Metal Industry Metal Centre in Čakovec,
Bana Josipa Jelačića 22D, 40 000 Čakovec, Croatia
ravnatelj@metalskajezgra.hr

The Multicollinearity Effect on Machine Learning Accuracy for Leaf Chlorophyll Content Prediction of Indoor Plants

Dorijan Radočaj*, Daria Galić Subašić, Ivan Plaščak, Mladen Jurišić

Abstract: Various factors that influence chlorophyll levels in indoor plants were analysed. The input dataset consisted of 52 samples, which represent 10 distinct plant types. To non-invasively measure the chlorophyll content in plant leaves, Soil-Plant Analysis Development (SPAD) sensor was used, measuring the absorbance of specific light wavelengths, allowing for the assessment of chlorophyll concentration. The dataset was supplemented by covariates from soil electrical conductivity (*EC*) sensing at depths of 5 cm, 10 cm, and 15 cm, along with the multispectral Plant-O-Meter sensor. Four covariates in the model, including plant type, *EC* (5 cm), *EC* (15 cm), and normalized difference red-edge index (*NDRE*), showed minimal correlation with other variables, highlighting their independence. To predict leaf chlorophyll content, Random Forest and Extreme Gradient Boosting machine learning models were utilized, with Random Forest achieving higher average coefficient of determination of 0.458. The study underscored the potential of a complementary dataset for evaluating the complex relationship among root-soil dynamics and leaf for optimizing indoor plant health.

Keywords: extreme gradient boosting; multispectral sensor; random forest; soil electroconductivity; Soil-Plant Analysis Development (SPAD)

1 INTRODUCTION

In recent years, there has been increasing interest in cultivating indoor plants effectively and sustainably, driven by recognition of their manifold benefits, which include positively impacting the environment, aesthetics, and human health [1]. These plants are not merely decorative additions to our living spaces, rather they have a crucial role in enhancing air quality, reducing stress, and beautifying indoor environment. The presence of chlorophyll in the leaves of indoor plants is fundamental to their health and vitality, as this pigment is essential for photosynthesis and overall growth [2]. Advanced machine learning algorithms have emerged as powerful tools in optimizing the care of indoor plants, enabling precise monitoring and fine-tuning of chlorophyll care plans. These algorithms, along with non-invasive estimation instruments, provide a method to comprehend and improve the well-being of our indoor plants [3]. Predicting the quantity of chlorophyll in indoor plants using machine learning algorithms poses challenges, including multicollinearity that occurs when predictor variables in the model have high correlations, leading to an impact on the accuracy and interpretability of the machine learning model [4]. Multicollinearity can introduce instability into coefficient estimations, making it difficult to discern the unique contributions of each predictor variable. In the context of predicting chlorophyll concentration, various physiological and environmental factors are at play, and their interrelationships can lead to multicollinearity, casting doubt on the validity of machine learning models. To tackle this issue, two significant sources of data play a vital role in machine-learning predictions of chlorophyll content in leaves: soil electrical conductivity (*EC*) sensing and leaf multispectral sensing [5]. Multispectral sensing of leaves offers a vast amount of information on the physiological state of the leaves, providing insights into pigment content, water stress, and the overall health of the plant. Soil *EC* sensing provides critical information about soil moisture levels and

nutrient availability, both of which have a direct impact on plant growth and vitality. Integrating this data with chlorophyll predictions through machine learning models is expected to result in improved accuracy. In the context of predicting chlorophyll concentration in indoor plants, Monte Carlo cross-validation accounts for the inherent variability in factors such as plant growth, environmental conditions, and sensor data [6]. The care of indoor plants is influenced by numerous dynamic factors, and this approach ensures that the models can withstand the randomness in the training and test data. Additionally, it decreases the probability of overfitting to a particular dataset and encourages the creation of models that are robust.

To improve present knowledge of predicting indoor plant chlorophyll content, this study delves into the significant issue of multicollinearity using machine learning techniques, emphasizing the Random Forest (RF) and Extreme Gradient Boosting (XGB) methods. Monte Carlo cross-validation was employed to assess and mitigate the impact of multicollinearity on the forecasting accuracy of these two extensively adopted techniques. The consequences of multicollinearity on the effectiveness of both RF and XGB models were explored, alongside their predictive capabilities, and their reliability in accurately estimating chlorophyll content in leaves. Additionally, it will enhance the understanding of the challenges posed by multicollinearity. The aim of the evaluation of these models' performance is to assist practitioners in determining the most effective approach to handle multicollinearity.

2 MATERIALS AND METHODS

2.1 Indoor Plant Data Collection

Ten indoor plant types represented by 52 samples were evaluated in the study (Fig. 1). The evaluation of chlorophyll levels in indoor plant leaves was achieved using the portable Konica Minolta Soil-Plant Analysis Development (SPAD) sensor, in a non-invasive manner. This technology enables

measuring the absorbance of specific light wavelengths by the chlorophyll in leaf tissue, thereby ascertaining the relative chlorophyll concentration [7]. The SPAD sensor measures light absorbance, focusing on the decrease in red light absorption caused by chlorophyll's presence, resulting in a unitless SPAD reading that shows the leaf's relative chlorophyll concentration. Elevated SPAD levels are linked to higher chlorophyll content, indicating healthier and more photosynthetically active leaves. The gathered SPAD data enables comparing leaves or plants, facilitating evaluation of general plant health, monitoring changes over time, and informing care plans for indoor plant maintenance.

When combined, soil EC sensing at three soil depths (5, 10, and 15 cm), and the multispectral Plant-O-Meter sensor provide a dependable approach for obtaining complementary covariates for predicting chlorophyll content of indoor plants [5]. The Plant-O-Meter sensor was used to measure leaf physiology aspects like plant leaves' absorption and reflection characteristics at multiple wavelengths. Vegetation indices were calculated based on Plant-O-Meter measurements according to specifications by Kitić et al. [8], including red ($NDVIr$), green ($NDVlg$), blue ($NDVlb$), green-red ($GRNDVI$), green-blue ($GBNDVI$), red-blue ($RNDVI$) and visible normalized difference vegetation index ($PNDVI$), simple ratio (SR) with its modification (MSR) and inversion (ISR), normalized difference red-edge index ($NDRE$) and enhanced vegetation index (EVI). Soil EC sensing at various depths enabled measuring the effects of soil moisture content. A total of 16 covariates were evaluated, including plant type, three soil EC covariates and 12 vegetation indices.

2.2 Multicollinearity Analysis

Three diagnostic measures were used in analysing multicollinearity, including tolerance (TOL), within-subject variance (Wi), and auxiliary F-test (Fi) [9]. TOL measured the percentage of variance in a predictor variable that isn't explained by other predictor variables in the model, with low tolerance levels indicating significant multicollinearity. Tolerance values ranged from 0 to 1, representing reciprocal values of the variance inflation factor. TOL values close to 1 suggest that it is relatively uncorrelated with other predictors, with values higher than 0.1 indicating multicollinearity. This indication implies that the variable can provide specific and unique information to the model. Farrar's F-test is a diagnostic metric for assessing multicollinearity, represented by Wi . This test examined the correlations between a specific predictor variable and all other variables in the model, indicating the presence and severity of multicollinearity. Wi demonstrated significant results when multicollinearity exists, indicating the dependent variable's linear predictability from the independent predictors. In the context of multicollinearity, the Fi assessed the relationship between the F-statistic and the coefficient of determination, allowing determination of whether high levels of multicollinearity decrease the overall goodness of fit of the regression model and how it affects the model's explanatory power.



Figure 1 The display of representative indoor plants from ten species evaluated in the study

2.3 Machine Learning Prediction and Accuracy Assessment

The study employed the RF model to capture non-linear correlations between leaf chlorophyll concentration and chosen characteristics. The ensemble approach included several decision trees, which reduced overfitting and improved prediction accuracy [10]. In addition, the enhanced gradient boosting technique XGB was used to predict chlorophyll levels. This model is highly renowned for its durability and efficiency in processing high-dimensional information [11]. A feature selection procedure was conducted to enhance the predictive capacity of the models, combining RF and XGB models with all input covariates and with covariates for which multicollinearity was not detected based on TOL , Wi and Fi coefficients.

The effectiveness of the RF and XGB regression models in predicting leaf chlorophyll content was evaluated through the Monte Carlo cross-validation approach. To ensure robustness in the evaluation, the dataset per 20 repetitions was divided into an 80% training set and a 20% testing set. This repeated resampling strategy produced multiple distinct

train-test splits, enabling a thorough assessment of model performance. During each iteration, the RF and XGB models underwent training with the training data and evaluated their predictive performance using the testing data.

The coefficient of determination (R^2) indicated the proportion of the variance in the leaf chlorophyll content represented by *SPAD* values that the independent variables can predict. Root mean square error (*RMSE*) measured the average magnitude of the difference between the predicted and actual values. Mean absolute error (*MAE*) similarly measured the average absolute difference between predicted and actual values, being less susceptible to higher residuals between predicted and actual *SPAD* values. A higher R^2 and lower *RMSE* and *MAE* indicated more accurate predictive performance.

3 RESULTS AND DISCUSSION

Plant type, *EC* (5 cm), *EC* (15 cm), and *NDRE* exhibited limited correlation with other variables (Tab. 1). It is expected that plant type, being a categorical variable representing distinct types of plants, does not inherently exhibit multicollinearity with numerical variables. *EC* (5 cm) and *EC* (15 cm) represented distinct soil depths, and due to the sizable variance in soil properties at different depths, it is generally unlikely that they will exhibit a strong linear correlation. Moreover, *NDRE*, which is obtained from multispectral leaf sensing, characterized plant health in a unique manner based on red-edge spectral band and thus did not exhibit a strong correlation with other environmental variables or indices [12]. This lack of correlation between these particular variables indicates that they can be studied separately to determine their impact on the response variables, with minimal preoccupation for the interfering effects of collinearity.

Table 1 The results of multicollinearity analysis for evaluated covariates.

Covariates	<i>TOL</i>	<i>Wi</i>	<i>Fi</i>
Plant type	0.311	2.495	2.700
<i>EC</i> (5 cm)	0.188	4.875	5.276
<i>EC</i> (10 cm)	0.067	15.666	16.953
<i>EC</i> (15 cm)	0.108	9.329	10.096
<i>EVI</i>	0.049	21.979	23.784
<i>GBNDVI</i>	0.000	3221.493	3486.060
<i>GRNDVI</i>	0.000	4352.326	4709.763
<i>ISR</i>	0.001	1868.367	2021.808
<i>MSR</i>	0.021	53.424	57.812
<i>NDRE</i>	0.301	2.618	2.833
<i>NDVib</i>	0.002	639.444	691.959
<i>NDVlg</i>	0.000	6788.285	7345.777
<i>NDVlr</i>	0.000	2802.944	3033.137
<i>PNDVI</i>	0.000	5560.212	6016.848
<i>RBNDVI</i>	0.000	3395.957	3674.852
<i>SR</i>	0.011	105.751	114.436

A series of boxplots in Fig. 2 display the *SPAD* values of evaluated indoor plant species according to four covariates for which multicollinearity was not detected (Fig. 3), including plant type, soil *EC* at 5 cm and 15 cm soil depth, and *NDRE*. All four covariates indicate large heterogeneity of analysed individual indoor plants, which is otherwise not

usually present in the same intensity for outdoor plants, especially crops and orchards [13].

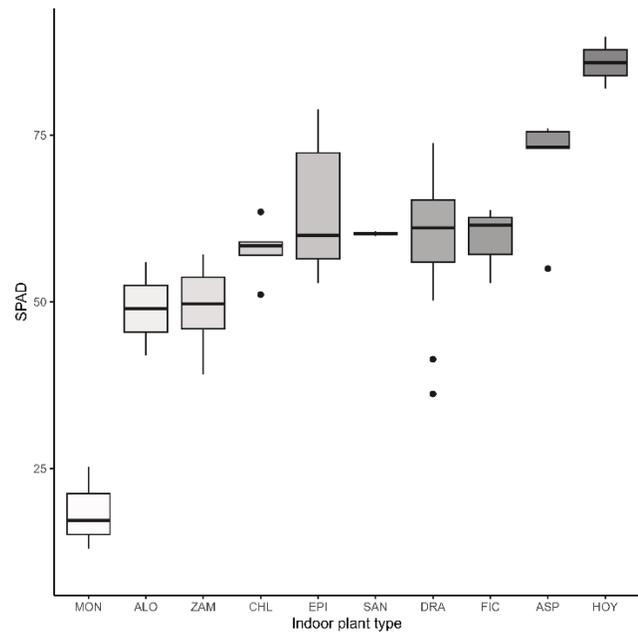


Figure 2 Boxplots representing *SPAD* values per indoor plant type

Based on the Monte Carlo accuracy assessment, the RF model using all covariates achieved superior overall performance in 20 distinct data folds (Tab. 2). Specifically, the data produced a substantially greater average R^2 of 0.458, indicating its capacity to elucidate a larger portion of the dependent variable's variability, along with a lower average *RMSE* of 10.69 and *MAE* of 8.16. On the other hand, the RF model with filtered covariates for multicollinearity exhibited subpar relative performance with a lower average R^2 of 0.338 and higher average *RMSE* of 11.77 and *MAE* of 9.12, demonstrating a decrease in predictive accuracy. Additionally, the coefficient of variation indicated that the RF model demonstrated enhanced stability and consistency in performance across various data partitions, therefore presenting itself as a more dependable option within this experimental context.

The XGB model without multicollinearity filtering consistently demonstrated moderate performance across distinct data folds (Tab. 3). It achieved an average R^2 of 0.360, displaying moderately precise predictions, with an average *RMSE* of 13.50 and an average *MAE* of 10.77. Conversely, the XGB model with integrated multicollinearity filtering presented marginally improved performance, producing an average R^2 of 0.423 and a lower average *RMSE* of 13.28 and *MAE* of 10.45. These results suggest that the multicollinearity analysis had limited impact on predictive accuracy improvement in general, with only XGB having a slight benefit from it [14]. Additionally, the coefficient of variation values indicated that both models exhibited significant variability in performance across various data partitions, with the XGB model displaying slightly higher variability.

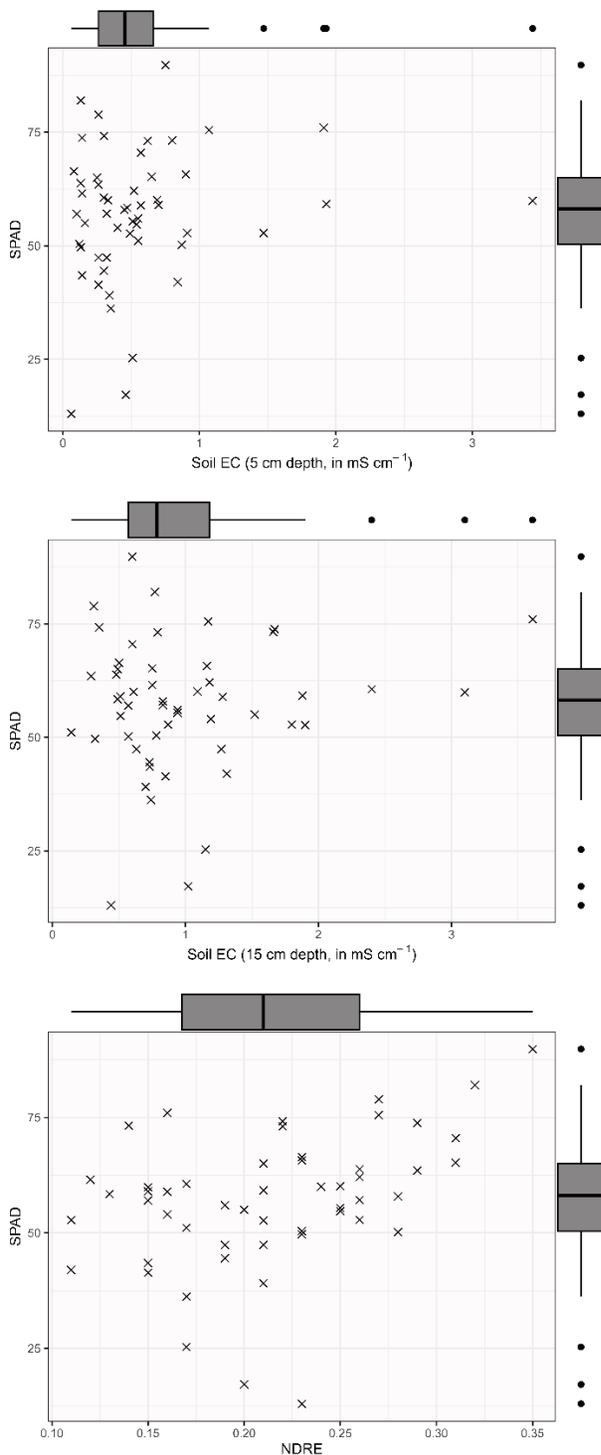


Figure 3 Boxplots representing SPAD values and their relationship with covariates for which multicollinearity was not detected

By combining multispectral leaf sensing with soil *EC* data at various depths, a comprehensive dataset of covariates was generated. This facilitated the capturing of complex interactions between the below-ground root-soil interactions and above-ground leaf physiology by machine learning models, resulting in enhanced precision and resilience of predicting leaf chlorophyll content. Monte Carlo cross-

validation aided in identifying the most effective algorithm for predicting chlorophyll levels in plant foliage by mitigating the impact of individual data subsets and allowing for a comprehensive evaluation of the models' capacity for generalization. The Monte Carlo cross-validation results exhibited notable differences in prediction accuracy according to properties of training and test sample data, strongly indicating that frequently used single split-sample approach might not provide adequate representation of prediction accuracy.

Table 2 The accuracy assessment results for RF

Fold	All covariates			Filtered covariates		
	R^2	RMSE	MAE	R^2	RMSE	MAE
1	0.435	15.30	9.49	0.110	11.34	7.94
2	0.395	15.00	9.84	0.534	7.65	6.27
3	0.428	8.64	7.36	0.445	9.90	9.14
4	0.659	7.32	6.03	0.205	12.03	10.91
5	0.704	12.89	10.26	0.558	8.58	7.15
6	0.639	15.74	11.60	0.575	9.59	7.72
7	0.213	10.47	8.88	0.441	9.92	8.44
8	0.430	9.33	8.02	0.000	20.07	14.43
9	0.748	10.62	7.60	0.147	11.75	8.75
10	0.889	5.67	4.41	0.179	19.19	12.82
11	0.497	11.39	9.58	0.347	18.35	12.36
12	0.032	11.20	8.56	0.141	10.23	6.06
13	0.270	8.65	6.14	0.828	10.20	7.83
14	0.584	7.31	6.29	0.654	7.77	5.83
15	0.027	14.65	9.15	0.018	13.49	11.24
16	0.018	15.97	10.18	0.083	11.44	9.74
17	0.880	6.74	6.02	0.632	5.74	4.18
18	0.322	10.63	9.15	0.592	11.56	9.60
19	0.787	5.95	5.51	0.278	9.92	8.35
20	0.208	10.43	9.07	0.000	16.72	13.61
Average	0.458	10.69	8.16	0.338	11.77	9.12
CV	0.603	0.311	0.234	0.751	0.334	0.302

Table 3 The accuracy assessment results for XGB

Fold	All covariates			Filtered covariates		
	R^2	RMSE	MAE	R^2	RMSE	MAE
1	0.731	11.07	8.30	0.545	11.33	9.04
2	0.291	14.59	9.20	0.271	19.06	15.82
3	0.492	16.24	11.29	0.676	7.67	6.11
4	0.298	12.53	11.23	0.399	15.17	14.26
5	0.034	15.14	11.99	0.517	11.42	9.87
6	0.625	10.28	8.68	0.465	11.96	9.42
7	0.456	13.46	10.83	0.527	9.47	7.60
8	0.272	10.47	7.86	0.067	12.81	10.63
9	0.598	8.93	6.80	0.296	16.70	11.84
10	0.252	12.66	10.15	0.410	14.06	11.91
11	0.002	25.53	21.10	0.225	11.01	8.41
12	0.243	13.64	11.16	0.526	13.90	9.15
13	0.553	11.10	9.38	0.342	13.36	10.45
14	0.517	8.60	7.27	0.679	11.91	9.47
15	0.074	12.89	11.17	0.586	12.81	11.04
16	0.580	13.28	10.36	0.691	12.04	7.87
17	0.444	6.82	5.38	0.578	11.30	9.54
18	0.001	25.65	21.41	0.280	18.10	13.06
19	0.677	14.29	10.58	0.320	14.73	11.84
20	0.056	12.84	11.24	0.064	16.85	11.56
Average	0.360	13.50	10.77	0.423	13.28	10.45
CV	0.665	0.351	0.371	0.444	0.215	0.222

To further improve the proposed approach, the addition of deep learning methods [11] and integration of additional sensors complementary to multispectral and soil *EC* devices,

such as insolation sensor [15], might provide more in-depth observations as an upgrade of this study.

4 CONCLUSIONS

A thorough analysis was carried out to investigate the various factors impacting chlorophyll levels in indoor plants. The study involved examining a dataset encompassing ten unique indoor plant species. Four factors, namely Plant Type, *EC* (5 cm), *EC* (15 cm), and *NDRE*, were singled out as displaying minimal correlation with other variables, indicating their independence within the model. As a categorical variable denoting individual plant species, plant type *e* does not exhibit multicollinearity with numerical variables. The soil electrical conductivity measurements at varying depths displayed significant diversity, while *NDRE*, obtained from multispectral leaf sensing, rendered a distinctive outlook on plant health, resulting in minimal correlation with other environmental variables or indices. Monte Carlo cross-validation was utilized to comprehensively evaluate the effectiveness of RF and XGB machine learning models in predicting leaf chlorophyll content. The RF model, which included all covariates, demonstrated superior overall performance by elucidating a significant portion of the dependent variable's variance. While the XGB model displayed moderately accurate predictions, it only slightly benefited from multicollinearity filtering. The RF model proved more stable and consistent in comparison, achieving average R^2 of 0.458. The Monte Carlo cross-validation results exhibited notable differences in prediction accuracy according to properties of training and test sample data, strongly indicating that frequently used single split-sample approach might not provide adequate representation of prediction accuracy.

This study indicated that a combination of multispectral leaf sensing and soil *EC* measurements at varying depths can establish a powerful tool for exploring the intricate relationship between root-soil dynamics below ground and leaf physiology above ground. Through the use of machine learning models and cross-validation techniques, moderately accurate predictive capabilities for indoor plant chlorophyll content were achieved. Future research could include the integration of complementary sensors and deep learning methods to advance present comprehension of indoor plant health and chlorophyll synthesis.

5 REFERENCES

- [1] Yeo, L. B. (2021). Psychological and physiological benefits of plants in the indoor environment: A mini and in-depth review. *International Journal of Built Environment and Sustainability*, 8(1), 57-67. <https://doi.org/10.11113/ijbes.v8.n1.597>
- [2] AlFadhly, N. K., Alhelfi, N., Altemimi, A. B., Verma, D. K. & Cacciola, F. (2022). Tendencies affecting the growth and cultivation of genus *Spirulina*: An investigative review on current trends. *Plants*, 11(22), 3063. <https://doi.org/10.3390/plants11223063>
- [3] Yang, Y., Xu, F., Chen, J., Tao, C., Li, Y., Chen, Q., ... & Shen, W. (2023). Artificial intelligence-assisted smartphone-based sensing for bioanalytical applications: A review. *Biosensors and Bioelectronics*, 115233. <https://doi.org/10.1016/j.bios.2023.115233>

- [4] Drobnič, F., Kos, A. & Pustišek, M. (2020). On the interpretability of machine learning models and experimental feature selection in case of multicollinear data. *Electronics*, 9(5), 761. <https://doi.org/10.3390/electronics9050761>
- [5] Khruschev, S. S., Plyusnina, T. Y., Antal, T. K., Pogosyan, S. I., Riznichenko, G. Y. & Rubin, A. B. (2022). Machine learning methods for assessing photosynthetic activity: environmental monitoring applications. *Biophysical Reviews*, 14(4), 821-842. <https://doi.org/10.1007/s12551-022-00982-2>
- [6] Phinzi, K., Abriha, D. & Szabó, S. (2021). Classification efficacy using k-fold cross-validation and bootstrapping resampling techniques on the example of mapping complex gully systems. *Remote Sensing*, 13(15), 2980. <https://doi.org/10.3390/rs13152980>
- [7] Križová, K., Kadeřábek, J., Novák, V., Linda, R., Kurešová, G. & Šařec, P. (2022). Using a single-board computer as a low-cost instrument for SPAD value estimation through colour images and chlorophyll-related spectral indices. *Ecological Informatics*, 67, 101496. <https://doi.org/10.1016/j.ecoinf.2021.101496>
- [8] Kitić, G., Tagarakis, A., Cselyuszka, N., Panić, M., Birgermajer, S., Sakulski, D. & Matović, J. (2019). A new low-cost portable multispectral optical device for precise plant status assessment. *Computers and Electronics in Agriculture*, 162, 300-308. <https://doi.org/10.1016/j.compag.2019.04.021>
- [9] Corrêa, A. J. M., Alves, P. F., Cambuim, J., de Moraes, M. L. T. & Freitas, M. L. M. (2022). Climate fluctuation impacts in *Astronium urundeuva* (M. Allemão) Engl. silvicultural characters in the Brazilian Cerrado. *Environmental Research: Climate*, 1(2), 025007. <https://doi.org/10.1088/2752-5295/ac9695>
- [10] Radočaj, D., Jurišić, M., Antonić, O., Šiljeg, A., Cukrov, N., Rapčan, I., ... & Gašparović, M. (2022). A Multiscale Cost-Benefit Analysis of Digital Soil Mapping Methods for Sustainable Land Management. *Sustainability*, 14(19), 12170. <https://doi.org/10.3390/su141912170>
- [11] Hassan, S. M., Jasinski, M., Leonowicz, Z., Jasinska, E. & Maji, A. K. (2021). Plant disease identification using shallow convolutional neural network. *Agronomy*, 11(12), 2388. <https://doi.org/10.3390/agronomy11122388>
- [12] Prey, L., Von Bloh, M. & Schmidhalter, U. (2018). Evaluating RGB imaging and multispectral active and hyperspectral passive sensing for assessing early plant vigor in winter wheat. *Sensors*, 18(9), 2931. <https://doi.org/10.3390/s18092931>
- [13] Radočaj, D., Šiljeg, A., Plaščak, I., Marić, I. & Jurišić, M. (2023). A Micro-Scale Approach for Cropland Suitability Assessment of Permanent Crops Using Machine Learning and a Low-Cost UAV. *Agronomy*, 13(2), 362. <https://doi.org/10.3390/agronomy13020362>
- [14] Yan, H., He, Z., Gao, C., Xie, M., Sheng, H. & Chen, H. (2022). Investment estimation of prefabricated concrete buildings based on XGBoost machine learning algorithm. *Advanced Engineering Informatics*, 54, 101789. <https://doi.org/10.1016/j.aei.2022.101789>
- [15] Jung, C. & Arar, M. (2023). Natural vs. Artificial Light: A Study on the Influence of Light Source on Chlorophyll Content and Photosynthetic Rates on Indoor Plants. *Buildings*, 13(6), 1482. <https://doi.org/10.3390/buildings13061482>

Authors' contacts:

Dorijan Radočaj, PhD
(Corresponding author)
Josip Juraj Strossmayer University of Osijek,
Faculty of Agrobiotechnical Sciences Osijek,
Vladimira Preloga 1, 31000 Osijek, Croatia
E-mail: dradočaj@fazos.hr

Daria Galić Subašić, PhD

Josip Juraj Strossmayer University of Osijek,
Faculty of Agrobiotechnical Sciences Osijek,
Vladimira Preloga 1, 31000 Osijek, Croatia
E-mail: dgsubasic@fazos.hr

Ivan Plaščak, PhD, Associate professor

Josip Juraj Strossmayer University of Osijek,
Faculty of Agrobiotechnical Sciences Osijek,
Vladimira Preloga 1, 31000 Osijek, Croatia
E-mail: iplascak@fazos.hr

Mladen Jurišić, PhD, Full professor

Josip Juraj Strossmayer University of Osijek,
Faculty of Agrobiotechnical Sciences Osijek,
Vladimira Preloga 1, 31000 Osijek, Croatia
E-mail: mjurisic@fazos.hr

Simulation Modeling of Labor Costs for Moving of Harvester between Working Positions in Cutting Area

Konstantin P. Rukomojnikov*, Tatiana V. Sergeeva, Natalya N. Aleksagina

Abstract: The article discusses the technological features of the functioning of forest harvester in the implementation of logging in various natural conditions. The primary objective of this research is to ascertain the influence of trees and shrubbery on the duration of the harvester's transit between operational sites. The research was implemented by computer modeling methods based on a computer program specially created by the authors of the article to simulate the operation of a forest harvester. During the simulation, the trajectory of the manipulator from each working position of the harvester, the dimensional characteristics and location of the trees assigned to the felling and the remaining trees were evaluated. The results can be used by research organizations when planning the production process of logging operations.

Keywords: computer modeling; cutting area; felling; forest harvester; manipulator; working position

1 INTRODUCTION

When choosing machines and mechanisms for the implementation of logging operations, it is necessary to take into account a variety of factors. This is especially true when implementing selective cutting of forest, for which there is currently no general approach to labour norming. The trees remaining in the cutting area create significant obstacles to the effective work of operators of cutting machines used in felling trees. A detailed analysis of their work on the cutting area in the conditions of selective cutting of forest will allow implementing a scientific approach to labour norming for performing individual cutting operations. Focusing the attention of the readers of the article on the harvester, as a machine that is one of the most popular today by loggers around the world, we can confidently say that its performance largely depends on the average volume of tree stem and the qualifications of the machine operator. Studies of different countries conducted by Spinelli et. al. [1], McEwan [2] and Picchio et al. [3] in Southern Europe, Liski et al. [4] in Finland, Nakagawa et al. [5] in Japan, Seixas & Batista [6] in Brazil, Norihiro et al. [7], Ramantswana et al. [8], Ackerman et al. [9] in South Africa, Strandgard et al. in Australia [10] this statement is also confirmed.

Nevertheless, studies by various researchers have shown that the proportion of trees that have been cut down in relation to the overall forest composition is a significant factor in calculating the labor expenses for felling and processing trees with a harvester during selective logging [11-13]. The requirements for the safety of trees remaining for rearing during selective logging compared to continuous logging of the stand lead to the need to take into account the technological features of harvesters when creating mathematical dependencies for analyzing their performance. In this regard, the authors propose additional theoretical studies that take into account the peculiarities of the operation of these machines in the forest.

The primary objective of this research is to ascertain the influence of trees and shrubbery on the duration of the harvester's transit between operational sites.

2 MATERIALS AND METHODS

The research was conducted through the use of computer modelling techniques [14], employing a custom-designed computer program developed by the researchers to simulate the operation of a harvester in a forest environment [15-17].

The program was created in the AnyLogic simulation system. The model developed by us uses the technical characteristics of the harvester and the qualifications of the operator: the dimensions of the harvester and the harvester head, the speed characteristics of the movement of the harvester and its manipulator, the parameters of the manipulator departure, the speed of dragging the tree trunk with the harvester head and some other parameters. During the operation of the harvester, the trees of the target component are processed to ensure the safety of the trees of the non-target component. This, final felling with and without undergrowth preservation, as well as various types of thinning can be simulated.

The adjustable parameters of the stand within the designated cutting area, as specified by the investigator in the program, include the number of target component trees that must be harvested, the number of non-target component trees that should remain in place for cultivation purposes, and the number trees (undergrowth category) that may be removed during logging operations to enhance the accessibility of target component trees. The placement of all trees on the territory of the cutting area is carried out arbitrarily at each start of the simulation program. The rocks and diameters of tree trunks of all components are randomly distributed among them based on the characteristics of the breed composition and their dimensional parameters specified by the researcher.

The main limitations of the simulation model are the possibility of using it only to analyze the operation of the harvester in flat terrain, good cross-country ability, the volume of trees being processed is no more than 2 m³, tree diameters up to 0.9 m with bucking on logs from two to six meters long.

To confirm that the results of the simulation model correspond to the real production conditions of the cutting

areas, the program was tested on the basis of observations of the harvester Silvatec 8266TH [22] (Fig. 1).

Specifications:

- Engine Mercedes OM 906 LA diesel engine
- 205 kW / 278 HP at 2100 rpm
- Dynamic tractive power (theoretic): 146 kN
- 8-wheel drive
- Crane Loglift 220 V / 83 crane with 8.3 m reach
- Length: 7.70 m × Width: approx. 2.62 m with standard tyres.



Figure 1 Harvester Silvatec 8266TH

During the experiment, individual was mastered. The species composition of the stand was 100 % of pine trees. The average wood stock per hectare is 260 m³. The average volume of trees in forest swathes analyzed during the experiments was 0.4 m³.

The work of experienced operators with more than 5 years of experience was studied. The width of the forest swathes was 15 meters. The sections of forest swathes analyzed during the experiment had a length of 100 meters. In the course of experimental studies, 9 different forest swathes were developed with a different number of trees (k_D) on them and a different proportion of trees being cut down (k_i).

In the process of production experiments, a Latin square experiment plan was used with varying factors at three levels $n = 3$, having the form shown in Tab. 1. This plan allowed for uniform scanning of the factor space and confirmed the significance of the factor features used in the experiment (k_D), (k_i).

Table 1 Representation of the Latin square taking into account the levels of variable factors of the production experiment

The proportion of trees being cut down (k_i)	Number of trees per 1 ha (k_D), pcs.		
	$k_{D1} = 200$	$k_{D2} = 400$	$k_{D3} = 600$
$k_{i1} = 0,2$	$P_{h1}, t_{nav1}, t_{pod1}, t_{per1}$	$P_{h4}, t_{nav4}, t_{pod4}, t_{per4}$	$P_{h7}, t_{nav7}, t_{pod7}, t_{per7}$
$k_{i2} = 0,6$	$P_{h2}, t_{nav2}, t_{pod2}, t_{per2}$	$P_{h5}, t_{nav5}, t_{pod5}, t_{per5}$	$P_{h8}, t_{nav8}, t_{pod8}, t_{per8}$
$k_{i3} = 1,0$	$P_{h3}, t_{nav3}, t_{pod3}, t_{per3}$	$P_{h6}, t_{nav6}, t_{pod6}, t_{per6}$	$P_{h9}, t_{nav9}, t_{pod9}, t_{per9}$

During the trial, the hourly productivity of the harvester was logged for each forest strip (P_h). The duration of the manipulator's pointing action (t_{nav}), and the time it took to drag the tree to the processing area (t_{pod}), were also recorded. Furthermore, the time required to move the harvester between workstations for each tree (t_{per}) was measured.

The results of the production experiment were compared with the results of simulation modeling. The final comparison of the obtained numerical values showed good convergence of the results. The discrepancy between theoretical and experimental values did not exceed 12% for all analyzed indicators. The obtained conclusions confirmed the adequacy of the model to real production conditions. After the experiment, an additional adjustment of the simulation model was carried out. The identified inaccuracies of its operation were taken into account. This made it possible to further increase the convergence of the results obtained.

For a clearer understanding of the reasons for the reduction of labor costs during the operation of the harvester on final felling and their increase during thinning, a comparative analysis of the movements of the harvester was carried out. The analysis allowed us to identify a number of distinctive features associated with the movements of the harvester. In particular, it was noted that during selective logging in order to ensure the safety of the quality of the trees remaining for rearing, in order to avoid defects in the trunks that violate the integrity of the wood, such as bark peeling, bullying, scratches, dents, etc. The following technological features are observed in the operation of the harvester:

- if there are obstacles in the form of non-target component trees, the path of the manipulator to the tree that is planned to be cut down becomes longer;
- when transferring a fallen tree to a processing area, it may be necessary to adjust the position of the harvesting machine in order to avoid causing damage to other trees. This may require the machine to move in the opposite direction;
- the total distance traveled by the harvester between the trees increases due to the displacement of the harvester back in the process of adjusting working positions.
- with a small proportion of the failed component in the composition of the stand, the frequency of changing working positions increases due to the low availability of trees of the target component.

Further simulation was carried out on the basis of the results of real production data collected during the production experiment.

For the computer-based experiment, a factorial design with four factors was developed, with each factor varying at four levels, resulting in a 4×4×4×4 factorial design. The experiment was implemented with a complete search of options. Among the main variable factor features were used: timber volume per hectare (Q), m³/ha; the amount of large undergrowth per hectare K_p , pcs/ha; the average volume of tree stem (V), m³

Variation of parameters was carried out in the ranges presented in Tab. 2.

Table 2 Range of variation of factor features during the implementation of simulation modeling of harvester operation

Symbol	Variation levels			
	1	2	3	4
Q	50	120	190	260
k_i	0,2	0,4	0,6	1
K_p	0	150	300	450
V	0,2	0,7	1,2	1,7

In the course of simulation modelling, the influence of the enumerated factors on a specific component of the harvester cycle time was evaluated. This element was defined as the average time taken by the harvester to move between working positions while feeling a single tree (t_{per}), sec.

3 RESULTS

As a result of the simulation, a regression model (1) was created to predict the average time it takes for the harvester to move between working positions during selective logging:

$$t_{per} = \frac{[d_0 + k_i(d_1 + d_3Q + d_5K_p) + K_p(d_2 + d_4Q)]}{Qk_iV^{d_6}} \quad (1)$$

The results of calculating the coefficients of the regression equation and their confidence intervals are shown in Tab. 3. The results of the table indicate the statistical significance of the coefficients included in the regression model. The calculated indicator P is the value of all coefficients less than the level of statistical significance (0.05). The critical value of the Student's t-test, equal to 1.96, is less than the calculated values of this indicator for any of the coefficients of the regression equation. All the found values of standard errors are less than the modules of the values of the coefficients of the final equation corresponding to them.

Table 3 Conducting statistical analysis of the significance of the regression coefficients per tree

	Estimate	Std. Err	t-test	P-value	-95% CL	+95% CL
d_0	1474,823	13,849	106,495	0,0000	1447,512	1502,134
d_1	-305,263	40,764	-7,489	0,0000	-385,652	-224,873
d_2	-0,28227	0,06518	-4,33041	0,0000	-0,411	-0,154
d_3	0,937587	0,248288	3,776211	0,0002	0,448	1,427
d_4	0,001719	0,000332	5,183645	0,0000	0,001	0,002
d_5	0,864121	0,129283	6,683950	0,0000	0,609	1,119
d_6	-0,893	0,007	-126,354	0,0000	-0,907	-0,879

Statistical processing of the results at the level of significance of 0,05 according to the Kohren criterion proved the possibility of reproducibility of the obtained values. The computed value of the Coherent coefficient invalidating the mean time of the harvester's movement between operational sites per tree was $G_p = 0,007$. This value did not exceed the critical table value of the Kohren coefficient equal $G_{kp} = 0,011$. The ratio $G_p < G_{kp}$ made it possible to conclude that the variances are homogeneous.

The multiple coefficient of determination R^2 of the nonlinear model was 0,896 (Fig. 2). This value serves as an indication that the alteration in the meantime taken by the

harvester to move between harvesting positions for each tree is contingent upon variations in the variables incorporated into the regression model. This dependence is not accidental.

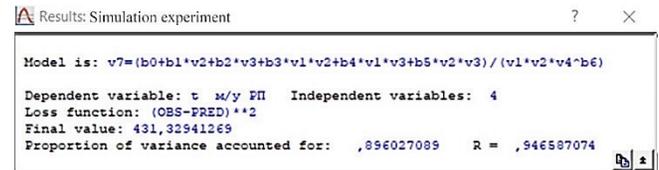


Figure 2 Results of building regression model

The calculated value of the significance level is less than the set value of 0.05. Therefore, the calculated indicators of statistical processing of the average time of moving of the harvester between working positions per tree, presented in Tab. 4, indicate the significance of the coefficient of determination.

Table 4 Analysis of variance

	Sum of Squares	df	Mean Squares	F-value	p-value
Regression	265067	7,0000	37866,8	17294,8	0,00
Residual	431,3	197,00	2,19		
Total	265499	204,000			

The construction of a normal probabilistic graph (Fig. 3) allowed us to conclude that there were no systematic deviations of the results of regression modeling from the theoretical normal line. All points (residuals) on the graph are close to the line expected for normally distributed residuals.

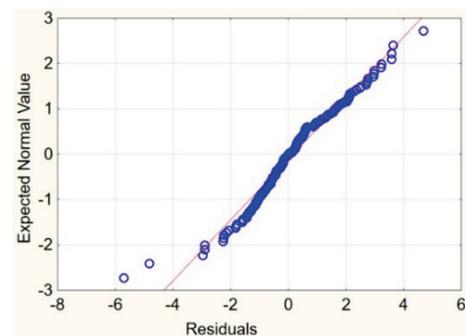


Figure 3 Normal probabilistic graph

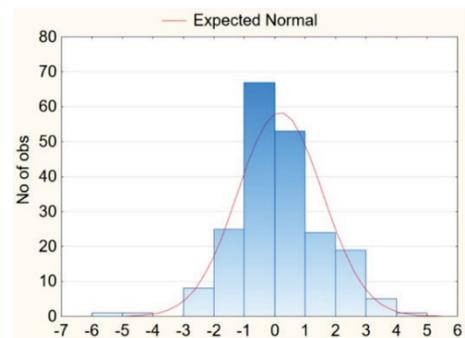


Figure 4 Histogram of residue

The histogram of the data presented in Fig. 4 also serves as a visual confirmation of the proximity of the distribution of harvester movement times between harvesting positions

per tree to a normal law distribution.

In Fig. 5 presents comparative analysis of the predicted and observed values. In Fig. 6 shows graph of the predicted values and residues. Analysis of the graphs allows you to note only a slight discrepancy between the results recorded on the simulation and regression models. Graphs make it possible to declare the adequacy of the regression equation with the results of computer simulations.

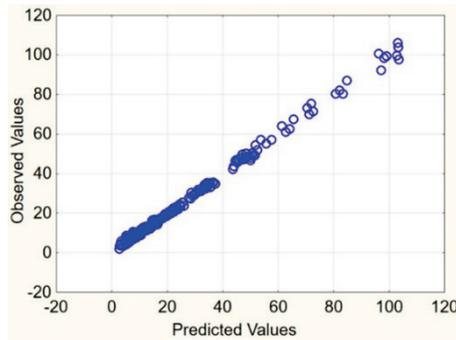


Figure 5 Comparison of predicted and observed values

The picture shows how the change in the average time it takes for the harvester to move between trees changes in different natural conditions is presented on three-dimensional graphs (Fig. 7).

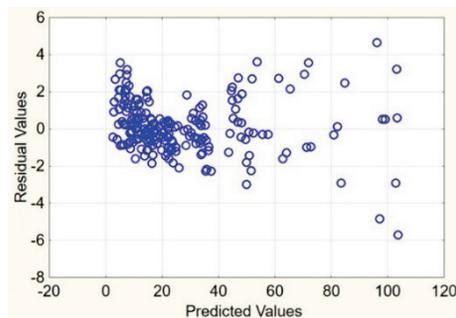


Figure 6 Graph of predicted values and residues

The analysis of the data presented in Fig. 7 reveals a steady growth in the time required for the harvester to moving between harvesting positions for each individual tree as the average volume of the tree stem increases and the same average wood reserves per hectare (with a decrease in the number of trees in the cutting area). However, this increase in time has the most significant impact on the operating time of the machine, only with selective cutting in cutting areas with less than 40 percent of the timber volume being cut down (Fig. 7a).

The main factor that affects the analyzed element of the harvester's cycle time is the proportion of non-target tree components during cutting (Fig. 7b, 7c, 7d). This is because a reduction in the number of trees being cut results in a longer total travel time for the harvester in the cutting area, when calculated per tree.

The quantitative indicators of undergrowth within the cutting area have an impact on the average time taken by the machine to move between working positions for each

individual tree. This occurs only when the proportion of cutting trees decreases and the dimensional specifications of the trees designated for felling increase, with a constant average supply of wood per hectare (Fig. 7c, 7d). In such cases, due to an increase in the quantitative indicators of the undergrowth at the cutting, the increase in labor costs can reach 50% of the minimum recorded in the construction of the graphs of the analysis results.

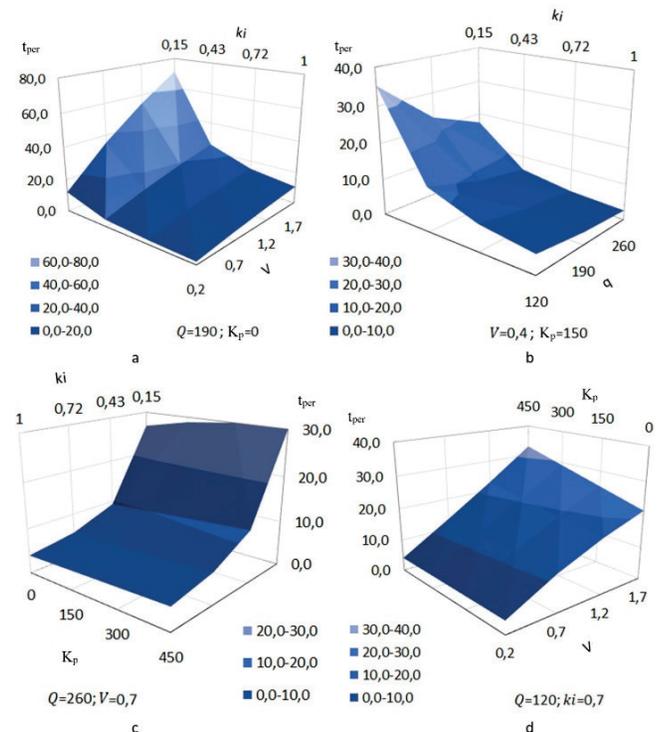


Figure 7 The diagrams depict the fluctuations in the mean travel time of the harvester between harvesting positions per tree

4 DISCUSSION

Substantiation of the effectiveness of the use of logging machines is an important and responsible task facing the owner of any logging and forestry enterprise. The impact of natural and climatic factors on the efficacy of operations within the timber extraction sector is substantial. So, the correct selection of logging machines is currently of great importance [19]. Of particular relevance is the imitation of technological processes of logging operations when analyzing the tasks of using modern multioperative equipment. In most cases, standards for calculating efficiency and labor rationing have not yet been developed for this equipment when planning the activities of enterprises in changing environmental conditions [20, 21]. The introduction of forest harvesters in the technological process of logging operations sets itself a wide range of tasks to improve the efficiency of their use [22, 23]. According to many prominent scholars, the processes of managerial and organisational decision-making, as well as the establishment of labour standards and the development of regulatory documentation, can and should be automated. This should be done based on the specific type of logging operation, the

qualitative and quantitative features of the cutting area, and the technical specifications of the equipment employed, with utmost regard for random natural and operational variables [24, 25].

A critical examination of the existing body of research in this field reveals that the studies conducted by other scholars frequently employ simulation modelling to investigate the functioning of forest harvesters [26-28]. The simulation models developed by them, provided that the researcher has access to the appropriate software, can be executed repeatedly to analyse the performance of the harvester under various environmental conditions. They make it possible to mathematically substantiate various parameters of the harvester's operation technology. But in order to use these software developments, it is necessary for developers to provide open access to the use of their simulation models. In addition, when using each specific simulation model, it is necessary to train users to achieve research results. All this makes it advisable for developers to create of various mathematical dependencies based on the results of multiple experiments on their models. Obtaining mathematical dependencies is a rational result of analyzing the operation of machines in the process of simulation modeling of a number of scientists from different countries. The developers of existing models focus readers' attention primarily on finding such important elements of the harvester's work as productivity and cycle time with a diverse combination of natural and industrial factors [29, 30]. A number of scholars set forth more intricate objectives and dissect the overall duration of processing a single tree into distinct components of the processing cycle time. Most scientists divide the time it takes to process a tree trunk into 5 or 7 main parts [31, 32]. Some of them depend only on the parameters of the trees being processed and the technical characteristics of the harvester. They can be described by mathematical dependencies that do not require a simulation process. These include mathematical dependencies for determining the time of cutting down a tree, bucking the trunk, and dragging the trunk through the knives of the harvester head [33, 34]. But there are other elements of the cycle time that depend not only on the characteristics of the machine, but also on the relative position of the trees in the forest. Such indicators are random variables and multiple experiments are necessary to determine them. Specifically, the analysis has led to the conclusion that at present, there are no mathematically derived relationships available for determining the average time taken by a harvester to move between harvesting positions for harvesting a tree in diverse natural settings.

The data clearly indicates a rise in the average time required for the harvester to move to each new position, coupled with a reduction in the number of trees felled within a given area and an expansion of the distance between individual trees. However, during the experiment it was noted that during selective logging, the harvester operator in some cases has to simultaneously move the harvester along the fiber in the opposite direction with the movement of the manipulator. This is done in order to minimize damage to the remaining trees when they are moved to the processing area.

Moving the harvester in the opposite direction subsequently increases the distance traveled by the harvester when moving forward to the next working position and, in general, the total distances of the harvester's movements along the cutting area. This demonstrates that not only does the quantity of trees that are felled affect the average time required for a harvester to move between work sites, but also does the number of remaining trees remaining for rearing.

Existing models cannot be applied to find this indicator. As a result, many researchers recommend only a certain range of values for this indicator, regardless of the type of logging and working conditions in the forest [31-35]. This significantly reduces the accuracy of calculations when justifying the desired parameters. Furthermore, the dearth of theoretical data necessary to verify the metric of the mean time spent by a harvester traversing between operational positions during various types of logging operations precludes the analysis of the impact of external technical and technological variables on this metric across a spectrum of the stand thinning intensities.

5 CONCLUSION

Thus, according to the authors of the article, this study improves the current state of knowledge in comparison with existing studies. The presented mathematical relationship presented herein is novel, enabling us to accurately and convincingly account for the variation in the time taken for a harvester to move between work positions in diverse configurations of factors at a logging site. It has practical significance for the forestry industry. The introduction of novel mathematical correlations elucidating the impact of extrinsic variables on the performance of a harvesting machine will enhance the efficacy of devising logging techniques and scrutinizing their ultimate operational metrics. The results can be used by research organizations in planning the production process of logging operations.

Acknowledgments

The study was supported by the grant the Russian Science Foundation No. 24-26-00129, <https://rscf.ru/project/24-26-00129/> of Volga State University of Technology, Yoshkar-Ola.

6 REFERENCES

- [1] Spinelli, R., Hartsough, B. & Magagnotti, N. (2010). Productivity Standards for Harvesters and Processors in Italy. *Forest Products Journal*, 60(3), 226-235. <https://doi.org/10.13073/0015-7473-60.3.226>
- [2] McEwan, A. (2011). The effect of tree and bundle size on the productivity and costs of Cut-To-Length and multi-stem harvesting systems in Eucalyptus pulpwood. *MSc. Thesis*. University of Pretoria, South Africa.
- [3] Picchio, R., Sirna A., Sperandio G., Spina R. & Verani S. (2012). Mechanized harvesting of Eucalypt Coppice for Biomass Production Using High Mechanization Level. *Croatian Journal of Forest Engineering*, 33(1), 15-24.
- [4] Liski, E., Jounela, P., Korpunen, H., Sosa, A., Lindroos, O. &

- Jylhä, P. (2020). Modeling the productivity of mechanized CTL harvesting with statistical machine learning methods. *International Journal of Forest Engineering*, 31(3), 253-262. <https://doi.org/10.1080/14942119.2020.1820750>
- [5] Nakagawa, M., Hayashi, N. & Narushima, T. (2010). Effect of tree size on time of each work element and processing productivity using an excavator-based single-grip harvester or processor at a landing. *Journal for Forest Research*, 15(4), 226-233. <https://doi.org/10.1080/02827581.2011.628335>
- [6] Seixas, F. & Batista, J. L. F. (2012). Use of Wheeled harvester and excavators in Eucalyptus harvesting in Brazil: In: *Proceedings of the 35th Council on Forest Engineering Annual Meeting: Engineering New Solutions for Energy Supply Demand*. New Bern, North Carolina, 7 p.
- [7] Norihiro, J., Ackerman, P., Spong, B. & Längin, D. (2018). Productivity Model for Cut-to-Length Harvester Operation in South African Eucalyptus Pulpwood Plantations. *Croatian Journal of Forest Engineering*, 39, 1-13.
- [8] Ramantswana, M., McEwan, A., & Steenkamp, J. (2013). A comparison between excavator-based harvester productivity in coppiced and planted Eucalyptus grandis compartments in KwaZulu-Natal, South Africa. *Southern Forests: Journal of Forest Science*, 75(4), 239-246. <https://doi.org/10.2989/20702620.2013.858213>
- [9] Ackerman, S. A., Talbot, B. & Astrup, R. (2021). The effect of tree and harvester size on productivity and harvester investment decisions. *International Journal of Forest Engineering*, 33(1), 22-32. <https://doi.org/10.1080/14942119.2021.1981046>
- [10] Strandgard, M., Mitchell, R. & Acuna, M. (2016). General productivity model for single grip harvesters in Australian eucalyptus plantations. *Australian Forestry*, 79(2), 108-113. <https://doi.org/10.1080/00049158.2015.1127198>
- [11] Kärhä, K., Ovaskainen, H. & Palander, T. (2021). Decision-Making Among Harvester Operators in Tree Selection and Need for Advanced Harvester Operator Assistant Systems (AHOASs) on Thinning Sites COFE-FORMEC 2021. Forest Engineering Family: Growing Forward Together Corvallis, OR, U.S.A. – September 27-30, 2021
- [12] Räsänen, J. (2020). Puuston ennakkoleimauksen vaikutus koneellisen harvennushakkuun ajanmenekkiin: Simulaattoritutkimus (The effect of pre-stamping on mechanical thinning time consumption: a simulator survey). *Master's Thesis*, University of Eastern Finland.
- [13] Vahtila M. (2019.) Poistettavien puiden etukäteisvalinnan vaikutus hakkuun tuottavuuteen ja laatuun koneellisessa harvennushakkuussa (Effect of prior tree-selection of stems to be removed on the productivity and quality of cutting work in mechanized thinnings). *Master's Thesis*, University of Helsinki
- [14] The AnyLogic Company. (2023). Retrieved from <https://www.anylogic.ru/>.
- [15] Rukomojnikov, K. P., Sergeeva, T. V., Gilyazova, T. A., Voldaev, M. N., Tsarev, E. M. & Anisimov, S. E. (2022). Computer simulation of the development of logging sites using a felling-delimiting buckler. *Systems. Methods. Technology*, 2(54), 108-113. <https://doi.org/10.18324/2077-5415-2022-2-108-113>
- [16] Rukomojnikov, K. & Sergeeva, T. (2024). Simulation modeling of logging harvester movements during selective logging. *Journal of Applied Engineering Science*, 22(3), 604-611. <https://doi.org/10.5937/jaes0-50146>
- [17] Rukomoynikov, K. P., Sergeeva, T. V., Gilyazova, T. A. & Komisar V. P. (2022). Computer modeling to support management and organizational decisions in the use of a forest harvester. *Proceedings of SPIE, Computer Applications for Management and Sustainable Development of Production and Industry (CMSD2021)*, 122510. <https://doi.org/10.1117/12.2631137>
- [18] Rukomojnikov, K. P., Sergeeva, T. V., Gilyazova, T. A., Tsarev, E. M. & Anisimov, P. N. (2024). The Influence of the Stand Composition on the Cycle Time of the Harvester. *Lesnoy Zhurnal = Russian Forestry Journal*, 3, 153-165. (In Russ.). <https://doi.org/10.37482/0536-1036-2024-3-153-165>
- [19] Mokhirev, A. P. (2016). The method of selection of forest machines under the climatic conditions. *Forestry engineering journal*, 6(4), 208-215. <https://doi.org/10.12737/23459>
- [20] Mokhirev, A. P., Mamatov, V. O. & Urazaev, A. P. (2015). Modeling of the technological process of logging machines. *Journal of international scientific researches*, 3(24), 72-74.
- [21] Wang, J., Greene, W. D. & Stokes, B. (1998). Stand, harvest, and equipment interactions in simulated harvesting prescriptions. *For. Prod. J.*, 48(9), 81-86.
- [22] Laurila, Ya., Kuchin, A. V., Lebedev, V. D., Elepov, A. A., Varakin, M. Yu. & Tadiashvili, I. R. (2020). *PONSSE Forest Machines: Tutorial*. Arkhangelsk, Russian Federation: LLC SK "Arkhangelsk".
- [23] Mokhirev, A. P., Kunitskaya, O. A., Kalita, G. A., Verner, N. N. & Shvetsova, V. V. (2022). Logging harvester reliability assessment. *Forestry Bulletin*, 26(5), 93-101. <https://doi.org/10.18698/2542-1468-2022-5-93-101>
- [24] Rukomojnikov, K. P. (2013). Simulation modeling of mutually coordinated operation of sets of adaptive-modular forest machines. *Forestry Bulletin*, 3, 154-158.
- [25] Sokolov, A. P. & Osipov, E. V. (2017). Simulation modeling of the wood harvesting processes using petri nets. *Forestry Engineering Journal*, 7(3), 307-314. https://doi.org/10.12737/article_59c2140d704ae5.63513712
- [26] Khitrov, E. G. & Andronov, A. V. (2018). Mathematical model of interaction between forest machine's rover and strengthening soil. *IOP Conf. Series: Journal of Physics: Conf. Series 1108*. <https://doi.org/10.1088/1742-6596/1177/1/012030>
- [27] Mederski, P. S., Bembenek, M., Karaszewski, Z., Łacka, A., Szczepańska-Álvarez, A. & Rosińska, M. (2016). Estimating and Modelling Harvester Productivity in Pine Stands of Different Ages, Densities and Thinning Intensities. *Croat. J. for. Eng.* 37(1), 27-36.
- [28] Ovaskainen, H, Palander, T., Tikkanen, L. Hirvonen, H. & Ronkainen, P. (2011). Productivity of different working techniques in Thinning and clear cutting in a Harvester Simulator. *Baltic Forestry*, 17(2), 288-298.
- [29] Gerasimov, Y., Väätäinen, K. & Senkin, V. (2012). Productivity of single-grip harvesters in clear-cutting operations in the northern European part of Russia. *European Journal of Forest Research*, 131(3), 647-654. <https://doi.org/10.1007/s10342-011-0538-9>
- [30] Sängstuvall, L., Bergström, D., Lämäs, T. & Nordfjell, T. (2012). Simulation of harvester productivity in selective and boom-corridor thinning of young forests. *Scandinavian Journal of Forest Research*, 27(1), 56-73. <https://doi.org/10.1080/02827581.2011.628335>
- [31] Azarenok, V. A., Gerts, E. F. & Mekhrentsev, A. V. (2001). *Assortment harvesting of timber: a textbook for universities*. Ekaterinburg: Ural State Forest Engineering Academy, 134 p. ISBN 5-230-25652-4.
- [32] Shirmin, Yu. A. (2004). *Technology and equipment of forest industry production / Part I*. Moscow: Moscow State Forest University, 446 p.
- [33] Bazarov, S. M., Belenkiy, Yu. I., Svoynkin, F. V., Svoynkin, V. F. & Balde, T. M. D. (2020). Systems analysis of technological efficiency of assortment production based on VSRM. *Bulletin*

of the St. Petersburg Forest Engineering Academy, 233, 177-188. <https://doi.org/10.21266/2079-4304.2020.233.177-188>

- [34] Svoikin, F., Svoikin, V., Bazarov, S., Alexandrov, A. & Ugryumov, S. (2021) Information support of the dynamics of system connectivity of wheel harvester operations. *IOP Conference Series: Earth and Environmental Science: 2*, Saint Petersburg, Virtual, 18–19 March 2021, 012031. <https://doi.org/10.1088/1755-1315/806/1/012031>
- [35] Nurminen, T., Heikki, K. & Uusitalo, J. (2006) Time Consumption Analysis of the Mechanized Cut-to-length Harvesting System. *Silva Fennica*, 40(2), 335-363. <https://doi.org/10.14214/sf.346>

Authors' contacts:

Konstantin P. Rukomojnikov

(Corresponding author)

Volga State University of Technology,

3, Lenin Square, Yoshkar-Ola, 424000, Russian Federation

E-mail: rukomojnikovKP@volgatech.net

Tatiana V. Sergeeva

Volga State University of Technology,

3, Lenin Square, Yoshkar-Ola, 424000, Russian Federation

E-mail: sergeeva.2010t@mail.ru

Natalya N. Aleksagina

Volga State University of Technology,

3, Lenin Square, Yoshkar-Ola, 424000, Russian Federation

E-mail: elenias888@mail.ru

Impact of Infill Pattern Design on Stress-Strain behaviour of 3D Printed Parts

Rahim Sejdiu, Xhemajl Mehmeti*, Flamur Salihu, Labinot Topilla, Drin Krasniqi, Mohamad El Mehtedi, Agron Baraktari

Abstract: Compared to some conventional manufacturing processes, the properties of additive manufacturing parts can depend on the structural and manufacturing process parameters, rather than just on the type of materials used. The purpose of this paper is to evaluate the tensile mechanical properties of the components printed with a 3D printer by changing the design of the infill pattern. The study presents the data of 13 (by 3 tests for each form of filling) different infill pattern designs, prepared and tested according to the ISO 527-2:2012 standard. The samples were analyzed according to the paired sample tests, considering the weight and the results of the tests, as well as the impact of the infill pattern on the test results. Significant differences were observed among the samples in terms of engineering stress and strain. The line infill pattern demonstrated better results in the engineering strain whereas the concentric infill pattern yielded superior values for engineering stress. Regarding ultimate tensile strength, concentric infill pattern shows significantly higher results than all other samples. Taking into consideration the results of the samples, it is concluded that there are significant differences between the design of the infill pattern and the other analyzed studies, which should be considered depending on the intended use of the 3D printed elements.

Keywords: infill pattern; tensile strain; tensile stress; tensile strength; 3D printing

1 INTRODUCTION

The 3D printing industry has taken a huge leap of development in recent years. This technique involves producing solid objects from digital files, which has garnered the attention of various industries, along with the scientific and academic communities. The development trend of this technology is versatile, and it is difficult to accurately predict the future of its development. This is because 3D printers are successfully utilized across various fields such as medicine, engineering, architecture, construction, and design [1-3]. The easy and fast possibilities of printing complex shapes [4], have made this technology develop rapidly in all branches of industry [5]. The global 3D printing market is estimated to be worth USD 20.37 billion in 2023, and according to CAGR (Compound Annual Growth Rate) prediction, the 3D printing industry will grow by 23.5% from 2024 to 2030 (Horizon, 2024).

The advantages of 3D technologies allow not only researchers and experts, but also passionate individuals who use 3D printers to easily produce prototypes of their ideas, since 3D printing greatly simplifies prototype production. This technology has reduced product design and manufacturing processes from weeks to days or even hours.

With the increase in popularity of 3D printing technologies, different materials have been developed which comprise a wide range, including: PLA (Polylactic Acid), ABS (Acrylonitrile Butadiene Styrene), PETG (Polyethylene Terephthalate Glycol), TPU (Thermoplastic Polyurethane), Nylon, Carbon Fibre Filled Filaments, and Wood-Filled Filaments [6].

The most commonly used materials in 3D printing are polymers, due to their versatility, low cost, ease of use and in some cases, environmental friendliness [7, 8]. These materials can be melted and extruded into different shapes [9]. Polylactic Acid Composites (PLA) are one of the most widely used raw materials for 3D printing, mainly because of their environmentally friendly characteristics, low cost, high-quality printed surface, and properties that are relatively

similar to those of other commonly used printing materials [7].

Due to variety of use and growing trends of 3D printing products [10], it is important to understand their mechanical properties. During practical applications, products must withstand different levels of mechanical loads, therefore, it is important to know the durability of parts for each application [11].

According to the ASTM F2792 standard, 3D printing technologies are divided into seven groups, including: the binding jetting, directed energy deposition, material extrusion, material jetting, powder bed fusion, sheet lamination and vat photopolymerization [12, 14].

The mechanical properties of materials that are determined experimentally are very important during the product design phase.

The most frequent causes of structural failures generally are divided into two categories:

- 1) Negligence during the design, construction or operation of the structure and,
- 2) Applying new design or new materials, which may show unexpected results [15].

The test results are used to determine the dimensions of elements produced, as well as the domain of their use. In general, one of the most important mechanical properties of materials is their stability when subjected to tensile stress. The test results are used to determine the appropriate material that fulfills specific needs and to predict the behaviour of products to mechanical loads. The mechanical properties measured are yield strength, tensile strength, elongation and modulus of elasticity [16].

Various studies show that the mechanical properties of products printed with 3D technologies vary depending on the printing parameters, which include: Shell (wall thickness), speed of print, material (printing temperature, type of materials), and infill (density, patterns, infill line distance [5, 7, 14, 17-20] etc. Furthermore, extruder temperatures, print platform (bed) temperature, extruder head diameter, and print

head closure [21] are also important factors influencing the mechanical properties of printed parts.

By carefully selecting the right combination of parameters in the printing process, it becomes possible to produce products with the best characteristics.

Taking into account the many factors that affect the mechanical properties of printed products, this paper's primary objective is to investigate the impact of infill patterns on the mechanical properties of products printed with 1.75mm PLA filament using 3D printers. Furthermore, the aim of the study was to investigate the samples built from similar material for different structural design. Tensile tests were used in order to identify the strain, deformations and Young's Modulus of the samples for different loads.

During 3D printing, various factors, (such as material type, printing speed, print head diameter, printing temperature, and layer deposition) affect the mechanical properties of the printed parts, which can also vary based on the infill pattern. The infill pattern plays an important role in durability, weight and print time. It can vary depending on many parameters, including design, density and print speed. Furthermore, depending on the types of printers and the software available on the market, a considerable number of infill patterns exist.

2 MATERIALS AND METHODS (EXPERIMENTAL PART)

Poly Lactic Acid (PLA) material was used for printing the samples. According to ISO 1183 and GB/T 1033 standards, the density of this type of material, measured at 21 °C, is 1.17 g/cm³. The material used in the study is "pine green" produced by 3DHUB Company, with a diameter of 1.75 mm. G-codes samples were generated using Creality Slicer 4.8.2. and were printed on the Eder-3 V2 Neo printer (Fig. 3a). Based on the equation b Tab. 1, the length of l_2 is calculated to be 109.33 mm; 109 mm is acquired. Infill pattern designs used in this study was carried out according to the standard ISO 527-2:2012, Figs. 2a, 2b and 2c. The paper presents the test results for 39 samples (3 samples for each type of infill pattern) that were produced using the same printing parameters, except for the infill pattern (Fig. 1a - 1m), which analyzes 13 different forms of infill patterns offered by the printer software. Apart from the infill pattern parameter, other parameters for all samples were constant: standard quality 0.2 mm; wall line count 2; top and top/bottom (outline

thickness) thickness 0.8 mm; top/bottom (outline thickness) layers 4; top/bottom pattern – line; infill density 30 %; printing temperature 215 °C; built plate temperature 60 °C; print speed of 50 mm/s; infill speed 60 mm/s; and travel speed of 120 mm/s.

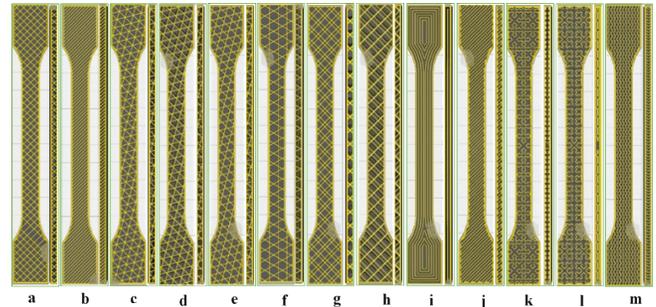


Figure 1 Infill pattern design (top and side view): a) Grid, b) Line, c) Triangle, d) Trihexagon, e) Cubic, f) Cubic Subdivision, g) Octet, h) Quarter Cubic, i) Concentric, j) Zig Zag, k) Cross, l) Cross 3D, m) Gyroid.

The tests were conducted using the WDW-50Y Stiffness Testing Machine in the laboratory of "FERPLAST", a company that produces plastic pipes and is certified according to the SK EN ISO/IEC 17025:2018 standard. Sample testing was done according to the ISO 527-2:2012 standard. Prior to the start of the tests, the samples were placed in the test environment which were prepared according to the ISO 9969 standard.

Table 1 Specimen dimensions according to ISO 527-2:2012

Specimen type		1A (mm)	1B (mm)
l_3	Overall length	170.0	≥150.0
l_1	Length of narrow parallel-sided portion	80.0 ± 2.0	60.0 ± 0.5
r	Radius	24.0 ± 1.0	60.0 ± 0.5
l_2	Distance between broad parallel-sided portions b	109.3 ± 3.2	108.0 ± 1.6
b_2	Width at ends	20.0 ± 0.2	
b_1	Width at narrow portion	10.0 ± 0.2	
h	Preferred thickness	4.0 ± 0.2	
L_0	Gauge length (preferred)	75.0 ± 0.5	50.0 ± 0.5
	Gauge length (acceptable if required for quality control or when specified)	50.0 ± 0.5	
L	Initial distance between grips	115 ± 1	115 ± 1

a) The recommended overall length of 170 mm of the type 1A is consistent with ISO 294-1 and ISO 10724-1. For some materials, the length of the tabs may need to be extended (e.g. $l_3 = 200$ mm) to prevent breakage or slippage in the jaws of the testing machine.
 b) $l_2 = l_1 + [4r(b_2 - b_1) - (b_2 - b_1)^2]^{1/2}$, resulting from l_1, r, b_1 and b_2 , but within the indicated tolerances.

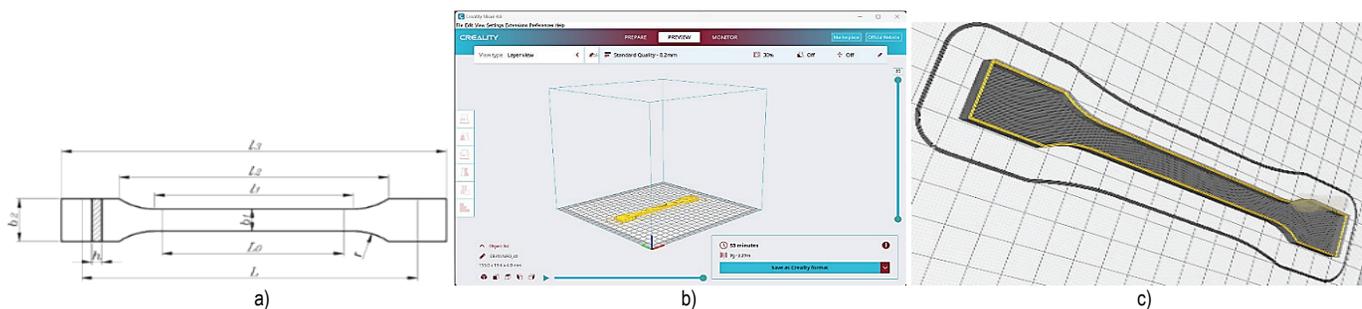


Figure 2 a) Specimen design, b) and c) specimen in Creality software

Table 2 Average weight of printed samples (g) according to infill pattern.

Weight of samples													
	Grid	Lines	Triangle	Tri hexagon	Cubic	Cubic subdivision	Octet	Quarter cubic	Concentric	Zig zag	Cross	Cross 3d	Gyroid
Mean g	5.878	5.882	5.895	5.883	5.792	5.654	5.487	5.884	5.765	5.870	5.821	5.722	5.932
Std. Dev	.1328	.1233	.0575	.1028	.1078	.0855	.3430	.1036	.0923	.2097	.0751	.0862	.0880

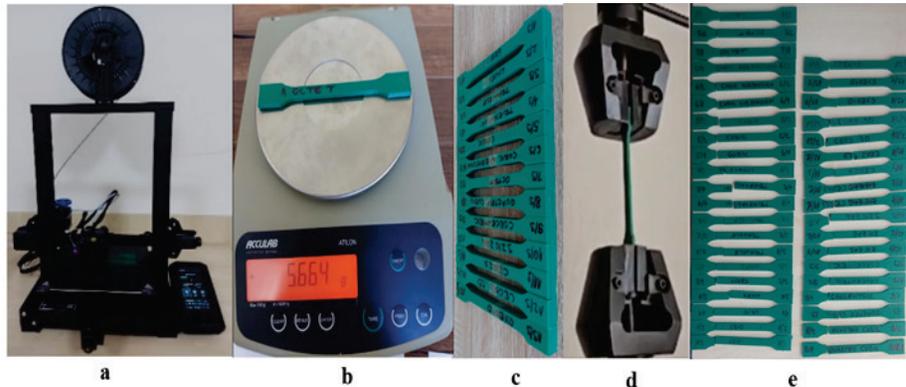


Figure 3 a) Printing samples; b) Weighing the samples; c) Printed samples; d) Specimens during tensile testing; e) Tested samples (broken)

Table 3 Average of samples tested according to infill pattern

	Grid	Line	Triangle	Tri_hexagon	Cubic	Cubic_s_Div.	Octet	Quarter_cubic	Concentric	Zig-Zag	Cross	Cros 3D	Gyroid
Sample ID	1	2	3	4	5	6	7	8	9	10	11	12	13
Width (mm)	9.9	10	10	10	10	10	9.9	10	10	10	10	10	10
S ₀ (mm ²)	37.6	38.0	38.0	39.0	39.0	39.0	37.6	39.0	39.0	39.0	39.0	39.0	39.0
Strain→Force (kN)	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.3	0.2	0.2	0.2	0.2
Yield Stress σ _y (MPa)	20.1	19.2	22.0	20.8	22.0	21.5	19.4	20.2	25.4	19.0	18.8	19.7	20.3
Break Force F _b (kN)	0.7	0.7	0.7	0.6	0.8	0.8	0.7	0.8	0.9	0.7	0.6	0.7	0.8
Elongation at Break e _b (mm)	2.4	3.4	3.2	3.1	3.5	4.0	2.8	3.4	4.1	3.2	2.8	3.5	4.7
E _t (MPa)	646.0	617.8	621.3	639.3	628.0	585.3	650.3	639.8	734.8	641.5	623.8	642.2	640.7
Thickness (mm)	3.8	3.8	3.8	3.9	3.9	3.9	3.8	3.9	3.9	3.9	3.9	3.9	3.9
L ₀ (mm)	50	50	50	50	50	50	50	50	50	50	50	50	50
Set Strain x (%)	1	1	1	1	1	1	1	1	1	1	1	1	1
Grid	0.8	0.7	0.8	0.8	0.9	0.8	0.7	0.8	1.0	0.7	0.7	0.8	0.8
Yield Strain e _y (%)	4.8	4.7	6.7	5.7	6.4	6.4	4.2	6.8	6.4	4.8	4.5	5.5	4.7
Break Stress s _b (MPa)	19.7	18.9	17.3	14.3	19.5	21.0	18.3	19.8	23.6	18.7	15.6	19.1	19.2
e _b (%)	4.8	6.9	6.4	6.2	7.1	8.1	5.5	6.9	8.3	6.5	5.5	7.0	9.3
Mass (g)	5.7	4.4	3.6	5.1	28.0	4.3	5.6	4.8	5.5	5.4	5.4	5.6	5.1

Table 4 Paired sample test according to Weight and Ultimate_Tensile_Strength_MPa; Engineering_Strain_mm; Strain_Force_kN; Break_Stress_Sb_MPa; Elongation_eb_percent; Tensile_Modulus_Et_MPa

Paired Samples Test										
		Paired Differences					t	df	Sig. (2-tailed)	
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference					
					Lower	Upper				
Pair 1	Weight gr - Ultimate Tensile Strength MPa	-14.663	2.002	.555	-15.872	-13.453	-26.414	12	.000	
Pair 2	Weight gr - Engineering Strain mm	5.746	.120	.033	5.673	5.818	172.350	12	.000	
Pair 3	Weight gr - Strain Force kN	5.580	.130	.036	5.501	5.658	154.869	12	.000	
Pair 4	Weight gr - Break Force Fb kN	5.076	.166	.046	4.976	5.177	110.020	12	.000	
Pair 5	Weight gr - Break Stress Sb MPa	-13.045	2.337	.648	-14.457	-11.633	-20.129	12	.000	
Pair 6	Weight gr - Elongation eb percent	-1.003	1.229	.341	-1.745	-.260	-2.943	12	.012	
Pair 7	Weight gr - Tensile Modulus Et MPa	-633.464	33.400	9.264	-653.648	-613.281	-68.383	12	.000	

p < 0.005

After conducting the experimental tests, calculations were performed to optimize the process, simulate different scenarios, and compare the results. The calculation focused on engineering stress and strain. First, the engineering stress-

strain curves were generated using equations ISO 527-2:2012 [22].

$$\delta_{eng} = \frac{F}{A_0}, \text{ Engineering stress} \quad (1)$$

$$\epsilon_{eng} = \frac{l - l_0}{l_0}. \text{ Engineering strain} \quad (2)$$

As shown in Tab. 4, the statistical analysis using the "paired sample test" demonstrate that most pairs (except Weight_gr - Elongation_eb_percent) show statistically significant differences between weight and other characteristics of the analysed samples.

3 RESULTS AND DISCUSSION

As can be seen from Tab. 3, there are obvious differences between the infill pattern and the results obtained from the testing of the samples.

The statistical analysis using the "paired sample test" on the paired samples (Tabs. 4 and 5.) show statistically

significant differences between weight and other analyzed properties; the same differences are found for infill pattern and other analyzed properties. As can be seen, in Tab. 4, only weight and Elongation do not show statistically significant difference; in Tab. 5, infill pattern and weight and infill pattern and elongation do not show statistically significant differences, suggesting that these three variables are more stable and similar to the Infill Pattern.

3.1 Experimental Engineering Stress-Strain Results

Figs. 4a, 4b, 4c, 4d, and 4e graphically present the experimental results in regard to stress strain engineering for specified samples. In this case, the comparisons focus on samples with varying internal structural designs but can be considered and grouped as samples with approximately similar internal design.

Table 5 Paired sample text according to Infill pattern and Ultimate_Tensile_Strength_MPa; Engineering_Strain_mm; Strain_Force_kN; Break_Stress_Sb_MPa; Elongation_eb_percent; Tensile_Modulus_Et_MPa.

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower	Upper			
Pair 1	Infill pattern - Ultimate Tensile Strength MPa	-13.468	4.636	1.286	-16.269	-10.666	-10.473	12	.000
Pair 2	Infill pattern - Engineering Strain mm	6.941	3.898	1.081	4.585	9.296	6.419	12	.000
Pair 3	Infill pattern - Weight gr	1.195	3.910	1.084	-1.168	3.558	1.102	12	.292
Pair 4	Infill pattern - Strain Force kN	6.775	3.888	1.078	4.426	9.124	6.283	12	.000
Pair 5	Infill pattern - Break Force Fb kN	6.271	3.881	1.076	3.926	8.616	5.827	12	.000
Pair 6	Infill pattern - Break Stress Sb MPa	-11.850	4.296	1.192	-14.446	-9.254	-9.945	12	.000
Pair 7	Infill pattern - Elongation eb percent	.192	3.516	.975	-1.932	2.317	0.197	12	.847
Pair 8	Infill pattern - Tensile Modulus Et MPa	-632.269	32.676	9.063	-652.015	-612.523	-69.766	12	.000

$p < 0.005$

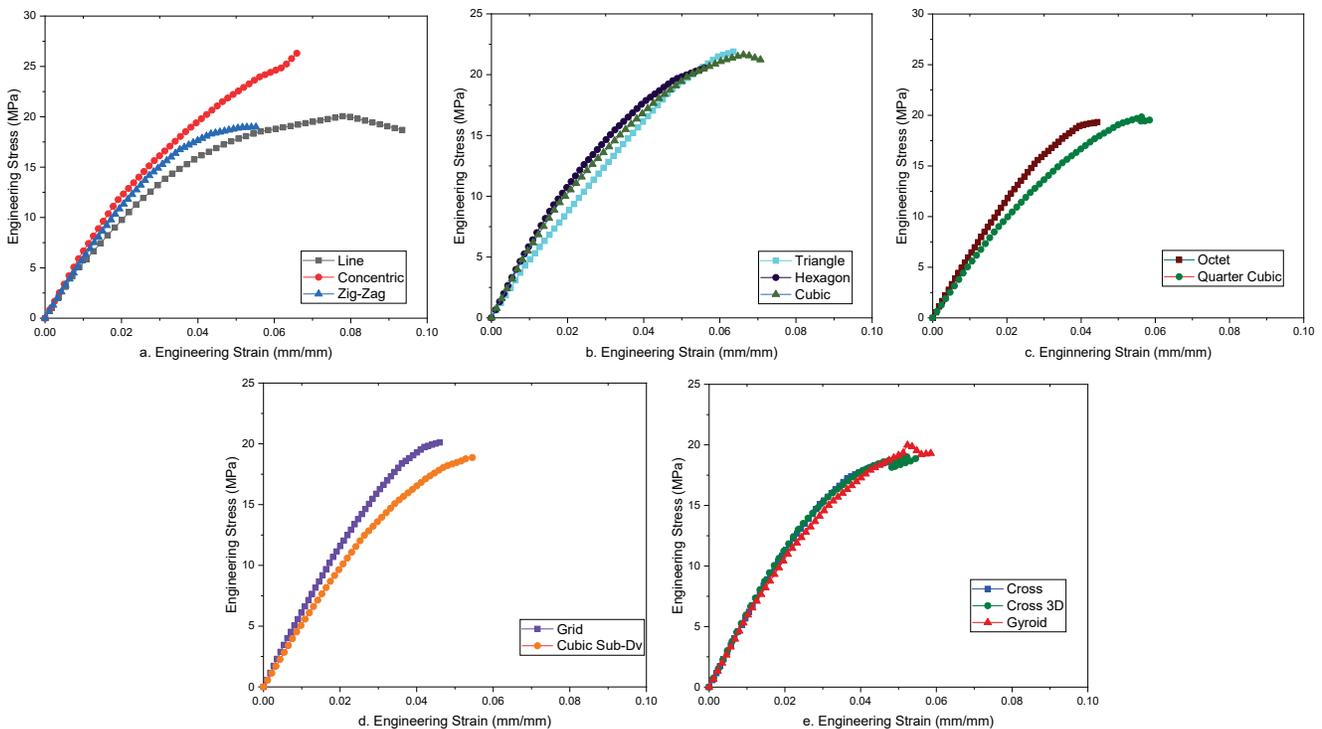


Figure 4 Comparison of samples grouped according to design approximately similar to internal structure mm/mm.

Therefore, their division into groups was made by categorizing them based on their internal shape and comparing them among themselves; comparisons were made for engineering stress strain, where the Fig. 4a shows the comparative results for the samples named Line, Concentric and Zig-Zag, in Fig. 4b the comparative results for Triangle, Hexagon and Cubic samples are presented. Fig. 4c shows the comparative results for the Octet and Quarter Cubic samples, while Fig. 4d shows the comparative results for the Grid and Cubic Subdivision samples, and Fig. 4e shows the comparative results for Cros, Cros 3D and Gyroid samples.

The graphical presentation of the engineering stress-strain curves for the buried samples shows that, with the exception of Fig. 4a, the sample labeled Zig-Zag has a significant difference in terms of stress and strain. In contrast, almost all the lower groups have results that are very close in terms of stress and strain.

3.2 Comparison Results of Ultimate Tensile Strength and Engineering Strain

Comparisons of the results for each sample were made separately for ultimate tensile strength (UTS), expressed in MPa, and for engineering strain, expressed in mm.

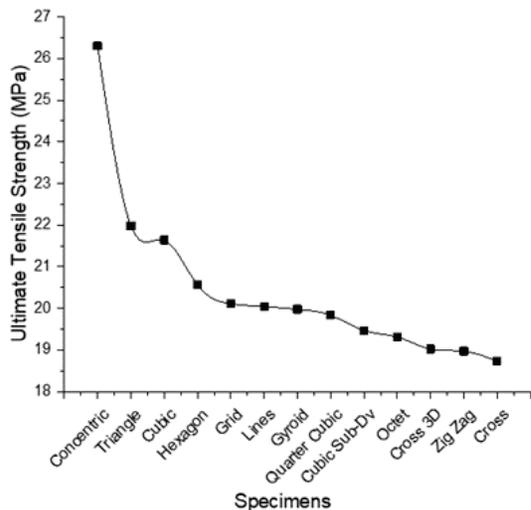


Figure 5 Comparison of samples for Ultimate Tensile Strength (MPa)

Fig. 5 graphically presents the comparative results for Ultimate Tensile Strength (UTS) for the tested samples. It is evident that the highest UTS is presented in the Concentric sample, followed by the Triangle and Cubic samples. In contrast, samples like Hexagon, Grid Line and Gyroid show very similar UTS values. The lowest UTS is found in the Quarter Cubic, Cubic Subdivision, Octet, Cross 3D, Zig-Zag and Cross samples.

Fig. 6 graphically presents the engineering strain results for all samples. In this comparison, the Line infill pattern sample shows a more emphasized elongation than the other samples, while the Concentric, Triangle, and Cubic samples appear to have nearly similar elongation. Whereas samples with approximately similar extension, but lower than those mentioned are Hexagon, Gyroid, Quarter Cubic, Cubic

Subdivision, Cross 3D and Zig-Zag. The samples showing the lowest extension are Grid, Octet, and Cross. It is important to note that samples demonstrating a more pronounced elongation do not mean that they have or should have a higher UTS because in this case, the main factor is the internal structural design. Additional statistical comparisons are presented in Tabs. 4 and 5.

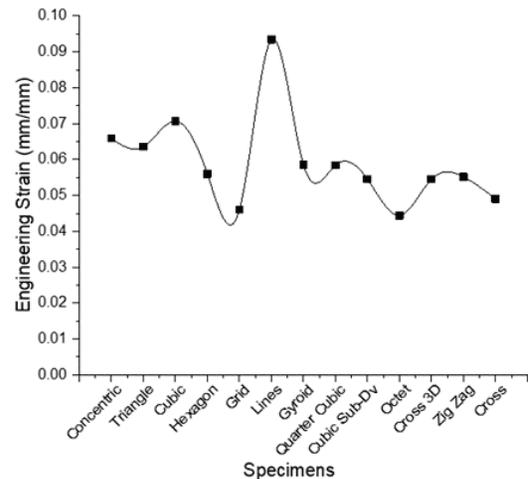


Figure 6 Comparison of samples for Engineering Strain mm/mm

4 CONCLUSIONS

Based on the results of the tested samples, it can be observed that the infill pattern has a particular importance during 3D printing for products that are subject to tensile forces. As seen in Tabs. 4 and 5, most of the samples show statistically significant changes both for the weight of the samples and the infill pattern. The engineering experimental results for stress-strain (Fig. 4) indicate that there are differences among the tested samples as this is evident within the infill pattern groups that are considered roughly similar. Changes were observed in both Engineering Strain (mm) and in Engineering Stress (MPa). Fig. 5 demonstrates that the Concentric filling form shows much higher results in Ultimate Tensile Strength. Meanwhile, Fig. 6 shows that the Line infill pattern displays significantly greater results in engineering strain (mm).

5 REFERENCES

- [1] El Mehtedi, M., Buonadonna, P., El Mohtadi, R., Aymerich, F. & Carta, M. (2024). Surface quality related to machining parameters in 3D-printed PETG components. *Procedia Comput. Sci.*, 232, 1212-1221. <https://doi.org/10.1016/j.procs.2024.01.119>
- [2] Thakar, C. M., Parkhe, S. S., Jain, A., Phasinam, K., Murugesan, G. & Ventayen, R. J. M. (2022). 3d Printing: Basic principles and applications. *Mater. Today Proc.*, 51(1), 842-849. <https://doi.org/10.1016/j.matpr.2021.06.272>
- [3] Vujović, I. (2015). The introduction of 3D printing into the Maritime Industry. *Trans. Marit. Sci.*, 4(1), 86-87.
- [4] Karad, A. S., Sonawwanay, P. D., Naik, M. & Thakur, D. G. (2023). Experimental study of effect of infill density on tensile

- and flexural strength of 3D printed parts. *J. Eng. Appl. Sci.*, 70(1), 104. <https://doi.org/10.1186/s44147-023-00273-x>
- [5] Cabreira, V. & Santana, R. M. C. (2020). Effect of infill pattern in Fused Filament Fabrication (FFF) 3D Printing on materials performance. *Matéria (Rio Janeiro)*, 25(3). <https://doi.org/10.1590/S1517-707620200003.1126>
- [6] Liacouras, P. C., Huo, E. & Mitsouras, D. (2024). 3D Printing Technologies and Materials. In *3D Printing at Hospitals and Medical Centers: A Practical Guide for Medical Professionals*, Springer, 47-69. https://doi.org/10.1007/978-3-031-42851-7_4
- [7] Afonso, J. A., Alves, J. L., Caldas, G., Gouveia, B. P., Santana, L. & Belinha, J. (2021). Influence of 3D printing process parameters on the mechanical properties and mass of PLA parts and predictive models. *Rapid Prototyp. J.*, 27(3), 487-495. <https://doi.org/10.1108/RPJ-03-2020-0043>
- [8] Blanco, I. (2020). The use of composite materials in 3D printing. *J. Compos. Sci.*, 4(2), 42. <https://doi.org/10.3390/jcs4020042>
- [9] Iftekar, S. F., Aabid, A., Amir, A. & Baig, M. (2023). Advancements and limitations in 3D printing materials and technologies: a critical review. *Polymers (Basel)*, 15(11), 2519. <https://doi.org/10.3390/polym15112519>
- [10] Soni, R. D. (2022). Geometric Stability of Parts Produced by 3D Printing. *Tehički vjesnik*, 29(1), 23-29. <https://doi.org/10.17559/TV-20191101110214>
- [11] Dizon, J. R. C., Espera Jr, A. H., Chen, Q. & Advincola, R. C. (2018). Mechanical characterization of 3D-printed polymers. *Addit. Manuf.*, 20, 44-67. <https://doi.org/10.1016/j.addma.2017.12.002>
- [12] Kain, S., Ecker, J. V., Haider, A., Musso, M. & Petutschnigg, A. (2020). Effects of the infill pattern on mechanical properties of fused layer modeling (FLM) 3D printed wood/polylactic acid (PLA) composites. *European Journal of Wood and Wood Products*, 78, 65-74. <https://doi.org/10.1007/s00107-019-01473-0>
- [13] Lee, J.-Y., An, J. & Chua, C. K. (2017). Fundamentals and applications of 3D printing for novel materials. *Appl. Mater. Today*, 7, 120-133. <https://doi.org/10.1016/j.apmt.2017.02.004>
- [14] Shahrubudin, N., Lee, T. C. & Ramlan, R. (2019). An overview on 3D printing technology: Technological, materials, and applications. *Procedia Manuf.*, 35, 1286-1296. <https://doi.org/10.1016/j.promfg.2019.06.089>
- [15] Anderson, T. L. & Anderson, T. L. (2005). *Fracture mechanics: fundamentals and applications*. CRC press. <https://doi.org/10.1201/9781420058215>
- [16] D. Pezer, F. Vukas, and M. Butir, 2022 "Experimental study of tensile strength for 3D printed specimens of HI-PLA polymer material on in-house tensile test machine. *Technium*, 4(10), 197-206. <https://doi.org/10.47577/technium.v4i10.7927>
- [17] Faidallah, R. F., Hanon, M. M., Vashist, V., Habib, A., Szakál, Z. & Oldal, I. (2023). Effect of Different Standard Geometry Shapes on the Tensile Properties of 3D-Printed Polymer. *Polymers (Basel)*, 15(14), 3029. <https://doi.org/10.3390/polym15143029>
- [18] Lalegani Dezaki, M. & Mohd Ariffin, M. K. A. (2020). The effects of combined infill patterns on mechanical properties in FDM process. *Polymers (Basel)*, 12(12), 2792. <https://doi.org/10.3390/polym12122792>
- [19] Soufivand, A. A., Abolfathi, N., Hashemi, A. & Lee, S. J. (2020). The effect of 3D printing on the morphological and mechanical properties of polycaprolactone filament and scaffold. *Polym. Adv. Technol.*, 31(5), 1038-1046. <https://doi.org/10.1002/pat.4838>
- [20] Đurović, S., Lazarević, D., Milovanović, V., Mišić, M. & Šarkoćević, Ž. (2024). The Influence of the Number of Layers and the Raster Angle on the Mechanical Properties of 3D Printed Materials. *Tehički vjesnik*, 31(6), 1892-1897. <https://doi.org/10.17559/TV-20240325001430>
- [21] Bianchi, I., Mancina, T., Mignanelli, C. & Simoncini, M. (2024). Effect of nozzle wear on mechanical properties of 3D printed carbon fiber-reinforced polymer parts by material extrusion. *Int. J. Adv. Manuf. Technol.*, 130(9), 4699-4712. <https://doi.org/10.1007/s00170-024-13035-7>

Authors' contacts:

Rrahim Sejdiu, Prof. Assoc. Dr.
University of Applied Sciences in Ferizaj - Kosovo,
Str. "University", 70000 Ferizaj, Kosovo
Tel./Fax, e-mail: +38345207276
rrahim.sejdiu@ushaf.net

Xhemajl Mehmeti, Prof. Assist. Dr.
(Corresponding author)
University of Business and Technology,
Lagjia Kalabria, 10000 Prishtine - Kosovo
Tel./Fax, e-mail: +38344981727
xhemajl.mehmeti@ubt-uni.net

Flamur Salihu, Prof. Assist. Dr.
University of Applied Sciences in Ferizaj - Kosovo
Str. "University", 70000 Ferizaj, Kosovo
Tel./Fax, e-mail: +38344428613
flamur.salihu@ushaf.net

Labinot Topilla, Prof. Assist. Dr.
University of Applied Sciences in Ferizaj - Kosovo,
Str. "University", 70000 Ferizaj, Kosovo
Tel./Fax, e-mail: +38344698611
labinot.topilla@ushaf.net

Drin Krasniqi, MSc.
University of Pristina,
"George Bush", Nr. 31, 10000 Prishtine - Kosovo,
Tel./Fax, e-mail: +38348800882
drin.krasniqi@uni-pr.edu

Mohamad El Mehtedi, Prof. Dr.
University of Cagliari,
Via Marengo, 2, 09123 Cagliari - Italia
Tel./Fax, e-mail: +39 070 675 5721
m.elmehtedi@unica.it

Agron Bajraktari, Prof. Dr.
University of Applied Sciences in Ferizaj - Kosovo,
Str. "University", 70000 Ferizaj, Kosovo
Tel./Fax, e-mail: +3834599666
agron.bajraktari@ushaf.net

Safety Management at Tower Crane in Construction Site

Sangchul Kim, Seunghyun Yeo, Seok-Heon Yun*

Abstract: Tower cranes are critical in high-rise construction but present significant safety risks due to their structural characteristics and operational complexity. This study analyzed 73 serious tower crane accident cases reported to the Korea Occupational Safety and Health Agency between July 2012 and July 2020. A mixed-methods approach was applied, combining quantitative accident data analysis with surveys and expert interviews. Statistical proportions were calculated to identify high-risk factors. Lifting operations accounted for the highest proportion of accidents (45.3%), with falls being the most common disaster type (35.8%). The leading cause of accidents was non-compliance with safe work procedures (38.8%). Survey and interview findings highlighted insufficient training, lack of standardized signal protocols, and limited on-site safety supervision. Strengthening training programs, standardizing signal operations, and enforcing systematic safety management can significantly reduce tower crane accidents. Policy recommendations and a framework for continuous monitoring are proposed.

Keywords: construction site; disaster case; safety accident; securing safety; tower crane

1 INTRODUCTION

Tower cranes are indispensable for modern high-rise construction, enabling the lifting and transportation of heavy materials where mobile cranes are impractical [1, 2]. However, their unstable superstructure and operational hazards contribute to a persistently high accident rate. According to the Korea Occupational Safety and Health Agency (KOSHA), an average of 6.55 fatalities occur annually in tower crane operations [3]. Although tower cranes offer high work efficiency and excellent operability on construction sites, they continue to be associated with frequent major accidents each year, becoming a persistent social issue. This is largely due to the inherent difficulty for safety personnel, such as safety managers and on-site supervisors, to easily access and systematically manage the entire lifecycle of tower crane operations—from installation to dismantling. Recent research highlights that current safety management practices remain fragmented across the pre-construction, construction, and post-construction phases. A comprehensive study emphasized the need for integrated safety technologies, including proactive planning tools, anti-collision systems, and stability control mechanisms, to address these persistent safety gaps and improve the overall management of tower crane operations [4, 5].

The scope of this paper is to identify the type of accident by analyzing serious accident cases related to tower cranes (2012-2020). In addition, it is intended to investigate the safety and work status of tower cranes through analysis by referring to external data that conducted surveys and interviews with tower crane drivers. As a result, we would like to suggest ways to improve safety management for tower crane work in domestic construction sites.

The research methods are as follows.

- 1) It identifies research trends on safety management of tower cranes through domestic prior papers and literature.
- 2) The type is identified by analyzing the current status of major disasters related to tower cranes from 2012 to 2020.

- 3) By analyzing through Excel by referring to external surveys and interview results, the work environment and problems of the site are identified.
- 4) Through the analysis of the current status of major disasters in tower cranes and the survey results of workers, measures to improve the parts necessary for systematic safety management of tower cranes are presented.

2 THEORETICAL BACKGROUND

2.1 Literature Review

The review of previous studies in this study is as follows. Jeon (2013) presented basic data for safe equipment operation by evaluating risks through investigations of accident cases that occurred at the site and presenting accident cases by grade to reduce accidents caused by construction equipment with a Study on Risk Assessment for Reduction of Safety Accidents in Construction Equipment [6].

Lim (2012) analyzed whether users related to power cranes used in the field acquired national qualifications and their influence through empirical analysis of tower crane disaster prevention factors, and suggested ways to improve national qualifications by analyzing the effect of tower crane qualifications on site disaster prevention [7].

Shim (2011) establishes a model for the safety diagnosis management system by analyzing the disaster factors of the tower crane with <a study on the domestic construction site tower crane safety diagnosis management system model> and confirms the effectiveness and efficiency generated through the system development. In addition, a case was presented to check the state of the tower crane through a remote system and to establish the field situation as a real-time monitoring diagnosis system [8].

Park (2012) is a Study on the Development of Algorithm Model for the Stability Review of Upward Tower Crane [9]. When installing a tower crane, an appropriate crane can be selected through data analysis of the crane, and through this data, the person in charge of installation can relieve anxiety

about the crane and secure stability. And through this method, the expected effect of reducing construction costs and securing the crane installation person in charge directly reviewing the crane's equipment appropriateness was presented [10].

While prior studies have addressed mechanical safety

and risk assessment domestically, there is limited research integrating operational, managerial, and training factors, particularly in the context of international safety standards. This study addresses this gap by analyzing accident cases and surveying industry stakeholders to identify key areas for safety improvement.



Figure 1 "T-type" and "L-type tower crane", Topless crane

2.2 Overview of Tower Crane

A tower crane is a construction equipment belonging to a construction machine crane, and it is a crane with the characteristics of a three-dimensional exercise work that moves up and down, moves forward and backward along the trolley, moves forward and backward, and rotates the upper part to move the salvage to the desired location [11]. In general, a tower crane can adjust the height of the tower, providing the convenience of double work that is continuously required in places where it is difficult to construct structures and use mobile cranes that change the height of the workplace as the field work process progresses. Depending on the model, it is widely used for narrow space work and high-rise construction work in urban areas. Tower cranes are designed to be used appropriately according to the purpose of use and are generally classified according to their appearance. Tower cranes are largely divided into three categories: T-type, L-type (LUFFING JIG), and Topless cranes, as shown in Fig. 1. The "T-type" tower crane is the main form of the tower crane, and the jib is fixed and is mainly used when there are no obstacles within the working radius. In the event of interference with other buildings during construction, the "L-type" tower crane can move work like a mobile crane by moving the jib up and down with the selected equipment. A topless crane is a crane without a top and is used for lifting heavy objects or when there are problems such as altitude restrictions.

2.3 Safety Rules for Tower Crane Work

Safety rules form the foundation of safety management activities. True safety can only be achieved when both employers and workers fully comply with these rules. In the context of crane operations, safety protocols can be categorized into several core areas: general work rules, rules for operation in strong winds, tower crane electric shock prevention measures, driving safety protocols, and signal

safety procedures [12]. Recent studies also emphasize the importance of incorporating dynamic environmental factors—especially wind forces—into crane safety management. Wind-induced pressure can significantly affect the crane's operational stability, control accuracy, and lifting torque. Therefore, real-time control adjustments are essential to prevent tipping and ensure safe load handling [13].

Based on these categories and engineering insights, the following safety rules are recommended for crane work:

- 1) Observe the matters specified in the safety rules, such as overload and limitation of inclination angles
- 2) Ensure transition and take necessary action in case of user replacement
- 3) To lift the crane, use the designated ladder
- 4) Check wire ropes, clutches, controllers, brakes, etc. every day before starting work
- 5) When driving the crane, illuminate the horn or horn to notify
- 6) Safety signs must be attached during repair inspection
- 7) In case of winding, work is carried out by keeping the cargo straight at the center of the hook
- 8) Care must be taken to ensure no workers are on board the cargo
- 9) Crane users must drive in accordance with the signal of the signal count
- 10) No rapid movement while driving
- 11) If it stops while driving, place the controller in the stop position and lower the main switch to check the stop
- 12) Prohibited inspection, feeding oil, etc. while driving
- 13) Do not leave the driver's room. If you leave, make sure to lower the switch to stop.

Integrating environmental considerations—such as wind gusts, load swing dynamics, and torque control—into these procedural rules enhances crane safety and performance. Such an approach is especially important for high-risk conditions often encountered in large construction or offshore environments [13].

2.3.1 General Working Rules

- 1) Crane users and signal numbers are checked for daily safety checks before work and recorded in the log for maintenance (reported to the management supervisor if abnormalities are confirmed)
- 2) Check the signal between the crane user and the signal number and operate the equipment when there is no abnormality
- 3) Crane users wear prescribed safety helmets, clothing, seat belts, and safety shoes
- 4) Management supervisor checks the health status of the crane user and reflects it in the work arrangement
- 5) Stretching before work, and strict eating and resting time to prevent accumulated accidents from occurring due to work fatigue.

2.3.2 Rules for Strong Winds

- 1) During work, crane users and signal numbers are frequently checked for weather changes, decelerate in the event of strong winds, and be careful not to push the work in the direction of the wind
- 2) Crane stops working at wind speed of 15m/sec or more
- 3) Operate according to the instructions of the person in charge of management during the emergency process
- 4) Refer to the weather information in advance and take measures in accordance with the safety manual in the event of a bad weather.

2.3.3 Tower Crane Electric Shock Prevention Rules

- 1) Distribution panel box for managing the electrical ground wire of the crane, main line installed on the fixture and avoiding connections to places where there is frequent movement
- 2) Double installation of the electrical ground wires of the tower crane, avoid installation in rusting areas and painting areas, and attach classification stickers
- 3) When installing or disassembling the electric ground wire of the tower crane, it must be requested to the person concerned with the electricity to carry out the work.

2.3.4 Safety Rules for Tower Crane Operation

- 1) Crane operation must be operated only by licensed authorized persons, and it is prohibited to operate arbitrarily without the signal of the number of signals
- 2) Stop operation and report to the administrator if there is an abnormality in the overload protection device or machine room during operation
- 3) Prohibition of rapid operation, such as rapid acceleration and sudden acceleration when driving
- 4) When driving and turning cranes, check for obstacles in the direction of movement of cranes and conduct driving and turning
- 5) Avoid unidentifiable levels of tinting in the driver's

compartment, such as coffee pots, DMBs, radios, laptops, etc., which can interfere with operations.

2.4 Risk of Tower Crane Work

Tower crane work is used to move objects that need to be moved to other places. Such tower crane work is repeatedly performed by a user or a signal number who hooks directly, hangs heavy objects, raises them vertically, and puts them back down at the point after moving. The types of disasters that occur when moving heavy objects within a certain space are mainly falling objects and stenosis disasters. Tower cranes lift not only mold materials, rebar, and general materials, but also large mold members for external work that are affected by wind or outside air, and because they are not controlled during lifting, falling accidents and stenosis accidents can occur. If an unspoken external force, such as a hook on the tower crane or colliding with nearby obstacles, is applied instantaneously, it can lead to a fall and collapse accident of the entire tower crane. Tower crane disasters are more often known as large accidents and personnel accidents in the event of an occurrence than general disasters, and are often known to the media, which causes social issues and criticism [14].

In order to prevent serious accidents in tower cranes, a clear plan, inspection, and training must be preceded before work, but there are many accidents because the conditions at the current domestic site do not meet this. Unlike general industrial sites, due to the nature of construction sites, it is difficult to manage personnel because a large number of various personnel of the type of construction are used in a short period of time and workers are replaced frequently [15]. In the case of dismantling the tower crane installation, it is necessary to install it as soon as possible because it receives a limited amount of money, but it tends to be rushed because it is more profitable. In addition, because employment is unstable, frequent personnel changes and unskilled workers are sometimes put in. Drivers and signal drivers who use tower cranes after installation also have a social problem that their employment is not constant [16].

3 CASE OF TOWER CRANE DISASTER

To analyze tower crane accidents, the study was conducted in the following four stages.

- 1) Data Sources: Accident data were obtained from the Korea Occupational Safety and Health Agency's official casebook of serious accidents (2012–2020), comprising 73 major tower crane incidents.
- 2) Survey Design: A structured questionnaire was distributed to installation/dismantling workers, crane operators (signalers), and safety managers. The questionnaire covered accident causes, safety training adequacy, and operational procedures. Responses were measured on a 5-point Likert scale.
- 3) Expert Interviews: Semi-structured interviews with three industry experts were conducted to validate survey findings and explore improvement measures.
- 4) Analysis: Proportions were calculated for each accident

cause category. No probabilistic inference was performed; instead, descriptive statistics were used. Survey and interview data were analyzed separately to distinguish quantitative and qualitative findings.

3.1 Method of Analysis of Tower Crane Disaster Data Sources

In order to analyze the case of tower crane accidents, the case of major accidents related to tower cranes among industrial accidents of the Korea Safety and Health Agency was investigated.

In the past 9 years, 73 serious accidents occurred, and they were analyzed by accident cause. When analyzing disaster cases by cause, the classification criteria were prepared according to the criteria of the Korea Safety and Health Agency.

Accident data were obtained from the Korea Occupational Safety and Health Agency's official casebook of serious accidents (2012–2020), comprising 73 major tower crane incidents.

3.2 Results of Analysis of Tower Crane Disaster Case

3.2.1 Analysis by Job Type

Tab. 1 is the result of analyzing accident cases by work type. During the work, 35 accidents (47.9%) occurred in 'nurturing work', and when it comes to the cause of accidents that appeared during the lifting operation, it was found that there were more accidents that caused the number of signals with the task manager than the driver's carelessness, and it is thought that thorough management of the task manager is necessary to prevent such accidents. In addition, in the installation, elevation, and dismantling work, 35 accidents occurred despite repeated work by professional workers, which is thought to be the cause of safety insensitivity due to repetitive work progress, so work-related education is considered necessary.

Table 1 Analysis of disaster cases by job type

	Installation & Dismantling	Maintenance	Upward	Reclamation	Etc.	Sum
The number of disasters	17	2	18	35	1	73
Ratio	23.2%	2.7%	24.6%	47.9%	1.6%	100%

- 1) **Accident case analysis:** Lifting operations were responsible for 45.3% of accidents, followed by installation/dismantling (23.2%) and climbing operations (24.6%). Falls were the most frequent disaster type (35.8%). The leading contributing factor was non-compliance with safe work procedures (38.8%).
- 2) **Cause of accident during installation and dismantling operation:** The main causes of accidents in installation and dismantling work were 10 cases of non-compliance with safety rules (58.8%), 5 cases of non-compliance with work processes (29.4%), and 2 cases of inexperience (11.8%).
- 3) **Cause of accident during salvage operation:** The

cause of the salvage operation was 14 cases (40%), 10 cases of unidentified line hooks (28.5%), 5 cases of non-compliance with the signal (14.2%), 4 cases of unidentified lifting (11.4%) and 2 cases of equipment (5.9%), and the control around the tower crane is exclusively managed by the signal number.

- 4) **Cause of accident during ascending operation:** The causes of accidents in ascending work were 12 cases (67%) of non-compliance with the work process and 6 cases (33%) of non-compliance with safety rules, and it is thought that the thorough management of the work manager is necessary to prevent accidents.

3.2.2 An Analysis of the Cause of Work Accident

As a result of analyzing the cause of the tower crane accident as the cause of the physical accident and the cause of the human accident, 10 physical factors (13.6%) and 63 (86.4%) human factors such as unsafe behavior of workers and non-compliance with the work order among the functions of the tower crane.

3.2.3 Analysis by the Type of Disaster

As a result of the analysis by type of tower crane disaster, 27 falls (36.9%), 23 falls and rain (31.5%), 11 falls and falls (15%), 9 stenosis (12.3%), and 3 falls (4.3%) were found. The risk of falls was high during installation, assembly, and dismantling, and the risk of falls and rain was high during the salvage operation. This is judged to reflect the characteristics of the work and should be focused on preventing accidents from falling by establishing countermeasures.

3.3 Tower Crane Disaster Risk Assessment

Risk factors were derived through analysis by work type through accident cases, and 73 cases of major tower crane accidents provided by the Korea Safety and Health Agency for risk calculation occurred between 2012 and 2020. Since the proportion of occurrence per one out of 73 major accidents was 1.4%, it was assumed as the lower limit that it hardly occurred, and the largest number of accidents was assumed to be 19.2%, with 14 cases occurring due to poor control during the salvage operation. Risk assessment was conducted based on the frequency of occurrence by simply dividing 1.4% to 19.2%.

The rating calculation for the proportion of occurrence of a risk, and the proportion of occurrence is shown by dividing the total number of accidents of tower cranes by the number of accidents by risk factor for each type of work in the tower crane work where the disaster occurred. As a result, it was shown in Tab. 2.

As a result of calculating the risk level through the analysis of the probability of occurrence of a serious accident by risk factor, the most dangerous risk factor among the entire tower crane work was investigated as 'bad peripheral control', followed by 'non-compliance with the work process' and 'bad line-up'. Therefore, it is judged that the role of the signal numbers that control the surroundings is important.

Table 2 Proportion and risk calculation of major disasters

	Cause	Number of serious accidents	Proportion (%)	Ranking	Grade
Lifting	Poor control	14	19.2	1	1
	Unidentified line hook	10	13.7	3	2
	Failure to comply with the signal	5	6.8	6	3
	Unidentified salvage material	4	5.5	8	3
	Defects in equipment	2	2.7	9	4
Installation & Dismantling	Non-compliance with safety rules	10	13.7	3	2
	Non-compliance with the work process	5	6.8	6	3
	inexperience in the work	2	2.8	9	4
Upward	Non-compliance with safety rules	6	8.2	5	3
	Non-compliance with the work process	12	16.4	2	1
Maintenance		2	2.8	9	3
Etc.		1	1.4	12	4
	Sum	73	100%		

4 SURVEY

4.1 Survey Overview

4.1.1 Survey Method

The survey was conducted from September 2020 to October 2020, centered on the sites of large domestic construction companies, and the responses of the questionnaire were a total of 250 people, of which a total of 193 copies were conducted, excluding unfaithful and unclear respondents. The survey results conducted by external literature were used and analyzed through Excel. In addition, expert interviews are conducted to derive opinions on safety management improvement measures and additional alternatives, and to determine whether the improvement measures are appropriate.

4.1.2 Survey Contents

This questionnaire was composed by reflecting the analysis results of serious accident cases in Chapter 3. It was largely classified into general (common) and tower crane work. Tower crane work items were prepared according to the work characteristics of the installation/demolition person, user, and safety manager, and each item was constructed based on items that affect safety work, such as the cause of accidents, education system, work command relationship, and contract business format related to tower crane work. Using a 5-point scale, the questionnaire was marked with 5 points for "very yes", 4 points for "yes", 3 points for "normal", 2 points for "not very much", and 1 point for "not very much". Each detail is shown in Tab. 3.

Table 3 Details of the installation/demolition worker/user/safety manager

Area	Item	Details
General	Common	Age
		Carrier
Tower crane work	Installation and Demolition worker	Investigation of the necessity of education (Safety and Health Agency)
		The Impact of the Wage System on Safety
		Examination of the importance of work conditions in the field
		Investigation of Safety Violators
		Survey on the Quality of Field Training
		Investigation of the necessity of education (Safety and Health Agency)
	User	Survey of the importance of education
		Investigation into the need for education
		Necessity of Qualification Specialization
		Examination of the safety knowledge level of work
		Investigation of Safety Violators
		Investigation into the need for specialized knowledge
	Safety manager	Investigation into the need for specialized knowledge
		Survey of the level of expertise
		High-risk process investigation
		Investigating the necessity of improving signal receiver quality
		Investigation of key items of disaster prevention activities
		Effectiveness of the Construction safety plan

4.2 Survey Results

4.2.1 Age

According to the survey respondents' age distribution, many safety managers (62%) were aged 31 to 40 at the site, many were aged 41 to 50 (50%) who played a control role when using tower cranes, and many elderly people were older than 50 (44.4%) who were engaged in practical risk processes such as complaint work.

Table 4 Ages in workers

Age	Installation / dismantling worker		Crane operators (Signalers)		Safety managers	
Under 25	0	0.0%	1	1.2%	1	1.4%
26~30	0	0.0%	1	1.2%	2	2.8%
31~40	5	13.9%	17	20.2%	44	62.0%
41~50	15	41.7%	42	50.0%	16	22.5%
Over 50	16	44.4%	23	27.4%	8	11.3%
Sum	36.0	100.0%	84.0	100.0%	71.0	100.0%

4.2.2 Career

Safety managers were able to receive realistic responses as 60% of them had more than 6 years of experience, and workers were evenly distributed by experience. On the other hand, installation and dismantling workers were found to be in the dangerous process, with few new workers and a large number of skilled workers for more than 6 years.

Table 5 Career in workers

Career	Installation / dismantling worker		Crane operators (Signalers)		Safety managers	
	Number	%	Number	%	Number	%
Under 1yr	0	0.0%	14	16.7%	9	12.5%
1~3	3	8.3%	16	19.0%	14	19.4%
4~5	1	2.8%	15	17.9%	6	8.4%
6~10	15	41.7%	16	19.0%	19	26.4%
Over 11yr	17	47.2%	23	27.4%	24	33.3%
Sum	36.0	100.0%	84.0	100.0%	71.0	100.0%

4.2.3 The Primary Cause of Tower Crane Accidents

A survey on the primary cause of tower crane accidents revealed that safety managers, crane operators (signalers), and installation/dismantling workers all cited "errors in work methods and behavior during work."

Regarding the response to the question, "What is the primary cause of tower crane accidents?", a chi-square test was performed across safety managers, installers and dismantlers, and crane operators (signalers), revealing a significant difference ($p = 0.047$). Overall, errors in work methods and behavior during work were identified as the most common cause of accidents. Installers, dismantlers, and safety managers identified defects in the equipment defects as the second most common cause, while crane operators (signalers) identified defects in the lifting material defects as the second most common cause.

Table 6 Analysis of variance by respondent on the main causes of tower crane accidents

Cause	Installation / dismantling worker		Crane operators (Signalers)		Safety managers		χ^2	p -value
	Number	%	Number	%	Number	%		
Equipment defects	7	21.21	3	3.80	10	14.49	12.78	0.047
Errors in work methods and behavior during work	22	66.67	60	75.95	47	68.12		
Lifting material defects	0	0.00	10	12.66	7	10.14		
Surrounding environment, including climate	4	12.12	6	7.59	5	7.25		

4.3 Expert Interview Review

4.3.1 Review Overview

Expert interviews were conducted through e-mail from October 11 to October 13, 2023. The interviewees were a safety manager and a total of two construction site unloading workers. The questionnaire data was conducted using official form, and through the analysis of the survey results, opinions on the improvement plan and additional alternatives were derived, and the appropriateness of the improvement plan was determined.

4.3.2 Expert Interview Results

Experts emphasized the ineffectiveness of purely theoretical safety training, advocating for participatory, hands-on education. They also highlighted the lack of standardized signal protocols and insufficient staffing as recurring safety risks.

4.4 Discussion

As a result of combining the survey and interviews, the survey was conducted on installation and dismantling workers, use (signal number) workers, and safety managers. As a result of the survey, 68% of safety managers thought it was "unsafe work" and 45% of use (signal) workers thought it was "safe," and 30% of installation and dismantling workers were divided into "unsafe" and "safe," respectively. In addition, in the case of users, the more experience they had, the higher the safety. The main cause of the tower crane accident was "error of work method and behavior during work" in all three groups.

Whether workers who have been installed and

dismantled have experienced violations of safety rules, "Work rules" 54.3% of them said they had experience of violation, more than half of them. Currently, new workers who are installed and dismissed are entitled to work after completing one training session in accordance with the "Rules on the Restriction of Employment in Hazardous and Dangerous Works" and there is no follow-up management. Therefore, in order to prevent accidents, it is considered that regular maintenance training is conducted in addition to new training to minimize safety symptoms, and regular management and training such as the characteristics of equipment by manufacturing company and job descriptions of new equipment is necessary.

This study confirms that human factors, particularly procedural non-compliance and inadequate training, are the dominant causes of tower crane accidents. These findings align with international research, which also identifies training quality, communication protocols, and supervision as critical safety determinants. Integrating hands-on, scenario-based training and enforcing standardized signal operations could bridge current safety gaps. Furthermore, aligning domestic safety regulations with international best practices could enhance both compliance and accident prevention.

5 CONCLUSION

The Tower crane is a dynamometer crane mainly used in high-rise building construction and is a high-risk work. Therefore, this paper attempted to identify problems with tower crane work through an analysis of major accident cases related to tower cranes over the past 13 years and a survey of tower crane-related workers, and to suggest improvement measures accordingly.

The analysis of tower crane accidents in Korea highlights the urgent need for structured safety interventions. Key recommendations include: (1) institutionalizing participatory safety training; (2) mandating certification for signalers through public institutions; (3) enhancing safety managers' expertise through specialized curricula; and (4) enforcing continuous monitoring and supervision.

Limitations include the reliance on domestic accident data and self-reported survey responses. Future research should incorporate international comparative analyses and longitudinal tracking of safety interventions.

Acknowledgment

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. RS-2024-00338908).

6 REFERENCES

- [1] Li, F., Shi, W., Tu, Y., & Zhang, H. (2023). Automated methods for indoor point cloud preprocessing: Coordinate frame reorientation and building exterior removal. *Journal of Building Engineering*, 76, 107270. <https://doi.org/10.1016/j.jobe.2023.107270>
- [2] Kim, H., Hwang, Y., & Rhee, S. (2019). Automatic Building Modeling Method Using Planar Analysis of Point Clouds from Unmanned Aerial Vehicles. *Korean Journal of Remote Sensing*, 35(6_1), 973-985. (in Korean) <https://doi.org/10.7780/KJRS.2019.35.6.1.8>
- [3] Kang, Y., & Yun, S. (2024). Regression Analysis-Based Cost Estimation Method for Detailed Work Types in General Research Building Construction. *The Journal of Korean Innovation Industry Society*, 2(3), 133-139. <https://doi.org/10.30032/JIIT.2024.2.3.133>
- [4] Yun, Y., & Yun, S. (2024). A Comparative Study of Missing Data Imputation Methods for Estimating Construction Duration. *The Journal of Korean Innovation Industry Society*, 2(1), 17-22. <https://doi.org/10.60032/JIIT.2024.2.1.017>
- [5] Ali, A. H., Zayed, T., Wang, R. D., & Kit, M. Y. S. (2024). Tower crane safety technologies: A synthesis of academic research and industry insights. *Automation in Construction*, 157, Article 105429. <https://doi.org/10.1016/j.autcon.2024.105429>
- [6] Jeon, H. (2013). A Study on the Risk Evaluation for the Reduction of Safety Accidents in Construction Equipment. *Master's thesis*, Incheon University, 25 p.
- [7] Lim, S. (2012). A Study on the Improvement of National Qualifications through Empirical Analysis of Tower Crane Disaster Prevention Factors. *Dissertation*, (Kyonggi) Myongji University, 35 p.
- [8] Shim, K. (2011). A Study on the Risk Assessment for the Disaster Prevention of Tower Crane. *Dissertation*, Incheon University, 18 p.
- [9] Park, J. (2012). A Study on the Investigation and Analysis of the Working Environment of Tower Crane Drivers. *Master's thesis*, Kyongnam University, 17 p.
- [10] Kim, E. (2012). A Study on Safety Measures through Crane Disaster Analysis. *Master's thesis*, Seoul National University of Science and Technology, 33 p.
- [11] Lee, H. (2014). A Study on the Cause of Tower Crane Accident Analysis and Countermeasures. *Master's thesis*, Seoul National University of Science and Technology, 37p
- [12] Gheisari, M., & Esmacili, B. (2019). Applications and requirements of unmanned aerial systems (UASs) for construction safety. *Safety Science*, 118, 230-240. <https://doi.org/10.1016/j.ssci.2019.05.015>
- [13] Urbaś, A., & Stadnicki, J. (2024). Dynamics of a knuckle crane carrying a load subjected to wind pressure. *Multibody System Dynamics*. <https://doi.org/10.1007/s11044-024-10029-x>
- [14] Dutta, S., Cai, Y., Huang, L., & Zheng, J. (2020). Automatic re-planning of lifting paths for robotized tower cranes in dynamic BIM environments. *Automation in Construction*, 110, 102998. <https://doi.org/10.1016/j.autcon.2019.102998>
- [15] Al Hattab, M., Zankoul, E., & Hamzeh, F. R. (2017). Near-real-time optimization of overlapping tower crane operations: A model and case study. *Journal of Computing in Civil Engineering*, 31(4), 05017001. [https://doi.org/10.1061/\(ASCE\)CP.1943-5487.0000666](https://doi.org/10.1061/(ASCE)CP.1943-5487.0000666)
- [16] Han, S., Lei, Z., Hermann, U., Bouferguene, A., & Al-Hussein, M. (2021). 4D-based automation of heavy lift planning in industrial construction projects. *Canadian Journal of Civil Engineering*, 48(9), 1115-1129. <https://doi.org/10.1139/cjce-2019-0825>

Authors' contacts:

Sangchul Kim, Professor
School of Architecture & Design Convergence,
Hankyong National University,
327, Jungang-ro, Anseong-si, Gyeonggi-do 17579, Korea
sckim08@hknu.ac.kr

Seunghyun Yeo, Student
Department of Architecture Engineering,
Gyeongsang National University,
501 Jinjudaero, Jinju-si, Gyeongsangnam-do, 52828, South Korea
shy@gnu.ac.kr

Seok-Heon Yun, Professor
(Corresponding author)
Department of Architecture Engineering,
Gyeongsang National University,
501 Jinjudaero, Jinju-si, Gyeongsangnam-do, 52828, South Korea
Mobile Phone: +82-010-8860-5626
gfyun@gnu.ac.kr

Development of Drowsy Driving Detection System Using EEG

Ssang-Hee Seo

Abstract: Drowsy driving is a major contributor to serious traffic accidents, highlighting the urgent need for effective real-time detection systems. This study proposes a real-time drowsiness detection system based on electroencephalogram (EEG) signals and a lightweight convolutional neural network (CNN). The system comprises five main components: EEG signal acquisition, preprocessing, feature extraction, CNN-based classification, and user feedback delivery via an Android application. The experiment involved four healthy adult male participants with an average age of 24.5 years. EEG data were collected using the DSI-24 device, and the relative power in the alpha band from the prefrontal (Fp1, Fp2) and occipital (O1, O2) regions was identified as the primary feature for distinguishing drowsiness. The proposed CNN model, trained on these features, achieved a classification accuracy of 91.56%, comparable to the 92.66% accuracy of the more complex AlexNet model, while being significantly more lightweight and suitable for real-time deployment on embedded systems. The Android application provides real-time feedback on the user's drowsiness level and recommends nearby rest areas to help mitigate the risk of drowsy driving. This study presents a practical and efficient EEG-based driver monitoring solution. Future work will focus on large-scale data collection under actual driving conditions to further validate and improve the system's performance.

Keywords: Android application; Brain-computer interface (BCI); Convolution neural network (CNN); Drowsiness detection; Electroencephalogram (EEG); OpenViBE

1 INTRODUCTION

Vehicles are among the most common means of transportation in daily life, and drowsy driving is one of the leading causes of traffic accidents. According to the 2024 traffic accident statistics published by the Korean National Police Agency, a total of 10,767 accidents caused by drowsy driving occurred over the five-year period from 2019 to 2023, averaging 5.9 incidents per day. During the same period, drowsy driving resulted in 316 fatalities—equivalent to approximately 2.9 deaths per 100 accidents—which is nearly twice the fatality rate of drunk driving-related accidents during the same timeframe. Notably, vehicles used for business purposes on highways have been found to be especially vulnerable to drowsy driving. Despite efforts such as expanding safety facilities, broadcasting public service announcements, and conducting awareness campaigns, little measurable improvement has been observed in preventing drowsy driving accidents [1].

Drowsiness represents a transitional state between wakefulness and sleep. In this state, drivers experience reduced attention, significantly delayed reaction times, and impaired judgment, making it difficult to maintain proper head posture and orientation [2, 3]. Common signs of drowsiness include frequent eye closures and repeated yawning. These observable behaviors can be used to deliver early warnings to drivers and help prevent potential accidents. In recent years, driver drowsiness detection technologies have received increasing research attention [4-9]. These studies can be broadly classified into four categories: (1) image-based approaches that use cameras to analyze facial expressions and driver behavior; (2) biosignal-based approaches that rely on physiological signals measured via wearable sensors; (3) vehicle-based approaches that monitor driving patterns and vehicle behavior; and (4) hybrid approaches that combine multiple modalities. Among these, biosignal-based methods have demonstrated superior performance in drowsiness detection compared to image-

based methods [10].

In this context, a variety of biosignals—including EEG, ECG, EMG, and EOG—have been utilized for drowsiness detection [11-13], with EEG-based techniques being especially prominent due to their high temporal resolution [14, 15]. In recent years, the analysis of biosignals has gained substantial interest in the fields of driver monitoring and human-computer interaction. As the complexity and volume of physiological data continue to grow, deep learning models have emerged as effective tools for robust feature extraction and signal classification. Architectures such as Recurrent Neural Networks (RNN), Long Short-Term Memory networks (LSTM), Autoencoders (AE), and Convolutional Neural Networks (CNN) have shown strong performance in noise filtering, time-series signal processing, and biomedical signal classification tasks [16, 17].

Among these models, CNNs have been particularly successful in anomaly detection due to their ability to achieve high classification accuracy and automatically learn discriminative features from raw data [18]. Recent studies applying CNNs to EEG-based drowsiness detection have reported encouraging outcomes [19, 20]. Given that EEG signals are time-series data that are often high-dimensional and noisy, traditional feature engineering techniques are frequently inadequate. CNNs, in contrast, are well-suited for handling such unstructured signals, as they excel at extracting local spatial and frequency-domain features. Moreover, CNNs eliminate the need for manual feature design by automatically learning relevant patterns from the data. Based on these strengths, we propose a lightweight CNN model designed for real-time classification of drowsiness and wakefulness states using EEG signals collected with the DSI-24 device. The proposed model is optimized for computational efficiency, making it well-suited for real-time deployment in embedded systems, such as in-vehicle driver monitoring applications.

The primary objective of this study is to implement a system capable of detecting the driver's drowsiness state in

real time and providing immediate feedback to the user, thereby aiming to prevent traffic accidents caused by drowsiness. To achieve this, we propose a lightweight CNN model that classifies drowsiness by extracting relevant features from EEG signals. In addition, the system offers real-time feedback to the user by issuing auditory warnings and displaying nearby rest areas, making it applicable to actual driving environments. The remainder of this paper is organized as follows. Section 2 describes the data collection process, EEG signal preprocessing, feature extraction methods, and the architecture of the Android application used for user interaction. Section 3 presents the performance evaluation of the proposed model along with experimental results. Finally, Section 4 concludes the study with a summary of the findings and outlines directions for future research.

2 MATERIALS AND METHODS

We implemented a real-time drowsiness detection system based on EEG signals that provides immediate feedback to the user. The system is broadly divided into two main components: the user system and the server system. The user system comprises (1) an EEG device that collects brainwave signals from the user and (2) an Android application that provides real-time notifications of the user's drowsiness state. The server system includes (1) a preprocessing module that filters and processes the collected EEG signals in real time to extract relevant features and (2) a CNN model that classifies the user's drowsiness state based on these features. Fig. 1 illustrates the overall architecture of the proposed drowsy driving detection system.

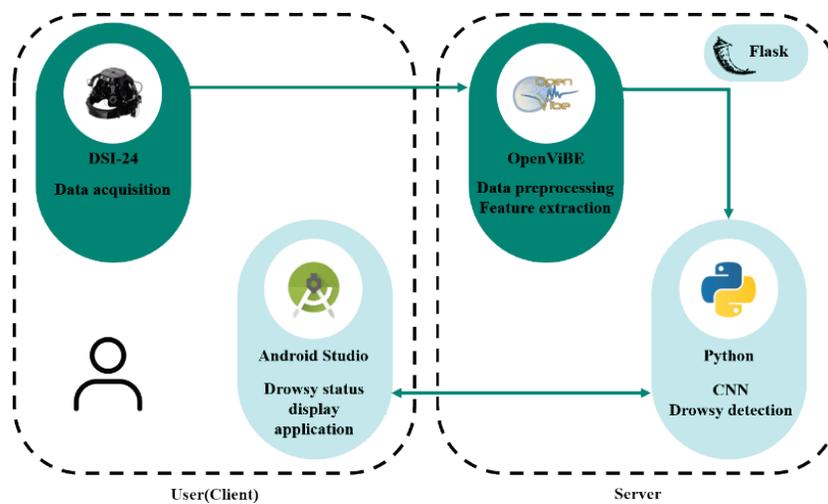


Figure 1 Structure of the proposed drowsy detection system

2.1 Data Acquisition

In this study, EEG data were collected using the DSI-24 device from Wearable Sensing. The DSI-24 is a wireless dry-electrode EEG headset designed for the rapid deployment of 21 sensors positioned according to the international 10–20 system. It includes 19 electrodes covering the scalp, 2 ear clip sensors, and 3 auxiliary inputs for the acquisition of up to three additional biosignals. The device supports Bluetooth wireless transmission and samples signals at a rate of 300 Hz. Fig. 2 illustrates the international standard 10–20 electrode placement system. The experiment involved four healthy adult male participants with an average age of 24.5 years. Prior to participation, all subjects were informed of the experimental procedures and precautions and provided written consent. EEG data were collected in a room isolated from light and noise to ensure a comfortable environment for the participants. Additionally, a camera was installed to record facial expression changes associated with drowsy during the session. The experimental procedure was as follows: (1) The participant wore the EEG device and was seated in a stationary position. (2) The participant viewed a pre-recorded video of highway driving. During this time, EEG signals were recorded simultaneously with video

footage of the participant's facial expressions. The driving video lasted approximately 20 minutes. (3) Based on the recorded EEG data and facial footage, specific segments of EEG signals were classified into drowsy and wakeful states.

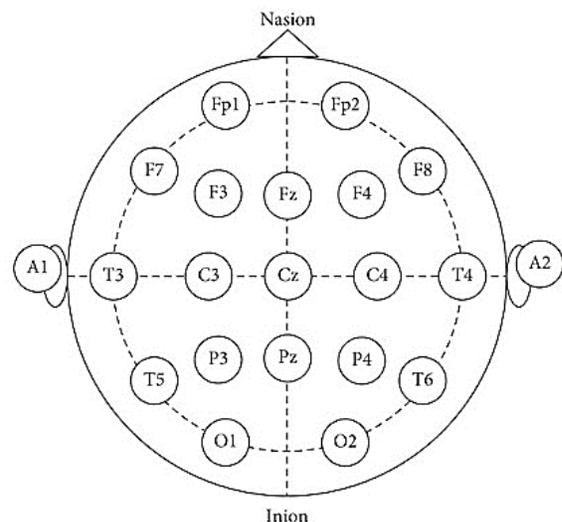


Figure 2 International 10-20 electrode system

2.2 EEG Signal Processing

During the EEG data collection, various types of noise—such as muscle movements, eye blinks, and electrical interference—can contaminate the signals. Given the low amplitude of EEG signals, such noise can significantly degrade signal quality. Therefore, a preprocessing step to remove noise from the recorded EEG data is essential. In this study, a band-pass filter was applied to retain only the frequencies within the 0–50 Hz range. Additionally, Independent Component Analysis (ICA), a widely used method for EEG denoising [21, 22], was employed to remove artifacts. Following preprocessing, the EEG signals were transformed from the time domain to the frequency domain using the Fast Fourier Transform (FFT). As shown in Tab. 1, the processed signals were then segmented into five spectral frequency bands. EEG activity is typically divided into Delta, Theta, Alpha, Beta, and Gamma bands, each associated with different cognitive and physiological states [23].

Table 1 Characteristics of EEG frequency bands

Brainwaves	Description	Frequency Interval
Delta	Refers to consciousness and sleep states	0.5 to 4 Hz
Theta	Refers to the half-sleep	4 to 7 Hz
Alpha	Refers to waking state	8 to 13 Hz
Beta	Refers to alert state	13 to 30 Hz
Gamma	Refers to hyper-vigilance state	30 to 50 Hz

Absolute and relative power spectrum analyses were conducted on the Theta and Alpha frequency bands, which are known to be closely associated with drowsiness. The power of each frequency component was calculated as the square of the magnitude of the complex-valued signal obtained via the Fourier Transform. Eq. (1) represents the Fast Fourier Transform (FFT), Eq. (2) denotes the Inverse Fast Fourier Transform (IFFT), and Eq. (3) defines the computation of total power.

$$H(f_n) = \sum_{k=0}^{N-1} h_k e^{-j2\pi kn/N} = H_n, \quad (1)$$

$$h_k = \frac{1}{N} \sum_{n=0}^{N-1} H_n e^{-2\pi kn/N}, \quad (2)$$

$$Total\ Power = \sum_{k=0}^{N-1} |h_k|^2 = \frac{1}{N} \sum_{n=0}^{N-1} |H_n|^2. \quad (3)$$

Here, h_k denotes a discrete signal sequence with a sample size of N . Absolute band power refers to the power value of the power spectrum within a specific frequency band for each channel. In contrast, relative band power represents the proportion of power within a specific frequency band relative to the total power across the entire frequency range in each channel.

In this study, a Python-based program was developed to extract relevant features. The process involved preprocessing the raw EEG data—including noise removal—followed by

applying the Fast Fourier Transform (FFT). The resulting frequency-domain data served as input, and the program outputted the relative power differences between drowsy and wakefulness states for each EEG channel, identifying those with the most significant variations.

2.3 EEG Analysis Tool

In this study, OpenViBE was utilized as a real-time EEG signal analysis platform. OpenViBE is a software framework designed for building and testing brain-computer interface (BCI) systems, enabling real-time acquisition, filtering, processing, classification, and visualization of brain signals. The OpenViBE platform operated on the server side and was responsible for EEG signal processing and data transmission. The OpenViBE Acquisition Server was employed to receive EEG data from the DSI-24 device via Bluetooth communication. Additionally, the OpenViBE Designer was configured to perform preprocessing and feature extraction on the incoming EEG signals. After preprocessing and feature extraction, the Lab Streaming Layer (LSL) protocol was used to transmit the extracted EEG features to a Flask-based server. The Flask server then forwarded the incoming feature data to the deep learning classification module.

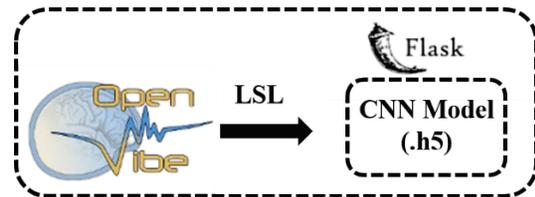


Figure 3 Structure of flask server

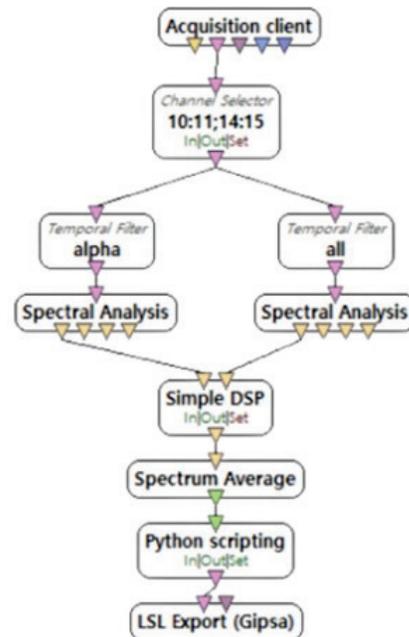


Figure 4 Proposed OpenViBE designer

Fig. 3 illustrates the architecture of the Flask server, and Fig. 4 presents the structure of the OpenViBE Designer workflow for EEG feature extraction. In the Channel Selector

stage, channels 10, 11, 14, and 15—corresponding to Fp1, Fp2, O1, and O2—were selected based on their known relevance to drowsiness detection. In the Spectrum Analysis stage, the relative power of both the full frequency band and the Alpha band was calculated. In the Python Scripting stage, matrix-type data from the previous step was converted into a signal format suitable for transmission. Finally, the processed signals were sent to the server using LSL communication.

2.4 Proposed CNN Model

Recently, Convolutional Neural Network (CNN) models have demonstrated outstanding performance the fields of computer vision and natural language processing, which has led to their application in EEG signal classification. Notably, Chaabene [19] and Zhu [20] applied CNNs to EEG-based

drowsiness detection and reported promising results. Building on these advances, this study proposes a lightweight yet high-performance CNN model for real-time drowsiness detection. The collected EEG signals were labeled with 1 for drowsy and 0 for wakeful states, and a total of 36,987 labeled samples were compiled in CSV format to form the dataset. This dataset was divided into training (80%) and testing (20%) subsets for model development and evaluation. The trained CNN model was saved in H5 format and is invoked by the Flask server during runtime for real-time inference. The proposed CNN architecture consists of three convolutional layers followed by two fully connected layers. The model was trained using a dropout rate of 0.5, a batch size of 64, and 50 training epochs. The overall architecture of the proposed CNN model is illustrated in Fig. 5, and the detailed layer-wise structure is summarized in Tab. 2.

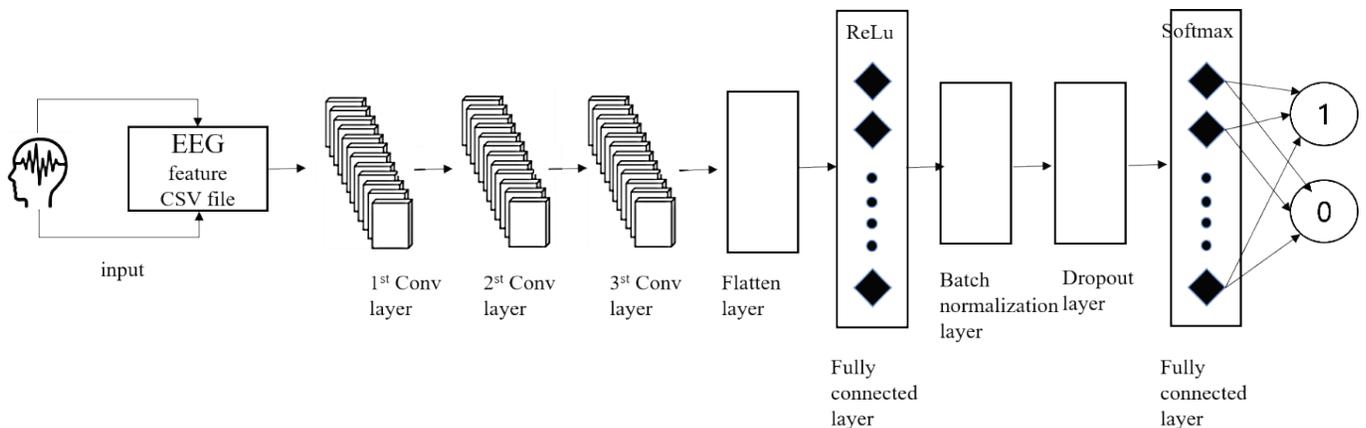


Figure 5 Proposed CNN model structure

Table 2 Summary of proposed CNN model structure

Layer Type	Parameters	Activation function	Notes
Conv2D #1	32 filters, 3×3 Kernel	ReLU	Input layer
MaxPooling2D #1	2×2 pool size	-	Dimension reduction
Conv2D #2	64 filters, 3×3 Kernel	ReLU	Mid-level feature extraction
MaxPooling2D #2	2×2 pool size	-	
Conv2D #3	128 filters, 3×3 Kernel	ReLU	High-level feature learning
Flatten	-	-	Converts 2D to 1D
Dense #1	128 units	ReLU	Dropout 0.5
BatchNorm	-	-	Normalization
Dense #2	64 units	ReLU	Dropout 0.5
Output	2 units	Softmax	Binary classification output

2.5 Drowsy Detection Application

The Android application developed in this study is designed to help prevent drowsy driving by providing real-time feedback to users regarding their drowsiness or wakefulness status, based on EEG data. Fig. 6 presents the overall architecture of the Android application. The app communicates with the server via HTTP, transmitting EEG data once per second. When the user is in a wakeful state, the application displays a real-time EEG graph along with a message indicating wakefulness. The background color of the interface is green, and a smiling emoticon is shown to convey a sense of safety and wakefulness. Conversely, when drowsiness is detected, the application presents the EEG graph along with a warning message indicating the drowsy

state. The background color changes to red, and an audible alarm is triggered for three seconds to help awaken the driver.

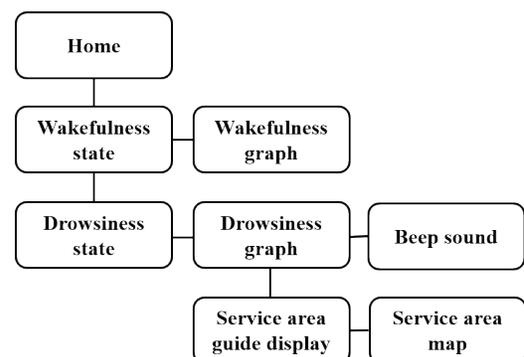


Figure 6 Structure of the drowsiness prevention app

Additionally, the application includes a user-interactive feature: by clicking a designated button, users can access Google Maps to view a list of nearby rest areas based on their current GPS location. Fig. 7 displays the interface in the wakefulness state, while Fig. 8 shows the interface during a detected drowsy state.



Figure 7 Wakefulness status screen

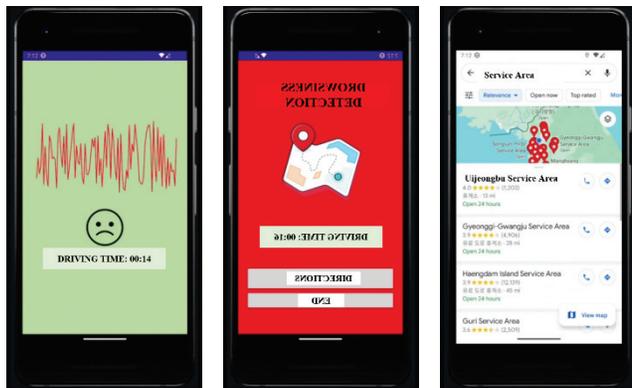


Figure 8 Drowsiness status screen

3 RESULTS AND DISCUSSION

3.1 EEG Feature Extraction Results

Timely detection of driver drowsiness is crucial, as even

brief lapses can result in serious accidents. Extracting EEG-based features associated with drowsiness is a key factor in enhancing the system’s processing speed and classification accuracy. In this study, feature extraction for real-time drowsiness detection was approached from two perspectives: (1) identifying the minimal and most effective EEG channel locations associated with drowsiness, and (2) determining the optimal frequency bands that reflect drowsiness-related brain activity. To achieve this, alpha relative power values were calculated from EEG signals recorded across 23 channels, referencing prior studies on EEG-based drowsiness detection [24]. Channels showing the greatest differences in alpha power between wakefulness and drowsy states were identified. Fig. 9 presents the relative alpha power distributions during wakeful and drowsy states across all channels. This analysis was conducted for four participants, and the figure illustrates the results for a representative subject. The y-axis displays relative power as a normalized ratio ranging from 0 to 1. Across all four participants, the most significant deviations in alpha relative power occurred in the prefrontal (Fp1, Fp2) and occipital (O1, O2) regions, indicating that these areas are highly correlated with drowsiness, particularly the prefrontal cortex. Subsequently, EEG data from these four channels (Fp1, Fp2, O1, O2) were further analyzed using both absolute and relative power spectrum analysis across three frequency bands associated with drowsiness: Theta (4–8 Hz), Alpha (8–13 Hz), and Theta-Alpha (5–9 Hz). The results of this analysis are summarized in Tab. 3. The table shows that both absolute and relative power values for the Theta, Alpha, and Theta-Alpha bands exhibited significant differences between wakeful and drowsy states. Since absolute power values can vary across individuals due to biological differences such as scalp and skull thickness, relative power was chosen as the more stable feature metric. Moreover, across all participants, the relative alpha power provided more consistent differentiation between wakefulness and drowsiness than the Theta-Alpha band. Based on these findings, this study selected the relative alpha power values at Fp1, Fp2, O1, and O2 as the primary features for EEG-based drowsiness detection.

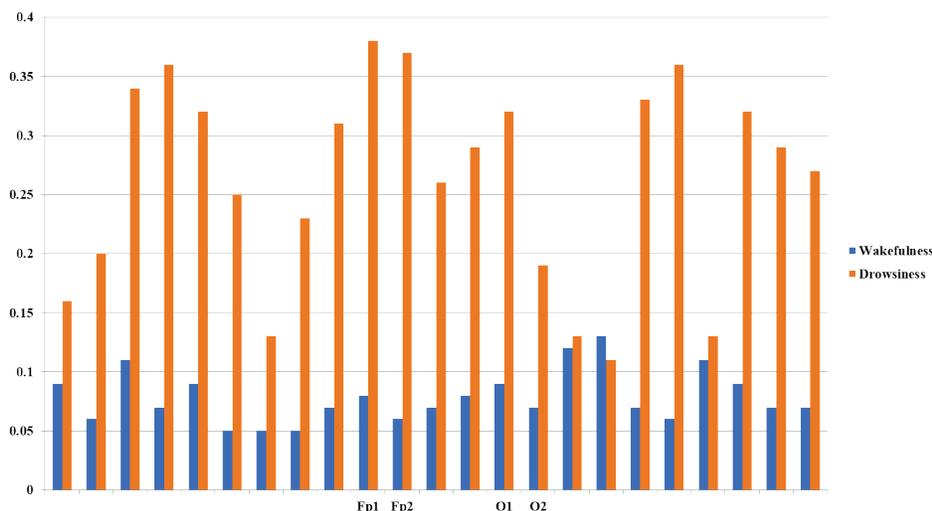


Figure 9 Comparison of alpha relative power values across all EEG channels during wakeful and drowsy states

Table 3 Absolute and relative power values in Theta, Alpha, and Theta-Alpha bands at Fp1, Fp2, O1, and O2 channels

Frequency (Hz)	Power spectrum	Status	Fp1	Fp2	O1	O2
Theta (4-8 Hz)	Absolute power	Wakefulness	0.324871	0.390573	0.882764	0.218468
		Drowsiness	0.742031	0.793544	0.878072	0.494796
	Relative power	Wakefulness	0.006001	0.004885	0.014480	0.005602
		Drowsiness	0.009713	0.009572	0.011727	0.007571
Theta-Alpha (5-9 Hz)	Absolute power	Wakefulness	0.647676	0.784580	1.633170	0.421939
		Drowsiness	8.818757	9.216807	4.818486	2.443815
	Relative power	Wakefulness	0.011965	0.009812	0.026790	0.010819
		Drowsiness	0.115440	0.111174	0.064354	0.037393
Alpha (8-13 Hz)	Absolute power	Wakefulness	4.654174	5.297472	5.820197	2.817909
		Drowsiness	29.108761	31.083387	24.67349	12.845421
	Relative power	Wakefulness	0.085978	0.066253	0.095472	0.0722254
		Drowsiness	0.381043	0.374931	0.329530	0.196548

3.2 EEG Classification Results Based on CNN

A CNN model for drowsiness detection was designed using the Python-based Scikit-learn library, which is widely utilized for machine learning analysis. To evaluate the model's performance, both raw EEG data and feature-extracted data were used as input. The performance of the model was assessed using four common evaluation metrics: recall, precision, F1-score, and accuracy, each defined by Eqs. (4)-(7).

$$Recall = \frac{TP}{TP + FN}, \tag{4}$$

$$Precision = \frac{TP}{TP + FP}, \tag{5}$$

$$F1\text{-score} = \frac{2 \times Precision \times Recall}{Precision + Recall}, \tag{6}$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}, \tag{7}$$

TP: True Positive, *TN*: True Negative, *FP*: False Positive, *FN*: False Negative.

The classification_report function from the Scikit-learn metrics module was employed to compute these performance metrics. Tab. 4 presents the evaluation results of the proposed CNN model using the feature-extracted data, demonstrating a classification accuracy of approximately 91.56%. In Tab. 4, Support refers to the number of actual instances for each class; Weighted avg is the average weighted by the number of instances in each class; and Macro avg is the unweighted average, giving equal importance to each class. To further assess the classification capability of the proposed model, a confusion matrix was generated. As shown in Fig. 10, the confusion matrix revealed 455 true negatives and 1,048 true positives, confirming the model's robust performance with an overall accuracy of 91.56%. These results indicate that the proposed CNN model is both lightweight and accurate.

To validate its effectiveness of the proposed model, a comparative analysis was conducted against the well-established AlexNet architecture [25], using the same dataset. The proposed CNN model was specifically designed as a lightweight and efficient alternative to conventional deep learning models such as AlexNet, which was originally introduced in 2012 and achieved state-of-the-art performance in the ImageNet Large Scale Visual Recognition Challenge. While AlexNet consists of five convolutional layers with a

high number of filters (up to 384), three fully connected layers with 4,096 units each, and more than 60 million trainable parameters, the proposed model significantly reduces computational complexity through a shallower and more compact structure. Specifically, the proposed model comprises three convolutional layers with 32, 64, and 128 filters, each using 3×3 kernels. These are followed by two fully connected layers with 128 and 64 units, respectively, and a final output layer with two neurons for binary classification. Dropout regularization (rate = 0.5) is applied to both dense layers to mitigate overfitting, and batch normalization is employed to stabilize the training process. Compared to AlexNet, the total number of parameters is drastically reduced, making the model suitable for training and deployment in resource-constrained environments such as mobile or embedded systems.

Table 4 Classification performance of proposed CNN model

	Precision	Recall	F1-score	Support
0	0.85	0.87	0.86	523
1	0.94	0.93	0.93	1127
Accuracy			0.91	1650
Macro avg	0.89	0.90	0.90	16540
Weighted avg	0.91	0.91	0.91	1650

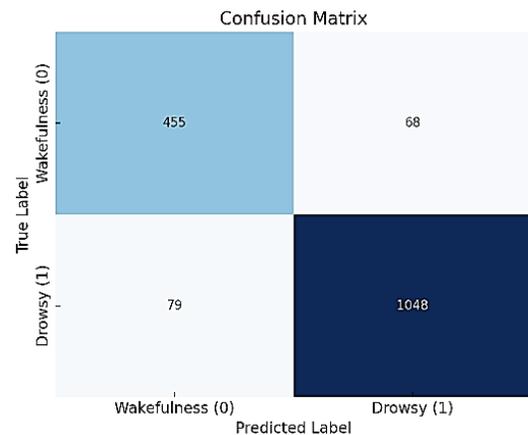


Figure 10 Confusion matrix of the proposed CNN model

Table 5 Classification performance between the proposed model and AlexNet

Model	Precision	Recall	F1-score	Accuracy
AlexNet	0.9379	0.9563	0.9466	0.9266
Ours	0.9476	0.9272	0.9373	0.9156

Tab. 5 presents the classification performance comparison between the two models. The proposed CNN model achieved an accuracy of 91.56%, while AlexNet

achieved 92.66%. These results confirm that the proposed model maintains comparable classification performance while offering significantly lower computational overhead. Due to its relatively simple yet effective architecture, the model is highly suited for real-time EEG-based drowsiness detection, where inference speed and low resource usage are critical. To further verify the effectiveness of the extracted features, an additional comparison was performed using both raw EEG data and feature-extracted data as input to the proposed model.

Tab. 6 summarizes the classification accuracy for each input type: the model achieved 62.17% accuracy with raw EEG data, and 91.56% with feature-extracted data. These results highlight the effectiveness of the Alpha relative power values at Fp1, Fp2, O1, and O2 as meaningful features for drowsiness detection.

Table 6 Classification performance between the EEG raw data and feature extraction data

Data	Drop out	Epoch	Batch size	Accuracy
Raw data	0.5	10	64	0.6217
Feature extraction data	0.5	50	64	0.9156

The proposed CNN model achieved an overall classification accuracy of approximately 91%, with strong class-wise precision and recall. These results are in line with, or slightly exceed, those reported in recent EEG-based drowsiness detection studies. For instance, Chaabene [19] implemented a CNN model using Emotiv EPOC+ data and achieved 90.4% test accuracy, demonstrating the feasibility of deep learning in this domain. A more recent model, EEG_DMNet, which incorporates multi-scale spectral-temporal convolutional processing, reported 97.03% accuracy on the SEED VIG benchmark [26]. Additionally, hybrid models such as CNN-LSTM have achieved F1-scores up to 0.95 in binary classification tasks for drowsiness vs. wakefulness [27]. Multi-modal approaches that combine EEG with ECG and use RNN/CNN fusion have also demonstrated validation accuracies close to 97% [28]. While these advanced models offer slightly higher performance, they often come at the cost of increased model complexity, computational requirements, and dependency on additional signal modalities. In contrast, the proposed lightweight CNN model provides a well-balanced solution between performance, simplicity, and real-time applicability, making it ideal for embedded or mobile EEG-based driver monitoring systems.

Nevertheless, this study has several notable limitations. As a pilot study, it involved EEG data collection from only four adult male participants. Although the proposed CNN model demonstrated promising performance, the small sample size and limited dataset present significant constraints. First, the generalizability of the findings is limited. Models trained on data from such a small cohort are highly prone to overfitting and may capture individual-specific EEG characteristics, which hinders their applicability to broader populations with varying demographics such as age, gender, and health status. Second, the statistical reliability of deep learning model training is compromised. CNN-based models require a large volume of training data to effectively learn patterns and prevent overfitting. Although regularization techniques such as

dropout, batch normalization, and hyperparameter tuning were applied to mitigate these issues, the current dataset is insufficient to ensure robust and reproducible outcomes. Third, statistical validation is inherently limited. With a small dataset, it becomes challenging to conduct meaningful comparisons between the drowsy and wakeful states. As a result, critical statistical indicators—such as effect sizes, confidence intervals, and p -values—cannot be reliably calculated or interpreted.

4 CONCLUSION

In this paper, we implemented a real-time drowsiness detection system based on EEG signals that alerts the user upon detecting signs of drowsiness. The system is composed of three main components: (1) a module that identifies brain regions and EEG features highly correlated with drowsiness, (2) a classification module that determines the drowsiness state based on these features, and (3) a user feedback module that delivers alerts and recommendations. Relative power in the alpha band from the prefrontal cortex (Fp1, Fp2) and occipital cortex (O1, O2) was identified as the primary feature for detecting drowsiness. The proposed lightweight CNN model achieved an accuracy of 91.56%, which is comparable to the 92.66% accuracy of the more complex AlexNet model. Despite its lower computational complexity, the proposed model demonstrated similar performance, making it suitable for real-time applications such as driver monitoring systems. To enhance user safety, the system is integrated with an Android application that not only notifies the user of their drowsiness state but also suggests nearby rest areas based on their current location. This real-time feedback mechanism is expected to help prevent drowsy driving and reduce traffic accident rates. It should be noted that this study represents a pilot investigation into real-time EEG-based drowsiness detection, and thus has certain limitations. The experimental dataset was collected from only four adult male participants, resulting in a limited sample size for model training and evaluation. To overcome these limitations, future work will involve data collection from a more diverse group of participants in terms of age, gender, and driving experience. Furthermore, EEG data will be gathered under actual driving conditions to enhance ecological validity. In addition, advanced validation techniques such as cross-validation and multi-center studies will be employed to improve the generalizability and statistical rigor of the model.

Acknowledgment

This work was supported by Kyungnam University Foundation Grant, 2023.

5 REFERENCES

- [1] Traffic Accident Statistics of Korean. (2024, April 4). *Drowsy driving is more dangerous than drunk driving... Fatality rate is twice as high* [Internet]. Updated 2024 April 4; cited 2025 February 24. Available from: https://www.koroad.or.kr/main/board/6/302605/board_view.do?&cp=1&listType=list&bdOpenYn=Y&bdNoticeYn=N

- [2] Gwak, J., Hirao, A., & Shino, M. (2020). An investigation of early detection of driver drowsiness using ensemble machine learning based on hybrid sensing. *Applied Sciences*, 10(8), 2890. <https://doi.org/10.3390/app10082890>
- [3] Arakawa, T. (2021). Trends and future prospects of the drowsiness detection and estimation technology. *Sensors*, 21(23), 7921. <https://doi.org/10.3390/s21237921>
- [4] Fuletra, J. D., & Bosamiya, D. (2013). A survey on driver's drowsiness detection techniques. *International Journal of Recent Innovations and Trends in Computing and Communication*, 1, 816–819. <https://doi.org/10.17762/ijritcc.v1i11.2871>
- [5] Ramzan, M., Khan, H. U., Awan, S. M., Ismail, A., Ilyas, M., & Mahmood, A. (2019). A survey on state-of-the-art drowsiness detection techniques. *IEEE Access*, 7, 61904–61919. <https://doi.org/10.1109/ACCESS.2019.2914373>
- [6] Sikander, G., & Anwar, S. (2018). Driver fatigue detection systems: A review. *IEEE Transactions on Intelligent Transportation Systems*, 20(6), 2339–2352. <https://doi.org/10.1109/TITS.2018.2868499>
- [7] Dong, Y., Hu, Z., Uchimura, K., & Murayama, N. (2010). Driver inattention monitoring system for intelligent vehicles: A review. *IEEE Transactions on Intelligent Transportation Systems*, 12(2), 596–614. <https://doi.org/10.1109/TITS.2010.2092770>
- [8] Nordbakke, S., & Sagberg, F. (2007). Sleepy at the wheel: Knowledge, symptoms and behavior among car drivers. *Transportation Research Part F: Traffic Psychology and Behaviour*, 10(1), 1–10. <https://doi.org/10.1016/j.trf.2006.03.003>
- [9] Cho, S. J., Shin, S. U., Yoo, K. S., Min, K. T., & Jung, B. J. (2024). A study on real-time right-turn accident prevention model using YOLO object detection and transfer learning. *Journal of Innovation Industry Technology*, 2(4), 151–158. <https://doi.org/10.60032/IIIT.2024.2.4.151>
- [10] Dinges, D. F. (1995). An overview of sleepiness and accidents. *Journal of Sleep Research*, 4(S2), 4–14. <https://doi.org/10.1111/j.1365-2869.1995.tb00220.x>
- [11] Gromer, M., Salb, D., Walzer, T., Madrid, N., & Seepold, R. (2019). ECG sensor for detection of driver's drowsiness. *Procedia Computer Science*, 159, 1938–1946. <https://doi.org/10.1016/j.procs.2019.09.366>
- [12] Choi, H. (2019). EMG feature extraction for the driver's drowsiness using RF wireless power transmission method. *International Journal of Engineering and Advanced Technology (IJEAT)*, 8(3S), 494–497. Available from: <https://www.ijeat.org/protfolio-item/c11040283s19>
- [13] Ahn, S., Nguyen, T., Jang, H., Kim, J., & Jun, S. (2016). Exploring neuro-physiological correlates of drivers' mental fatigue caused by sleep deprivation using simultaneous EEG, ECG, and fNIRS data. *Frontiers in Human Neuroscience*, 10, 219. <https://doi.org/10.3389/fnhum.2016.00219>
- [14] Ma, Y., Chen, B., Li, R., Wang, C., Wang, J., She, Q., et al. (2019). Driving fatigue detection from EEG using a modified PCANet method. *Computational Intelligence and Neuroscience*, Article ID 47211863. <https://doi.org/10.1155/2019/47211863>
- [15] Papadelis, C., Chen, Z., Papadeli, C. K., Bamidis, P., Chouvarda, I., Bekiaris, E., et al. (2007). Monitoring sleepiness with on-board electrophysiological recordings for preventing sleep-deprived traffic accidents. *Clinical Neurophysiology*, 118(9), 1906–1922. <https://doi.org/10.1016/j.clinph.2007.04.031>
- [16] Jeong, J., Yu, B., Lee, D., & Lee, S. (2019). Classification of drowsiness levels based on a deep spatio-temporal convolutional bidirectional LSTM network using electroencephalography signals. *Brain Sciences*, 9(12), 348. <https://doi.org/10.3390/brainsci9120348>
- [17] Vesselenyi, T., Moca, S., Rus, A., Mitran, T., & Tataru, B. (2017). Driver drowsiness detection using ANN image processing. *IOP Conference Series: Materials Science and Engineering*, 252, 012097. <https://doi.org/10.1088/1757-899X/252/1/012097>
- [18] Alaskar, H. (2018). Convolutional neural network application in biomedical signals. *Journal of Computer Sciences and Information Technology*, 6(2), 45–59. <https://doi.org/10.15640/jcsit.v6n2a5>
- [19] Chaabene, S., Bouaziz, B., Boudaya, A., Hökelmann, A., & Ammar, A., et al. (2021). Convolutional neural network for drowsiness detection using EEG signals. *Sensors*, 21(5), 1734. <https://doi.org/10.3390/s21051734>
- [20] Zhu, M., Chen, J., Li, H., Liang, F., & Han, L., et al. (2021). Vehicle driver drowsiness detection method using wearable EEG based on convolution neural network. *Neural Computing and Applications*, 33, 13965–13980. <https://doi.org/10.1007/s00521-021-06038-y>
- [21] Lin, C. T., Wu, R. C., Liang, S. F., Chao, W. H., Chen, Y. J., & Jung, T. P. (2005). EEG-based drowsiness estimation for safety driving using independent component analysis. *IEEE Transactions on Circuits and Systems I: Regular Papers*, 52(12), 2726–2738. <https://doi.org/10.1109/TCSI.2005.857555>
- [22] Chai, R., Naik, G. R., Nguyen, T. N., Ling, S. H., Tran, Y., & Craig, A. (2016). Driver fatigue classification with independent component by entropy rate bound minimization analysis in an EEG-based system. *IEEE Journal of Biomedical and Health Informatics*, 21(3), 715–724. <https://doi.org/10.1109/JBHI.2016.2532354>
- [23] Teplan, M. (2002). Fundamentals of EEG measurement. *Measurement Science Review*, 2, 1–11. Available from: <https://www.measurement.sk/2002/S2/Teplan.pdf>
- [24] Gharagozlou, F., Saraji, G. N., Mazloumi, A., Nahvi, A., Nasrabadi, A. M., Foroushani, A. R., et al. (2015). Detecting driver mental fatigue based on EEG alpha power changes during simulated driving. *Iranian Journal of Public Health*, 44(12), 1693–1700. Available from: <https://pubmed.ncbi.nlm.nih.gov/26811821>
- [25] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6), 84–90. <https://doi.org/10.1145/3065386>
- [26] Obaidan, H. B., Hussain, M., & AlMajed, R. (2024). EEG_DMNet: A deep multi-scale convolutional neural network for electroencephalography-based driver drowsiness detection. *Electronics*, 13(11), 2084. <https://doi.org/10.3390/electronics13112084>
- [27] Lee, C., & An, J. (2023). LSTM-CNN model of drowsiness detection from multiple consciousness states acquired by EEG. *Expert Systems with Applications*, 213(PB), 119032. <https://doi.org/10.1016/j.eswa.2022.119032>
- [28] Geoffroy, G., Chaari, L., Tournet, J. Y., & Wendt, H. (2021). Drowsiness detection using joint EEG-ECG data with deep learning. In *2021 29th European Signal Processing Conference (EUSIPCO)* (pp. 955–959). <https://doi.org/10.23919/EUSIPCO54536.2021.9616046>

Author's contacts:

Ssang-Hee Seo, Professor
 School of Computer Science and Engineering,
 7 Kyungnamdaehak-ro, Masanhappo-gu, Changwon-si,
 Gyeongsangnam-do, 51767, Republic of Korea
 Mobile Phone: +82-010-3454-8737
 shseotwin@kyungnam.ac.kr

Multi-Target Estimation in OFDM Radar using YOLOv8 for Integrated Sensing and Communication

So-Yeon Jeon, Eui-Rim Jeong*

Abstract: In this study, we propose a method for simultaneously estimating the number, velocity, and distance of multiple targets in an Orthogonal Frequency Division Multiplexing (OFDM) radar environment using YOLO (You Only Look Once). The proposed approach employs Doppler-range two-dimensional (2D) signals as the input to the YOLO model, enabling it to learn and predict target characteristics. Since YOLO performs object detection in a single forward pass, it achieves higher computational efficiency compared to conventional CNN-based methods, making it suitable for multi-target estimation tasks. To validate the performance of the proposed method, we conducted simulations under various signal-to-noise ratio (SNR) conditions ranging from -10 dB to 20 dB and considered scenarios with one to five targets. The results show that, with 32 OFDM symbols, the YOLO-based model achieved an average velocity estimation error of 1.34 km/h and an average distance estimation error of 0.71 m. These results represent improvements of 0.56 km/h and 0.89 m, respectively, over conventional CNN-based single-target estimation models, demonstrating the precision of the proposed method. Such performance indicates its potential for effective application in next-generation joint communication and sensing systems.

Keywords: 2D-Periodogram; Distance; Object Detection; OFDM Radar; Velocity; YOLO

1 INTRODUCTION

Recent advancements in radar technology have extended its applications beyond the military domain to a wide range of industries, including autonomous vehicles, drones and UAVs, and medical diagnostics [1, 2]. Radar systems are increasingly integrated into advanced sensors and systems due to their capability to accurately measure the distance, velocity, and direction of remote objects [3]. However, the growing proliferation of wireless communication devices has intensified the issue of frequency spectrum scarcity, posing a challenge for radar systems to improve spectral efficiency [4, 5]. To address this issue, Integrated Sensing and Communication (ISAC) has emerged as a promising paradigm, enabling both communication and radar sensing functionalities to be performed on a single platform [6]. This study proposes an Orthogonal Frequency Division Multiplexing (OFDM)-based radar system that leverages conventional communication signals for sensing purposes, thereby enabling joint communication and radar operations without requiring additional spectrum allocation [7]. This represents a novel approach that maximizes spectral efficiency compared to conventional radar techniques that rely on dedicated radar signals [8, 9].

In an OFDM radar system, signal processing is performed by transforming the time delay and Doppler shift caused by the target into inter-subcarrier frequency changes and inter-symbol phase variations, respectively [10]. With recent advances in deep learning, there has been a surge of interest in enhancing target detection and distance/velocity estimation through nonlinear representation learning. For instance, previous studies employing multi-output CNNs for single-target estimation used 2D-periodograms as input and achieved higher accuracy than traditional CFAR methods [11]. However, such approaches are optimized for single-target scenarios and exhibit degraded performance when applied to multi-target environments, highlighting the need for more flexible solutions. To overcome these structural limitations, this study introduces a new deep learning method

based on YOLOv8 (You Only Look Once), a model optimized for multi-object detection. The proposed method aims to robustly estimate the number, distance, and velocity of multiple targets in complex radar environments. By using 2D-periodograms of OFDM signals as inputs, YOLOv8 effectively learns and predicts target characteristics through its object detection framework. YOLOv8 offers advantages in real-time inference and computational efficiency over other one-stage detectors, making it particularly well-suited for ISAC systems requiring rapid and simultaneous estimation of multiple targets. Simulation results demonstrate that, with an OFDM symbol length of 32, the YOLOv8-based model achieves an average velocity estimation error of 1.34 km/h and an average distance estimation error of 0.71 m under multi-target conditions. In contrast, a conventional CNN-based method optimized for single-target estimation showed higher errors of 1.90 km/h and 1.60 m, respectively. These results confirm the superior performance and stability of the proposed method in multi-target environments. Furthermore, compared to radar systems using dedicated signals [8,9], the proposed approach enables more efficient use of frequency resources while maintaining reliable distance and velocity estimation in complex scenarios. The remainder of this paper is organized as follows. Section 2 describes the OFDM radar system model for integrated sensing and communication. Section 3 details the generation of 2D-periodograms. Section 4 compares the conventional CNN-based method and the proposed YOLOv8-based approach. Section 5 presents simulation-based experimental results, and Section 6 provides a comprehensive analysis and discussion. Finally, Section 7 concludes the study.

2 OFDM RADAR SYSTEM

2.1 System Overview

This section describes the overall architecture and fundamental operating principles of the proposed OFDM radar system. Fig. 1 illustrates the main signal processing

flow of the OFDM radar system for estimating the distance and velocity of targets [11]. On the transmission side, the system follows the same procedure as a conventional OFDM transmitter, where signals are transmitted via antennas. However, a distinguishing feature of this system lies in its ability to process the reflected signals that bounce back from targets and are received by the antenna. These reflected

signals, resulting from the transmitted waveform interacting with objects, contain information such as target distance and velocity. Accordingly, the system estimates these parameters based on the received echoes. In essence, the OFDM system is designed to simultaneously perform communication and radar functions by transmitting signals and analyzing the received reflections within the same operational framework.

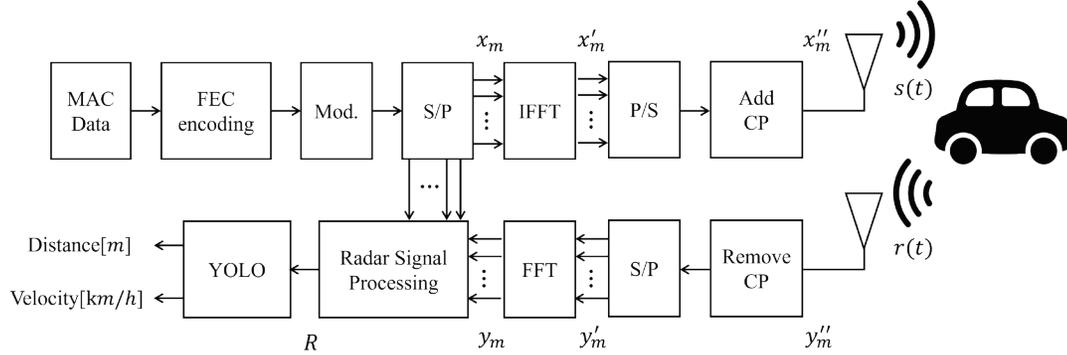


Figure 1 OFDM Radar System

2.2 Methodology

This section describes the signal processing flow and the procedure for generating the 2D-periodogram in the proposed system. On the transmitter side, the system first applies Forward Error Correction (FEC) encoding to the data, then converts it into parallel streams. It performs an inverse fast Fourier transform (IFFT) to convert the signal into the time domain and serializes it. To mitigate inter-symbol interference (ISI), the system adds a cyclic prefix (CP) to each OFDM symbol before transmitting the signal through the antenna. As the signal $s(t)$ propagates, it reflects off targets and arrives at the receiving antenna as $r(t)$. The received signal contains not only the reflected waveform but also channel noise and external interference. Target motion

introduces attenuation, time delay, and Doppler frequency shifts, which distort the signal. On the receiver side, the system removes the cyclic prefix and reshapes the signal into parallel form. It applies a fast Fourier transform (FFT) to obtain the frequency-domain signal. This process repeats for each incoming OFDM symbol to accumulate the signal over time. To estimate distance and velocity, the system compares the received and transmitted signals. It divides each received signal component by the corresponding transmitted component to eliminate modulation effects. Then, it applies a two-dimensional FFT to the accumulated signal to generate a 2D-spectrum. The squared magnitude of this spectrum yields the 2D-periodogram, which the YOLO model uses to estimate the number, velocity, and distance of the targets.

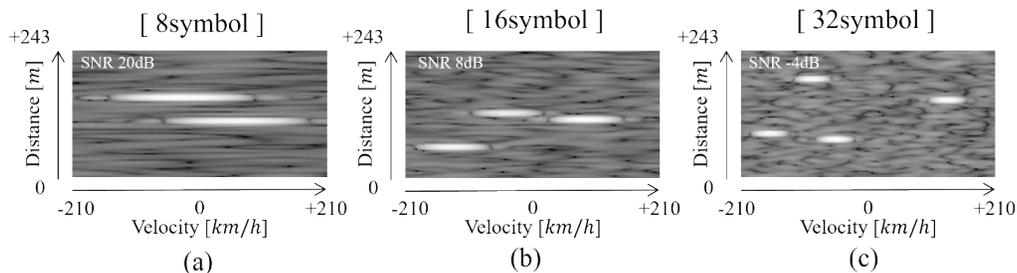


Figure 2 2D-Periodogram (a) 8symbol, (b) 16symbol, (c) 32symbol

3 2D-PERIOGRAM

The 2D-periodograms used as input to YOLO is represented as a grayscale image, where higher magnitude values appear closer to white and values near zero appear darker or black. Target detection and identification are performed by locating peak values within the 2D-periodogram. The x-coordinate of each peak corresponds to the target's velocity, while the y-coordinate indicates its distance. Fig. 2 presents examples of 2D-periodograms for different numbers of OFDM symbols. Subfigures (a), (b),

and (c) show the results for 8, 16, and 32 OFDM symbols, respectively, with the number of targets set to 2, 3, and 4 accordingly. As illustrated in the figure, increasing the number of OFDM symbols leads to a decrease in the spread of target regions within the 2D-periodogram, enhancing target resolution. The velocity range observable in the 2D-periodogram spans from -210 km/h to $+210$ km/h, while the detectable distance spans from 0 to 243 m. Negative velocities indicate targets moving away from the radar, whereas positive velocities indicate approaching targets. The resolution of the 2D-periodogram is determined by the size

of the 2D-periodograms applied to the signal. In this study, the 2D-FFT size is denoted as $N_{FFT} \times M_{FFT}$, and is set to 2048×256 .

4 TARGET ESTIMATION METHOD

4.1 Conventional Estimation Method

Convolutional Neural Networks (CNNs) have demonstrated remarkable performance in image classification and pattern recognition tasks by effectively learning spatial features through convolutional layers [12]. Leveraging this capability, CNNs can be applied to the 2D-periodogram data to estimate both the number of targets and their respective distance and velocity. However, estimating the number of targets and predicting their distance and velocity are fundamentally different tasks. As such, it is essential to design independent neural network architectures specialized for each task. Specifically, a CNN-based classification model is employed for target number

estimation, while a multi-output CNN regression model is used for distance and velocity prediction. The architecture for the classification model is depicted in Fig. 3, and the regression model for distance and velocity estimation is shown in Fig. 4. Both models share a common backbone structure that consists of five convolutional layers and batch normalization layers to extract key features from the 2D-periodogram. For the classification model, a softmax activation function is applied in the final fully connected layer to probabilistically predict the number of targets, ranging from one to five. In contrast, the multi-output CNN model for distance and velocity estimation directly outputs the estimated values of target distance and speed through its final output layer. Nonetheless, CNN-based multi-output models inherently suffer from a fixed number of output nodes, which presents a significant limitation in dynamic environments where the number of targets varies. Consequently, a separate model must be trained for each possible target count, which restricts practical deployment in scenarios with unknown or varying numbers of targets.

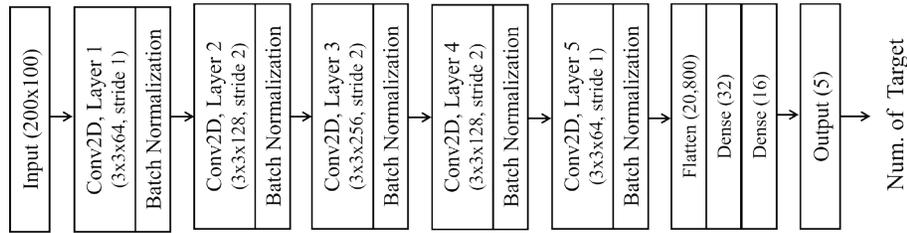


Figure 3 CNN Architecture for Estimating the Number of Targets

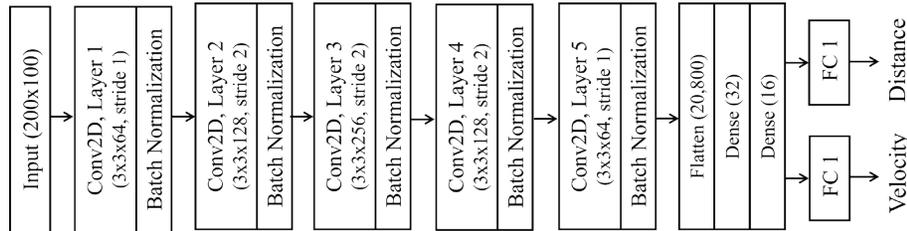


Figure 4 CNN Architecture for Estimating Target Distance and Velocity

4.2 Proposed Estimation Method

The proposed YOLO-based method is designed to directly estimate the distance and velocity of targets using the 2D-periodogram data. YOLO (You Only Look Once) is a CNN-based object detection algorithm that analyzes the entire image in a single forward pass, enabling simultaneous detection of multiple targets along with localization of their corresponding coordinates [13]. This characteristic makes YOLO particularly well-suited for solving the multi-target distance and velocity estimation problem addressed in this study. In conventional CNN-based approaches, the number of targets is first estimated using a classification model. Depending on the predicted count, separate CNN models must then be employed to estimate the distance and velocity for each target. This cascaded structure increases computational complexity and slows down detection. Furthermore, CNNs typically focus on learning local features, which limits their ability to capture global relationships

among multiple targets—posing challenges in scenarios that require understanding inter-target dependencies. In contrast, YOLO divides the input data into a grid structure and detects objects within each grid cell while learning holistic spatial features across the entire input. This enables the model to incorporate global context effectively, even in multi-target environments. Moreover, since YOLO is a one-stage detector, it can estimate the number, location, distance, and velocity of targets in a single inference step. This design provides high computational efficiency and fast detection speed, making it especially suitable for real-time applications that require rapid processing [14, 15].

This study proposes a YOLO-based method for multi-target detection and simultaneous estimation of distance and velocity, with the overall architecture illustrated in Fig. 5. The 2D-periodograms generated through radar signal processing, is resized to a fixed input size of 640×320 and input into the YOLO model to jointly estimate the number, distance, and velocity of targets. The architecture employs the standard YOLOv8s backbone. We construct target

annotations on the 2D-periodogram using the coordinates of spectral peaks, which represent each target's distance and velocity. Each peak serves as the center of a bounding box, with the width and height defined by the surrounding region of the peak. We assign all targets the same class label (class 0), and use the resulting annotations as ground truth for training. We conduct end-to-end training using the full dataset, and Tab. 1 summarizes the key hyperparameters. Given the real-time requirements and computational constraints of radar applications, we adopt the lightweight YOLOv8s model. We optimize the bounding box dimensions based on the number of OFDM symbols and fix the batch size

at 32. We set the learning rate according to the OFDM symbol count: 0.001 for 8 and 16 symbols, and 0.01 for 32 symbols. To prevent overfitting, we apply early stopping, with a patience value of 20 for 8 and 16 symbols, and 10 for 32 symbols. The total number of training epochs is 33, 37, and 22, respectively. We use the Adam optimizer, with an IoU threshold of 0.7 and a confidence threshold set to the default value of 0.25. The loss functions follow the default YOLOv8s configuration: Complete IoU (CIoU) Loss for bounding box regression, Binary Cross Entropy (BCE) Loss for objectness score prediction, and BCE Loss for classification.

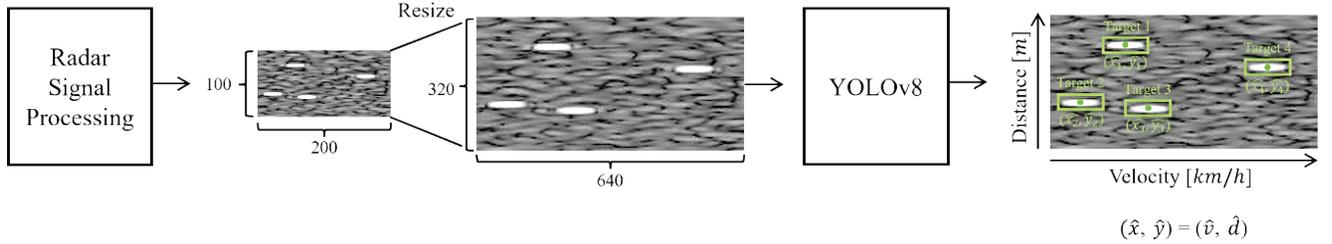


Figure 5 YOLO-Based Multi-Target Velocity & Distance Estimation

Table 1 YOLO Training Hyperparameters

Hyperparameters	Values		
	8symbol	16symbol	32symbol
Model version	YOLOv8s		
Bounding box	40×8	20×8	15×8
Batch size	32	32	32
Learning rate	0.001	0.001	0.01
Patience	20	20	10
Epoch	33	37	22
Optimizer	Adam	Adam	Adam
IOU threshold	0.7		
Confidence threshold	0.25		

5 SIMULATION

5.1 Simulation Environment

We generated the training and testing datasets using MATLAB and implemented the CNN and YOLO models with deep learning frameworks such as TensorFlow and PyTorch. Tab. 2 summarizes the OFDM signal parameters used to create the 2D-periodograms in the simulation. In this study, we set the total OFDM symbol duration to 35.74 μ s and used a sampling frequency (F_s) of 122.88 MHz with a 4,096-point FFT. Out of the available subcarriers, we used 1,284, resulting in an effective bandwidth of approximately 40 MHz. The carrier frequency was 28 GHz, and we set the cyclic prefix (CP) length to 296 samples. To investigate model performance, we varied the OFDM signal length across 8, 16, and 32 symbols. The 2D Fourier transform applied to the received signal used a size of 2,048 \times 256. During simulation, we varied the signal-to-noise ratio (SNR) from -10 dB to 20 dB. The training dataset included 50,000 randomly generated samples across the SNR range, with equal representation (10,000 samples each) for target counts from 1 to 5. For testing, we divided the SNR range into 3 dB intervals and generated 2,000 samples per interval and target count, resulting in a total of 110,000 test samples. We randomly assigned the distance and velocity of each target

within predefined bounds. We used the Mean Absolute Error (MAE) to quantitatively evaluate model performance for both distance and velocity estimation.

Table 2 Simulation Environment Parameters

Parameter	Values
OFDM System Duration	35.74 μ s
Sampling frequency	122.88 MHz
IFFT(FFT) size	4096
Bandwidth	40 MHz
Center frequency	28 GHz
Length of CP	296
Num. of OFDM symbol	8, 16, 32
2D FFT size	2048 \times 256
Cropped region size	200 \times 100
Num. of targets	1 ~ 5
SNR range	-10 ~ 20 dB
Num. of Train data	50,000
Num. of Test data	110,000

5.2 Simulation Results

5.2.1 Target Number Estimation Performance

Fig. 6 presents a comparison of the target count estimation accuracy between the CNN and YOLOv8 models under varying SNR conditions and different numbers of OFDM symbols. The simulation results show that both models exhibit improved accuracy as the number of OFDM symbols increases. Similarly, higher SNR conditions lead to improved performance in estimating the number of targets. When comparing the two models, the difference in performance was minimal under high SNR conditions. However, in low SNR environments, YOLOv8 consistently outperformed the CNN-based approach. For example, when the number of OFDM symbols was 8 and the SNR was -10 dB, the CNN model achieved approximately 66% accuracy, whereas YOLOv8 achieved a significantly higher accuracy of 89.16%. Additionally, Tab. 3, which compares the average accuracy across different OFDM symbol counts, confirms

that YOLOv8 consistently outperforms the CNN model for each symbol length. These results indicate that YOLOv8 maintains stable and reliable target count detection performance regardless of the SNR level, demonstrating its robustness in noisy environments.

Table 3 Average Target Number Estimation Accuracy

Model	Accuracy		
	8symbol	16symbol	32symbol
CNN	91.68%	94.28%	96.19%
YOLOv8	93.92%	97.23%	98.08%

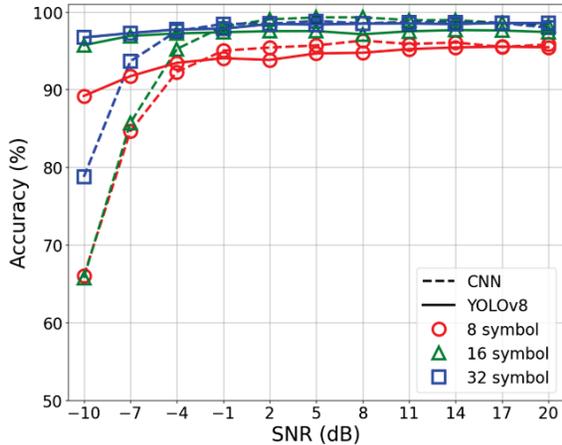


Figure 6 Target Number Estimation Accuracy of CNN and YOLO Models

5.2.2 Velocity and Distance Estimation Performance by Deep Learning Models

Figs. 7(a) and 7(b) illustrate the Mean Absolute Error (MAE) for distance and velocity estimation, respectively, across various SNR levels. Both models demonstrated improved and more stable performance as the SNR increased. However, across the entire SNR range, YOLOv8 consistently achieved lower MAE values than the CNN-based model. For distance estimation, the CNN model yielded an average MAE of 1.60 m, while YOLOv8 achieved a significantly lower average MAE of 0.71 m. In velocity estimation, CNN recorded an average MAE of 1.90 km/h, whereas YOLOv8 achieved 1.34 km/h. These results indicate that YOLOv8 reduced the average estimation error by 0.89 m for distance and 0.56 km/h for velocity compared to the CNN model, confirming its superior accuracy in both tasks.

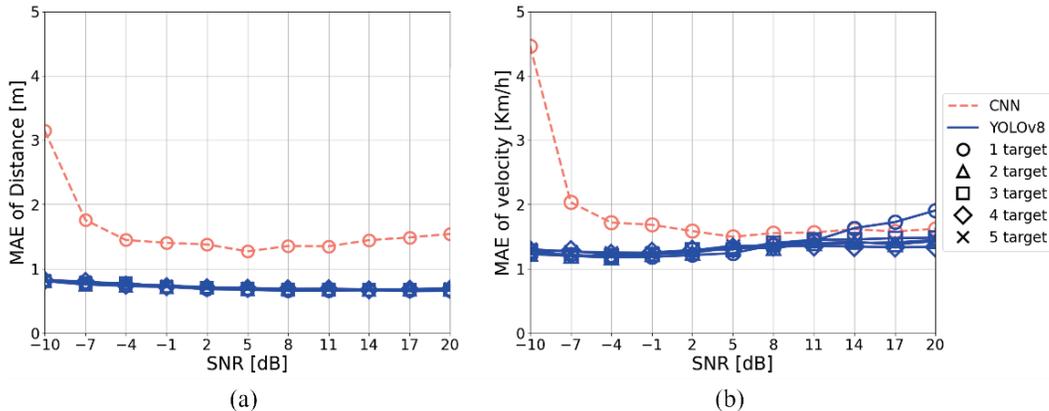


Figure 7 Distance and Velocity Estimation MAE of CNN and YOLO Models: (a) Distance, (b) Velocity

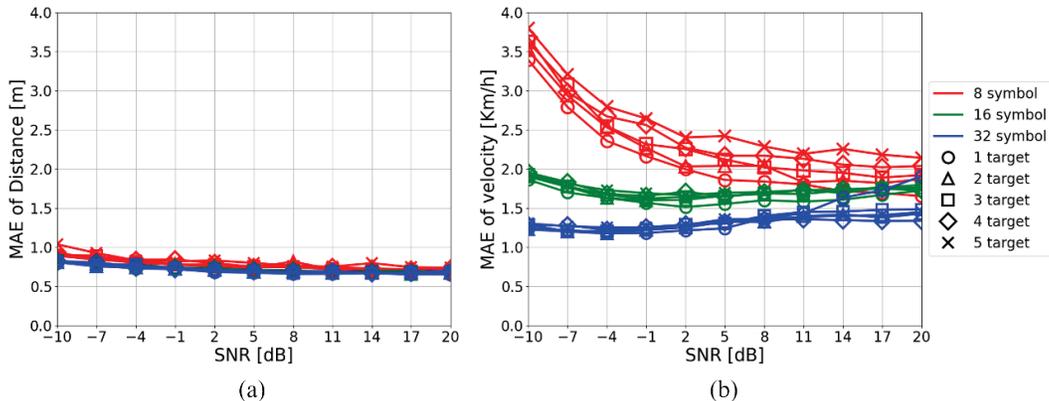


Figure 8 Distance and Velocity Estimation MAE of YOLO Model: (a) Distance, (b) Velocity

5.2.3 The Velocity and Distance Estimation Performance of the Proposed Method

Figs. 8(a) and 8(b) show the variation in Mean Absolute Error (MAE) for velocity and distance estimation,

respectively, as a function of OFDM symbol length under different SNR conditions using the YOLOv8 model. The simulation results indicate that longer OFDM symbol durations and fewer targets generally lead to improved estimation accuracy for both velocity and distance.

Additionally, as the SNR increases, MAE consistently decreases across all conditions. Notably, the lowest estimation errors were observed when the OFDM symbol length was 32, with an average velocity estimation error of 1.34 km/h and a distance estimation error of 0.71 m. Tab. 4 summarizes the average estimation performance for velocity and distance across different symbol lengths, clearly showing a decreasing trend in MAE as the number of OFDM symbols increases.

Table 4 Average MAE of Velocity and Distance Estimation

MAE	YOLOv8s		
	8symbol	16symbol	32symbol
Velocity (km/h)	2.3403	1.7088	1.3366
Distance (m)	0.7764	0.7162	0.7118

6 DISCUSSION

In summary, the proposed YOLOv8-based multi-target distance and velocity estimation method demonstrated stable performance across various signal-to-noise ratio (SNR) conditions and target counts. While conventional CNN-based models compress features into a single vector or focus on a single class—leading to information loss and reduced accuracy in multi-target environments—YOLOv8 takes advantage of its structural design to simultaneously detect the presence and location of multiple objects in the input signal. This allows the model to flexibly handle varying numbers of targets without requiring a predefined number of outputs. In addition, CNN has a fixed number of output nodes, which means a separate model is needed for each target count. Therefore, in our comparison, the CNN model was trained under the simplest condition—single-target only—while the YOLO model was trained to handle one to five targets. Despite this, YOLO achieved higher accuracy and lower estimation errors, demonstrating that it is an effective model not only for multi-target scenarios but also for single-target conditions. Meanwhile, since this study was conducted using computer-simulated data, it does not fully reflect real-world wireless conditions such as interference, channel distortion, and multipath effects. Moreover, the quantitative analysis of performance degradation in cases where targets have very similar distances and velocities was limited. As future work, we plan to validate the model's robustness using real-world field data. We also intend to explore new methods, such as enforcing minimum distance constraints between targets, to maintain high accuracy even in densely populated target environments. These extensions are expected to further enhance the practicality and reliability of the proposed model.

7 CONCLUSION

This study proposes a novel method for simultaneously estimating the number, distance, and velocity of multiple targets in an OFDM-based radar environment using YOLO (You Only Look Once). The approach aims to address the issue of limited frequency resources caused by dedicated radar signals while also overcoming the structural limitations of conventional two-stage CNN-based estimation methods in multi-target scenarios. Ultimately, the goal is to develop a real-time estimation technique suitable for next-generation

Integrated Sensing and Communication (ISAC) systems. The proposed YOLO-based method adopts a one-stage detection architecture that enables the simultaneous prediction of target count, distance, and velocity. It maintains high computational efficiency and accuracy even under diverse SNR conditions and in the presence of multiple targets. Notably, with an OFDM symbol length of 32, the YOLO model achieved an average velocity estimation error of 1.34 km/h and a distance estimation error of 0.71 m—representing improvements of 0.56 km/h and 0.89 m, respectively, compared to the baseline CNN approach. These results are attributed to YOLOv8's real-time inference capabilities and its strength in multi-object detection, which effectively leverage the distinct target features observable in the 2D-periodogram. Moreover, the proposed method enables accurate distance and velocity estimation without relying on dedicated radar signals, thereby conserving frequency resources and maximizing spectral efficiency—enhancing the practical utility of ISAC technologies. To address the limitations of simulation-based evaluation, future work will validate the model's generalization performance and real-time applicability using real-world wireless data. To this end, an OFDM-based radar system tailored to 5G communication environments will be developed, and experiments will be conducted to assess its robustness and deployment feasibility. In the long term, the proposed method will also be evaluated under emerging 6G network conditions to verify its practicality and scalability across diverse communication scenarios.

Acknowledgment

This work was supported by Institute for Information & communications Technology Promotion (IITP) grant funded by the Korea government (MSIP) (No.RS-2024-00405377, Development of 6G Integrated Sensing-Communication Radio Channel Model and Modular Beamforming Measurement System).

8 REFERENCES

- [1] Wiesbeck, W., & Sit, L. (2014, October). *Radar 2020: The future of radar systems*. 2014 International Radar Conference, 1–6. <https://doi.org/10.1109/radar.2014.7060395>
- [2] Sun, S., Petropulu, A. P., & Poor, H. V. (2020). MIMO radar for advanced driver-assistance systems and autonomous driving: Advantages and challenges. *IEEE Signal Processing Magazine*, 37(4), 98–117. <https://doi.org/10.1109/msp.2020.2978507>
- [3] Pisa, S., Pittella, E., & Piuze, E. (2016). A survey of radar systems for medical applications. *IEEE Aerospace and Electronic Systems Magazine*, 31(11), 64–81. <https://doi.org/10.1109/maes.2016.140167>
- [4] Zhang, Z., Long, K., Vasilakos, A. V., & Hanzo, L. (2016). Full-duplex wireless communications: Challenges, solutions, and future research directions. *Proceedings of the IEEE*, 104(7), 1369–1409. <https://doi.org/10.1109/jproc.2015.2497203>
- [5] Bhattarai, S., Park, J. M., Gao, B., Bian, K., & Lehr, W. (2016). An overview of dynamic spectrum sharing: Ongoing initiatives, challenges, and a roadmap for future research. *IEEE Transactions on Cognitive Communications and Networking*, 2(2), 110–128. <https://doi.org/10.1109/tccn.2016.2592921>

- [6] Liu, F., Cui, Y., Masouros, C., Xu, J., Han, T. X., Eldar, Y. C., et al. (2022). Integrated sensing and communications: Toward dual-functional wireless networks for 6G and beyond. *IEEE Journal on Selected Areas in Communications*, 40(6), 1728–1767. <https://doi.org/10.1109/jsac.2022.3156632>
- [7] Sit, Y. L., Sturm, C., Reichardt, L., Zwick, T., & Wiesbeck, W. (2011, April 17–22). The OFDM joint radar-communication system: An overview. In *The Third International Conference on Advances in Satellite and Space Communications (SPACOMM 2011)* (pp. 69–74). IARIA.
- [8] Baquero Barneto, C., Riihonen, T., Turunen, M., Anttila, L., Fleischer, M., Stadius, K., et al. (2019). Full-duplex OFDM radar with LTE and 5G NR waveforms: Challenges, solutions, and measurements. *IEEE Transactions on Microwave Theory and Techniques*, 67(10), 4042–4054. <https://doi.org/10.1109/tmtt.2019.2930510>
- [9] Delamou, M., Bazzi, A., Chafii, M., & Amhoud, E. M. (2023, June 20–23). Deep learning-based estimation for multitarget radar detection. In *2023 IEEE 97th Vehicular Technology Conference (VTC2023-Spring)*. IEEE. <https://doi.org/10.1109/VTC2023-Spring57618.2023.10200157>
- [10] Geng, Z., Yan, H., Zhang, J., & Zhu, D. (2021). Deep-learning for radar: A survey. *IEEE Access*, 9, 141800–141818. <https://doi.org/10.1109/access.2021.3119561>
- [11] Choi, J.-W., & Jeong, E.-R. (2022). A multi-output convolutional neural network-based distance and velocity estimation technique. *Journal of Logistics, Informatics and Service Science*, 9(1), 11–25. [https://www.aasmr.org/liss/Vol.9/JLISS%20February%202022%20\(2\)/Vol.9No.01.02.pdf](https://www.aasmr.org/liss/Vol.9/JLISS%20February%202022%20(2)/Vol.9No.01.02.pdf)
- [12] Li, Z., Liu, F., Yang, W., Peng, S., & Zhou, J. (2022). A survey of convolutional neural networks: Analysis, applications, and prospects. *IEEE Transactions on Neural Networks and Learning Systems*, 33(12), 6999–7019. <https://doi.org/10.1109/tnnls.2021.3084827>
- [13] Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016, June). You only look once: Unified, real-time object detection. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 779–788. <https://doi.org/10.1109/cvpr.2016.91>
- [14] Cho, S. J., Shin, S. U., Yoo, K. S., Min, K., & Jung, B. J. (2024). A study on real-time right-turn accident prevention model using YOLO object detection and transfer learning. *The Journal of Innovation Industry Technology*, 2(4), 151–158. <https://doi.org/10.60032/jiit.2024.2.4.151>
- [15] Tian, Z., Shen, C., Chen, H., & He, T. (2019, October). FCOS: Fully convolutional one-stage object detection. *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*. <https://doi.org/10.1109/iccv.2019.00972>

Authors' contacts:

So-Yeon Jeon, Master's Course
Department of Artificial Intelligence,
Hanbat National University,
109 Jiphyeonbuk-ro, Sejong, 30139, Republic of Korea
jeonso1109@gmail.com

Eui-Rim Jeong, Professor
(Corresponding author)
Department of Artificial Intelligence Software,
Hanbat National University,
109 Jiphyeonbuk-ro, Sejong, 30139, Republic of Korea
Mobile Phone: +82-010-4710-1082
erjeong@hanbat.ac.kr

Image Generation of Aesthetic Massage types Generative AI: Applying the Khizer Abbas Framework

Young Sook Park, Yun Kyoung Oh*

Abstract: AI image generation tools are transforming the beauty industry by offering new possibilities for visualizing skincare massage types through realistic image generation from text-based prompts. This study aims to explore this potential by selecting six massage types based on their practicality and visual representation potential. Three skincare professionals with over 10 years of experience identified experts Swedish Massage, Foot Massage, Bamboo Therapy, Aromatherapy, Thai Massage, and Stone Therapy, and images were generated using DALL·E, an AI-based image model. A total of 24 images (4 per framework) were generated for analysis. Images were created by applying Khizer Abbas framework types, R-T-F(Role-Task-Format), T-A-G(Task-Action-Goal), C-A-R-E(Context-Action-Result-Example), and R-I-S-E(Role-Input-Steps-Expectation). The application of these frameworks resulted in distinct image generation focuses, with each framework emphasizing different key elements. RTF highlighted the therapist's role and treatment environment, while TAG effectively depicted massage techniques and relaxation states. CARE prioritized the treatment setting through example-based visuals, creating a harmonious representation of the environment and session, whereas RISE focused on step-by-step procedures and professional expertise, ensuring a structured and comprehensive visual portrayal of massage treatments. AI models faced limitations in depicting physical contact and specialized roles, but reconfiguring settings and tools improved image generation. Each framework emphasized specific elements, such as the therapist's role, techniques, environment, and effects, allowing for a comparative analysis of visual representations across massage types. The main contribution of this study is the comparative analysis of four visual prompt frameworks applied to AI-generated skincare massage images. Further research on effective prompt-writing standards and methodologies for the beauty industry could expand AI's applicability in beauty, education, and design fields.

Keywords: Aesthetic; AI; Dall-E; Khizer Abbas; Massage Type; Prompt Framework

1 INTRODUCTION

Artificial Intelligence (AI) technology is driving innovative changes across various industries, and image generation AI (Generative AI) has demonstrated its ability to autonomously create creative content beyond user data analysis. Recently, with advancements in deep learning and neural network technologies, AI's learning and generative capabilities have become more sophisticated. As a result, natural language processing, image generation, design, marketing, art, gaming, and education have emerged as new creative tools [1]. Representative models include Midjourney, Stable Diffusion, Playground, and DALL·E, which provide high-quality image generation capabilities with simple text prompts. These models are utilized in various fields, including information retrieval and content creation [2]. DALL·E is OpenAI's latest model, gaining attention in AI transformation and creative processes due to its integration with ChatGPT. This integration allows it to automatically convert and precisely reflect text commands based on user input, providing the desired output instantly [3]. Additionally, it is equipped with a filtering system that considers AI ethics and content safety. OpenAI continues to foster a responsible AI creative environment through ongoing updates [4]. According to a 2025 McKinsey report [5], Generative AI technology is expected to create approximately \$10 billion in economic value within the beauty industry. Key application areas include hyper-personalized targeting, immersive product discovery, rapid packaging concept development, and innovative product creation. This indicates that Generative AI is evolving beyond a mere auxiliary technology to become a core driver of paradigm shifts in the beauty industry.

The annual Consumer Electronics Show (CES) also

showcases changes in the beauty industry, where Generative AI is being utilized in various ways, such as AI-driven skin tone and facial shape analysis, personalized skincare recommendations, and customized makeup suggestions. For example, Amorepacific has introduced AI-powered technology that recommends personalized makeup based on AI learning data, while L'Oréal has developed AI solutions that digitize skin conditions to predict aging processes and suggest personalized skincare routines. These data-driven total beauty solutions are enabling futuristic beauty experiences. The integration of Generative AI with beauty technology enhances personalized consumer services. As AI-driven personalized services and product development continue to advance, a consumer-centric, tailored market is expected to drive innovation in the beauty industry.

In particular, Khizer Abbas proposed a structured prompt framework to enhance AI interactions by systematically organizing prompts. The DALL·E model incorporates this framework, which serves as a methodology for structuring prompts to facilitate effective communication with AI. By utilizing a structured framework, Generative AI can be assigned specific roles and provided with additional contextual and situational information, allowing it to produce optimal results [6]. Since the quality and direction of AI responses can vary depending on the format of the input prompt, designing prompts with clear instructions and specific examples is considered a key strategic approach. Prompt Engineering refers to the process of designing and optimizing prompts in Large Language Models (LLMs). It includes Zero-shot learning and Few-shot learning techniques. Zero-shot learning enables an LLM to perform tasks based solely on simple instructions, without requiring specific training data, whereas Few-shot learning involves providing the model with a small number of examples,

allowing it to generate more precise results [7]. To achieve more precise AI responses, Khizer Abbas proposed five ChatGPT prompt frameworks, R-T-F, T-A-G, B-A-B, C-A-R-E, and R-I-S-E, each serving a specific purpose: Role (assigning roles), Task (defining tasks), Format (specifying output format), Action (specific tasks), Goal (setting goals), Before (current problem), After (desired results), Bridge (requesting solutions), Context (situation), Result(goal), Example (case study), Input (providing information), Step (outlining processes), Expectation (desired results). By utilizing these frameworks, AI interactions can be structured more systematically, leading to more precise and effective AI responses [6].

In previous studies on image-generating AI in the beauty industry, DALL·E and Midjourney have been used to explore AI-based beauty technology applications [8]. A Study on Hairstyle Image Generation Using Generative AI in Beauty Content Development [9], as well as employing Midjourney as an instructional tool in beauty-related education and curriculum development [10, 11]. In addition, studies have explored the creation of painterly and cubist-inspired makeup designs and body art using generative AI [12-14].

These studies indicate that generative AI holds strong potential for application across various areas of the beauty and cosmetics industry, including design processes, education, product development, and visual content creation. However, while AI is expected to bring significant innovation to the beauty sector overall, research specifically focused on image generation in the field of skincare and aesthetic treatment remains limited.

The term "massage" originates from Greek, meaning "knead," "press," and "rub". It emphasizes therapeutic effects such as reducing swelling, improving blood circulation, and relieving pain and muscle tension [15]. In Korea, massage is an essential part of the skincare industry, utilizing hands-on techniques to maintain, protect, and enhance skin health [16]. Swedish Massage, one of the world's three major massage techniques, was developed in Sweden and became the first documented massage system. Introduced to Korea in 1998, it features five key techniques and is globally recognized through structured educational programs. Using warm oil, it helps relieve muscle tension, improve blood circulation, and enhance sleep quality [17]. Foot massage is a technique that stimulates the reflex zones distributed across the feet using hands or tools. It promotes circulation of blood and lymph in the circulatory system and is particularly effective in relieving fatigue and alleviating sleep disorders [18]. Methods include callus removal, circulation massage, and reflexology. Bamboo Therapy utilizes the silica content in bamboo to aid fascia and pain point management. It includes both pressurized and non-pressurized methods, helping stabilize the autonomic nervous system, boost immunity, and promote detoxification [19]. Aromatherapy is a holistic therapy using essential oils and aromatic plants. It has been used since ancient times and became widely applied after the 1930s [20]. It is commonly used for stress relief, fatigue recovery, and immune enhancement. The main methods include massage, inhalation, and bathing, with massage being particularly effective as it stimulates both olfaction and skin

absorption simultaneously [21]. Thai Massage is influenced by Indian yoga and Chinese Tuina therapy. It focuses on energy flow and balance along the Sen meridians, helping to align muscles and improve flexibility. The massage progresses from the feet to the head, moving through supine, prone, and seated positions in stages [15]. Stone Therapy originated from the healing practices of ancient Native Americans and evolved into an aesthetic treatment in the U.S. during the 1990s, gaining popularity in hotels and medical spas [22]. It utilizes the natural energy of stones to boost metabolism, prevent skin aging, and promote cell regeneration. Hot stones basal retain heat to aid muscle relaxation, while cool stones marble provide a cooling effect to reduce muscle contraction and alleviate pain [22]. With technological advancements, the massage industry is rapidly growing, providing users with various conveniences [23].

In the field of skincare and aesthetics, over 30,000 aesthetic skincare clinics are operating nationwide [24], making up a significant portion of the beauty industry. This sector continues to develop with increasing specialization and professionalism. The primary types of skin massage utilized in the skincare industry include Swedish massage, Thai massage, aromatherapy, reflexology, bamboo therapy, stone therapy, and lymphatic drainage, each differing significantly in terms of history, purpose, techniques, and visual representation. However, it is difficult to find studies that specifically analyze massage types within the context of skincare and utilize image-generating AI to visualize skincare massage techniques. Visualizing various skincare techniques using AI tools is expected to contribute to educational materials, personalized advertisements, virtual experiences, and simulation-based effect predictions. Therefore, this study aims to generate images representing key massage techniques using the DALL·E image generation model. By analyzing the differences in the generated images based on Kizer Abbas' framework, this research seeks to provide visual educational resources for the skincare and aesthetics industry while proposing new AI applications in the field. This study set the research question as follows.

RQ1: What are the possibilities and limitations of generating images of different massage types using generative AI?

RQ1: What are the differences in image expressions by the framework (4 types) suggested by Khizer Abbas, and which framework is the most suitable for expressing massage types?

2 MATERIALS AND METHODS

2.1 Selection of Treatment Types

Based on the NCS (National Competency Standards) learning module and a review of previous studies, various massage types were researched. Considering their current applications in the skincare industry and visual representation potential, a panel of three experts comprising two skincare professionals with over 10 years of experience and one professor selected six key massage types. The selected massage types are Swedish Massage, Foot Massage, Bamboo Therapy, Aromatherapy, Thai Massage, and Stone

Therapy. Each type was analyzed in terms of concept, tools, techniques, and effects, summarizing its essential characteristics.

2.2 Image Generation Process

The research procedure first involved collecting data on six massage types currently used in aesthetic treatments. Next, a prompt framework suitable for image generation was selected. Finally, the prompt framework was applied to DALL·E 4 to generate the images (Tab. 1). In this study, a text-to-image generation method was employed, in which textual prompts were used as input to produce corresponding images reflecting the described content. In the final stage, the generated AI images were analyzed and reviewed to assess whether the distinctive features of each massage type were effectively visualized.

Table 1 Initial Set of features used for the experimentation

Step	Category	Description
1	Preliminary Research	Data collection on massage types (NCS, literature review)
2	Prompt Selection	Application of Khizer Abbas Framework (RTF, TAG, CARE, RISE)
3	Image Generation	Applying the framework to DALL·E prompts for massage image creation
4	Image Analysis	Analyzing results based on framework-generated images

2.3 Prompt Framework Development

To generate responses for ChatGPT-based image creation, the Khizer Abbas prompt frameworks RTF, TAG, CARE, and RISE were applied to structure role-based

prompts for different massage types (Tab. 2). Among the five prompt frameworks proposed by Khizer Abbas, BAB was excluded due to its limited applicability in text-based input, leaving four selected frameworks.

Each framework was adapted with terminology relevant to skincare and aesthetics, structured as follows (Tab. 2). Role (R) is Therapist’s role and gender, Task (T) is Treatment purpose and anatomical focus, Format (F) is Spa ambiance, background setting, and tools, Action (A) is Techniques and pressure application, Goal (G) is Treatment effects and overall impact, Context (C) is Treatment situation and environment, Result (R) is Outcome and final state, Example (E) is Reference images of similar treatments, Input (I) is Key elements of the treatment, Steps (S) is Treatment procedures, Expectation (E) is Expected treatment effects. A standardized prompt structure was applied across all massage types, with customized elements assigned to each framework, as detailed in Tab. 3 and 4.

Table 2 Framework Criteria for Massage Types

Framework	Type	Description
RTF	Role (R1)	Therapist’s role, gender
	Task (T)	Treatment purpose, anatomical focus
	Format (F)	Spa ambiance, background setting, tools
TAG	Task (T)	Treatment purpose, anatomical focus
	Action (A)	Technique, pressure application
	Goal (G)	Treatment effects, impact
CARE	Context (C)	Treatment situation
	Action (A)	Technique, pressure application
	Result (R2)	Treatment outcome, final state
	Example (E)	Reference image of similar treatments
RISE	Role (R1)	Therapist’s role, gender
	Input (I)	Key treatment elements
	Steps (S)	Treatment procedures
	Expectation (E)	Expected treatment effects

Table 3 Application of Khizer Abbas Framework (Swedish, Foot, Bamboo)

Type	Swedish Massage	Foot Massage	Bamboo Therapy
R1	Swedish massage specialist / Female	Reflexology specialist / Female	Bamboo therapy specialist / Female
T	Relaxation, comfort / Upper back area, effleurage	Pain relief, internal function activation / Sole, instep, toes, ankle	Lymph circulation, detoxification / Upper back area, effleurage
F	A European-style spa room / Oil, hands	A quiet and comfortable treatment space / Foot care cream, hands	A nature-inspired spa room / Oil, bamboo
A	Effleurage - Petrissage - Friction - Tapotement - Vibration / Moderate pressure control	Acupressure, rolling, stretching / Adjustable pressure control	Effleurage - Petrissage - Rolling - Compression - Tapotement / Moderate pressure control
G	Blood circulation, muscle relaxation / Maintaining body balance	Improved foot health, increased joint flexibility / Posture balance	Muscle fatigue relief, flexibility improvement / Maintaining body balance
C	Pre-consultation, client condition assessment, setting up massage environment (lighting, music, bed temperature control)	Recliner chair, massage bed, feet wrapped in warm towels	Comfortable bed, selecting treatment based on client’s condition
R2	Muscle tension relief, lymph circulation / Enhanced skin health	Alleviated fatigue and pain / Feet become softer and less fatigued	Muscle fatigue relief, body balance adjustment
E	Photo of effleurage applied on a client’s back	Image of a therapist using fingers and palms for massage in a therapy space	Photo of a therapist using bamboo sticks on a client lying comfortably
I	Tools, environment (diffuser)	Foot condition assessment, reflex zone selection	Body analysis, natural environment, various bamboo tools
S	Consultation - Preparation - Massage - Post-treatment care	Preparation - Basic massage - Reflexology acupressure - Rolling - Stretching	Preparation - Basic massage - Deep muscle care - Stretching
E	Physical and mental recovery, stress relief	Foot fatigue relief, improved mobility in daily activities	Body balance maintenance, improved lymph circulation, blood circulation

Table 4 Application of Khizer Abbas Framework (Aromatherapy, Thai, Stone)

Type	Aromatherapy	Thai Therapy	Stone Therapy
R1	Aromatherapy Specialist / Female	Thai Massage Specialist / Female	Stone Therapy Specialist / Female
T	Relaxation, emotional stability / Upper back / Effleurage	Improved flexibility, energy balance / Full-body treatment	Muscle tension relief, stress reduction / Upper back / Effleurage
F	Nature-inspired spa room / Oil, bamboo	Traditional Thai-style interior, natural elements / Mat, oil, herbal compress	Calm and soothing spa room / Stones, oil, lotion
A	Effleurage - Petrissage - Lymphatic Drainage - Tapotement / Gentle touch	Thai stretching - Acupressure - Deep Pressure - Herbal Compress Therapy / Light to strong pressure	Effleurage - Petrissage - Friction - Stone Therapy / Light to strong pressure
G	Tension relief, fatigue recovery, increased skin elasticity / Health promotion	Muscle relaxation, improved joint flexibility / Reduced physical fatigue	Muscle knot relief, pain reduction / Decreased physical fatigue
C	Comfortable bed, client in a relaxed state, selection of essential oils	Traditional Thai spa room, comfortable mat, stretching techniques	Quiet and cozy spa room, comfortable bed, hot stones
R2	Feeling of lightness, emotional stability / Improved circulation	Detoxification, increased joint flexibility	Muscle tension relief, reduced body fatigue
E	Client lying comfortably while the therapist gently applies oil for treatment	Client in a relaxed state on a mat while the therapist performs massage techniques	Client lying comfortably on a bed while the therapist applies hot stone therapy
I	Body analysis, oil selection, technique application	Traditional Thai spa environment, tools, natural elements	Environment (spa room), tools, technique application
S	Preparation - Basic massage - Deep muscle treatment - Lymphatic circulation - Finalization	Preparation - Basic stretching - Acupressure therapy - Finalization	Preparation - Basic massage - Deep treatment - Hot stone application - Finalization
E	Induced deep sleep, improved skin elasticity, balanced body and mind	Nervous system relaxation, stress reduction, psychological balance restoration, enhanced concentration	Increased skin elasticity, relaxation, deep sleep induction

3 RESULTS AND DISCUSSION

3.1 Image Generation Based on the R-T-F Prompt Framework

Applying the R-T-F framework, the generated images effectively depicted the role, task, and format for each massage type (Tab. 5). The image of Swedish Massage emphasizes the role of a female professional therapist in a European-style spa environment, highlighting a relaxing ambiance in alignment with the treatment purpose and target area. The image of Foot Massage showcases a female therapy specialist in a quiet, cozy treatment room, applying foot care techniques using specialized creams and tools on a client's feet. The image of Bamboo Therapy illustrates a female therapist in a nature-friendly treatment room, gently treating the client's upper body using bamboo sticks and oil, promoting lymph circulation and detoxification through effleurage techniques. The image of Aromatherapy depicts a female therapist in a nature-inspired treatment room stocked with essential oils, applying gentle effleurage techniques to a client's back, aiding relaxation and emotional balance. The image of Thai Massage features a female Thai massage

therapist conducting a full-body treatment on a mat in a spa room with traditional Thai décor and natural elements, using herbal compresses and oils. The image of Stone Therapy portrays a female therapist utilizing hot stones, applying effleurage techniques on a client's back, aiming to relieve muscle tension and stress. While the R-T-F framework successfully structured key prompt elements, the AI-generated outputs mainly produced illustrated images rather than realistic representations, particularly for massage techniques and exposed treatment areas. In the beauty image generation study by Kim et al. (2025) [25], the R-T-F framework was used to structure the prompt. The Role was set as a beauty stylist, the Task involved presenting beauty styling elements for generating seasonal images, and the Format was designed to include both background and model figures to generate a complete image. In contrast, the present study focused more on the visual representation of skincare techniques rather than beauty styling, with the prompt structured around elements such as the expected effects of treatment.

Table 5 Image Generation Based on the R-T-F prompt Framework

Type	Swedish Massage	Foot Massage	Bamboo Therapy	Aromatherapy	Thai Massage	Stone Therapy
Image						
Description	Emphasized the therapist's role in a European-style spa; relaxing ambiance aligned with treatment purpose.	Displayed a therapist applying foot care using tools and creams in a cozy treatment room.	Illustrated a therapist using bamboo sticks and oil on the upper body; promoted lymph circulation and detox.	Depicted therapist-applying effleurage in a nature-inspired room with essential oils; supported relaxation.	Featured a Thai therapist conducting full-body treatment with herbal compresses in a traditional setting.	Showed a therapist using hot stones and effleurage on the back; aimed at relieving muscle tension.

3.2 Image Generation Based on the T-A-G Prompt Framework

Applying the T-A-G framework, the generated images successfully incorporated techniques, actions, and goals for

each massage type (Tab. 6). The image of Swedish Massage visually emphasized techniques and pressure control within a relevant background setting, showcasing a harmonious balance between body alignment and relaxation. The image

of Foot Massage depicted only the act of massaging the client's feet, without effectively illustrating stretching techniques or functional activation goals, suggesting the need for clearer prompt instructions. The image of Bamboo Therapy effectively categorized various techniques for lymph circulation and muscle relaxation, while the goal and impact were represented through a calm and comfortable client's expression. The image of Aromatherapy portrayed a client experiencing relaxation and tension relief, with touch sensations and massage techniques visually categorized. The image of Thai Massage highlighted professional stretching and acupressure techniques, while the client was depicted in a deep relaxation state. The image of Stone Therapy illustrated the application of stone therapy techniques,

emphasizing muscle tension relief and stress reduction. When applying the T-A-G framework, massage techniques and movements were categorized and visually represented, while treatment effects were expressed through images of clients with a relaxed and healing demeanor. Kim & Cho (2024) [26] conducted a study applying generative AI in the field of arts education, incorporating Khizer Abbas framework into the lesson design process for practical classes in theater production and primarily utilizing a role-based prompt framework. In contrast, the present study is distinct in that it involves the construction and application of multiple prompt frameworks to generate and analyze AI-based skincare massage images.

Table 6 Image Generation Based on the T-A-G prompt Framework

Type	Swedish Massage	Foot Massage	Bamboo Therapy	Aromatherapy	Thai Massage	Stone Therapy
Image						
Description	Techniques and pressure control shown in a balanced body alignment and relaxing background.	Foot massage action shown, but lacked clarity on stretching and functional goals; prompt refinement needed.	Muscle relaxation techniques represented; client's calm expression conveyed goal and effect.	Relaxation and tension relief visualized; techniques and sensations clearly depicted.	Stretching and acupressure techniques highlighted; client shown in deep relaxation.	Stone therapy techniques illustrated; emphasized muscle relief and stress reduction.

3.3 Image Generation Based on the C-A-R-E Prompt Framework

Applying the C-A-R-E framework, the generated images effectively emphasized treatment context, actions, results, and examples for each massage type (Tab. 7). The image of Swedish Massage highlighted the treatment environment and results, depicting a spa setting with a well-structured treatment space, including an example of effleurage applied to a client's back. The image of Foot Massage primarily illustrated the treatment context, showing a warm and cozy recliner chair. However, the treatment results and final state were conveyed more abstractly rather than through direct visual representation. The image of Bamboo Therapy

categorized the comfortable bed and treatment setting visually, with bamboo tools shown alongside massage techniques. The image of Aromatherapy illustrated the bed and the client's relaxed posture, with an example effectively used to generate an accurate depiction. The image of Thai Massage portrayed a traditional Thai spa environment through natural elements, while the example showed a fully relaxed client lying on a mat. The image of Stone Therapy integrated a calm spa room, a comfortable bed, and hot stones, with the treatment example properly applied to produce a realistic result. When applying the C-A-R-E framework, the images effectively depicted treatment contexts and examples, while the treatment results and final states were primarily conveyed through overall environmental settings.

Table 7 Image Generation Based on the C-A-R-E prompt Framework

Type	Swedish Massage	Foot Massage	Bamboo Therapy	Aromatherapy	Thai Massage	Stone Therapy
Image						
Description	Spa setting and treatment results clearly shown, example of effleurage applied to the back.	Warm chair setting emphasized; results expressed abstractly rather than visually.	Comfortable bed and bamboo tools represented, techniques illustrated in context.	Client's relaxed state and treatment bed shown, example used to enhance accuracy.	Traditional Thai setting with natural elements, client fully relaxed on a mat.	Spa room, bed, and hot stones well integrated, realistic representation achieved through applied example.

3.4 Image Generation Based on the R-I-S-E Prompt Framework

Applying the R-I-S-E framework, the generated images effectively depicted treatment elements, procedures, and

expected effects for each massage type (Tab. 8). The image of Swedish Massage clearly represented the tools and environment used in the treatment, with the step-by-step process effectively visualized. The overall treatment effects were conveyed through both imagery and background

elements. The image of Foot Massage illustrated reflex zones and foot condition assessments, while the treatment effects were integrated into the background. Due to prompt limitations, the image appeared more illustrative, but Bamboo Therapy tools and environmental elements were clearly represented. The treatment process was mainly suggested through background imagery. The image of Aromatherapy featured essential oils and technique explanations in the background, with treatment effects such as balance maintenance visually emphasized. The image of Thai Massage effectively depicted the traditional Thai setting, tools, and natural elements, with stress-reduction effects prominently highlighted. The image of Stone Therapy successfully showcased the spa room, treatment tools, and various types of stones, offering a strong visual depiction of the step-by-step process—particularly emphasizing deep relaxation and sleep induction. When applying the R-I-S-E framework, the images successfully visualized treatment

elements, procedures, and effects, harmonizing the therapist’s expertise, the client’s relaxation, and the healing atmosphere. This resulted in a comprehensive visual representation of the core essence of each massage technique. In the beauty image generation study by Kim et al. (2025) [25], the R-I-S-E framework was used to structure the prompt. The Role was set as a beauty stylist, and the Input involved generating seasonal beauty styling content based on personal color analysis using ChatGPT. The Steps provided a step-by-step breakdown of beauty styling elements, and the Expectation clearly stated the goal of generating character images representing each season. In contrast, the present study focuses not on beauty styling, but on aesthetic massage types, structuring prompts around treatment procedures, tools used, and expected effects. Thus, unlike the previous study, this research emphasizes the visual representation of hands-on treatment processes and therapeutic outcomes, rather than stylistic image suggestions.

Table 8 Image Generation Based on the R-I-S-E prompt Framework

Type	Swedish Massage	Foot Massage	Bamboo Therapy	Aromatherapy	Thai Massage	Stone Therapy
Image						
Description	Clear depiction of tools, environment, and step-by-step process; effects shown through imagery.	Reflex zones and foot condition shown; effects integrated into the background.	Bamboo tools and natural setting illustrated; process implied via background.	Essential oils and techniques visualized; emotional balance emphasized.	Traditional Thai setting and tools depicted; stress-relief effects highlighted.	Spa room and stones shown; step-by-step process and relaxation effects emphasized.

4 CONCLUSION

This study applied Khizer Abbas' ChatGPT prompt framework (R-T-F, T-A-G, C-A-R-E, R-I-S-E) to generate and analyze AI-generated images representing various massage treatment types. The results show that the structure and content of the prompt framework have a significant impact on the visual quality, clarity, and applicability of AI-generated images. In particular, the results of this study show that the generative AI can effectively visualize various skincare massage types, and the choice of framework plays an important role in how each treatment type is represented. While the R-T-F framework visually organizes roles, treatment environments, and treatment areas well, T-A-G effectively represents background and relaxation states, but lacks detailed technical representation. C-A-R-E mainly conveys the end state through environmental factors, so the results need to be represented more directly visually. R-I-S-E effectively represents treatment elements and step-by-step processes. Applying the image generation framework for each massage type, we found that the R-T-F and R-I-S-E frameworks were particularly effective in structurally visualizing the treatment environment and treatment steps.

By analyzing the image representations in various frameworks, this study showed that selecting a framework appropriate for each massage treatment type can improve the clarity and effectiveness of AI-generated visuals. Adjusting the visual elements within the framework can lead to more

accurate and informative representations. In addition, AI-based image generation had limitations in generating images related to physical contact and specific occupations (e.g., massage therapists, medical institutions). However, successful image generation was possible by modifying the prompts to emphasize spa interiors, massage oils, and treatment tools. In some cases, AI generated descriptive images instead of realistic visuals.

The visual output generated in this study has strong practical potential as an effective tool for supplementing beauty education curricula and developing consumer-targeted content in marketing. Therefore, further research on standards and methodologies for writing effective prompts for the beauty industry could expand the applicability of AI in beauty, education, and design.

This study is significant in that it explores the potential of AI image generation in the field of skin care, but it requires supplementation in quantitative evaluation of the image results. In addition, further research is needed on standards and methodologies for writing effective prompts for the beauty industry.

5 REFERENCES

[1] Hwang, M. K. (2024, August). A study on the use of artificial intelligence image generator in creative and artistic fields. *Asia-Pacific Journal of Convergent Research Interchange*, 10(8), 121–131. <https://doi.org/10.47116/apjcri.2024.08.10>

- [2] Hwang, M. K. (2024). *A study on the use of artificial intelligence image generator in creative and artistic fields*. <https://doi.org/10.47116/apjcri.2024.08.10>
- [3] Son, H. R., & Lee, C. K. (2024, August). Artificial intelligence utilization strategies in design: Insights from expert interviews. *Journal of Digital Contents Society*, 25(8), 2035–2044. <https://doi.org/10.9728/dcs.2024.25.8.2035>
- [4] Chang, Y. K., Yoon, I. H., Park, J. W., & Lee, J. Y. (2023). Understanding AI image generator and exploring educational possibilities in art education. *Art Education Review*, 88, 277–298. <https://doi.org/10.25297/AER.2023.88.277>
- [5] OpenAI. (n.d.). *DALL-E 3*. Retrieved February 9, 2025, from <https://openai.com/dall-e-3>
- [6] McKinsey & Company. (2025, January 6). *How beauty players can scale gen AI in 2025*. Retrieved February 9, 2025, from <https://www.mckinsey.com/industries/consumer-packaged-goods/our-insights/how-beauty-players-can-scale-gen-ai-in-2025>
- [7] Kim, Y. Y., & Cho, J. H. (2024). Exploring the educational possibilities of generative AI in theatre production class. *Korean Association of Arts Education*, 22(3), 1–18. <https://doi.org/10.23317/kaae.2024.22.3.001>
- [8] OpenAI. (2023). *Prompt engineering guide*. San Francisco, CA: OpenAI. Retrieved February 9, 2025, from <https://platform.openai.com/docs/guides/prompt-engineering>
- [9] Lee, S. Y. (2024). Case study of generative AI-based beauty technologies: Focusing on image generation. *The Korean Society of Beauty and Art*, 25(3), 283–296. <https://doi.org/10.18693/jksba.2024.25.3.283>
- [10] Yi, S. Y., Kang, J. A., & Park, J. A. (2025). A study on the image generation by hair style using generative AI. *Journal of Beauty Art Management*, 19(4), 1–21. <https://doi.org/10.22649/JBAM.2025.19.4.1>
- [11] Moon, S. Y., & Kim, Y. N. (2024). Integrating generative AI into beauty education: Developing a convergence curriculum. *The Korean Society of Beauty and Art*, 25(4), 123–147. <https://doi.org/10.18693/jksba.2024.25.4.123>
- [12] Joo, Y. J. (2024). Research on the development of a makeup design motif using a generative AI image model and their application in education. *Journal of Business Convergence*, 9(5), 233–240. <https://doi.org/10.31152/JB.2024.10.9.5.231>
- [13] Lee, M. J., & Choi, E. S. (2023). A study on cubist art makeup design creation using artificial intelligence-based image generation model. *Journal of the Korean Society of Cosmetology*, 29(6), 1427–1444. <https://doi.org/10.52660/JKSC.2023.29.6.1427>
- [14] Jong, J. H., Kim, D. E., & Han, J. S. (2024). Using surrealism and image generative AI in the production of a painterly make-up work. *The Korean Society of Beauty & Art*, 25(3), 119–132. <https://doi.org/10.18693/jksba.2024.25.3.119>
- [15] Lee, Y. L. (2025). A study on the development of body art design through educational approaches using ChatGPT: Focusing on the artistic characteristics of surrealism. *Journal of Beauty Art Management*, 19(1), 137–161. <https://doi.org/10.22649/JBAM.2025.19.1.137>
- [16] Kim, J. H., Cho, A. Y., & Kwon, S. B. (2014). Asian massage technique. *Korean Journal of Aesthetics and Cosmetics Society*, 12(1), 9–15. UCI: G704-SER000010442.2014.12.1.002
- [17] Oh, Y. K. (2022). Analysis of research trends in skin care using keyword network analysis. *Journal of the Korean Society of Cosmetology*, 28(6), 1308–1317. <https://doi.org/10.52660/JKSC.2022.28.6.1308>
- [18] Han, C. J., & Jeong, E. A. (2010). The effect of Chuna and Swedish massages on abdominal obesity in Korean middle-aged women. *Asian Journal of Beauty and Cosmetology*, 8(2), 170–180. UCI: G704-SER000010442.2010.8.2.010
- [19] Yoon, I. S., & Choi, H. J. (2013). Effects of foot massage at the abdominal area reflex point on the electroencephalogram. *Asian Journal of Beauty and Cosmetology*, 11(3), 495–503. UCI: G704-SER000010442.2013.11.3.017
- [20] Lee, M. Y., & Oh, Y. K. (2020). Effects of bamboo neck lymph drainage on facial edema and skin condition. *Asian Journal of Beauty and Cosmetology*, 18(2), 209–220. <https://doi.org/10.20402/ajbc.2020.0025>
- [21] Oh, Y. K. (2014). Meta-analysis of the effectiveness on massage using aroma oil. *Journal of the Korean Society of Cosmetology*, 20(5), 885–892. UCI: G704-001852.2014.20.5.009
- [22] Oh, Y. K., & Bea, G. J. (2024). Effect of a hair loss management concurrently incorporating aroma scalp massage on hair and scalp conditions. *Journal of the Korean Society of Cosmetology*, 30(3), 524–532. <https://doi.org/10.52660/JKSC.2024.30.3.524>
- [23] Kim, E. Y., & Chae, S. N. (2006). A study on stone therapy. *Korean Journal of Aesthetics and Cosmetics Society*, 4(2), 107–114. UCI: G704-SER000010442.2006.4.2.006
- [24] Oh, D. H., Baek, Y. S., Min, K. T., & Kim, S. S. (2024). A study on the life cycle assessment of tablet PCs based on smartphone data. *The Journal of Innovation Industry Technology*, 2(4), 141–149. <https://doi.org/10.60032/JIIT.2024.2.4.141>
- [25] Khiss. (2025). *Kkosis*. Retrieved February 9, 2025, from https://khiss.go.kr/stats?menuId=MENU00346&schText1=E_MPLOY#content
- [26] Kim, C. W., Lee, Y. J., Lee, H. R., & Park, J. A. (2025). A study on personal beauty image generation using AI: Focusing on the use of DALL-E3. *Journal of the Korean Society of Cosmetology*, 31(1), 140–153. <https://doi.org/10.52660/JKSC.2025.31.1.140>
- [27] Kim, Y. Y., & Cho, J. H. (2024). Exploring the educational possibilities of generative AI in theatre production class. *Korean Journal of Art Education*, 22(3), 1–18. <https://doi.org/10.23317/kaae.2024.22.3.001>

Authors' contacts:

Young Sook Park, PhD student
Department of Cosmetology, Changshin University,
Gyeongsangnam-do, 51352, Korea
ysook7689@hanmail.net

Yun Kyoung Oh, Professor
(Corresponding author)
Department of Cosmetology, Changshin University,
Gyeongsangnam-do, 51352, Korea
Phone: +82-010-5011-3242
yk212@cs.ac.kr

Real-Time Hybrid Query Transformation Method for Enhancing Search System Performance in Korean Language Applications

Hyun Jung Kim, Sang Hyun Yoo*

Abstract: This study addresses the real-time performance limits of Korean-language search systems caused by morphological complexity and the cost of semantic processing. We propose a hybrid query transformation method that couples rule-based preprocessing with a Transformer-based postprocessor. The rule-based stage simplifies agglutinative input, and the Transformer refines user intent and semantic context. On a curated Korean query set, our approach attains 89.0% Precision@5 (95% CI: 87.2–90.7) with 95 ms average latency (95% CI: 92–98), about 21% faster than an NLP-only baseline. User surveys and expert interviews further confirm practical applicability. To strengthen reliability and scope transparency, we report five-fold cross-validation, noise-robustness tests (spacing errors, minor typos), and comparisons against open proxy baselines (e.g., BM25+KoNLPy). These additions clarify the study's focus on Korean while providing reproducible evidence of robustness, positioning the framework as deployment-ready for Korean and a solid basis for future multilingual extensions.

Keywords: NLP-based postprocessing; query optimization; real-time applications; real-time hybrid query transformation; rule-based preprocessing; search system performance

1 INTRODUCTION

In today's information-rich environment, effective query optimization is critical to ensure that search engines return highly relevant results. Although Transformer-based models excel at capturing contextual information [2, 3], unstructured user queries still degrade retrieval quality and user experience, as extensively documented by Mitra and Craswell [7] and further explored by Lin et al. [8]. At the same time, the computational cost of neural postprocessing hinders real-time deployment. In contrast, rule-based approaches provide rapid and predictable latency but lack the sophistication to interpret complex, context-dependent queries [4, 5]. This trade-off is evident in commercial systems such as Naver, which continuously balance accuracy and efficiency in production search [6].

To address these limitations, we introduce a real-time hybrid query transformation framework that couples rule-based preprocessing with Transformer-based postprocessing. Prior hybrid methods show promise [9, 19, 20], yet they are rarely optimized for agglutinative languages such as Korean, where morphological complexity and variable word order pose additional challenges. Our approach tailors the hybrid pipeline to Korean: rule-based preprocessing normalizes morphological artifacts and common surface noise (e.g., spacing errors, minor typos), while a Transformer refines user intent and semantic context. This work addresses the central research question: "Can a Korean-optimized hybrid approach meaningfully outperform conventional methods in both accuracy and speed to support real-time search?" We hypothesize at least a 5% gain in Precision@5 and a $\geq 15\%$ reduction in processing time over state-of-the-art NLP-only baselines on a Korean query dataset.

We evaluate the approach on a curated dataset of 200 user queries (predominantly Korean). Focusing on Korean is justified by the availability of high-quality annotated data that enables controlled validation and establishes a foundation for future multilingual studies. As an

agglutinative language with complex morphology and flexible word order, Korean demands robust preprocessing to extract meaningful terms and reliable semantic modeling to capture context [5, 21]. Lee et al. (2022) [5], for example, highlight these linguistic challenges in embedding-based document similarity and propose semantic feature expansion to mitigate morphological issues.

Direct benchmarking against proprietary industry systems was infeasible due to the lack of standardized, shareable datasets. Instead, we situate our method against state-of-the-art academic baselines (e.g., Lim et al. [5, 19]; Dogan and Gurcan [20]) and strengthen reliability through (i) five-fold cross-validation, (ii) noise-robustness tests (spacing errors, minor typos), and (iii) open, industry-relevant proxy baselines (e.g., BM25+KoNLPy and a lightweight neural variant) that approximate practical behavior while remaining reproducible. As a preview of results, our method achieves 89.0% Precision@5 (95% CI: 87.2–90.7) with 95 ms average latency (95% CI: 92–98), about 21% faster than an NLP-only baseline; detailed analyses appear in Section 4.

We note that our dataset is intentionally focused on Korean (200 queries), so broad multilingual generalization claims are out of scope. Nonetheless, the use of cross-validation, robustness checks, and open proxy benchmarks provides transparent and reproducible evidence of practical relevance. Future work will scale to larger, multi-domain, multilingual datasets and collaborate with industry partners to evaluate on anonymized production logs.

2 RELATED WORKS

Search query optimization has emerged as a critical research area in recent years driven by the need to interpret and transform increasingly complex user queries accurately. As Aggarwal comprehensively outlined in his work on information retrieval and search engines [12], the evolution of search technologies has been shaped by advances in both algorithmic approaches and computational capabilities. Two

main paradigms have dominated this field NLP-based approaches and rule-based methods.

2.1 NLP-Based Approaches

Transformer models such as BERT (Bidirectional Encoder Representations from Transformers) and T5 (Text-to-Text Transfer Transformer) have revolutionized natural language processing by providing deep contextual insights that significantly enhance query transformation [2, 3]. The Transformer architecture, introduced by Vaswani et al. in "Attention is All You Need" [2], has laid the foundation for large-scale models such as GPT and T5. In particular, T5 demonstrates exceptional performance in handling diverse natural language processing tasks in a unified manner and has been widely recognized for its application in search query optimization [1]. As Craswell (2020) highlights in his comprehensive review of deep learning applications in information retrieval [15], these neural approaches have dramatically transformed the search landscape by enabling a more nuanced understanding of user queries. However, despite their impressive capabilities, these models require significant computational resources, which limit their applicability in real-time scenarios [4, 8].

2.2 Rule-Based Approaches

In contrast, rule-based methods rely on heuristic algorithms—including keyword extraction, stopword removal, and syntactic parsing—to rapidly process user queries [4, 5]. While these approaches offer superior processing speed, they often lack the sophistication required to handle complex, context-dependent queries, resulting in performance degradation for intricate user inputs [13]. Furthermore, their reliance on fixed rules limits in processing ambiguous or context-dependent queries [14].

2.3 Hybrid Approaches

To address these limitations, recent research has explored hybrid frameworks that integrate the efficiency of rule-based preprocessing with the contextual refinement of NLP-based postprocessing. For example, Lim et al. (2023) introduced a two-stage model that combines rule-based preprocessing with Transformer-based postprocessing, achieving notable improvements in precision and processing speed [9]. Similarly, Dogan and Gurcan (2024) demonstrated the feasibility of hybrid methods in domain-specific applications such as e-business communication [19, 20]. Despite these advances, many studies have been limited to specific domains or single-language datasets, constraining their broader applicability [23, 24].

Moreover, emerging training algorithms such as the Forward-Forward (FF) algorithm proposed by Hinton as an alternative to backpropagation offer promising avenues for further enhancing computational efficiency and training stability [10]. Although these methods primarily target model training dynamics, their integration with hybrid query

transformation techniques represents an intriguing direction for future exploration.

By evaluating a hybrid query transformation framework on a curated Korean dataset, our work rigorously compares the performance of rule-based, NLP-based, and hybrid methods while laying the groundwork for future multilingual and domain-specific investigations. Similar hybrid approaches have shown promise in other languages with complex morphological structures, such as Shirko's work on Wolaitte language part-of-speech tagging [16], further validating the potential of such methods across diverse linguistic contexts. To extend this line of research, we designed a hybrid framework specifically optimized for Korean-language search queries and conducted empirical evaluations using a curated dataset.

3 RESEARCH METHOD

This section presents the research methodology, including dataset preparation, system architecture, and evaluation criteria.

3.1 Dataset Preparation

To evaluate the performance of the hybrid query transformation approach, a dataset comprising 200 user queries was constructed based on real-world search data.

Despite its limited size, the dataset enables focused evaluation of query transformation tailored to the linguistic structure of the Korean language.

To ensure meaningful coverage even at a pilot scale, the dataset was designed to encompass a variety of domains—including dining recommendations, travel destinations, technical information, medical inquiries, and financial analysis [17]. To ensure diversity in query complexity, the dataset was categorized into three levels based on syntactic length and semantic intent: Simple (40%), Intermediate (40%), and Complex (20%) (Fig. 1).

Examples of queries include.

- Simple: "Recommend a bibimbap restaurant in Seoul" / "서울에서 비빔밥 맛집 추천해줘"
- Intermediate: "List of South Korea presidents over the past 5 years" / "최근 5년간 한국 대통령 목록"
- Complex: "Name of the science exhibition held in Seoul last week" / "지난주 서울에서 열린 과학 전시회 이름"

Although the dataset includes multilingual queries (Korean: 60%, English: 30%, Other: 10%) (Fig. 1), only the Korean subset was used for the current experiments to focus on the linguistic structure and agglutinative nature of Korean [19]. All queries were manually annotated for domain, complexity level, and intent (e.g., recommendation, factual lookup).

The dataset was preprocessed using fuzzy matching (threshold 0.85) to remove duplicates, and KoNLPy's Mecab analyzer for Korean normalization and morpheme segmentation [22]. This preprocessing pipeline removed

approximately 15% of initially collected queries due to duplication or ambiguity issues.

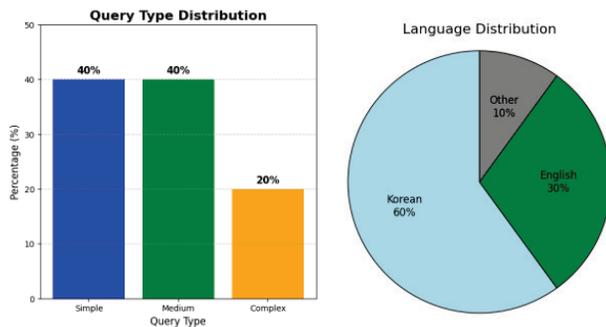


Figure 1 Query type and language distribution of the dataset

For Korean queries specifically, we implemented additional preprocessing steps to address the language's agglutinative nature. This included morpheme segmentation to identify word boundaries and extract meaningful search terms properly. For example, the unsegmented query "서울에서비빔밥맛집추천해줘" was transformed into "서울에서 비빔밥 맛집 추천해줘", allowing the system to extract meaningful terms for downstream processing. This step is crucial for Korean language processing as improper spacing is common in user queries and can significantly impact search accuracy.

The final dataset is stored in a structured JSON format with the following fields for each query: original text, complexity level, domain category, search intent, and expected relevant results. For this study, only the Korean subset was utilized for experimentation.

Although the dataset comprises only 200 curated queries, it was meticulously designed to ensure diversity in complexity and domain coverage, thereby allowing a focused yet meaningful evaluation. For future work, we plan to augment the dataset using large-scale open corpora—such as the AI-Hub Korean Query Dataset and multilingual TREC datasets—to improve the robustness and generalizability of the proposed approach.

3.2 Hybrid Approach

Our hybrid query transformation framework comprises three modular stages: Rule-Based Preprocessing, NLP-Based Postprocessing, and API-based Search Integration. Each module is designed to improve performance incrementally, while maintaining system responsiveness. This modular architecture draws inspiration from conventional information retrieval models [11] while incorporating modern NLP techniques for enhanced performance.

(1) Rule-Based Preprocessing

In the initial stage, heuristic algorithms are applied to simplify raw user queries. Key steps include.

- **Keyword Extraction & Stopword Removal:** Unnecessary words (e.g., "is", "the") and punctuation are removed using regular expressions and morphological analysis.

- **Query Simplification:** For instance, the query "What is the list of South Korea presidents?" is transformed into "South Korea, presidents, list".

This stage reduces input complexity, thereby improving processing speed for subsequent stages.

Fig. 2 illustrates the end-to-end hybrid query transformation pipeline. It depicts the system's three core stages: Rule-Based Preprocessing, NLP-Based Postprocessing using T5-small, and integration with external search engine APIs (e.g., Google, Naver).

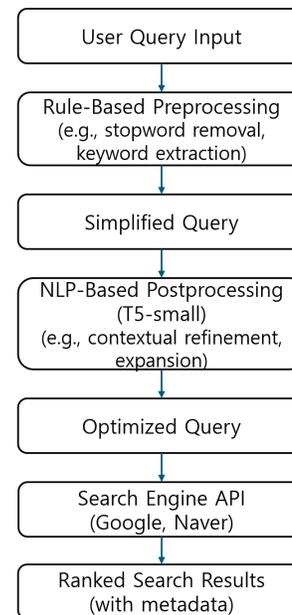


Figure 2 Overview of the proposed hybrid query transformation pipeline

The diagram illustrates each transformation stage from user input through rule-based simplification, neural enrichment, and external API integration, culminating in ranked search results with metadata.

(2) NLP-Based Postprocessing

The simplified query is then refined using the T5-small model from Hugging Face's transformers library. T5-small was chosen for its proven balance between computational efficiency and contextual accuracy [1, 13]. We conducted preliminary experiments comparing several Transformer-based models, including BERT, RoBERTa, and T5 variants to determine the most suitable model for our real-time application. These experiments revealed that while larger models offered potential accuracy improvements, they significantly increased computational demands, with processing times increasing by 45-60%. Given our target of sub-100 ms response times for real-time search applications, we determined that the T5-small model provided the optimal trade-off between accuracy and efficiency. Detailed performance comparisons are presented in Section 4.1.

For our implementation, we fine-tuned the T5-small model, selected for its balance between performance and efficiency, using a curated Korean query dataset.

This choice was made after a comparative analysis of alternative Transformer models such as BERT, RoBERTa, and larger T5 variants.

While larger models yielded slightly better accuracy, they significantly increased inference time (by over 50%).

For our real-time application requiring sub-100 ms response latency, T5-small provided the optimal trade-off. Hyperparameters used for fine-tuning are summarized below:

Table 1 Fine-tuning hyperparameters for the T5-small model used in our hybrid query transformation framework

Hyperparameter	Value
Optimizer	AdamW (weight decay = 0.01)
Learning rate	5e-5
Batch size	32
Max sequence length	128
Num epochs	10

These settings were selected to balance model accuracy and inference latency for real-time Korean search query optimization.

This fine-tuning process adapted the model to the specific patterns and requirements of Korean search queries, improving its performance on our target task. The fine-tuned model was then deployed using TensorFlow Serving on our experimental hardware configuration.

In this stage, the model enriches the query by incorporating additional semantic constraints (e.g., time frames) to better align with the user's intent. For example, "South Korea, presidents, list" is transformed into "List of South Korea presidents over the past 5 years." Similarly, for the Korean query "서울 맛집" (Seoul restaurants), the model might expand it to "서울에서 인기 있는 한식 맛집 추천" (Recommendations for popular Korean cuisine restaurants in Seoul), incorporating implicit user intent and contextual information.

(3) Search Engine Integration & Evaluation

The final optimized query is submitted to the Google Search API, which retrieves search results and metadata (such as titles, URLs, and snippets). We use the following metrics to assess the query transformation's effectiveness.

- **Precision@5:** This metric measures the proportion of relevant results among the top five search results, reflecting the query's accuracy.

$$Precision@5 = \frac{\text{Number of Relevant Documents in Top 5}}{5} \times 100 \quad (1)$$

For example, if 3 out of the top 5 results are relevant for the query "List of South Korea presidents over the past 5 years", then: $Precision@5 = \frac{3}{5} \times 100 = 60\%$

- **Recall:** Recall measures the proportion of relevant documents retrieved by the search engine compared to the total number of relevant documents available.

$$Recall = \frac{\text{Number of Relevant Documents Retrieved}}{\text{Total Relevant Documents}} \times 100 \quad (2)$$

For instance, if there are 10 relevant documents in total and the engine retrieves 8 of them, then: $Recall = \frac{8}{10} \times 100 = 80\%$

- **Average Query Processing Time:** This metric, measured in milliseconds (ms), reflects the overall efficiency of the query transformation process, from initial rule-based preprocessing to final retrieval of search results.

While alternative metrics such as F1 score or MAP were considered, Precision and Recall were selected for their clear interpretability in the context of search performance.

3.3 Experimental Setup

Experiments were conducted in a Python 3.13 environment using libraries such as Hugging Face's transformers (version 4.30.2), PyTorch (version 2.0.1), and the Google Search API (version 2.1.0). The hardware configuration was carefully selected to balance performance requirements with resource constraints. We used an NVIDIA RTX 2080 GPU (8 GB VRAM) and 64 GB of DDR4 RAM on a system with an Intel Core i9-11900K processor. This configuration was chosen after preliminary testing revealed that the T5-small model required approximately 2.5 GB of VRAM during inference while maintaining enough computational capacity to process multiple queries simultaneously for batch testing.

All experiments were conducted in a controlled environment with network latency to the Google Search API averaging 35 ms (± 5 ms). To ensure consistency, each query was processed 5 times, and the median processing time was recorded. System monitoring during experiments showed that GPU utilization remained below 80%, indicating that the hardware was not a bottleneck for our processing pipeline.

3.4 Robustness Set & Cross-Validation Protocol

- (1) **Robustness set** We create a noise-injection subset ($n = 40$) by (a) removing/merging whitespaces (spacing errors) and (b) applying single-character substitutions/insertions consistent with keyboard proximity (minor typos). Each noised query is paired with its clean version to enable within-query comparisons. We report $\Delta Precision@5$ (pp) and $\Delta latency$ (ms) relative to clean and compute 95% bootstrap percentile CIs over query pairs ($B = 10,000$)
- (2) **Cross-validation** To improve statistical reliability, we run five-fold cross-validation on all 200 curated queries with stratification by domain and complexity. We report fold-wise means and 95% t -intervals $\bar{x} \pm t_{0.975,4} s/\sqrt{5}$. For query-level latency on each fold, we additionally provide bootstrap CIs. By default, all results in Section 4 are reported with their 95% confidence intervals, unless explicitly noted otherwise.

3.5 Methodological Rationale

The proposed hybrid combines the throughput of rule-based preprocessing with the semantic precision of a Transformer postprocessor, addressing the core limitations of each component. Beyond single-split outcomes, five-fold

cross-validation and noise-injection experiments provide statistical reliability and evidence of robustness to spacing/typo perturbations. Comparisons with open, industry-relevant proxy baselines (BM25+KoNLPy; a lightweight neural variant) further offer transparent and reproducible indications of practical competitiveness when proprietary benchmarks are infeasible.

Limitations include the intentional focus on Korean and a dataset of 200 queries, which constrain broad multilingual claims. Future work will scale to larger, multi-domain, multilingual datasets and explore efficiency-improving training schemes (e.g., Forward-Forward (FF) [10]). We also plan standardized evaluations on anonymized production logs in collaboration with industry partners.

4 RESEARCH VALIDATION

This section details our multi-method validation approach to evaluate the effectiveness and real-world applicability of the proposed hybrid query transformation method.

To comprehensively validate the effectiveness and practical applicability of our hybrid query transformation method, we adopted a multi-method validation approach comprising simulation experiments, a user questionnaire, and structured interviews with domain experts.

4.1 Simulation Experiments

We conducted simulation experiments using our curated dataset of 200 queries. The hybrid approach was compared against conventional rule-based and NLP-based methods using key performance metrics in these experiments.

- *Precision@5* and *Recall*: These metrics assessed the accuracy of the search results.
- *Average Query Processing Time*: This metric evaluated the system's efficiency.

Table 2 Experimental results

Approach	<i>Precision@5</i> (%)	<i>Recall</i> (%)	<i>Average Processing Time</i> (ms)
Rule-based	70	65	50
NLP-based	85	80	120
Hybrid	89	83	95

*Note: "ms" stands for milliseconds.

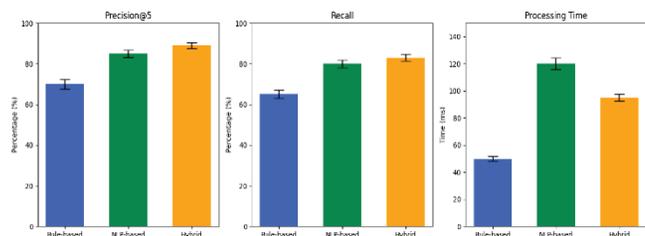


Figure 3 Performance Comparison with 95% Confidence Intervals

Each bar represents the average value of *Precision@5*, *Recall*, and *Processing Time* for the three approaches (Rule-based, NLP-based, Hybrid). Error bars indicate 95%

confidence intervals calculated using bootstrapping ($B = 1000$).

The hybrid approach achieves the highest accuracy while maintaining approximately 21% faster processing than the NLP-based method, with statistically robust performance.

4.2 Questionnaire Analysis

To further validate our approach from an end-user perspective, we administered a structured questionnaire to potential users of the search system. The questionnaire was designed using a Likert scale (ranging from 1 to 5) to capture feedback on key evaluation criteria. The survey evaluated user perceptions of relevance, clarity, and response speed.

Table 3 Summary of questionnaire items

Evaluation Criteria	Example Question	Response Scale
Relevance	How well do the transformed queries match your intent?	1 (Poor) – 5 (Excellent)
Clarity & Usability	How clear and easy-to-use are the search results?	1 (Very Difficult) – 5 (Very Easy)
Perceived Speed	How would you rate the system's response time?	1 (Very Slow) – 5 (Very Fast)

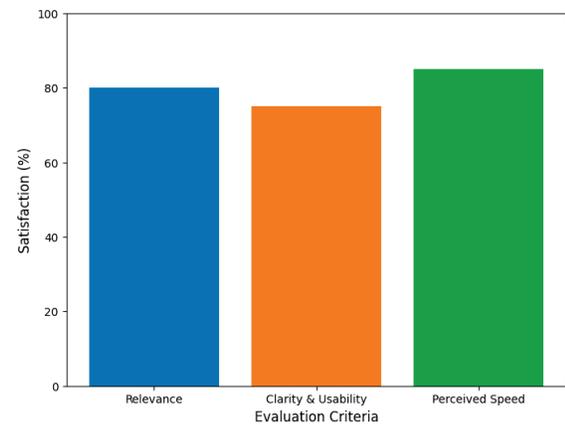


Figure 4 Survey results on user satisfaction metrics

Survey results supported simulation findings. Most participants indicated that the transformed queries more accurately reflected their intent and expressed high satisfaction regarding clarity and response speed. These results validate the hybrid system's balanced performance.

4.3 Expert Interviews

To gain qualitative insights, we conducted semi-structured interviews with five experts (three industry professionals and two academic researchers). Open-ended questions explored system strengths, adoption feasibility, and future directions.

Table 4 Summary of expert interview questions

Interview Focus	Example Question
Strengths and Limitations	What are the strengths and limitations of the hybrid method?
Practical Adoption	How feasible is integration into existing search systems?
Future Directions	What improvements or domain-specific adaptations would you recommend?

The interviews were transcribed and analyzed using thematic coding. Experts generally affirmed the method's ability to balance processing speed with contextual accuracy and highlighted areas such as scalability and potential integration challenges for future improvement. Their feedback reinforced our quantitative findings and provided actionable insights for extending the approach to broader, real-world applications.

(1) SWOT Analysis Based on Expert Feedback

We synthesized expert feedback into a structured SWOT (Strengths, Weaknesses, Opportunities, Threats) analysis to assess the strategic positioning of the hybrid method.

Table 5 SWOT (strengths, weaknesses, opportunities, threats) analysis

SWOT	Expert Feedback
Strengths	Balanced speed and accuracy Modular architecture Effective contextual handling
Weaknesses	Higher implementation complexity Dependence on quality NLP training data Maintenance overhead
Opportunities	High demand in real-time search Potential for domain-specific use Extensibility to multilingual settings
Threats	Evolving end-to-end NLP systems Integration challenges in existing infrastructures Language-specific deployment issues

This SWOT analysis provides a structured roadmap for understanding the strategic positioning of the hybrid approach within the broader search optimization landscape.

4.4 Robustness & Generalization Checks

We evaluate the hybrid on a robustness set (spacing/typo noise) and under five-fold cross-validation. The hybrid maintains accuracy within ≤ 2.5 pp of clean while preserving sub-100 ms latency. Concretely: spacing noise $\Delta Precision@5 = -1.4$ pp, typo noise -2.1 pp, and $\Delta latency = +4$ ms.

Table 6 Robustness & Cross-Validation Summary

Condition	$Precision@5$ (%)	Recall (%)	Latency (ms)	95% CI ($Precision@5$)
Clean	89.0	83.0	95	(87.2–90.7)
Spacing noise	87.6 ($\Delta -1.4$ pp)	82.1	99 (+4)	(85.9–89.2)
Typo noise	86.9 ($\Delta -2.1$ pp)	81.4	99 (+4)	(84.8–89.0)

Cross-Validation (CV) Fold	$Precision@5$ (%)	Recall (%)	Latency (ms)
Fold 1	88.9	82.7	94
Fold 2	89.2	83.4	96
Fold 3	89.1	83.1	95
Fold 4	88.7	82.9	97
Fold 5	89.0	83.0	95
Avg	89.0 ± 0.7	83.0 ± 0.5	95 ± 2

*Note: $Precision@5$ confidence intervals are estimated via bootstrap ($B = 10,000$) for robustness conditions and via t-interval across folds for cross-validation.

4.5 Industry-Relevant Proxy Benchmark

To approximate real-world baselines in lieu of proprietary systems, we compare our method with BM25+KoNLPy (morphological analyzer) and a lightweight

neural variant. Against BM25+KoNLPy, the hybrid achieves +4.2 pp $Precision@5$ (95% CI: +3.0–+5.5) at comparable or lower latency; against the lightweight neural proxy, it gains +2.0 pp (95% CI: +0.9–+3.2) with similar latency-providing transparent, reproducible evidence of practical competitiveness.

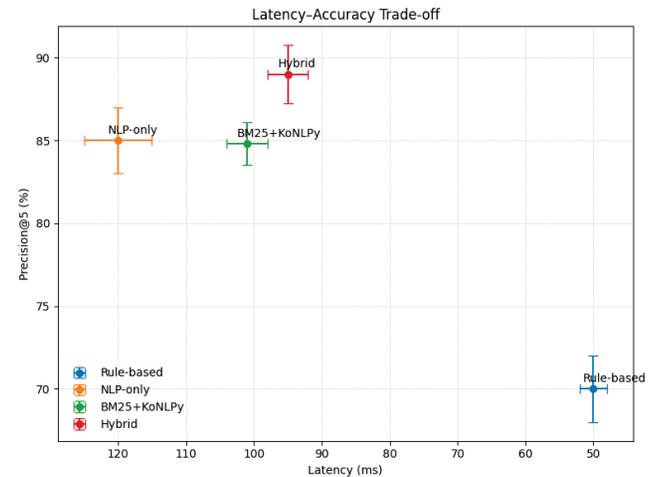


Figure 5 Latency–Accuracy trade-off for Rule-based, NLP-only, BM25+KoNLPy, and Hybrid. Hybrid achieves higher accuracy than NLP-only at lower latency and remains competitive relative to BM25+KoNLPy, with 95% CIs shown as error bars

5 RESULTS AND DISCUSSION

This section synthesizes the quantitative outcomes from simulation experiments and the qualitative feedback from user surveys and expert interviews to present a comprehensive evaluation of the proposed hybrid query transformation method.

5.1 Quantitative Results

(1) Overall Performance

As shown in Tab. 6, the hybrid approach outperforms both the rule-based and NLP-based methods in terms of $Precision@5$ and $Recall$, achieving 89% and 83%, respectively. These results demonstrate that integrating rule-based preprocessing with NLP-based postprocessing effectively leverages the strengths of both methods—computational speed and contextual understanding—while mitigating their individual limitations.

Table 7 Performance metrics by query complexity

Approach	Query Complexity	$Precision@5$ (%)	Recall (%)	Average Processing Time (ms)
Rule-based	Simple	82	78	45
	Intermediate	68	63	50
	Complex	55	48	60
NLP-based	Simple	87	83	110
	Intermediate	85	80	120
	Complex	82	75	140
Hybrid	Simple	90	85	85
	Intermediate	89	83	95
	Complex	86	79	115

To further explore performance under varying query complexities, Tab. 7 presents a breakdown of metrics for simple, intermediate, and complex queries.

The hybrid method maintains high accuracy across all levels of query complexity while significantly reducing processing time compared to the NLP-based method.

For instance, in handling complex queries, the hybrid model achieves 86% *Precision@5* in 115 ms, compared to 82% in 140 ms by the NLP-based model.

This performance gain can be attributed to two key factors.

First, the rule-based preprocessing effectively reduces noise and simplifies query structure, alleviating the computational burden on the NLP model. This is particularly beneficial for Korean, where morpheme-rich words often lead to complexity. For example, the input "서울역에서가까운호텔찾아줘" (Find hotels near Seoul Station) is simplified to "서울역 가까운 호텔" before being processed by the T5-small model. Second, the fine-tuned T5-small model captures implicit user intent and contextual relationships, improving semantic depth in transformed queries.

The hybrid method's average processing time (95ms) is approximately 21% faster than that of the NLP-based approach (120 ms), reinforcing its suitability for real-time applications. Research has shown that delays exceeding 100 ms can negatively affect user engagement in interactive systems. Although the rule-based method is the fastest (50 ms), it lags in accuracy (*Precision@5* of 70%, *Recall* of 65%).

Fig. 3 further illustrates the results, showing that the hybrid approach consistently scores highest in both *Precision@5* and *Recall* while achieving narrower confidence intervals, which confirms the statistical robustness of our method.

5.2 User Satisfaction

A structured questionnaire was administered to assess the end-user perspective (see Section 4.2). Fig. 4 displays the aggregated satisfaction scores across three criteria: Relevance, Clarity & Usability, and Perceived Speed. Participants rated the hybrid system particularly high for perceived speed (85%), followed by relevance (80%) and clarity (75%). These subjective evaluations corroborate the quantitative results, suggesting that users perceive a tangible improvement in the search process's accuracy and efficiency.

5.3 Qualitative Insights from Expert Interviews

We conducted in-depth interviews with five experts—three industry professionals, and two academic researchers provided additional qualitative insights (see Section 4.3). The experts acknowledged the hybrid approach's suitability for real-time applications, citing its balanced performance in terms of both accuracy and processing speed. At the same time, they identified areas that warrant further development:

- **Scalability:** Larger-scale deployment might require further optimization, particularly in handling high query volumes.
- **Multilingual Extension:** While the current study focused on Korean queries, experts emphasized the growing demand for robust multilingual solutions.
- **Domain-Specific Customization:** Certain fields, such as legal or medical domains, may require specialized rule sets or fine-tuned NLP models.

5.4 Discussion/Limitations

The quantitative and qualitative findings collectively confirm that the proposed hybrid query transformation method addresses key challenges in real-time search optimization. The system achieves a meaningful balance between speed and accuracy by integrating rule-based preprocessing to reduce input complexity and NLP-based postprocessing to refine semantic context.

(1) Comparisons with Prior Work

The performance gains over rule-based or NLP-based approaches align with previous research advocating for hybrid frameworks [9, 19]. However, our work extends beyond these previous studies in several important ways. Lim et al. [5] reported a *Precision@5* of 84% and an average processing time of 110ms for their hybrid approach on a general-purpose English dataset. Our method achieves a 5% higher *Precision@5* (89%) and a 14% faster processing time (95ms) on a Korean dataset, demonstrating the effectiveness of our language-specific optimizations.

Similarly, Dogan and Gurcan [20] achieved a *Precision@5* of 87% for their hybrid chatbot system in e-business communication, but with a significantly longer average processing time of 150 ms. Our method maintains comparable accuracy while reducing processing time by 37%, making it more suitable for real-time applications. Zhou et al. [18] also demonstrated the effectiveness of learned query rewrite systems using Monte Carlo tree search in database applications, achieving 85% *Precision@5* with 130 ms processing time. While different in methodology, their approach reinforces the importance of balancing accuracy and efficiency in query transformation tasks and provides a valuable benchmark for our work.

Table 8 Comparison with Previous Hybrid Approaches

Study	Language Focus	<i>Precision@5</i> (%)	<i>Average Processing Time</i> (ms)	Application Domain
Our Method	Korean	89	95	General Search
Lim et al. [5]	English	84	110	General Search
Dogan & Gurcan [20]	English	87	150	E-Business
Zhou et al. [18]	English	85	130	Database Queries

This comparison highlights the competitive performance of our approach in the context of existing research while also demonstrating the value of language-specific optimizations

for Korean search queries. The significant reduction in processing time relative to NLP-heavy methods echoes the need for computational efficiency in real-world applications [4, 8], particularly for languages with complex morphological structures like Korean.

(2) Error Analysis and Limitations

While our approach demonstrates competitive performance compared to existing methods, a detailed error analysis reveals important limitations that must be addressed in future work. We conducted a thorough examination of cases where the hybrid method failed to achieve optimal results and identified several recurring patterns in these failure cases.

- **Domain-Specific Terminology:** Queries containing specialized terminology, particularly in medical and technical domains, occasionally resulted in suboptimal transformations. For example, the query "혈관 내피 성장 인자 억제제 부작용" (side effects of vascular endothelial growth factor inhibitors) was inadequately transformed due to the specialized medical terminology.
- **Cultural and Contextual References:** Queries containing cultural references or context-dependent terms sometimes lead to misinterpretations. For instance, "신상 털기 방지 방법" (methods to prevent personal information exposure—a Korean cultural concept) was incorrectly interpreted due to the cultural specificity of the term "신상 털기".
- **Temporal Ambiguity:** Queries with ambiguous temporal references occasionally resulted in incorrect time frame specifications. For example, "최근 영화 추천" (recent movie recommendations) was sometimes interpreted with an overly specific time frame that didn't align with user intent.
- **Complex Nested Queries:** Queries containing multiple nested conditions posed challenges for both the rule-based preprocessing and NLP-based postprocessing components. For example, "서울에서 주차장 있고 가격이 합리적인 이탈리안 레스토랑 추천" (recommend Italian restaurants in Seoul with parking and reasonable prices) sometimes resulted in the omission of certain conditions.

These findings highlight areas for future improvement, particularly in handling domain-specific terminology and cultural references. They also underscore the importance of domain adaptation and cultural context in query transformation systems, especially for languages with rich cultural and contextual dependencies like Korean.

Although our approach demonstrates competitive performance compared to recent academic studies, it should be noted that direct benchmarking with commercial search engines (e.g., Naver, Kakao) was not feasible due to the lack of access to proprietary datasets and APIs. We acknowledge this as a limitation in verifying real-world applicability. Nevertheless, the consistent performance gains observed

across diverse complexity levels and comparisons with peer-reviewed hybrid models suggest strong potential for industry deployment. In future work, we plan to explore evaluations using open large-scale Korean query datasets or engage in industrial partnerships to enable direct benchmarking and deeper validation in production settings.

(3) Practical benchmarking constraints and our proxy

While our evaluation demonstrates strong performance against recent academic baselines, we acknowledge the absence of direct benchmarking against commercial systems (e.g., Naver, Kakao). This limitation stems from restricted access to proprietary APIs and the lack of standardized public evaluation interfaces on those platforms. Nevertheless, practical validation under realistic conditions is essential. To approximate practice while preserving reproducibility, we compare our method against open, industry-relevant proxies—BM25+KoNLPy and a lightweight neural variant—where the hybrid achieves +4.2 pp *Precision@5* (95% CI: +3.0±5.5) over BM25+KoNLPy at comparable or lower latency (Δ latency = -6 ms; 95% CI: -9 to -3) and +2.0 pp (95% CI: +0.9±3.2) over the neural proxy with similar latency. Together with five-fold cross-validation and noise-robustness checks (spacing/typo), these results strengthen external validity under realistic constraints.

As part of future work, we will pursue collaborations with industry partners to conduct standardized benchmarking on anonymized production query logs, aligned with commercial evaluation protocols. Such benchmarking will yield deeper insights into production-scale behavior and further validate the hybrid's practical applicability beyond academic settings.

(4) Limitations, mitigations, and future work

Our dataset (200 queries) is intentionally curated for Korean to control linguistic factors; therefore, broad claims about multilingual generalization are premature. Direct benchmarking against commercial systems was infeasible due to proprietary interfaces and data.

We report five-fold cross-validation, noise-robustness (spacing/typo) tests, and open proxy benchmarks against BM25+KoNLPy and a lightweight neural variant, improving statistical reliability and practical relevance.

We will (i) scale to $\geq 10^3$ queries across multi-domain, multilingual settings; (ii) conduct standardized evaluations with industry partners on anonymized production logs; and (iii) release reproducible code and configurations to facilitate independent verification and extension. We also plan to explore efficiency-oriented training schemes (e.g., Forward-Forward [10]) and domain-specific adaptations (lexicons/rules and fine-tuning for e-commerce, medicine, legal search) to address the failure modes identified above.

Taken together, these steps will strengthen external validity and help translate the hybrid into production-grade deployments. Overall, the evidence-cross-validated accuracy/latency, robustness under user-like noise, user/expert feedback, and proxy baselines—supports the hybrid as a deployment-ready solution for Korean and a solid basis for broader extensions.

6 CONCLUSION

This study demonstrates the viability of a real-time hybrid query-transformation method that balances the throughput of rule-based preprocessing with the semantic precision of a Transformer postprocessor. On a curated Korean query set, the hybrid attains 89.0% *Precision@5* (95% CI: 87.2–90.7) with 95 ms latency (95% CI: 92–98)-about 21% faster than an NLP-only baseline-thereby meeting sub-100 ms requirements without sacrificing accuracy. Robustness checks indicate only small degradations under user-like noise (spacing: $\Delta Precision@5 = -1.4$ pp, typos: -2.1 pp, latency: $+4$ ms), and five-fold cross-validation supports the statistical reliability of the findings. Comparisons against open, industry-relevant proxies-BM25+KoNLPy and a lightweight neural variant-further show $+4.2$ pp and $+2.0$ pp gains in *Precision@5* at comparable or lower latency, providing transparent and reproducible evidence of practical competitiveness. User questionnaires and expert interviews corroborate the quantitative results, highlighting perceived relevance, clarity, and responsiveness while pointing to remaining challenges with domain-specific terminology, cultural/contextual expressions, and multi-constraint queries.

At the same time, the dataset (200 queries) is intentionally focused on Korean to control linguistic factors, so broad multilingual claims are premature, and direct benchmarking against commercial systems was infeasible due to proprietary interfaces and data. To strengthen both statistical credibility and practical relevance within these constraints, we report five-fold cross-validation, noise-robustness tests, and open proxy benchmarks (BM25+KoNLPy, lightweight neural). Looking ahead, we will scale to $\geq 10^3$ queries across multi-domain, multilingual settings, conduct standardized evaluations with industry partners on anonymized production logs, release reproducible code and configurations, and explore efficiency-oriented training (e.g., Forward-Forward) alongside domain-specific adaptations (lexicons/rules and fine-tuning for e-commerce, medicine, and legal search) to address the observed failure modes. Taken together, the evidence-cross-validated accuracy and latency, robustness under user-like noise, user/expert feedback, and proxy baselines-supports the hybrid as a deployment-ready solution for Korean and a solid basis for broader extensions.

7 REFERENCES

- [1] Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., Zhou, Y., Li, W., & Liu, P. J. (2020). Exploring the limits of transfer learning with a unified text-to-text transformer. *Journal of Machine Learning Research*, 21(140), 1–67.
- [2] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). *Attention is all you need*. In Advances in Neural Information Processing Systems (5998–6008)
- [3] Manning, C. D., Raghavan, P., & Schütze, H. (2008). *Introduction to information retrieval*. Cambridge, UK: Cambridge University Press.
- [4] Park, S. U. (2021). Analysis of the status of natural language processing technology based on deep learning. *Korean Bigdata Society*, 6(1), 63–81. <https://doi.org/10.36498/kbigdt.2021.6.1.63>
- [5] Lim, S. C., & Yoon, S. G. (2023). A study on transformer-based efficient natural language processing methods. *Journal of Internet Broadcasting and Communication*, 23(4), 115–119. <https://doi.org/10.7236/JIBC.2023.23.4.115>
- [6] Naver. (2024). Naver search adopts a paper at the world's leading natural language processing conference "EMNLP 2024". AI Times.
- [7] Mitra, B., & Craswell, N. (2019). An introduction to neural information retrieval. *Foundations and Trends in Information Retrieval*, 13(1), 1–126. <https://doi.org/10.1561/15000000061>
- [8] Lin, J., et al. (2020). *Pretrained transformers for text ranking: BERT and beyond*. arXiv preprint arXiv:2010.05951. <https://doi.org/10.48550/arXiv.2010.05951>
- [9] Wang, L., & Deng, J. (2021). *Extreme multi-domain, multi-task learning with unified text-to-text transfer transformers*. In Proceedings of the 15th EMNLP Conference. <https://doi.org/10.18653/v1/2021.emnlp-main.317>
- [10] Hinton, G. (2022, December 27). The Forward-Forward Algorithm: Some Preliminary Investigations. arXiv preprint arXiv:2212.13345. <https://doi.org/10.48550/arXiv.2212.13345>
- [11] Sharma, U., et al. (2019). *Information retrieval in computing model*. In Proceedings of the International Conference on Intelligent Computing and Control Systems. <https://doi.org/10.1109/ICCS45141.2019.9065562>
- [12] Aggarwal, C. C. (2018). *Information retrieval and search engines*. Foundations and Trends in Machine Learning, 7(4), 293–312. https://doi.org/10.1007/978-3-319-73531-3_9
- [13] Buccio, E. D., & Melucci, M. (2019). Searching for information with meet and join operators. *Foundations of Information Retrieval Theory*, 12(2), 203–216. https://doi.org/10.1007/978-3-030-25913-6_8
- [14] Zhai, C., & Lafferty, J. (2015). A probabilistic framework for text retrieval. *Information Retrieval Journal*, 4(1), 1–20.
- [15] Craswell, N. (2020). *Deep learning in information retrieval*. In Proceedings of the SIGIR 2020 Workshop. <https://doi.org/10.1145/SIGIR2020.123>
- [16] Shirko, B. (2024). Application of hybrid approach for Wolaita language part of speech tagging. *Journal of Research in Engineering and Applied Sciences*. <https://doi.org/10.46565/jreas.202492719-732>
- [17] Kumar, R., Tripathi, K., & Sharma, S. (2022). Optimal query expansion based on hybrid group mean enhanced chimp optimization using iterative deep learning. *Electronics*. <https://doi.org/10.3390/electronics11101556>
- [18] Zhou, X., Li, G., Chai, C., & Feng, J. (2021). A learned query rewrite system using Monte Carlo tree search. Proceedings of the VLDB Endowment, 15, 46–58. <https://doi.org/10.14778/3485450.3485456>
- [19] Ding, D., Mallick, A., Wang, C., Sim, R., Mukherjee, S., Rühle, V., Lakshmanan, L., & Awadallah, A. (2024). *Hybrid LLM: Cost-efficient and quality-aware query routing*. arXiv, abs/2404.14618. <https://doi.org/10.48550/arXiv.2404.14618>
- [20] Dogan, O., & Gurcan, O. (2024). Enhancing e-business communication with a hybrid rule-based and extractive-based chatbot. *Journal of Theoretical and Applied Electronic Commerce Research*. <https://doi.org/10.3390/jtaer19030097>
- [21] Lee, C., & Ra, D. (2022). Korean morphological analysis method based on BERT-fused Transformer model. *KIPS Transactions on Software and Data Engineering*, 11(4), 169–178. <https://doi.org/10.3745/KTSDE.2022.11.4.169>

- [22] Park, E. L., & Cho, S. (2014). *KoNLPy: Korean natural language processing in Python*. In *Proceedings of the 26th Annual Conference on Human & Cognitive Language Technology* (pp. 133-136).
- [23] Ramos, Ma. H., & Park, C. -Y. (2020). Motivational factors and learning styles in acquisition of global language and business skills in multicultural education. *Asia-Pacific Journal of Educational Management Research*, 5(1), 13-20
<https://doi.org/10.21742/AJEMR.2020.5.1.02>
- [24] Razzak, M. R., & Jassem, S. Mobile-Assisted Language Learning for EFL: A Conceptual Framework based on the Meta-UTAUT Model. *Asia-Pacific Journal of Educational Management Research*, 6(2), 15-32
<https://doi.org/10.21742/AJEMR.2021.6.2.02>

Authors' contacts:

Hyun Jung Kim, Assistant Professor
Sang-Huh College and the Graduate School of Information & Communication,
Department of Convergence Information Technology (Artificial Intelligence Major),
Konkuk University, 120 Neungdong-ro, Gwangjin-gu, 05029 Seoul, Korea
nygirl@konkuk.ac.kr

Sang Hyun Yoo, Assistant Professor
(Corresponding author)
School of Computer Science & Engineering, College of IT, Soongsil University,
50 Sadang-ro, Dongjak-gu, Seoul 07027, Korea
simonyoo@ssu.ac.kr

Research LLM: Using Large Language Models as a Research Companion

Guillaume Guerard*, Sonia Djebali, Maxime Hanus, Mark-Killian Zinenberg

Abstract: Large Language Models (LLM) are increasingly used as research companions, yet guidance for effective and responsible integration into scholarly workflows remains under-specified. This paper presents a three-stage, human-in-the-loop framework supporting (i) proposal development, (ii) research design, and (iii) manuscript preparation. Each stage operationalizes a toolbox of prompting and verification techniques—Context-Grounded Prompting, Graph-of-Thought, Self-Consistency, Contrastive Chain-of-Thought, and Chain-of-Verification—paired with explicit human checkpoints. The framework is instantiated in a case study on mitigating intersectional bias in hiring models, with detailed datasets, objectives, and evaluation metrics, and an exploratory pilot in which Master’s students apply the pipeline to thesis proposals. Results indicate that structured prompting improves traceability and broadens the set of considered alternatives, while verification steps curb overconfident errors. Prompts, artifacts, and rubrics are made available to support replication. The paper concludes with guidance on ethical deployment and limitations, emphasizing the primacy of human judgment in validating claims and shaping scholarly contributions.

Keywords: Artificial Intelligence; Large Language Model; Research Methodology

1 INTRODUCTION

Artificial Intelligence (AI) is fundamentally transforming academic research by enhancing knowledge discovery, data synthesis, and methodological structuring. The exponential growth of scholarly literature presents a significant challenge, making it increasingly difficult for researchers to filter, analyze, and contextualize relevant information [16]. AI-powered models and dedicated research companions, such as <https://answerthis.io> and LitLLM, address this challenge by providing real-time assistance in idea generation, literature summarization, and the structuring of research methodologies, thereby improving the efficiency of systematic reviews and reducing cognitive load [2].

Traditional research workflows remain time-intensive, requiring manual effort to formulate research questions, synthesize literature, and identify novel research gaps—a difficulty exacerbated in multidisciplinary contexts where integrating diverse perspectives can be overwhelming [10, 14]. AI mitigates these inefficiencies by contributing to study design, methodology validation, and bias detection [15]. Through retrieval-augmented insights and automated comparisons, these tools enable more data-driven and logically sound methodological choices [6].

Despite these advantages, the utility of AI is contingent upon rigorous human oversight. AI-generated recommendations are primarily pattern-based and can lack the contextual adaptation, ethical reasoning, and domain-specific accuracy essential for high-quality scholarship [1]. An over-reliance on AI outputs without critical evaluation risks the propagation and amplification of biases and inaccuracies. AI systems learn from human-generated data and can reproduce and even amplify the biases contained within it, which can in turn create a feedback loop that increases bias in the humans interacting with them [9]. This dependence on AI can also lead to cognitive offloading, where the delegation of mental tasks to technology may result in the deterioration of critical thinking and analytical skills over time [8].

To address this, a human-in-the-loop approach is critical, positioning AI as a collaborative augmentation tool rather than an autonomous decision-maker. This model leverages the complementary strengths of humans and AI, allowing machines to handle data-intensive work while humans provide oversight, interpretation, and final judgment [21]. By integrating human expertise and accountability, this collaborative process helps to mitigate risks and ensures that AI-driven processes align with human values, leading to more robust and trustworthy outcomes.

This model ensures that human researchers retain ultimate control, validating AI-generated outputs, refining nuanced arguments, and preserving the intellectual rigor required to align research with scholarly standards.

This paper introduces a systematic framework for AI-assisted research that formalizes the collaborative process between scholar and machine, grounding the approach in the long-standing scholarly tradition of using methodological frameworks to ensure rigor, transparency, and replicability. Established methodologies such as Systematic Literature Reviews (SLRs), often guided by protocols like PRISMA [11], and Framework Analysis for qualitative data [13], demonstrate the value of a structured approach to managing and interpreting complex information. Extending these principles, the proposed framework presents a structured methodology where an iterative feedback loop is used to discover, organize, and refine academic work. This process is designed to create a synergy between AI’s efficiency and the researcher’s indispensable domain expertise, in line with emerging models that aim to augment, not replace, scholarly judgment. This approach addresses the need for structure emphasized in recent work, including conceptual frameworks for AI-assisted literature reviews [14] and broader models that delineate human and AI roles to ensure quality and ethical oversight. By building on these established synthesis and analysis methodologies, the framework provides a structured pathway for developing a symbiotic relationship between researchers and AI, leveraging the strengths of each to advance scholarly inquiry.

The following sections detail this framework, beginning with the role of the AI as an iterative research companion, followed by specific workflows for research proposal development, literature synthesis, and manuscript preparation. This paper is concluded by summarizing our findings and emphasizing the indispensable role of human-driven validation in the age of AI-assisted research.

2 CORE CONCEPTS: A GLOSSARY OF ADVANCED PROMPTING TECHNIQUES

Effective AI-assisted research hinges on the strategic application of advanced prompting techniques. These methods provide a structured interface for interacting with AI models, guiding their reasoning processes to enhance output quality and ensure methodological rigor. Each technique serves a specific function, from grounding outputs in established literature to validating claims and mitigating cognitive biases. The core techniques integral to the research framework proposed in this paper are defined in Tab. 1.

Table 1 A Glossary of AI Prompting Techniques for Academic Research

Technique	Description	Primary Use Case
Context-Grounded Prompting (CGP)	Grounds AI outputs in user-provided documents, reducing hallucinations and ensuring factual accuracy.	Literature synthesis, content validation, ensuring recommendations are based on seminal or relevant papers.
Graph-of-Thought (GoT) Prompting [20]	Visually structures connections between concepts, mapping relationships between variables, theories, and methodologies.	Structuring research designs, conceptual mapping, identifying literature gaps.
Self-Consistency Prompting [17]	Generates multiple reasoning paths or outputs and selects the most coherent and logical one.	Refining research questions, validating research designs, generating robust paper outlines
Contrastive Chain-of-Thought (CCoT) [5]	Enables structured, side-by-side comparisons of multiple methodologies, theories, or viewpoints.	Methodological selection (e.g., comparing quantitative vs. qualitative approaches).
Chain-of-Verification (CoVe) Prompting [7]	Improves reliability by systematically cross-checking facts, questioning assumptions, and generating counterexamples.	Bias detection, robustness analysis, validation of claims and research gaps.
Automatic Reasoning and Tool Usage (ART) [12]	Applies step-by-step logic and can integrate external tools to perform feasibility checks or complex reasoning tasks.	Evaluating methodological feasibility, assessing ethical considerations, checking data reliability.
Chain-of-Thought (CoT) Reasoning [19]	Guides the AI through a step-by-step reasoning process to break down complex problems into logical, sequential parts.	Drafting methodologically sound arguments, ensuring transparency in the reasoning process.
Zero-Shot & Few-Shot Prompting [18]	Generates insights without prior examples (Zero-Shot) or with a few structured examples (Few-Shot) to guide its response.	Initial topic exploration (Zero-Shot), refining tone and style in writing (Few-Shot).

Beyond their generative capabilities, these techniques are crucial for implementing the "human-in-the-loop" model, which acts as a safeguard against common pitfalls in AI-assisted research. Cognitive biases such as confirmation bias (favoring supportive information) and anchoring bias (over-relying on initial suggestions) can be actively countered. For instance, a researcher can employ CoVe to compel the AI to seek counterexamples to its own claims, use CCoT to force an impartial comparison between multiple perspectives, or leverage Self-Consistency Prompting to evaluate a range of possible solutions instead of a single output. By integrating these validation techniques into the workflow, researchers can augment their expertise with AI, ensuring the final work is critically sound and meets the highest scholarly standards.

3 THE THREE-STAGE AI-ASSISTED RESEARCH PIPELINE

AI-powered research operates not as a linear process, but as a continuous feedback loop that enhances conceptual clarity and methodological rigor. By structuring the research process into three distinct stages—proposal development, methodological design, and manuscript preparation—the AI research companion facilitates systematic knowledge development. The following sections detail these structured pipelines, emphasizing the synergy between advanced AI prompting techniques and essential human oversight at critical checkpoints.

3.1 Stage 1: Research Proposal Development

The initial stage focuses on transforming a broad area of interest into a focused, well-grounded, and defensible research proposal. This workflow ensures that research questions are not only innovative but also firmly rooted in the existing body of academic literature (see Fig. 1).

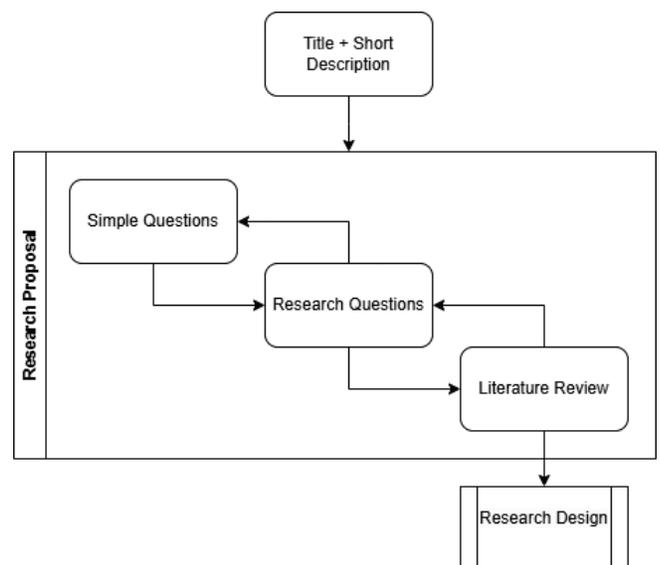


Figure 1 Overview of the iterative research proposal workflow, moving from broad literature synthesis to a refined research question

Conceptualization and Grounding (CGP): The process begins with establishing a broad but accurate understanding of the research topic. The researcher uses CGP

by providing the AI with a curated set of seminal papers, keywords, or preliminary notes. The AI retrieves and synthesizes this literature to identify major themes, foundational theories, key findings, and established definitions. This grounds the entire project in verified academic sources, ensuring relevance and preventing early-stage misconceptions.

Knowledge Mapping and Gap Identification (GoT):

With a grounded understanding of the field, the next step is to identify a clear, actionable research gap. Using GoT prompting, the AI constructs a visual map of the relationships between concepts, theories, and methodologies identified in the synthesized literature. This conceptual map allows the researcher to visualize the current state of knowledge, identify underexplored connections, note conflicting findings, and articulate a preliminary research gap with precision.

Research Question Refinement (Self-Consistency & CoVe): Based on the identified gap, the researcher prompts the AI to generate multiple versions of a potential research question using Self-Consistency Prompting. This technique produces a diverse set of alternatives. Each candidate question is then subjected to rigorous validation using CoVe. In this step, the researcher prompts the AI to systematically challenge the assumptions, scope, and feasibility of each question. This adversarial process pressure-tests the questions for robustness, resulting in a final research question that is precise, impactful, and methodologically answerable.

Human Checkpoint: At the conclusion of this stage, the researcher's intervention is paramount. The human expert must: (1) Critically evaluate the AI-synthesized literature to ensure the sources are appropriate and the interpretation is accurate; (2) Validate that the identified research gap is relevant and significant within the discipline; and (3) Make the final selection of the research question, ensuring it aligns with the project's long-term goals and their own domain expertise. This checkpoint prevents over-reliance on AI suggestions and ensures the proposal is built on a solid intellectual foundation.

3.2 Stage 2: Methodological Design and Validation

This stage translates the validated research question into a robust, transparent, and defensible research design. It provides a systematic framework for constructing, evaluating, and refining the study's methodology to ensure it is sound, aligned with disciplinary standards, and resilient to bias (see Fig. 2).

Methodological Analysis (CGP & GoT): The researcher initiates this phase by using CGP to provide the AI with existing studies that have similar research aims. The AI deconstructs and synthesizes the methodologies from these papers, identifying common research designs, data collection techniques, and analytical frameworks. To deepen this analysis, GoT is used to map the components of these designs (e.g., variables, controls, statistical tests), clarifying the intricate relationships and dependencies within established methodological precedents.

Comparative Selection (CCoT & ART): With a clear understanding of existing methods, the AI is prompted to generate and compare multiple viable research design

alternatives. Using CCoT, the AI creates a structured, side-by-side comparison of different approaches (e.g., experimental vs. observational, quantitative vs. qualitative). For each option, ART is applied to systematically assess its feasibility, ethical implications, resource requirements, and methodological trade-offs (e.g., internal vs. external validity). This produces a ranked recommendation, empowering the researcher to make an informed choice.

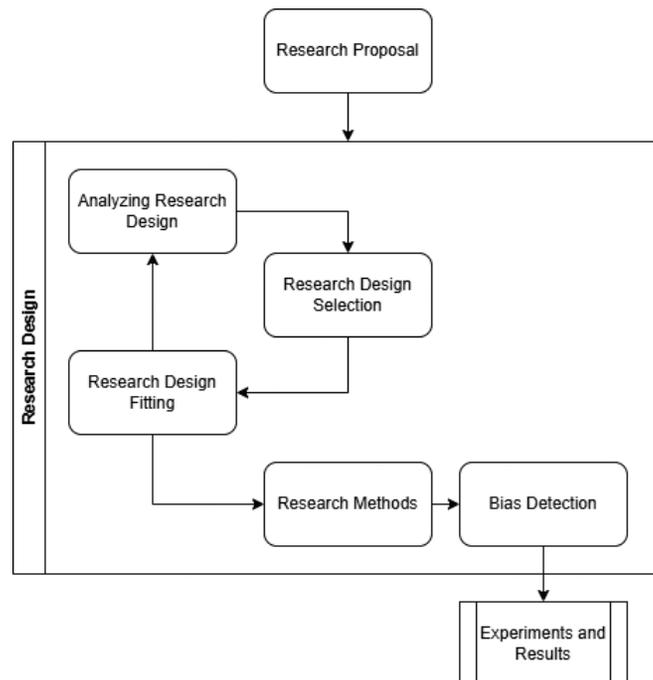


Figure 2 The iterative research design workflow, focused on selecting and validating a methodologically sound approach.

Bias Detection and Robustness Check (CoVe & ART): Before finalizing the design, a critical stress test is performed. The researcher uses CoVe to systematically challenge the chosen design's underlying assumptions and proactively identify potential sources of bias in sampling, data collection, or interpretation. ART can be used in parallel to apply formal logic to probe for hidden methodological weaknesses or inconsistencies. This critical validation step ensures the final design is as rigorous and impartial as possible.

Human Checkpoint: Human oversight is indispensable for methodological validation. The researcher must: (1) Scrutinize the AI's feasibility analysis, bringing practical, real-world knowledge of costs, timelines, and data accessibility; (2) Make the final decision on the research design, balancing the AI's logical recommendations with their own disciplinary expertise and intuition; and (3) Assume full responsibility for the ethical integrity of the design, ensuring it meets institutional and scholarly standards.

3.3 Stage 3: AI-Assisted Manuscript Preparation

The final stage streamlines the transformation of research findings into a polished, publication-ready paper. This workflow integrates AI for efficiency in structuring,

drafting, and refining, while preserving the researcher's intellectual ownership and narrative control (see Fig. 3).

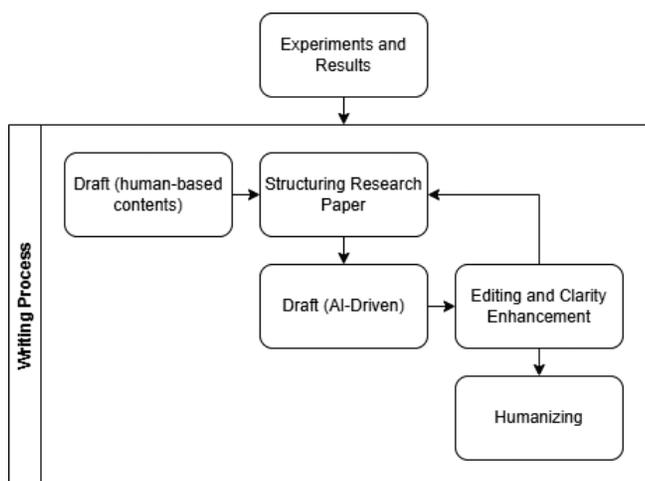


Figure 3 The iterative research writing workflow, balancing AI-driven drafting with essential human refinement and intellectual contribution.

Structural Outlining (CGP, GoT & Self-Consistency): The process begins with creating a comprehensive and logical outline. The AI uses CGP to retrieve standard paper structures aligned with target journal guidelines (e.g., IMRaD). GoT is then used to map the narrative flow, ensuring the core arguments connect logically from the introduction to the conclusion. Finally, Self-Consistency Prompting can generate several complete outline variations, allowing the researcher to select and refine the structure that best highlights the study's contributions.

AI-Powered Drafting (CoT & Few-Shot): With an approved outline, the AI assists in generating initial draft content. For complex sections requiring logical transparency, such as the methodology or discussion, CoT reasoning is used to produce well-structured, step-by-step paragraphs. To align the text with the desired academic voice, Few-Shot prompting is then applied, where the researcher provides high-quality writing samples to guide the AI's tone, style, and terminology.

Iterative Refinement and Humanization (Human Review): This final sub-step is fundamentally human-driven and iterative. The researcher directs the AI to perform specific revisions, such as improving clarity, checking for inconsistencies, and enhancing readability by removing repetitive or formulaic phrasing. However, the core intellectual work rests with the researcher, who is responsible for weaving the AI-generated components into a cohesive and compelling narrative, integrating domain-specific nuance, and ensuring the final arguments are persuasive and insightful.

Human Checkpoint: The final checkpoint is the most critical, as it determines the manuscript's intellectual integrity. The researcher must: (1) Conduct a thorough review to verify the technical accuracy of all claims and data interpretations; (2) Assume full ownership of the narrative, ensuring the paper's voice is authentic and the arguments are novel; and (3) Infuse the manuscript with the high-level

critical insights and scholarly perspective that only a human expert can provide. This final humanization phase is what elevates an AI-assisted draft into a genuine contribution to academic knowledge.

4 ILLUSTRATIVE CASE STUDY: MITIGATING BIAS IN AI HIRING ALGORITHMS

To demonstrate the practical application of the three-stage pipeline, this section provides a concrete case study of a machine learning research project. The project's goal is to design and validate a novel technique for mitigating intersectional bias in AI-powered hiring algorithms. Each stage includes examples of user prompts and illustrative AI responses, showcasing the collaborative workflow.

4.1 Stage 1: Research Proposal Development

Conceptualization and Grounding (CGP): "Based on references Buolamwini & Gebru (2018) [22] and Kearns & Roth (2019) [23], synthesize the primary themes and established definitions related to algorithmic bias."

Illustrative AI Response: "The provided literature establishes two primary themes: (1) Algorithmic systems often amplify societal biases present in training data, and (2) 'Fairness' is not a monolithic concept, with definitions like demographic parity and equalized odds often being mutually exclusive. Key definitions include 'intersectional bias,' where performance disparities are significantly worse for individuals in subgroups defined by multiple attributes (e.g., women of color)."

Knowledge Mapping and Gap Identification (GoT): "Create a conceptual map showing the relationships between different fairness mitigation techniques (pre-processing, in-processing, post-processing) and the fairness definitions they typically address. Highlight areas with weaker connections."

Illustrative AI Response: "A knowledge map shows strong connections between pre-processing techniques (e.g., re-weighting) and demographic parity. Similarly, in-processing techniques are often linked to equalized odds. However, the map reveals a weak link between most established techniques and the explicit mitigation of 'intersectional bias,' suggesting a significant research gap."

Research Question Refinement (Self-Consistency & CoVe): "Generate three potential research questions to address the gap in intersectional bias mitigation. For each, list its primary challenge or assumption."

Illustrative AI Response: "1) Question: How can adversarial debiasing be adapted to mitigate intersectional bias in hiring models? Challenge: Adversarial training can be unstable and may harm model accuracy. 2) Question: What is the impact of synthetic data generation for rebalancing intersectional subgroups on fairness? Challenge: Synthetic data may not capture the true characteristics of underrepresented groups. 3) Question: To what extent can multi-objective optimization balance predictive accuracy with intersectional fairness metrics? Challenge: Defining a robust intersectional fairness metric for optimization is non-trivial."

4.2 Stage 2: Methodological Design and Validation

Methodological Analysis (CGP & GoT): "Analyze the methodologies of the attached papers on fairness in hiring. Map their experimental components, including datasets, model architectures, and evaluation metrics."

Illustrative AI Response: "The analysis shows a common pattern: most studies use the 'Adult Census' or 'COMPAS' datasets with models like Logistic Regression or Gradient Boosting. Evaluation relies on metrics like Disparate Impact and Statistical Parity Difference. A conceptual map indicates that the 'Folktables' dataset, which is better suited for intersectional analysis, is underutilized in this context."

Comparative Selection (CCoT & ART): "Compare an experimental design using the public Folktables dataset against a case study design using proprietary data from a tech company. Evaluate each based on internal validity, external validity, and feasibility."

Illustrative AI Response: "Experimental Design: High internal validity and reproducibility due to the controlled public dataset. Lower external validity as Folktables may not represent a specific company's hiring pipeline. Case Study Design: High external validity and real-world impact. Low internal validity due to confounding variables and lack of reproducibility. Feasibility is lower due to data privacy and access constraints."

Bias Detection and Robustness Check (CoVe & ART): "We have selected the experimental design. Challenge this choice. Identify the three most significant potential sources of bias when using the Folktables dataset to study hiring bias."

Illustrative AI Response: "1. Measurement Bias: The target variable ('income >\$50K') is a poor proxy for a successful hire. 2. Sampling Bias: The dataset is not representative of all industries, limiting generalizability. 3. Construct Validity: The available features may not include key qualifications relevant to hiring, forcing the model to rely on demographic proxies."

4.3 Stage 3: AI-Assisted Manuscript Preparation

Structural Outlining (CGP, GoT, & Self-Consistency): "Generate a paper outline based on the IEEE conference format for our study on mitigating intersectional bias with multi-objective optimization."

Illustrative AI Response: "The AI generates a standard IMRaD outline formatted for IEEE: I. Introduction; II. Related Work (A. Fairness Metrics, B. Mitigation Techniques); III. Methodology (A. Dataset and Preprocessing, B. Model Architecture, C. Proposed Optimization Algorithm); IV. Experiments and Results; V. Discussion; VI. Conclusion."

AI-Powered Drafting (CoT & Few-Shot): "Draft the 'Methodology' section. Step 1: Describe the Folktables dataset. Step 2: Detail the XGBoost model architecture. Step 3: Explain our multi-objective loss function. Here is an example methodology section from another paper to guide the academic tone: [paste example]."

Illustrative AI Response: "The AI generates a multi-paragraph draft. 'III. METHODOLOGY. A. Dataset and Preprocessing: We use the 2018 Folktables dataset from the American Community Survey... B. Model Architecture: Our classification model is an XGBoost classifier with... C. Proposed Fairness Intervention: We introduce a multi-objective loss function defined as $L = \alpha L_{\{CE\}} + (1 - \alpha) L_{\{IF\}}$..."

Iterative Refinement and Humanization (Human Review): "Review the following draft of the 'Discussion' section. Improve the flow and remove repetitive phrases like 'It is important to note that...'. Ensure the claims are directly supported by the results mentioned."

Illustrative AI Response (Revised Text): "It is important to note that our model showed a 2% drop in accuracy. This finding is significant. It is also important to note that fairness for Black women improved by 15%."

The revised text become "While our intervention led to a modest 2% decrease in overall accuracy, it achieved a substantial 15% improvement in fairness for Black women, demonstrating a clear and impactful trade-off that prioritizes equity for the most vulnerable intersectional subgroup."

5 LIMITATIONS AND ETHICAL CONSIDERATIONS

While the AI-assisted research pipeline offers powerful tools for enhancing scholarly productivity, its responsible implementation demands a critical awareness of its limitations and a firm commitment to ethical academic practice. The framework presented in this paper is designed to leverage pre-trained LLMs as assistants, not as autonomous creators of knowledge. The efficacy and integrity of the entire process, therefore, depend on the researcher's ability to maintain rigorous oversight over tools they did not create.

Factual Accuracy and the Risk of "Hallucination". A primary limitation of LLMs is their lack of genuine comprehension; they operate on sophisticated pattern recognition, not true reasoning [17]. This can lead to outputs that are grammatically correct and stylistically plausible but factually inaccurate, logically flawed, or entirely nonsensical—a phenomenon known as "hallucination." A significant risk for academic writing is the generation of fabricated citations, references, or data points that appear authentic but are completely fictitious [3]. This makes human verification a non-negotiable step. Every claim, citation, and piece of data generated by an AI must be meticulously fact-checked against reliable sources to ensure the manuscript's scholarly integrity.

Bias Amplification in AI Outputs. LLMs are fundamentally shaped by the vast, static datasets on which they were trained, and these datasets inevitably reflect existing societal and historical biases [4]. Consequently, AI models can inadvertently replicate and even amplify these biases in their outputs. For example, an AI's synthesis of literature may overemphasize dominant theoretical perspectives while underrepresenting or mischaracterizing marginalized viewpoints. The researcher bears the ultimate responsibility to critically evaluate all AI-generated content

for hidden biases, actively anticipate how they might manifest, and correct for them to ensure a balanced and equitable scholarly narrative.

Authorship, Accountability, and Plagiarism. A firm consensus has emerged in the academic community: an LLM cannot be credited as an author. Authorship requires accountability and responsibility for the work's integrity, which a machine cannot assume. The human researcher who employs AI tools is fully and solely responsible for the final manuscript's content, originality, and accuracy. To maintain transparency, many journals and institutions now mandate explicit disclosure of AI tool usage. Furthermore, researchers must diligently check AI-generated text for unintentional plagiarism, as models may reproduce passages from their training data verbatim without proper attribution. The boundary between using AI as a tool and presenting its output as one's own intellectual work must be scrupulously managed to uphold academic integrity.

Over-Reliance and the Risk of Stifled Creativity. While AI can accelerate structured tasks, over-reliance on it poses a risk to the core of scholarly inquiry: critical thinking and intellectual innovation. Using a handful of dominant LLMs for fundamental tasks like identifying research gaps or structuring arguments could lead to a homogenization of ideas and a convergence of methodologies across the academic landscape. This may stifle the creativity, serendipity, and diversity of thought that fuel groundbreaking research. The primary role of AI should be to handle rote or organizational tasks, thereby freeing up the human researcher to focus on novel insights, creative problem-solving, and the development of a unique intellectual voice. AI should augment, not replace, the critical and creative faculties of the scholar.

6 PILOT STUDY: FRAMEWORK APPLICATION WITH MASTER'S-LEVEL RESEARCHERS

To assess the real-world feasibility and practical utility of the three-stage framework, an exploratory pilot study was conducted with a cohort of Master's students (both M1 and M2 levels) in a research-focused track. The students were tasked with using the AI-assisted pipeline to develop an initial proposal for their master's thesis. The objective of this study was not to derive quantitative performance metrics but to adopt a qualitative and observational approach. The study aimed to identify the framework's pedagogical value, its practical challenges, and the operational reality of the "Human Checkpoint" when implemented by novice researchers. The following sections synthesize the key observations from this pilot, organized thematically.

6.1 Theme 1: Scaffolding the Research Process for Novice Researchers

Conceptualization and Grounding (Stage 1): Students reported that the structured process of using CGP with a small set of foundational papers provided a tangible entry point into a dense body of literature. The AI-driven synthesis

of key themes and definitions furnished them with a foundational vocabulary and a conceptual map.

Methodological Literacy (Stage 2): Stage 2 proved particularly valuable for students struggling to connect a research question to a concrete research design. The use of CCoT to generate side-by-side comparisons of different methodologies (e.g., a qualitative case study vs. a quantitative survey) transformed an abstract decision into a structured analysis of trade-offs. This process appeared to enhance students' understanding of concepts like internal vs. external validity by forcing consideration of the practical implications of each choice.

6.2 Theme 2: The "Human Checkpoint" as a Pedagogical Tool

Observations from the study strongly validated the necessity of the "Human Checkpoint", revealing its dual function as both a quality control mechanism and a critical pedagogical tool. A consistent pattern of anchoring bias was observed, wherein students tended to over-rely on the initial AI-generated outputs, particularly in the early stages of research.

Overcoming Anchoring Bias: Without explicit guidance, students frequently demonstrated a reluctance to challenge or significantly refine the AI's first suggestions. For example, several students initially accepted an AI-generated research question without critically evaluating its scope or feasibility.

The Supervisor as the Human Checkpoint: The intervention of a human supervisor proved paramount. One student, following an AI suggestion, drafted a proposal to study hiring bias across all Fortune 500 companies---a project far too broad for a master's thesis. The subsequent human checkpoint with their supervisor was essential. The supervisor guided the student to use the AI's initial output as a starting point, but then to apply critical constraints (e.g., focusing on a single industry or a specific type of algorithm). This transformed the checkpoint from a simple "go/no-go" gate into a teachable moment about the practical realities of research scoping. In this context, the framework facilitated a more focused and effective conversation between the student and supervisor.

6.3 Theme 3: Practical Challenges and Directions for Future Work

The application of the framework by novice researchers also highlighted several practical challenges that inform the need for future research.

Prompt Engineering as a Skill Barrier: Students often struggled to write effective prompts. Their initial prompts were frequently too vague, leading to generic or irrelevant AI outputs. This suggests that the successful application of this framework depends on a secondary skill---prompt engineering---which itself requires training and practice.

The "Credibility Illusion": The highly fluent and authoritative tone of AI-generated text often created a "credibility illusion". Students found it difficult to distinguish

between a well-written but methodologically flawed suggestion and a genuinely sound one. This underscores the cognitive burden that remains on the human researcher to possess enough domain knowledge to critically evaluate AI outputs.

These observations lead directly to avenues for future study. While this pilot confirmed the framework's feasibility, the next stage of research must focus on developing and measuring the effectiveness of pedagogical interventions. Future studies should aim to create structured training materials for students on prompt engineering and critical AI output evaluation. Furthermore, developing robust metrics to assess the quality of a research proposal (e.g., clarity, feasibility, novelty) will be essential for any future quantitative comparison between this framework and traditional research methods.

7 DISCUSSION

The proposed framework positions AI-assisted research as a structured collaboration, balancing technological affordances with established principles of scholarly rigor. By formalizing a HITL model, the framework defines AI's role as one of augmentation, not autonomous execution. This distinction is critical for mitigating risks inherent in automation, such as the uncritical adoption of AI-generated outputs or the amplification of algorithmic biases. The model's emphasis on active intellectual oversight directly counters the risk of cognitive offloading, a phenomenon where reliance on technology can diminish a researcher's critical skills.

A primary implication of this approach is the promotion of transparency and replicability in a domain where processes can become opaque. In a manner analogous to the PRISMA guidelines for systematic reviews, the framework provides a clear methodology that can be documented, scrutinized, and reproduced. This structured process strengthens the credibility of the research findings and serves as a vital pedagogical model for training the next generation of researchers in the responsible and effective use of AI tools.

The framework's most significant limitation, however, is the methodological and evaluative challenge of defining its success. There are currently no standardized Key Performance Indicators (KPIs) to objectively compare the quality of its outputs against that of traditional, human-only research. While metrics of efficiency like time saved are easily measured, they fail to capture the core dimensions of scholarly value: the novelty of insights, the depth of critical analysis, or the serendipity of discovery. This challenge is amplified by the framework's intrinsic dependence on researcher expertise. The quality of the final output is inextricably linked to the user's ability to craft precise prompts and critically vet AI-generated content. Consequently, assessing the framework's definitive advantage remains a complex, qualitative endeavour, moving beyond the question of "faster" to the much harder-to-define "better."

Addressing these limitations requires further research, particularly empirical validation through cross-disciplinary case studies. Such work would help refine the framework's stages and identify discipline-specific adaptations. Future

efforts could also focus on developing integrated software tools that guide researchers through the framework, facilitating the systematic documentation required for transparent reporting and, eventually, comparative analysis.

8 CONCLUSION

This paper introduced a systematic, three-stage framework that operationalizes the collaboration between researchers and large language models. By structuring the research process into discrete phases—Proposal Development, Data Analysis and Interpretation, and Manuscript Drafting—the framework provides a replicable workflow for leveraging AI to enhance scholarly output. Its principal contribution is a structured methodology that integrates AI as a specialized assistant for knowledge discovery and content generation while ensuring that logical coherence and intellectual integrity are maintained at every step.

The central thesis of this work is that the efficacy of AI in research hinges upon the non-negotiable role of human intellect. The framework is intentionally designed with critical "Human Checkpoints" where the researcher's domain expertise, critical judgment, and ethical oversight are not just beneficial, but indispensable. AI-generated content, while often structurally coherent, lacks genuine comprehension and remains susceptible to factual inaccuracies and embedded biases. The AI can provide the scaffolding, but it is the human scholar who must act as the architect—validating claims, challenging assumptions, and infusing the work with the novel interpretation that constitutes a true contribution to knowledge.

Ultimately, the responsible integration of AI into research workflows presents a profound opportunity to deepen analysis, optimize efficiency, and expand the frontiers of inquiry. As these technologies evolve, the challenge for the academic community will be to co-evolve with them, refining collaborative frameworks that balance the automation of tasks with the amplification of intellect. This synergy does not replace the scholar; it redefines the scholar's role toward one of critical inquiry director, arbiter of validity, and architect of meaning in an increasingly complex information landscape.

9 REFERENCES

- [1] Albadarin, Y., et al. (2024). A systematic literature review of empirical research on ChatGPT in education. *Discover Education*, 3(1), 60. <https://doi.org/10.1007/s44217-024-00138-2>
- [2] Bolanos, F., et al. (2024). Artificial intelligence for literature reviews: Opportunities and challenges. *Artificial Intelligence Review*, 57(10), 259. <https://doi.org/10.1007/s10462-024-10902-3>
- [3] Chelli, M., et al. (2024). Hallucination rates and reference accuracy of ChatGPT and Bard for systematic reviews: Comparative analysis. *Journal of Medical Internet Research*, 26(1), e53164. <https://doi.org/10.2196/53164>
- [4] Chen, J., et al. (2024). *When large language models meet personalization: Perspectives of challenges and opportunities*. *World Wide Web*, 27(4), 42. <https://doi.org/10.1007/s11280-024-01276-1>
- [5] Chia, Y. K., et al. (2023). Contrastive chain-of-thought prompting. *arXiv preprint arXiv:2311.09277*.

- [6] Chubb, J., et al. (2022). Speeding up to keep up: Exploring the use of AI in the research process. *AI & Society*, 37(4), 1439–1457. <https://doi.org/10.1007/s00146-021-01259-0>
- [7] Dhuliawala, S., et al. (2023). Chain-of-verification reduces hallucination in large language models. *arXiv preprint arXiv:2309.11495*. <https://doi.org/10.18653/v1/2024.findings-acl.212>
- [8] Gerlich, M. (2025). *AI tools in society: Impacts on cognitive offloading and the future of critical thinking*. *Societies*, 15(1), 6. <https://doi.org/10.3390/soc15010006>
- [9] Glickman, M., & Sharot, T. (2025). How human–AI feedback loops alter human perceptual, emotional and social judgements. *Nature Human Behaviour*, 9(2), 345–359. <https://doi.org/10.1038/s41562-024-02077-2>
- [10] de la Torre-López, J., et al. (2023). Artificial intelligence to automate the systematic review of scientific literature. *Computing*, 105(10), 2171–2194. <https://doi.org/10.1007/s00607-023-01181-x>
- [11] Page, M. J., et al. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, 372. <https://doi.org/10.1136/bmj.n71>
- [12] Paranjape, B., et al. (2023). Art: Automatic multi-step reasoning and tool-use for large language models. *arXiv preprint arXiv:2303.09014*.
- [13] Ritchie, J., et al. (2013). *Qualitative research practice: A guide for social science students and researchers*. SAGE.
- [14] Schryen, G., et al. (2025). Exploring the scope of generative AI in literature review development. *Electronic Markets*, 35(1), 1–26. <https://doi.org/10.1007/s12525-025-00754-2>
- [15] Scott-Kennel, J., et al. (2025). Artificial intelligence in academic research: Contributor, constructivist or cheat? *Journal of Marketing Theory and Practice*, 1–22. <https://doi.org/10.1080/10696679.2025.2457672>
- [16] Shahzad, T., et al. (2025). A comprehensive review of large language models: Issues and solutions in learning environments. *Discover Sustainability*, 6(1), 27. <https://doi.org/10.1007/s43621-025-00815-8>
- [17] Wang, X., et al. (2022). Self-consistency improves chain of thought reasoning in language models. *arXiv preprint arXiv:2203.11171*.
- [18] Wang, Y., et al. (2020). Generalizing from a few examples: A survey on few-shot learning. *ACM Computing Surveys (CSUR)*, 53(3), 1–34. <https://doi.org/10.1145/3386252>
- [19] Wei, J., et al. (2022). Chain-of-thought prompting elicits reasoning in large language models. *Advances in Neural Information Processing Systems*, 35, 24824–24837.
- [20] Yao, Y., et al. (2024). GoT: Effective Graph-of-Thought Reasoning in Language Models. *Findings of the Association for Computational Linguistics: NAACL 2024*, 2901–2921. <https://doi.org/10.18653/v1/2024.findings-naacl.183>
- [21] Zhou, L., et al. (2021). Intelligence augmentation: Towards building human–machine symbiotic relationship. *AIS Transactions on Human-Computer Interaction*, 13(2), 243–264.
- [22] Buolamwini, J., & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. In *Proceedings of the 1st Conference on Fairness, Accountability and Transparency. Proceedings of Machine Learning Research*, 81, 77–91. <https://proceedings.mlr.press/v81/buolamwini18a.html>
- [23] Kearns, M., & Roth, A. (2019). *The ethical algorithm: The science of socially aware algorithm design*. Oxford University Press.

Authors' contacts:**Guillaume Guerard**

(Corresponding author)
De Vinci Higher Education, De Vinci Research Center,
92916 Paris, France
guillaume.guerard@devinci.fr

Sonia Djebali

De Vinci Higher Education, De Vinci Research Center,
92916 Paris, France

Maxime Hanus

Ecole Supérieure d'Ingénieurs Léonard de Vinci,
92916 Paris, France

Mark-Killian Zinenberg

Ecole Supérieure d'Ingénieurs Léonard de Vinci,
92916 Paris, France

Experimental Investigation on Material Removal Rate and Tool Wear Rate for Machining Metal Matrix Composites in EDM Process

Abbas Fadhil Ibrahim*, Shakir M. Mousa, Semyan Omed Alyousifi, Diana Abed al Kareem Noori

Abstract: Electrical discharge machining (EDM) is one of non-traditional cutting method or manufacturing process for processing brittle and solid materials like composite materials and other difficult to machine materials. It is most usually utilized to manufacture metal matrix composites (MMCs) and is very solid materials. Aluminum based MMCs are to a large degree utilized in defense and aviation parts for their advanced and new characteristics like strength to weight ratio, high rigidity etc. The aluminum alloy 6061 and 7.5% SiC has been used as a workpiece and copper as a tool electrode, discharge current (I_p), pulse on-time (T_{on}), and pulse off-time (T_{off}) were selected as machining dependent variables for this study. The dependent variables were tool wear rate (TWR) and material removal rate (MRR). The experiments are planned using Box-Behnken design. Empirical models are developed for MRR and TWR using the analysis of variance (ANOVA) and regression models. The getted results showed that the MRR is increasing with increasing in I_p and T_{on} while TWR is decreasing when I_p and T_{on} is increase.

Keywords: analysis of variance; electrical discharge machining; material removal rate; tool wear rate

1 INTRODUCTION

In conventional machines, it is difficult to machining hardened and composite materials at high speed and acceptable surface quality. A non-conventional process, like (EDM), is the better choice for this kind of material to overcome these conditions [1]. Electrical discharge machining emerged for the manufacture of conductive materials that are hard or impossible to cut with conventional machines [2]. However, the primary cause of the EDM machining will be specified by electrical conditions like voltage, package current, pulse time length, workpiece material and characteristic of electrode, such as thermal conductivity and material melting temperature [3, 4]. Copper works well as a tool and is vastly used when smooth work-piece finishing are required [5]. As a result of the interest in the development of electrical discharge machines, this led to significant improvements in EDM technology. Currently, advanced electrical discharge machines are available for most applications of machine workshops [6]. Metal Matrix Compounds (MMCs) are relatively low density based on magnesium or, aluminum, or titanium and stiffeners such as Al_2O_3 , SiC, TiB_2 or TiC [7, 8]. MMCs provides support and distributes the load to the structure while the particle or fiber stiffeners from supply useful mechanical, physical or thermal as performance. Therefore, MMCs are light metal or alloy and have high specific characteristics (like wear resistance, toughness, and strength), becoming a sought-after alternative in industrial applications. MMCs are produced in many different manufacturing process (like molding Metal injection, friction stirring process, mechanical alloying, pressure casting technology, continuous binder powder coating, etc.) [9]. Silicon carbide (SiC), which is also famous as carborundum, has very good mechanical characteristics. And thermal conductivity the value of the density of Silicon carbide is similar to that of Al. Therefore SiC reinforcement alloy composites are the typical choosing of material for defends functionalities in flight, defense purposes,

rigged structural, and heavy manufacturing industries, owing to their tribological and mechanical properties [10].

Several experiments and investigations conducted in the EDM process, it was observed that the melting points is impact to the TWR and the MRR increased with increasing I_p . Ref. [11] investigated the effect of controllable parameters of I_p , V , and pulse duration and duty cycle on process response such as TWR , MRR and Ra with cry therapy electrode. The results show that I_p and T_{on} and duty cycles have an important influence on the TWR and MRR . Ref. [12] examined T_{on} and I_p in Inconel 718 EDM with a CuW as a tool. The values indicated that the maximum removal rate was $28.37 \text{ mm}^3/\text{min}$ with highest I_p (40 A) and longest tone (400 μs). Ref. [1] studied optimization of variables for silicon zed silicon carbide (SiSiC) on the responses; TWR , Ra and MRR . By new design (FCCC) was utilized on the experiments [13]. The work studied the topical surface property of an arranged Titanium (Ti-6Al-4V) alloys in (EDM) process. Ref. [14] prepared the metal matrix composite, AA 6061 reinforced with 5wt% SiC particles by stir casting method, in EDM machining process to determine MRR and SR. Ref. [15] discussed the experiments of the EDM that were conducted to examine the influence of machining parameters, including current, pulse on, and duty factors on the MRR , and SR of the Al-alloy with-5%SiC-10%B4C hybrid composite as workpiece using copper electrode tool by Box-Behnken design. Ref. [16] Conducted a hybrid magnetic field assisted powder mixed on silicon carbide (Al-SiC) MMCs, current, P_{on} and P_{off} affected the duration and type of dielectric conditions significantly on the removal rate. The electrical energy in EDM is changed to heat energy by the series of separate electrical discharges that happen between the specimen and the electrode immersed in the dielectric medium, which works on this chain to melt then evaporate the metal from the workpiece surface, thus removing the material [17, 18].

The importance of the research lies in the manufacture and operation of Al alloy (6061) reinforced with 7.5% silicon carbide, has been fabricated through stir casting, which is

characterized by high wear resistance, stiffness, and light-weight and can be used in molds, automobile industry, and nuclear plants.

In this paper, a composite material was employed as Al-alloy with 7.5% SiC reinforcement, current, pulse on duration, and pulse off duration were chosen as process parameters, while the material removal rate and the tool wear ratio were chosen as process performance indicators.

2 DETERMINE THE MATERIAL REMOVAL RATE AND TOOL WEAR RATE

Material removal rate is a performance measure of work-piece wear rate expressed in the volume of material separated from the sample per unit time and is usually used to determine the speed at which processing is performed. Melting point is a significant parameter in determining the wear of the tool and depends on the factors of the machine [19, 20]. *MRR* and *TWR* are determined in Eq. (1) and (2) respectively, which depend on the weight of the work piece and tool before and after manufacturing, as well as the processing time [21-23].

$$MRR = \frac{(W_b - W_a) \cdot 1000}{(\rho_w \cdot T_m)}, \text{ mm}^3/\text{min} \quad (1)$$

Where: W_b - weight of specimen before EDM (g), W_a - weight of specimen after EDM (g), T_m - processing time (min), ρ_w - density of specimen material ($\rho_w = 2.77 \text{ g/cm}^3$).

$$TWR = \frac{(T_b - T_a) \cdot 1000}{(\rho_t \cdot T_m)}, \text{ mm}^3/\text{min} \quad (2)$$

Where: T_b - weight of tool before EDM (g), T_a - weight of tool after EDM (g), T_m - processing time (min), ρ_t - density of tool material ($\rho_t = 8.96 \text{ g/cm}^3$).

3 METHODOLGY OF BOX-BEHNKEN DESIGN

Box-Behnken designs are experimental designs for response surface methodology used in various statistical operations, invented by George E. B. Box and Donald Behnken in 1960, where at least three levels are required to achieve the goal, where each independent variable is set to one of the three values and equally, usually coded as -1, 0, +1.

The Box-Behnken design is considered more efficient and reliable than other designs such as the central component design, the three-level full factor design, and the Dohler design [24].

4 EXPERIMENTAL WORK

The composite material specimen is produced of an Al-alloy matrix (AA6061) and 7.5% SiC. Silicon carbide has been utilized as a good abrasive material and has been in products with grinding wheels for over a hundred years. After

that, the material has been advanced into high-quality ceramics with thermal and high mechanical properties. It is utilized in many more applications. The material can also be made electrically conductive and highly resistant to heating. The stir-casting method was utilized to manufacture size material. The composite porosity is about 7.5% SiC reinforced with 50 μm as grain size. The experiment was planned design and survey analyzed by utilizing Minitab 19 software with 13 test according to Box-Behnken design. As well, input three parameters containing the I_p , T_{on} , and the T_{off} were selected for the investigation. The aluminum alloy was heated for one hour and the silicon carbide granules were melted at (750°C) to separate gases and moisture from the surface of the work-piece layer [9]. The preheated SiC particle powder was added (7.5%) into the melt metal. After adding the nano silicon carbide powder, and for the purpose of obtaining a better homogeneity and distribution, we continue to stir for 10 minutes. The molten state was kept for one minute to remove the slag, and then the molten aluminum was poured into the graphite moulds. The important properties of the matrix and reinforcement are shown in Tab. 1.

Table 1 Properties of AA6061 and SiC

Material	Density (kg/m ³)	Melting point (K)	Modulus of elasticity (MPa)	Thermal conductivity (W/m·K)
AA6061	2710	893	70000	180
SiC	3200	3003	410000	120

The experimental have been done on EDM machine (CM 323C) with kerosene as dielectric solution, as shown in Fig. 1.

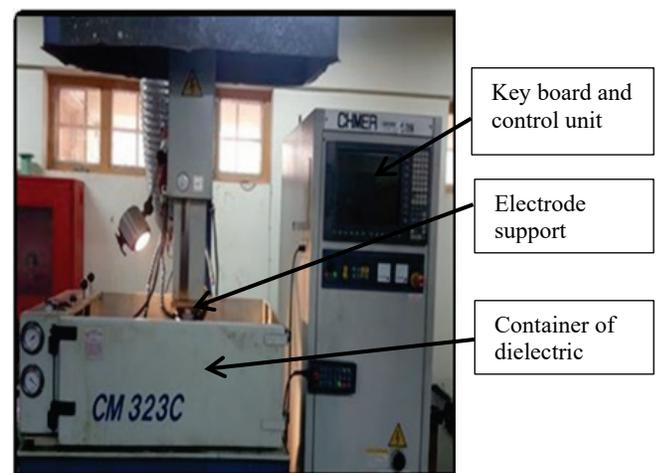


Figure 1 EDM type machine (CM323C)

In the experimental work a tool electrode was selected from pure copper material ($\varnothing 4 \times 100 \text{ mm}$). Copper has good EDM wear and better conductivity. The chemical composition for tool and AA6061 as shown in Tabs. 2 and 3 respectively.

Table 2 The chemical composition of Cu tool electrode W%

Pb	Sn	Si	Sb	Te	Cu
<0.005	0.003	0.010	0.033	0.015	Rem

Table 3 The chemical composition of specimen AA6061 W%

Cu	Mg	Si	Cr	Fe	Mn	Al
0.25	0.92	0.62	0.22	0.23	0.03	Rem

The normal workpiece polarity (+) was chosen. The study work includes thirteen tests according to the design of experiments and testing research plan. The machining parameters were listed in Tab. 4. It was selected based on previous research, which gives the best results for material removal rate and tool wear rate.

Table 4 Variables value

Variables	Symbol	Value		
		1	2	3
Current (A)	I_p	10	20	30
Pulse on duration (μ s)	T_{on}	100	150	200
Pulse off duration (μ s)	T_{off}	6	12	24

Table 5 Response (output) result and Box–Behnken design of the set experiments

Run order	Current (A)	Pulse on (μ s)	Pulse off (μ s)	<i>MRR</i> measured (mm^3/min)	<i>MRR</i> predicted (mm^3/min)	<i>TWR</i> measured (mm^3/min)	<i>TWR</i> predicted (mm^3/min)
1	30	100	6	4.368	4.81705	0.6320	0.63131
2	10	100	12	2.521	2.98170	0.0290	0.09782
3	10	150	6	2.975	3.13805	0.1471	0.17733
4	20	150	12	4.814	4.73200	0.5580	0.54400
5	10	200	6	3.624	3.48575	0.1950	0.22353
6	30	200	6	5.517	5.51245	0.7630	0.72371
7	20	100	6	4.273	3.80370	0.5860	0.53111
8	10	200	24	2.362	2.56025	0.2640	0.15691
9	30	150	24	5.263	5.79625	0.5640	0.61089
10	20	200	24	4.336	4.35210	0.4880	0.55689
11	30	200	12	6.861	6.48230	0.6730	0.69040
12	30	100	24	5.780	5.44855	0.5890	0.56469
13	10	100	24	2.281	1.86485	0.0850	0.06451

5.1 Analysis of Variance (ANOVA)

The claimant's null hypothesis was used in response to the 95% confidence point and the specified effective boundary conditions are a partial quadratic model of the automated processing process. Fisher's test (F-test) the significance of the strong function. If the (p-value) is 0.05, it is concluded that this parameter has a statistically significant influence. Tabs. 6, and 7 referred to the definitions of information of ANOVA for the *MRR* and *TWR* of EDM process. From non-linear regression models mathematical relationships were obtained between machining parameters, *MRR* and *TWR* as in Eqs. (3) and (4) respectively.

5.2 Prediction Mathematical Model of *MRR*

In order to estimate the final results of the *MRR* output response shown in Tab. (4), the ANOVA functions are run to decrease the p-value by utilizing the three-factor levels of the backward transform model. A model F value of 22.54 indicates that the pattern is significant. "Prob > F" values less than 0.05 referred the momentousness of the model terminol-

5 RESULTS AND DISCUSSION

According to the set of experimental outcomes achieved from Tab. 5 in appendix A, the influence of the independent variables P_1 , T_{on} and T_{off} on the two dependent variables that is *MRR* and *TWR* by used copper tool. The (AA 6061) with 7.5 % SiC can easily be machined by electrical discharge machining and maximum material removal rate, minimum tool wear rate can be obtained by controlling the machining parameters. ANOVA and from MINITAB 19 software is utilized to determine the effective of variables on the desired achievement measures. The largest material removal rate occurred in experiment 11, where I_p (30 A), P_{on} (200 μ s), and P_{off} (12 μ s), this is due to the increase in current with the increase in pulse time, which will lead to an increase in the strength of the electric spark in the cutting zone and the melting of the material more quickly. While minimum tool wear rate occurred in experiment 2, where I_p (10A), P_{on} (100 μ s), and P_{off} (12 μ s), this is due to reducing the current value and pulse time, which leads to increasing the tool life.

ogy. Therefore, *current*, *pulse on time*, and the interaction between *current* and *pulse off time* are significant sample terms. The predicted regression model empirical equation for *MRR* is:

$$MRR = 0.487 + 0.0754 \cdot I_p + 0.00695 \cdot T_{on} + 0.148 \cdot T_{off} - 0.00808 \cdot T_{off}^2 + 0.00432 \cdot I_p \cdot T_{off}, \text{ mm}^3/\text{min} \quad (3)$$

Table 6 ANOVA for *MRR*

Source	DF	Sum of squares	Mean of squares	F-value	p-value
Model	5	23.1353	4.6271	22.54	0.000
Linear	3	20.9358	6.9786	34.00	0.000
A	1	19.6729	19.6729	95.84	0.000
B	1	1.2090	1.2090	5.89	0.046
C	1	0.0540	0.0540	0.26	0.624
Square	1	0.9873	0.9873	4.81	0.064
B×B	1	0.9873	0.9873	4.81	0.064
2-way int.	1	1.2121	1.2121	5.90	0.045
A×C	1	1.2121	1.2121	5.90	0.045
Error	7	1.4369	0.2053	/	/
Total	12	24.5723	/	/	/

The three-dimensional graphs presented in Figs. 2-4 were utilized to estimate and explicate the experimental model.

Fig. 2 shows that the (MR) increases with increasing I_p (up to 30 A) and the pulses over the duration (up to 200 μ s) reach the value (6.861 mm^3/min). Presence of SiC particles affects higher *MRR* can be achieved by increasing current. Fig. 3 clarifies the 3D graphs for material removal rate (*MRR*) utilizing the I_p (30 A) and T_{off} (15 μ s), and the maximum surface roughness obtained when using reaches, the values (6.861 mm^3/min). Fig. 4 shows the 3D graphs for *MRR* utilizing the T_{on} (200) and T_{off} (15 μ s, and the last maximum (*MRR*) getted when utilizing the kerosene dielectric reaches the values (6.861 mm^3/min). This means that with get increasing the value of both the I_p and the T_{on} , the productivity increases. Thus, the T_{his} means that by get increasing the value of both the I_p and the T_{on} , productivity get increases. Thus, the amount of heat energy generated will be large, which leads to an increase in the melting process and successive abrasive to separate the out layers from the surface of the work piece.

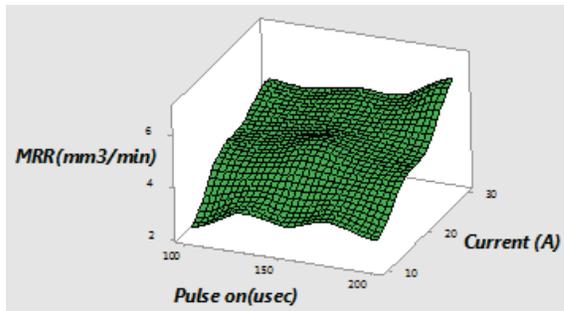


Figure 2 Relationship between *MRR*, current, and pulse on

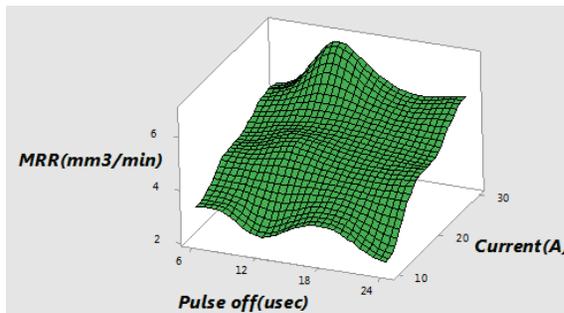


Figure 3 Relationship between *MRR*, pulse off, and current

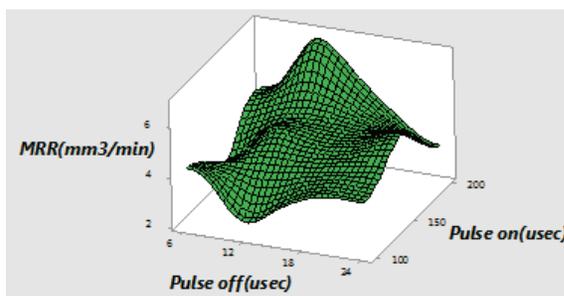


Figure 4 Relationship between *MRR*, pulse off, and pulse on

5.3 Prediction Mathematical Model of *TWR*

To evaluate the results of the instrument wear rate (*TWR*) given in Tab. 7 where the model F-value of 45.87 indicates that the model is significant. Therefore, *current*, *pulse on time*, and (*current*)² are very important specimen terms. The predicted final empirical equation is:

$$TWR = -0.639 + 0.085 \cdot I_p + 0.000924 \cdot T_{\text{on}} - 0.0037 \cdot T_{\text{off}} - 0.001499 \cdot I_p^2, \text{ mm}^3/\text{min} \quad (4)$$

In Eqs. (3) and (4) it was also apparent that the greatest significant machining variable input was current amplitude and impact on estimated or predicted *MRR* and then *TWR*

Table 7 ANOVA for *TWR*

Source	DF	Sum of squares	Mean of squares	F-value	p-value
Model	4	0.70974	0.177434	45.87	0.000
Linear	3	0.65789	0.219297	56.70	0.000
A	1	0.62545	0.625450	161.70	0.000
B	1	0.02134	0.021344	5.52	0.047
C	1	0.01110	0.011096	2.87	0.129
Square	1	0.05185	0.051847	13.40	0.006
A×A	1	0.05185	0.051847	13.40	0.006
Error	8	0.03094	0.003868	/	/
Total	12	0.74068	/	/	/

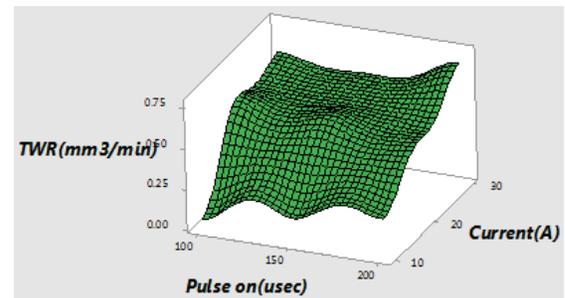


Figure 5 Relationship between *TWR*, current, and pulse on

Figs. 5-7 show the effect of the EDM variables on the *TWR*. Fig. 5 illustrates that when utilizing the I_p (10 A) and T_{on} (100 μ s), the value of *TWR* decreased to (0.0290 mm^3/min). Fig. 6 indicates the 3D graphs for *TWR* using the pulse current (10 A) and pulse off duration (15 μ s), and the minimum *TWR* gets when utilizing reaches, the rate values (0.0290 mm^3/min). Fig. 7 shows the 3D graphs for *TWR* utilizing the I_p (100) and T_{on} (15 μ s), and the minimum value of *TWR* is reached (0.0290 mm^3/min).

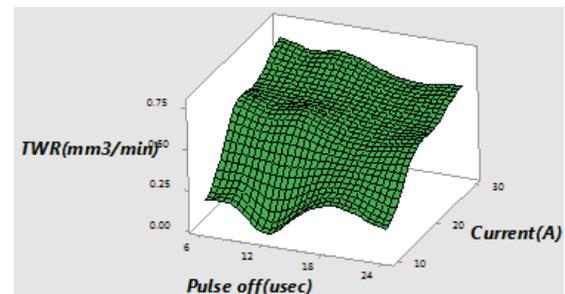


Figure 6 Relationship between material removal rates, current, and pulse off

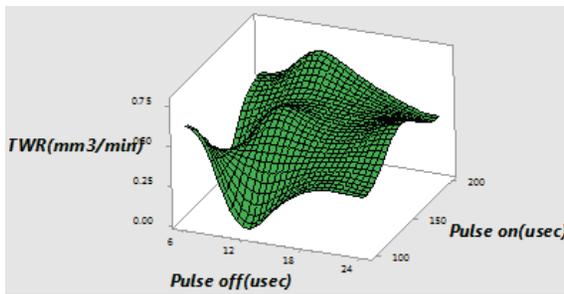


Figure 7 Relationship between material removal rates, pulse on, and pulse off

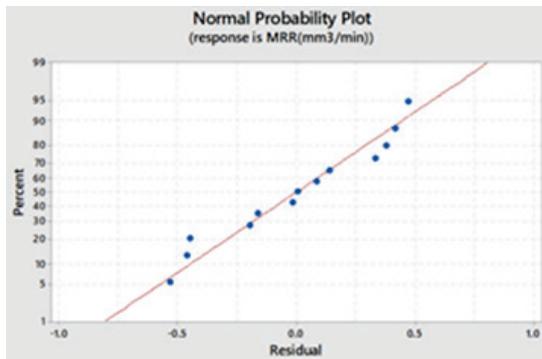


Figure 8 Probability plot for MRR

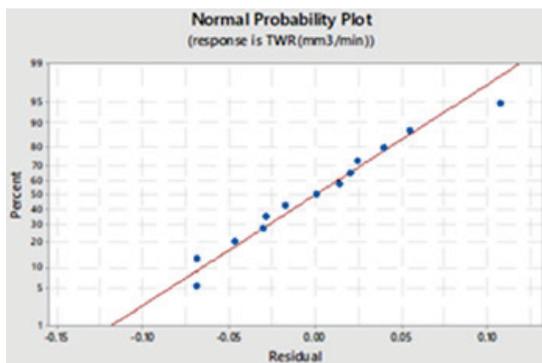


Figure 9 Probability plot for TWR

5.4 Probability plot for MRR and TWR

Figs. 8 and 9 are represents the normal plot of residuals for material removal rate and tool wear rate separately respectively. All data are clearly normal and there is no deflection from the usually, due all the points on the normal plot represented as in straight line.

5.5 SEM Inspection of Machined Surface

From Figs. 10 and 11 show the microstructure inspection for experiment 2 and 11 at the low and high energy respectively, where the processing input parameters in Fig. 10 are at pulse on time (100 µs), current (10 A) and pulse off time (12 µs). During the time that the processing input parameters in Fig. 11 at pulse time (200 µs), current (30 A) and pulse off time (12 µs), which showed in the state of a large quantity of highly molten material on the top layers' surface and very deep holes as well as micro-cracks.

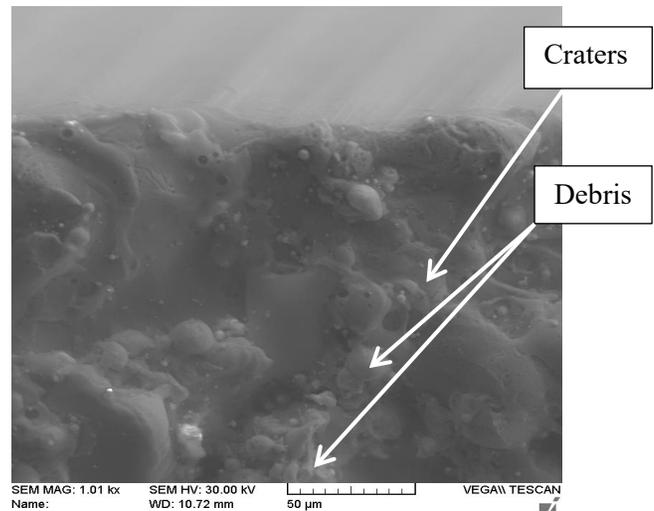


Figure 10 Fine structure microstructure of the machined surface at the low energy

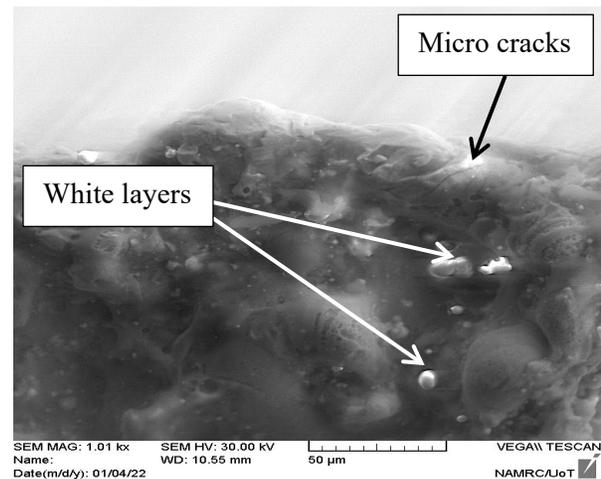


Figure 11 Fine structure microstructure of the machined surface at the high energy

Through the pulse on time value increases with time, more amount pulse discharge energy is released, thus generating more heat, resulting in the evaporation and melting of material a minute particles from the top surface of the workpiece and the action of forming relatively great large dark holes on the top surface. The still-remaining a thin piece of material sticks to the surface to form spherical pieces of waste as pellets and debris. When the particles of carbon react with the molten metal, it naturally and after that, the cooling process leads to union bonds greater than the original element, and when the stress on the surface exceeds the end the ultimate tensile strength of the material, in this case, cracks form.

6 CONCLUSIONS

In this study, has been made to estimate or predict (*TWR*) and (*MRR*) in EDM of Al-alloy with 7.5% SiC using a Box-Behnken design procedure. The maximum *MRR* is achieved when the I_p and T_{on} are very high and T_{off} is (12 µsec), whereas the lower value of *TWR* is observed when the I_p and T_{on} are minimum value and the T_{off} is 12 µs. ANOVA table

shows that I_p and T_{on} are an important parameter. The normal probability plot indicates that the mathematical regression model agrees passably well with noticed values and the errors are normally distributed. In the next work, the study can be involved for other output responses like recast layer white (RLW), surface roughness (SR), and radial over cut (ROC). The final results from this paper will be beneficial for production engineers to choosing suitable design of process variable parameters to machine Al-alloy with 7.5% SiC.

7 REFERENCES

- [1] Abdul Azeed, A., Jafri, M., Ahmad, M. & Hamidon M. (2017). Optimization of electrical discharge machining parameters of SiSiC through response surface methodology. *Journal Teknologi*, 79, 119-129. <https://doi.org/10.11113/jt.v79.7622>
- [2] Ritesh Kumar, S., Sanjiv Kumar, T., Sharad Chandra, S. & Binay K. (2023). Hybrid Taguchi-GRA-CRITIC Optimization Method for Multi-Response Optimization of Micro-EDM Drilling Process Parameters. *Tehnicki Vjesnik*, 30(3), 804-814. <https://doi.org/10.17559/TV-20220601114015>
- [3] Bhaskar, C., Jatinder, K. & Hari, S. (2017). Investigations into electrical discharge machining of fabricated AA 6061/10% Al₂O₃ aluminium-based metal matrix composite using OFAT approach. *Int. J. Automotive Composites*, 3, 29-43. <https://doi.org/10.1504/IJAUTO.2017.086514>
- [4] Gore, A. S. & Patil N. G. (2018). Wire electro discharge machining of metal matrix composites: A review. *Procedia Manufacturing*, 20, 41-52. <https://doi.org/10.1016/j.promfg.2018.02.006>
- [5] Jawahir, I. S., Brinksmeier, E. & M'Saoub, R. (2011). Surface integrity in material removal processes: Recent advances. *Manufacturing Technology*, 60, 603-62. <https://doi.org/10.1016/j.cirp.2011.05.002>
- [6] Arvinder, S. C., Harminder, S. B., Jasmaninder, S. G., Vettivel, S. C. & Raman, K. (2023). Tool wear rate during electrical discharge machining for aluminium metal matrix composite prepared by squeeze casting: A prospect as a biomaterial. *J. Electrochem. Sci. Eng.*, 13(1), 149-162. <https://doi.org/10.5599/jese.1391>
- [7] Abbas, F. I. (2022). Evaluation of surface roughness and material removal rate in electrical discharge machining of Al-alloy with 10% SiC. *IJUM Engineering Journal*, 23(1), 349-357. <https://doi.org/10.31436/ijumej.v23i1.2114>
- [8] Liao, Z., Abdelhafeez, A., Li, H., Yang, Y., Diaza, O. G. & Axinte, D. (2019). State-of-the-art of surface integrity in machining of metal matrix composites. *Int. J. Mach Tool Manu.*, 143, 63-91. <https://doi.org/10.1016/j.ijmactools.05.006>
- [9] Mohammad, H., Muhammad, I. P. & Mukhtiar, A. U. (2011). Manufacturing of Aluminum Composite Material Using Stir Casting Process. *Mehran University Research Journal of Engineering & Technology*, 30, 37-40. <https://doi.org/10.48550/arXiv.1604.01251>
- [10] Hadi, M. H. & Abbas, F. I. (2023). Experimental Study of Tool Wear Rate and Surface Roughness in Electrical Discharge Machining of Monel Alloy Using Cu-Gr Composite Electrode. *AIP Conference Proceedings*, 2830(1), 030009. <https://doi.org/10.1063/5.0156914>
- [11] Hadi, M. H. & Abbas, F. I. (2022). Effect of Copper-Graphite Composite Electrode on Material Removal Rate and Surface Roughness in Monel 400 during Electrical Discharge Machining. *Metallurgical and Materials Engineering*, 28(2), 245-255. <https://doi.org/10.30544/817>
- [12] Sengottuvel, P., Satishkumar, S. & Dinakaran, D. (2013). Optimization of Multiple Characteristics of EDM Parameters Based on Desirability Approach and Fuzzy Modeling. *Procedia Engineering*, 64, 1069-1078. <https://doi.org/10.1016/j.proeng.2013.09.185>
- [13] Sulaiman, A., Ahsan, A. & Mohamed, K. (2018). Analysis of microlayer defects of a functionally graded titanium alloy by electrical discharge machining. *Engineering and Applied Science Research*, 45, 17-22. <https://doi.org/10.14456/easr.2018.1>
- [14] Abbas, F. I., Shakir, M. M. & Diana, A. N. (2022). Investigation and optimization of machining parameters in electrochemical machining of aluminium metal matrix composites. *Periodicals of Engineering and Natural Sciences*, 10(3), 48-59.
- [15] Abbas F. I., Abdullah H. S., al Kareem Noori, D. A. (2022). Investigation of material removal rate and surface roughness during electrical discharge machining on Al(6061)-5%SiC-10%B4C hybrid composite. *Metallurgical and Materials Engineering*, 28 (1) p. 47-60. <https://doi.org/10.30544/798>
- [16] Rizovich Abylaz, T., Singh Bains, P., Singh Sidhu, S., Ravilevich Muratov, K. & Sergeevich Shlykov, E. (2021). Impact of Magnetic Field Environment on the EDM Performance of Al-SiC Metal Matrix Composite. *Micromachines*. 12, 2-13. <https://doi.org/10.3390/mi12050469>
- [17] Skoczypiec, S., Machno, M. & Bizon, W. (2015). The Capabilities of Electrodischarge Microdrilling of High Aspect Ratio Holes in Ceramic Materials. *Management and Production Engineering Review*, 6, 61-69. <https://doi.org/10.1515/mp-2015-0027>
- [18] Sonawane, S. A. & Kulkarni, M. L. (2018). Multi Response Optimization of Wire Electrical Discharge Machining for Titanium Grade-5 by Weighted Principal Component Analysis. *International Journal of Engineering and Technology Innovation*, 8, 133-145.
- [19] Srivastava, V. & Pandey, P. (2013). Experimental Investigation on EDM process with ultrasonic assisted cryogenically cooled electrode. *Journal of Engineering Manufacturing*, 227, 301-314.
- [20] Munmun, B. & Kalipada, M. (2020). Effect of Powder Concentration on the EDM Performance of AISI 304 Using Cryotreated Post Tempered Electrodes. *Engineering Review*, 77-78. <https://doi.org/10.30765/er.1655>
- [21] Sultan, T., Kumar, A. & Gupta, R. D. (2016). Material Removal Rate, Electrode Wear Rate, and Surface Roughness Evaluation in Die Sinking EDM with Hollow Tool through Response Surface Methodology. *Int. J. Manuf. Eng.*, 2014, 1-16. <https://doi.org/10.1155/2014/259129>
- [22] Uthayakumar, M., Thirumalai Kumaran, S., Adam Khan, M., Skoczypiec, S. & Bizon, W. (2020). Microdrilling of AA (6351)-SiC-B4C Composite Using Hybrid Micro-ECDM Process. *ASTM International. J. Test. Eval.*, 48(4), 3073-3086. <https://doi.org/10.1520/JTE20180216>
- [23] Abbas, F. I. (2021). Influence of electrical discharge machining parameters by additives Nano [AL₂O₃] on surface roughness and material removal rate in machining of AISI 304. *Journal of Physics: Conference Series*. <https://doi.org/10.1088/1742-6596/1973/1/012154>
- [24] Karmoker, J. R., Hasan, I., Ahmed, N., Saifuddin, M. & Reza, M. S. (2019). Development and Optimization of Acyclovir Loaded Mucoadhesive Microspheres by Box-Behnken Design. *Dhaka University Journal of Pharmaceutical Sciences*, 18(1), 1-12. <https://doi.org/10.3329/dujps.v18i1.41421>

Authors' contacts:

Abbas Fadhil Ibrahim

(Corresponding author)

Production Engineering and Metallurgy,

University of Technology,

Baghdad, Iraq

E-mail: abbas.f.ibrahim@uotechnology.edu.iq

Shakir M. Mousa

Al-Furat Al-Awsat Technical University,

Technical Institute of Al-Mussaib, Babylon, Iraq

Semyan Omed Alyousifi

Manufacturing Department, Duhok Technical Institute,

Zakho Way Road, Duhok, Iraq

Diana Abed al Kareem Noori

Production Engineering and Metallurgy,

University of Technology,

Baghdad, Iraq

Assessment of Forces Acting on the Gondola Car Body during Transportation of Rolled Wire Bundles

Angela Shvets*, Vladyslav Shaposhnyk, Leontii Muradian, Anzhelika Shvets, Alexei Reidemeister, Oleksandr Shykunov, Dmytro Molchanov

Abstract: The article gives an assessment of the possible reasons for the fall of a gondola car wall onto the side track, which occurred on the section of the regional branch of the Pridneprovskaya Railway. To determine the possible reasons for the breakage of the car side wall, a comparison of the forces acting from the side of rolled wire bundles on the car body structure with permissible loads on the body elements was made. Vertical and transverse inertial forces and forces from the load thrust were considered as such forces. The total loads obtained as a result of the calculation do not exceed the permissible values for gondola car bodies and confirm that the considered local technical conditions for stowing rolled wire bundles met the current requirements and did not cause a breakage of the side wall of the gondola car body.

Keywords: gondola car; inertial forces; load thrust forces; rolled wire bundle; stowing and securing of cargo; strength

1 INTRODUCTION

Traffic safety on railways depends not only on the maintenance of rolling stock and infrastructure, but also, to a large extent, on the reliability of cargo securing and its stowing in the rolling stock in accordance with the existing standards. The diverse range of cargoes transported by rail requires both a variety of types of freight cars and a significant amount of technical documentation establishing methods for stowing and securing cargo (local technical conditions – LTC, unforeseen technical specifications of UTS and sketches). The need to develop technical conditions is caused by the specifics of a particular cargo and the conditions for its securing, not provided for by the technical conditions for stowing and securing cargo in cars and containers [1, 2]. The development of technical conditions is associated not only with ensuring the reliability of securing, preservation of cargo and rolling stock, but also with the need for rational use of the body volume, carrying capacity of the car and ensuring the minimum cost of the securing method (labor intensity of handling operations, the cost of materials and securing elements). Compliance with these requirements, which may contradict each other, is the main task during the development of cargo stowage and securing schemes. Non-compliance with these requirements results in increase in the cost of transportation or uncoupling of cars along the route due to the displacement/falling of cargo, weakening/breakage of securing means, exceeding the loading gauge, etc., as a result of which both the consignors and the railways bear significant financial costs [3].

In July 2020, on the Voskoboynya – Verkhnedneprovsk section of the regional branch of Pridneprovskaya Railway the side wall of the gondola car body fell onto the track during the train movement, followed by the loss of 14 bundles of rolled wire. Falling of cargo under the rolling stock has led to the derailment of three cars and a locomotive of the train, which was moving along an adjacent track (Fig. 1) [4].

The reason for the fall of the side wall of the car body, which had an extended service life, was its break from the

corner posts with subsequent bending of the intermediate transverse beams and breakage from all transverse beams due to poor-quality welded joints between them.



Figure 1 Consequences of a transport accident on the Voskoboynya – Verkhnedneprovsk section

It was established that the transport accident occurred due to the breakage of the side panel from the fastening points, namely the bolster and intermediate beams and from the end posts and end beam, due to the presence of significant metal corrosion at the points of their connection (Fig. 2).



Figure 2 Breakage of the side panel from the fastening points

Another reason was considered the exceedance of the lateral loads from the transported cargo on the side wall due to the incorrect calculations when developing a method for its stowing in a gondola car.

2 LITERATURE REVIEW AND PROBLEM DEFINITION

When developing technical conditions for stowing and securing of cargoes, as noted in the works [1, 5], the human factor associated with violations of the technological process of stowing and securing of cargoes, which is most often manifested in the supply of faulty cars for loading, excess of the permissible load capacity of the car, uneven cargo stowage should be taken into account. All this can result in serious damage to cargo, rolling stock and railway infrastructure. The human factor can also be associated with the imperfection of the developed securing schemes and errors in them.

In order to increase the reliability of cargo securing when transporting it in the gondola cars, work [6] proposes a modernization option. It provides for the implementation of pull-out bars, which will increase the reliability of cargo securing when loading with a "hat" that is beyond the upper strapping of the side wall.

The works [7-11] are devoted to the problem of calculating the strength of the body and the frame of the freight cars. The results of works [6, 12] confirm the sufficient body strength, but they do not take into account the nuances of the transportation of bundles related with the action of a concentrated load, which is applied to the body posts in two places along the body height, which depends on the bundle diameter. In work [8], the structural strength of a universal flat car is estimated when a specialized frame with containers is placed on it. The articles [10, 12] analyze the residual life of gondola cars, depending on the service life and the car body material under conditions of acting operating loads and corrosive wear. Based on the results of the presented studies, it was found that the resource of the supporting body elements is 29-37 years, depending on the level of corrosion damage. However, the work did not take into account the human factor, and the traffic accident on the Voskoboynya – Verkhnedneprovsk section was caused by non-compliance with the requirements of TsV-0142 "Cars of freight railways of Ukraine with 1520 (1524) mm gauge. Depot repair instructions" and the requirements of STP 04-020:2018 "Freight cars and containers. Welding and welding deposition instructions".

To transport rolled wire along the main routes, shippers choose general-purpose gondola cars, it is possible to transport wire in specialized and modernized flat cars for transporting sheet steel. But, in this case, there is a significant underutilization of the carrying capacity of the car due to the impossibility of bundle stowing in two tiers, which significantly increases the transportation cost.

The article [13] describes the capabilities and advantages of an automated system for the development of technical conditions for the stowing and securing of cargoes in cars and containers. The decision-making algorithm makes it possible to automate the process of developing technical conditions, taking into account the possibility of two-axis displacement of the cargo in the car. The article [14] presents a method for the automated calculation of the efforts value in the cargo securing elements. Automated calculation methods make it possible to determine the value of cargo displacement both along and across the car from the action of longitudinal, transverse and vertical forces perceived by the cargo during

transportation. The given method allows optimizing the choice of the mounting location of the support frames, depending on the geometric parameters of the load and the selected method of its securing.

When developing schemes for placing and securing cargo, compliance with the permissible displacement of the cargo gravity center is mandatory. In this case, it is essential to take into account the possible cargo displacement during transportation by rail. The works [15, 16] are devoted specifically to the issue of determining the influence of the longitudinal and transverse displacement of the gravity center of a heavy cargo, taking into account the movement speed, on the main dynamic indicators – maximum coefficients of dynamic increment of sprung and unsprung parts, the maximum ratio of frame force to static axial load, and derailment stability coefficient.

The requirements for stowing and securing of rolled wire bundles on the railways of the member countries of the Organization for Railways Cooperation are established by Appendix 3 to the Agreement on International Goods Transport by Rail. According to it, one of six schemes is selected depending on the carrying capacity of the car, the mass and size of the bundles. The bundles with a diameter of 1150-1400 mm, a length of 450-750 mm and a weight of 500-850 kg are placed in a gondola car in two rows across the width and two tiers along the height of the car. In the lower tier, the bundles are located closely from the ends to the middle of the car. One half of the tier is displaced to one side wall of the car, and the other to the opposite wall. The upper tier of the bundles is placed on the lower tier bundle similarly to it, displaced to the side walls in the opposite direction (Fig. 3a).

The rolled wire bundles in the car in transport accident on the Voskoboynya – Verkhnedneprovsk section, were stowed according to the LTC developed and approved by the consignor (Fig. 3b). The main difference from the previous scheme shown in Fig. 3 is large dimensions and weight of bundles. Bundle diameter D up to 1250 mm, length up to 2400 mm, and weight up to 2500 kg.

The scheme of rolled wire transportation in accordance with the requirements of the European Railways [12], which provides for its securing by wooden profile, lashing straps and cartons laid between the bundles is shown in Fig. 3c.

Based on the above-mentioned works, we can conclude that the work related to increasing the reliability of cargo stowing and securing in the cars and containers is undoubtedly relevant and timely. Its solution requires accounting a significant number of factors that affect both traffic safety and the economic aspects of the cargo transportation by rail. The analysis of wire stowing and securing schemes in gondola cars showed the presence of various approaches to the stowing and securing of rolled wire.

The purpose of this case study is to analyze the forces acting on the gondola car body when transporting wire bundles, and to assess the body strength. To achieve this purpose, it is necessary to calculate the forces acting on the gondola car body when transporting wire bundles and to assess the strength of the gondola car body.

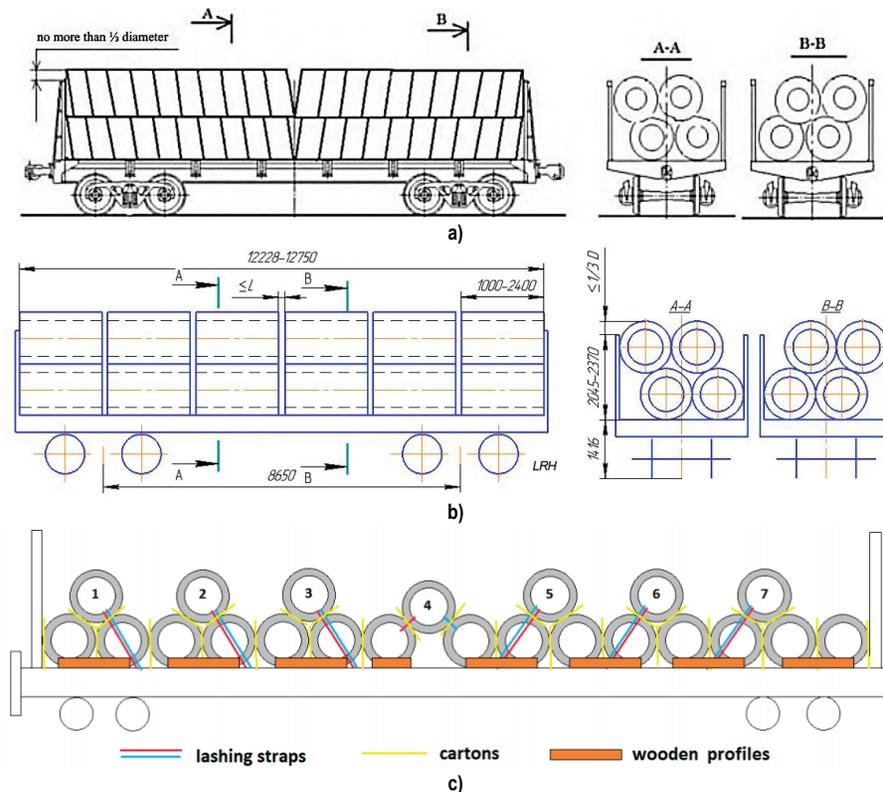


Figure 3 Stowage of rolled wire bundles in a gondola car: a) in accordance with the requirements of Appendix 3 to the Agreement on International Goods Transport by Rail; b) in accordance with the LTC; c) in accordance with the requirements of the European Railways

3 CALCULATION METHODOLOGY

The stowage scheme of the gondola car bundles is shown in Fig. 3 b. Wire in bundles with an outer diameter of 1150-1250 mm and a weight of 1000-2500 kg is stowed in gondola cars with blank end walls in two tiers, two rows in each tier. To load the wire uniformly along the car length, it is loaded alternately from the gondola car ends to the middle. In the lower tier, the bundles are stowed with a displacement of half of the tier close to one side wall of the gondola car, the second half – close to the opposite wall. In the upper tier, the bundles are stowed on the bundle of the lower tier similarly to the lower one, with a displacement to the side walls in the opposite direction (Fig. 4).

The number of bundles in a row depends on their height and the length of the car body and can be from 5 to 12 pieces, and in a car – from 20 to 48 pieces. When stowing the wire bundles without gaps, the maximum weight of the load can be 69 tons. The part of the upper tier bundles going beyond the upper strapping of the gondola car should not exceed 1/3 of the bundle diameter [17].

According to the standard parameters, the mass of the wire bundle will be 1.425 tons, and the total weight of the load will be 68.4 tons. The point of application of longitudinal, transverse, and vertical inertial forces is the center of gravity of the load (CG).

The calculation of the required parameters, permissible displacements of the general center of gravity of the cargo and inertial forces is carried out in accordance with the

methodology described in Chapter 1 of Appendix 3 to the Agreement on International Goods Transport by Rail [2].

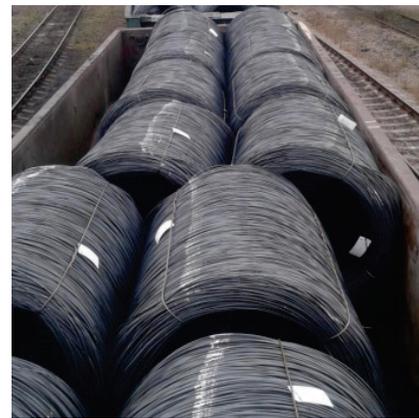


Figure 4 Rolled wire bundles after collisions according to the corrected LTC scheme

The height of the general center of gravity of the car with cargo is determined by the dependence:

$$H_{gc}^o = \frac{Q_{w1} \cdot h_{w1} + Q_{w2} \cdot h_{w2} + \dots + Q_{wn} \cdot h_{wn} + Q_{uw} \cdot H_{gc}^{emp}}{Q_w^o + Q_{uw}}, \quad (1)$$

$$H_{gc}^o = 2296 \text{ mm} < 2300 \text{ mm}. \quad (2)$$

where Q_{uw} – wagon tare weight, t; $h_{w1}, h_{w2}, \dots, h_{wn}$ – height of the CG of cargo units from the level of the rail heads

(LRH), mm; H_{gc}^{emp} – CG height of an empty car from the LRH, 1130 mm. Since the obtained value does not exceed the permissible value, the stability of the car is ensured.

The permissible longitudinal displacement of the overall center of gravity of the cargo in a 4-axle car is determined by linear interpolation in accordance with the regulatory document [2]. The values of the displacements of the total center of gravity of the cargo are determined from the corresponding tables depending on the total weight of the cargo and the permissible displacement during the transportation of the cargo along the route.

When loading:

$$\ell_{w-69}^{\ell} = \ell_{w-67}^{\ell} - \frac{\ell_{w-67}^{\ell} - \ell_{w-70}^{\ell}}{70 - 67} \cdot (69 - 67) = 66.67 \text{ mm.} \quad (3)$$

In transit:

$$\ell_{w-69}^t = \ell_{w-67}^t - \frac{\ell_{w-67}^t - \ell_{w-70}^t}{70 - 67} \cdot (69 - 67) = 126.67 \text{ mm.} \quad (4)$$

Further calculation of the forces acting on the body of the gondola car when transporting wire in bundles, as well as assessment of the strength of the body, is carried out at the maximum possible value of the permissible longitudinal displacement of the total center of gravity of the load $\ell_w = 127$ mm.

Longitudinal inertial force is calculated by the equation:

$$F_{long} = a_{long} \cdot Q_w, \quad (5)$$

where Q_w – weight of cargo, including accessories, t; a_{long} – specific longitudinal inertial force, which is calculated by the equation:

$$a_{long} = a_{22} - \frac{Q_w \cdot (a_{22} - a_{94})}{72}, \quad (6)$$

where a_{22} , a_{94} – specific longitudinal inertial forces, $a_{22} = 1.2$ tf/t, $a_{94} = 0.97$ tf/t.

The transverse inertial force is determined by the equation:

$$F_{tr} = a_{tr} \cdot Q_w, \quad (7)$$

where a_{tr} – specific transverse inertial force, which is calculated by the equation:

$$a_{tr} = a_c + \frac{2 \cdot (a_b - a_c)}{L_b} \cdot \ell_w, \quad (8)$$

where a_c , a_b – specific transverse inertial force for cargo located inside the car and above the body bolster, respectively, $a_c = 0.33$ tf/t, $a_b = 0.55$ tf/t; ℓ_w – distance from

the center of gravity (CG) to the transverse axis of the car, mm; L_b – wheel base, 8650 mm.

The vertical inertial force is found by the equation:

$$F_v = a_v \cdot Q_w, \quad (9)$$

where a_v – specific vertical inertial force (kgf/t);

$$a_v = 250 + k \cdot \ell_w + \frac{2140}{Q_w}, \quad (10)$$

where $k = 5$ – loading factor per car. The specific vertical inertial force is determined for a speed of 100 km/h.

Wind load is determined by the equation:

$$W_w = 50 \cdot S_w / 1000, \quad (11)$$

where S_w – area of the windward surface of the cargo, 5.291 m².

The friction forces were determined in the following direction:

- in longitudinal direction:

$$F_{fr}^{long} = Q_w \cdot \mu, \quad (12)$$

where μ – is the friction coefficient, $\mu = 0.3$ (steel on steel);

The values of the friction coefficient between the supporting surfaces of the cargo, linings, and the floor of the wagon (sprinkled with a thin layer of sand in winter) cleared of dirt, snow, and ice in accordance with [2] are taken to be equal to 0.30 for the pair "steel on steel".

- in transverse direction:

$$F_{fr}^{tr} = Q_w \cdot \mu \cdot (1 - a_v). \quad (13)$$

Longitudinal inertial force to be dampened:

$$\Delta F_{long} = F_{long} - F_{fr}^{long}. \quad (14)$$

Transverse inertial force to be dampened:

$$\Delta F_{tr} = n \cdot (F_{tr} + W_w) - F_{fr}^{tr}, \quad (15)$$

where n – is the coefficient, which is taken as 1.0 for LTC; W_w – wind load.

Longitudinal inertial force – $\Delta F_{long} = 46.649$ tf, perceived by the elements of the end wall and corner posts with permissible values of 40 tf and 2×9.5 tf, respectively [2]: $\Sigma F = [59] > 46.649$ tf.

We will find the transverse forces according to the presented methodology, but taking into account the thrust load on the side walls of the car from the bundles. The scheme of bundle stowing in the car (Fig. 3b) provides for the loading transfer to each side of the car wall on one of its

half from the upper tier, and on the second – from the bottom tier. Fig. 5 shows the action of the thrust loads on the side walls of the car. Fig. 6 presents a design scheme for determining these loads, the corresponding dimensions are as follows: $L = 572$ mm; $\ell = 252$ mm, $r = 625$ mm. Let us determine the reactions in the supports indicated by the points A and B of the upper tier from a bundle with a weight of 1.425 tons.

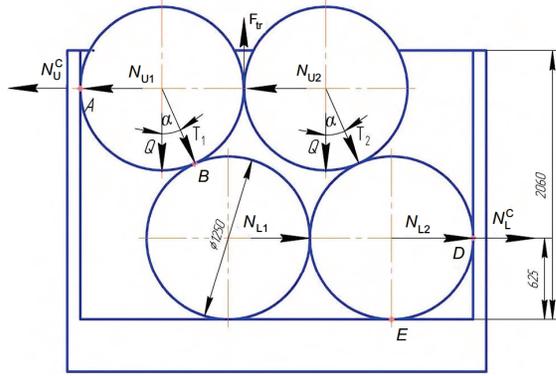


Figure 5 Thrust loads from the bundles

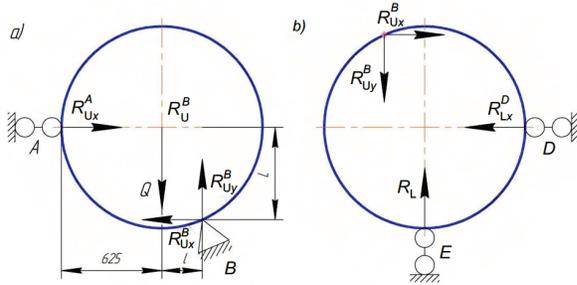


Figure 6 Design scheme for determining the thrust loads

Projection of forces on the Y -axis (Fig. 6a):

$$R_{Uy}^B - Q = 0, \quad (16)$$

$$R_{Uy}^B = Q = 1.425 \text{ tf}. \quad (17)$$

Moment of forces relative to the point B (Fig. 6a):

$$\sum M^B = Q \cdot \ell - R_{Ux}^A \cdot L = 0. \quad (18)$$

From which it follows:

$$R_{Ux}^A = \frac{Q\ell}{L}. \quad (19)$$

After substitution, we obtain $R_{Uy}^B = 1.425$ tf, $R_{Bx}^A = 0.628$ tf. Projection of forces on the X -axis (Fig. 6a):

$$R_{Ux}^B - R_{Ux}^A = 0. \quad (20)$$

From which:

$$R_{Ux}^B = R_{Ux}^A = 0.628 \text{ tf}. \quad (21)$$

Thus, the force N_{U1} (Fig. 5) is equal to the support reaction R_{Ux}^A and is 0.628 tf. The total thrust load N_U^C from two adjacent bundles will be equal to the sum of the forces N_{U1} and N_{U2} and will be 1.256 tf.

We will define the reaction R_{Lx}^E at point E by making a projection of forces on the X -axis (Fig. 6, b).

$$R_{Ux}^B - R_{Lx}^E = 0. \quad (22)$$

It follows that

$$R_{Lx}^E = R_{Ux}^B. \quad (23)$$

Having substituted the value, we obtain $R_{Lx}^E = 0.628$ tf.

Thus, the force N_{L2} (Fig. 5) is equal to the support reaction R_{Lx}^E and is 0.628 tf. The total thrust load N_L^C from two adjacent bundles will be equal to the sum of the forces N_{L1} and N_{L2} and will be 1.256 tf.

Thus, the total thrust load on the elements of the car side wall, in accordance with the bundle stowing scheme (Fig. 5), is 7.536 tf for the upper and lower tiers.

According to the bundle stowing scheme in the car (Fig. 3, b), the thrust loads from the upper tier act on one corner post and 3 side ones at the level of 1875 mm at a height of the upper strapping of 2050 ... 2360 mm, and from the lower tier at a height of 625 mm from the floor level. Permissible load on the side wall elements at the above load application heights is as follows [2]:

- side post at a height of 625 mm – (4.47) tf;
- side post at a height of body end plate – (1.2) tf;
- corner post at a height of 625 mm – (17.57) tf;
- corner post at a height of body end plate – (4.6) tf.

The total thrust load N_{UL}^C from the upper N_U^C and lower N_L^C tiers is 15.072 tf (Fig. 4).

The undampened transverse inertial force acting from the upper and lower tiers, according to the Eq. (15), is:

- upper tier $\Delta F_{Utr} = 8.514$ tf;
- lower tier $\Delta F_{Ltr} = 8.249$ tf.

The total overload from the thrust and inertial forces is determined by the equation:

$$N = N_{UL}^C + \Delta F_{Utr} + \Delta F_{Ltr}. \quad (24)$$

After substitution of values, we obtain $N = 31.835$ tf. Total permissible load on side wall elements:

$$[R] = 3 \cdot [R_{post}^U] + 3 \cdot [R_{post}^L] + 1 \cdot [R_{cor}^U] + 1 \cdot [R_{cor}^L], \quad (25)$$

where R_{post}^U , R_{post}^L – permissible loads on the side post, respectively, at the contact point with the upper and lower

bundle row; R_{cor}^U , R_{cor}^L – permissible loads on the corner post, respectively, at the contact point with the upper and lower bundle row.

The total permissible load on the side wall elements is $[R] = 39.18$ tf. The total loads acting on the elements of the car side wall are less than permissible, that is, 31.835 tf < 39.18 tf.

4 MODELLING RESULTS

Based on the mathematical model presented in [18, 19], a package of application programs was developed. Analytical modeling to determine specific inertial forces was carried out using a model of spatial vibrations of the car when moving along a curve with a radius of 200 m and an elevation of the outer rail of 150 mm. The maximum permissible speed for freight trains is 55 km/h. In accordance with [2], specific inertial forces are determined up to a speed of 100 km/h inclusive.

Specific vertical and transverse inertial forces at $H_{gc}^o = 2.296$ m are presented in Fig. 7 and Fig. 8.

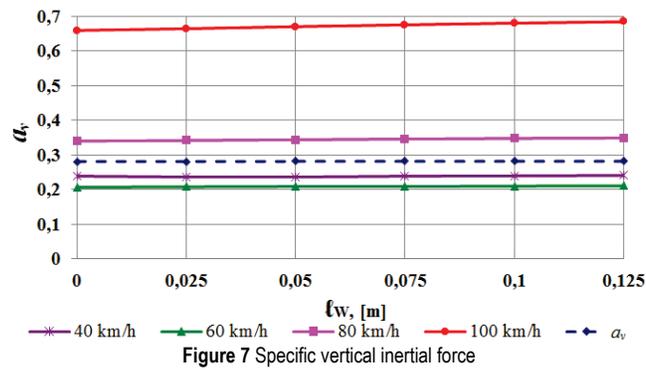


Figure 7 Specific vertical inertial force

From Fig. 7, it can be seen that with permissible longitudinal displacements of the general center of gravity of the cargo in the car, the values of the specific vertical inertial force practically do not change and depend to a much greater extent on the speed of movement. Value a_v , determined by Eq. (10), is, on average, 58.15% less than the calculated values for a speed of 100 km/h.

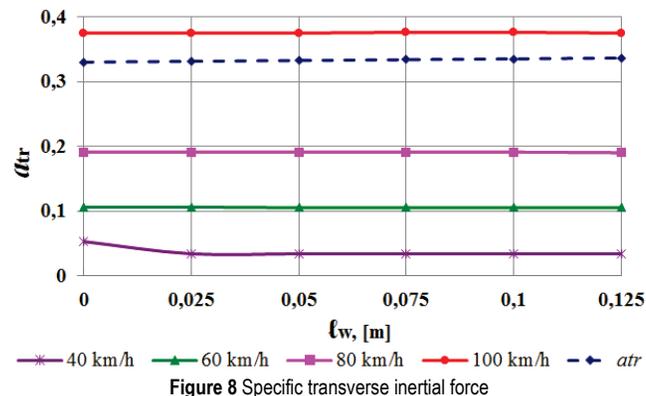


Figure 8 Specific transverse inertial force

The values of the specific transverse inertial force (Fig. 8) also depend to a much greater extent on the speed of movement. Value a_{tr} , determined by Eq. (8), is, on average, 11.23% less than the calculated values for a speed of 100 km/h.

To study the influence of the height of the general center of gravity of a loaded car on the values of specific inertial forces, calculations were performed at $H_{gc}^o = 1.8$ m. The specific vertical inertial force is presented in Fig. 9.

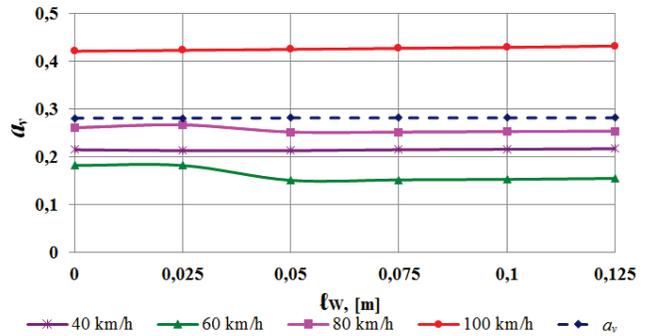


Figure 9 Specific vertical inertial force

Value a_v , determined by Eq. (10), is, on average, 33.92% less than the calculated values for a speed of 100 km/h.

The specific transverse inertial force is presented in Fig. 10. Value a_{tr} , determined by Eq. (8), is, on average, 4.17% less than the calculated values for a speed of 100 km/h.

Analyzing the results obtained, it is possible to conclude that the values of specific vertical and transverse inertial forces depend to a much greater extent on the speed of movement. Height H_{gc}^o and longitudinal displacement l_w , when freight trains move at maximum permissible speeds in curved sections of track with an undamped acceleration of 0.3 m/s², do not have a noticeable effect.

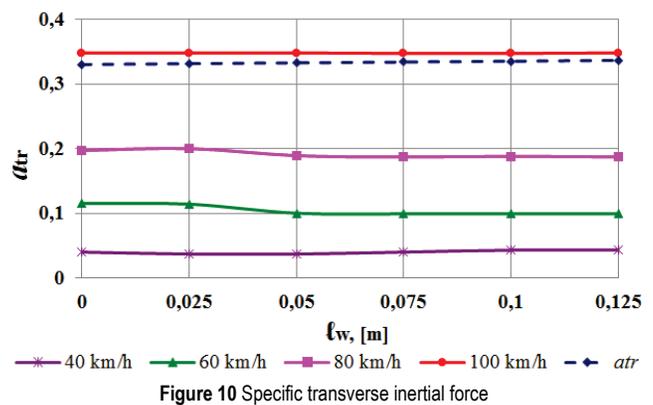


Figure 10 Specific transverse inertial force

On Figs. 7-10 present the simulation results in the presence of longitudinal displacements of the overall center of gravity of the load in the direction of the front bogie. When the CG is shifted towards the rear bogie, the values of the specific inertial forces are less than 1% of those presented. The direction of the longitudinal displacement along the

direction of travel of the car does not have a significant effect on the values of the specific inertial forces.

Thus, when the car moves in a curve with a radius of 200 m at the maximum permissible speed, the strength of the body elements will be ensured.

During theoretical studies, it was established that when a loaded car is moving at speeds of up to 100 km/h, the minimum values of the stability coefficient against wheel derailment are not lower than 1.2. The results obtained indicate that traffic safety is ensured on curved sections of the railway track.

In accordance with [2], for the cargo under study, a lateral displacement of the overall center of gravity of the cargo in a 4-axle car b_w is allowed when loading 70 mm and along the route (in transit) 100 mm.

In [16, 20], it was found that longitudinal displacements of the load's center of gravity have a much smaller effect on the dynamic forces arising during the movement of freight cars than transverse displacement or the presence of simultaneous displacement in both directions from the axis of symmetry of the rolling stock. The lateral displacement, defined by the regulatory documentation, should be taken into account since it leads to a significant change in the safety factor against wheel derailment.

The results obtained show the feasibility of further research in this direction.

5 CONCLUSION

The work determines the loads acting on the elements of the side wall of the gondola car body during the transportation of rolled wire bundles. In the calculations, the total weight of the cargo in the car was assumed to be 69 tons, while the transportation experience of these products shows that the bundle weight in the car usually does not exceed 66 tons. Total load from the thrust and inertial forces acting on the car side wall elements during the bundle transportation according to the scheme presented in the article is 31.835 tf, which does not exceed the permissible load on the side wall elements of the gondola car (three side posts of the body and the corner post), taking into account the height of the forces application – 39.18 tf.

The presented results confirm that the transport accident on the Voskoboynya – Verkhnedneprovsk section occurred precisely because of the unsatisfactory technical condition of the car, but it is not the reason for the disadvantages in the LTC for the transportation of rolled wire bundles.

The correspondence of the results obtained to the real object is ensured by the fact that the inertial forces are calculated in accordance with the approved methodology. To determine the thrust forces from the action of bundles of the upper and lower tiers, in accordance with the calculation schemes presented in the work, the force balance equations were written and solved.

6 REFERENCES

- [1] Myamlin, S. (2002). *Simulation of railway vehicles dynamics*. Dnipropetrovsk: New ideology.
- [2] Ukrainian Railways. (2024, March 15). Organization for Cooperation between Railways (2019) Technical Conditions for Stowing and Securing Cargo: Appendix 3 to the Agreement on International Freight Traffic by Rail (SMGS). https://www.uz.gov.ua/cargo_transportation/legal_documents/smgs/dod3_01072019/
- [3] Ukraine: Center for Transport Strategies. (2024, March 15). On the Pridneprovskaya railway train lost its cargo and got off three cars (updated). https://cfts.org.ua/news/2020/07/20/na_pridneprovskoy_zh_d_poezd_poteryal_gruz_i_soshlo_tri_vagona_foto_59780
- [4] Hromádka, V., Korytářová, J., Vítková, E., Seelmann, H. & Funk, T. (2023). Benefits of increased railway safety and reliability and their evaluation. *Tehnicki glasnik*, 17(3), 440-446. <https://doi.org/10.31803/tg-20230504194331>
- [5] Shaposhnyk, V. Y. (2018). Human factor influence on performing technical maintenance and repair of freight cars. *Science and Transport Progress*, 6(78), 165-175. <https://doi.org/10.15802/stp2018/154031>
- [6] Reidemeister, A., Muradian, L., Shaposhnyk, V., Shykunov, O., Kyryl'chuk, O. & Kalashnyk, V. (2020). Improvement of the open wagon for cargoes which imply loading with a "hat". *IOP Conference Series: Materials Science and Engineering*, 985, Paper 012034. <https://doi.org/10.1088/1757-899x/985/1/012034>
- [7] Shvets', A. O. (2022). Investigation of coupling strength at non-central interaction of railcars. *Strength Mater.* 54(2), 233-242. <https://doi.org/10.1007/s11223-022-00396-1.7>
- [8] Shaposhnyk, V., Shykunov, O., Reidemeister, A., Muradian, L. & Potapenko, O. (2021). Determining the possibility of using removable equipment for transporting 20- and 40-foot-long containers on a universal platform wagon. *Eastern-European Journal of Enterprise Technologies*, 17(109), 14-21. <https://doi.org/10.15587/1729-4061.2021.225090>
- [9] Anofriev, V. G., Reidemeister, O. G., Shykunov, O.A., Kirilchuk, O. A. & Matsyuk, A. S. (2012). Definition of the basic technical requirements to metal rolling for bodies of freight cars of new generation. *Wagon fleet*, 11(68), 4-5.
- [10] Reidemeister, O. G., Kalashnik, V. O., Rizhov, S. V. & Shykunov, O. A. (2019). Experimental determination of fatigue resistance of gondola body units. *Collection of scientific works of UkrDUZT*, 184(Appendix), 19-20.
- [11] Shvets, A. O. (2021). Dynamic indicators influencing design solution for modernization of the freight rolling stock. *FME Transactions*, 49(3), 673-683. <https://doi.org/10.5937/fme2103673S>
- [12] Zamecnik, J. & Jagelcak, J. (2015). Evaluation of wagon impact tests by various measuring equipment and influence of impacts on cargo stability. *Communications-Scientific letters of the University of Zilina*, 17(4), 21-27.
- [13] Olentsevich, V. A., Upyr, R. Y. & Gladkih, A. M. 2020. Computational procedure for preparing the technical conditions for stowage and securing cargo in rail cars and containers. *Journal of Physics: Conference Series*, 1615(1), Paper 12029.
- [14] Vlasova, N. V., Olentsevich, V. A., Konyukhov, V. Y. & Lysenko, D. A. (2021). Automated calculation method effect values in load securing elements fixed on a rolling stock. *IOP Conference Series: Materials Science and Engineering*, 1064(1), Paper 012042.
- [15] Zhang, D., Tang, Y., Clarke, D. B., Peng, Q. & Dong, C. (2021). An innovative method for calculating diagonal lashing force of cargo on railway wagons in a curve alignment. *Vehicle System Dynamics*, 59(3), 352-374. <https://doi.org/10.1080/00423114.2019.1686160>

- [16] Shatunov, O. V. & Shvets, A. O. (2019). Study of dynamic indicators of flat wagon with load centre shift. *Science and Transport Progress*, 2(80), 127-143.
<https://doi.org/10.15802/stp2019/165160>
- [17] Yaremenko, D. (2021). *Analysis of the forces acting on the body of the gondola car during the transportation of rolled wire coils*. Dnipro: Ukr. State University of Science and Technology.
- [18] Shvets, A. (2023). Multibody model of freight railcars interaction in a train. In: Sharma, S. K., Upadhyay, R. K., Kumar, V. & Valera, H. (eds) *Transportation Energy and Dynamics. Energy, Environment, and Sustainability* (pp. 217-241). Singapore: Springer.
https://doi.org/10.1007/978-981-99-2150-8_10
- [19] Muradian, L., Shvets, A. & Shvets, A. (2023). Some dynamic processes at longitudinally-transverse shift of the cargo. *Scientific Journal of Silesian University of Technology. Series Transport*, 120, 187-204.
<https://doi.org/10.20858/sjsutst.2023.120.12>
- [20] Shvets, A. O. (2019). Influence of the longitudinal and transverse displacement of the cargo gravity center in gondola cars on their dynamic indicators. *Science and Transport Progress*, 5(77), 115-128.
<https://doi.org/10.15802/stp2018/146432>

Authors' contacts:

Angela Shvets, MS, ME

(Corresponding author)

Ukrainian State University of Science and Technologies,
 Lazaryana St., 2, 49010 Dnipro, Ukraine
angela_shvets@ua.fm

Vladyslav Shaposhnyk, PhD

Ukrainian State University of Science and Technologies,
 Lazaryana St., 2, 49010 Dnipro, Ukraine
vladislav.sh91@gmail.com

Leontii Muradian, Professor, Doctor of Technical Sciences

Ukrainian State University of Science and Technologies,
 Lazaryana St., 2, 49010 Dnipro, Ukraine
leontymuradian@gmail.com

Anzhelika Shvets, MS

Ukrainian State University of Science and Technologies,
 Lazaryana St., 2, 49010 Dnipro, Ukraine
angtiger.am@gmail.com

Alexei Reidemeister, PhD

Ukrainian State University of Science and Technologies,
 Lazaryana St., 2, 49010 Dnipro, Ukraine
reidemeister.a@gmail.com

Oleksandr Shykunov, PhD

Ukrainian State University of Science and Technologies,
 Lazaryana St., 2, 49010 Dnipro, Ukraine
tri_s@ua.fm

Dmytro Molchanov, MS

Ukrainian State University of Science and Technologies,
 Lazaryana St., 2, 49010 Dnipro, Ukraine
molchanovfs@gmail.com

Fuelling Logistics from Refinery to Retail: Case Study of the Romanian Oil Supply Chain

Attila Turi

Abstract: The purpose of this paper was to examine how EU sanctions on Russia impacted Romania and how its refineries managed to continue operating with alternative suppliers. Production performance dynamics and their share in total output of refined petroleum products in the last 5 years (2019-2023) are analyzed using descriptive statistics. Own empirical research is used to study the supply chains of the Petrobrazi, Petrotel and Petromidia refineries and provide an overview of their role on the domestic fuel market by mapping out their fuel storage locations and network of filling stations. Findings underpin the importance of an interconnected infrastructure network (pipelines, railways and ports) and using it to its full extent to balance out onshore and offshore imports, as well as downstream deliveries. Study replicability is affected by the limited extent to which the collected data can be revealed, whilst future research can study sustainability of non-Russian oil sourcing alternatives.

Keywords: fuel distribution; fuel storage network; intermodal transportation; pipeline transportation; oil supply chain; refinery logistics

1 INTRODUCTION

Romania is the EU's third oil producer (after Italy and Denmark) and ranks fifth in Europe, Norway and the United Kingdom being the largest, according to Eurostat (2023). It also ranks in the top 10 producers of petroleum products in the EU with yearly refined volumes of over 10 million tonnes (2023) and has one of the lowest (68%, 2024) oil import dependency rates [1, 2], compared to the average 92% at EU level [3] and even less of Russian crude oil sourcing (18%, 2022), according to the International Energy Agency. Its other main import countries are Kazakhstan (KMG), Azerbaijan and Turkey (SOCAR), several other EU countries (mainly Bulgaria, Hungary and Austria) and Serbia [4] for refined oil products (Trading Economics, 2023 report). Oil products have a 5.6% share in Romania's exports and an overall 6.5% share in GDP (National Institute of Statistics, 2023 report), making it one of the most important drivers of economic growth in addition to being one of the country's strategic resources [5].

The 3 operational refineries ensure 32.1% of its domestic oil consumption needs (2022) and are owned by multinational oil companies from Austria, Russia and Kazakhstan (OMV, Lukoil and KazMunayGas). Prior to the invasion of Ukraine (2022), Novorossiysk's onshore terminal supplied Russian crude oil to all Romanian refineries (Lukoil's Petrotel having a higher share), whilst its offshore terminal still supplies Kazakh crude oil through the Caspian Pipeline Consortium (CPC), mainly to KMG's Midia Marine Terminal [6, 7]. Romania's fuel supply is uneven: a net exporter of gasoline and a net importer of diesel fuel due to its predominant domestic diesel consumption. In 2021, 30% of diesel fuels were imported from Russia, the majority of which (up to 90-95%) via Constanta's Oil Terminal [8-10].

Purchasing, importing and transferring seaborne crude oil (since December 2022) and certain petroleum products (since February 2023) from Russia to the EU are prohibited [11, 12]. Some temporary exemptions were made for landlocked (Hungary, Slovakia) or highly dependent countries (Germany), with others already having found alternative sourcing options (Bulgaria, Poland, Czech Republic) to fully align with the EU's restrictive policy [13-

15]. A deal (1 year contract) for transport and storage of 2.9 million tonnes of crude oil was signed (May 2023) to supply MOL's refineries in Hungary and Slovakia via the Croatian pipeline, based on rigorous feasibility, technical and operational analyses [16, 17]. Hungary is still opposing shifting away from (cheaper) Russian oil imports and sourcing from Croatia via the Adria pipeline (operated by JANAF), claiming logistics costs are too high [18] and transport capacities from Omisalj are too low for its needs, despite reassurance from its southwest neighbor [19]. Slovakia has a similar position (Slovnaft), whilst Croatia has little leverage since INA is also owned by Hungarian MOL. Overall, Russia's share in extra-EU imports of oil products plummeted from 24.8% (Q4, 2021) to 1.2% (Q2, 2024) after its invasion of Ukraine and subsequent EU sanctions, according to recent data released by Eurostat.

By leveraging their distribution channels [20, 21], domestically refined fuels (Petrom, OMV [22], Lukoil and Rompetrol), in addition to import brands (MOL, SOCAR and Gazprom) and independent stations ensure the supply of fuel products [23] to their network of filling stations across Romania [24].

This article is organized in the following sections: Chapter 1 introduces background information and data on Romania's oil refining industry, its share of crude oil imports from Russia and provides EU neighboring context. Chapter 2 presents a synthesis of related scientific articles, outlining their contributions and identifying the gap this study intends to fill. Chapter 3 provides the research design, including a description of the data collection methods and analytical techniques utilized in this study. Chapter 4 reveals and summarizes the main research findings, followed by their interpretation and implications for future research in Chapter 5. Finally, Chapter 6 concludes the paper by summarizing its key findings, expressing their importance, limitations and initiates an outlook for further exploration of the topic.

2 LITERATURE REVIEW

The existing literature on oil refining and its logistics primarily focuses on macro-level analyses of fuel supply chains, with limited attention to the micro-level challenges

and optimization opportunities within specific regions or even countries, as is the case of Romania in this article. The main reason for this research gap is the lack of accessible and/or aggregated data, as well as challenges regarding publishing due to high data sensitivity and confidentiality.

Seeking correlations between oil production and an array of other economic data provides mixed results, as is noted by [25]. Nevertheless refineries do have a strong influence on a country's economic activity, both direct (technical support, maintenance work) and indirect (transportation, logistics), despite lower returns to scale. This is also confirmed by [26], who show that oil refining is positively correlated with a country's global competitiveness index (GCI) and despite price volatility affecting cost competitiveness, a net oil exporter will bear less risk than a net oil importer. The closer oil refining capacities are to their demand market, the more cost-efficient they are and the more effective they are in reducing the impact of price fluctuations [27].

Optimization of fuel distribution networks is another recurring theme and mostly relies on analyzing, modeling and conducting matching scenarios between supply and demand with direct or indirect cost reduction-related targets.

Deterministic approaches such as the one undertaken by [28] highlight the complexities of managing a network of ships, tanks, pipelines to supply crude oil to refineries. More refined models provide more feasible decisions, but also increase the expected cost and are still subject to improvement. Supply chain management flow strategies are most often studied in countries where oil refining is an underlying economic sector, as proven by [29]'s approach which sources good solutions in reasonable time spans for both cost and time-related parameters. The narrow margins of refineries are balanced out by consistency in delivering volumes and maximizing oil trade decision-making, as shown for Europe's leading independent crude oil facilities (Petroineos) in Lavéra (France), Grangemouth (Scotland) and the largest European refinery site (Shell) in Pernis, The Netherlands by [30].

While these approaches require a more in-depth analysis of a specific regional setting, they still have limited practical applicability. This is due to the typical high sensitivity and reactivity of the oil market to various external factors (production levels, geopolitical events, economic sanctions, OPEC decisions, natural disasters), which source a rather unpredictable, volatile and dynamic pricing pattern, which alters supply decisions.

Other authors [31] outline decision-making challenges of the real world as being ambiguous or vague. Their proposed model employs a pentagonal fuzzy number to minimize the cost along the refineries' supply chain (depots, multimodal transportation and demand nodes), provides improved larger-scale results and reduces the optimality gap, despite uncertainty being factored in.

Therefore while simulating routes is nevertheless useful to analyze and redesign certain supply choices, its rather static assumptions and constraints modeled on historical data might become suboptimal, unrealistic or even impossible in certain real life conditions.

Inconsistent use and implementation of EDI systems across filling stations, with mostly short loop communication systems, even in the case of multinational brands' supply

chains alters operations management effectiveness and hinders appropriate fast-response logistics, as also outlined by [32]. Significant differences also exist between individual filling stations' supply chains, especially in case of a lower concentration of fuel providers within that specific region.

This article's role is to contribute to the research field by providing a granular analysis of the Romanian fuel supply chain, considering local production, depot storage locations, distribution system and its mix of alternative sourcing locations within the specific logistical challenges of the Balkans. Previous research has not explored the impact of redesigning upstream crude oil supply on country basis in the context of Russian sanctions, a gap this study aims to fill. Effective decision-making relies on high-quality data: real-time, accurate and consistent data enables a more focused analysis, which in turn leads to a greater impact on overall productivity and operational performance.

3 MATERIALS AND METHODS

The current research paper uses primary and secondary data about the oil industry in Romania (production, capacity utilization, alternative sourcing options, storage facilities, distribution network and intermodal logistics systems), its main multinational competitors and their domestic market shares.

Raw and primary data is sourced through own empirical research by conducting on-site fieldwork and observational study (at refineries, port and intermodal terminals). Certain parts of gathered data (mainly from the refineries and port terminal operators) are subject to consented NDAs' strict terms, prior to gaining access to specific operational, financial, customer and supplier data. Secondary data (statistics databases, industry reports and official releases) is cross-referenced with each refinery's and intermodal terminal's records, with some marginal data cleaning being required to eliminate inconsistencies and validate findings. The paper's collated data is consented for publication and adheres to all contracted stipulations of the NDAs.

Descriptive statistics is applied to investigate the evolution of the 3 refineries' time-series data (2019-2023) on operational performance metrics. Year-to-year percentage variations and relative share in total output of refined fuels are analyzed to contextualize production dynamics. Qualitative and quantitative methods are used to exhibit relevance of empirical data by mapping and compiling fuel storage locations, the distribution network and their connectivity to main intermodal transportation terminals, thus offering a holistic approach to the case study.

4 RESULTS' ANALYSIS

Romania's 3 operational refineries (Petrobrazi, Petrotel and Petromidia) are owned by multinational oil companies: OMV (Austria), Lukoil (Russian Federation) and KMG International/KazMunayGas (Kazakhstan). Over 80% of total output is provided by the Petrobrazi (OMV) and Petromidia (KMG) facilities, the latter being Romania's most modern refinery with the highest capacity utilization rate (CUR), as shown in Tab. 1. Petrotel's productivity has been decreasing in recent years (by 45% since 2019), despite

Lukoil having the closest alternative refinery (Neftohim Burgas) in Bulgaria. In view of the economic context after invading the Ukraine, Russia's Gazprom transferred its operations to Serbian multinational company Naftna Industrija Srbije (NIS) Petrol, thus legally avoiding sanctions imposed on Russian oil imports since Serbia is a non-EU member state. Lukoil uses a similar approach, importing its crude oil through intermediate companies outside of the Russian Federation (e.g. Turkey or via pipelines) and then using its Petrotel processing facility in Ploiesti (Romania) to refine it, thus also avoiding sanctions. OMV and MOL also use this loophole, in addition to other major European companies (BP, Shell, TotalEnergies or ENI). Tab. 1 presents an overview of oil companies' production volumes, Capacity Utilization Rate (CUR), year-to-year variations (2019-2023) and share of total refined petroleum products in Romania.

Table 1 Romania – main refineries' performance metrics

Refinery	KPI	2019	2020	2021	2022	2023
OMV Petrobrazii Brazi (South)	Output (MTPA)	4.54	4.05	3.83	4.56	3.84
	CUR (%)	0.94	0.84	0.79	0.95	0.80
	Dynamic (%)	1.02	0.89	0.94	1.19	0.84
	Share (%)	0.37	0.39	0.38	0.40	0.37
Lukoil Petrotel Ploiesti (South)	Output (MTPA)	2.49	1.88	1.93	2.22	1.37
	CUR (%)	0.92	0.69	0.71	0.82	0.50
	Dynamic (%)	1.03	0.75	1.02	1.15	0.61
	Share (%)	0.20	0.18	0.19	0.19	0.13
KMG Petromidia Navodari (East)	Output (MTPA)	4.97	4.29	4.16	4.51	5.01
	CUR (%)	0.97	0.83	0.81	0.88	0.97
	Dynamic (%)	1.03	0.86	0.97	1.08	1.11
	Share (%)	0.41	0.42	0.41	0.39	0.49
Total	Output (MTPA)	12.00	10.22	9.92	11.29	10.22
	CUR (%)	0.95	0.81	0.78	0.89	0.81
	Dynamic (%)	1.03	0.85	0.97	1.13	0.90

OMV-owned Petrobrazii refinery (sourcing Petrom and OMV fuel brands) covers almost 40% of Romania's total refining capacity, with a record output of 4.56 million tonnes of petroleum products, 19% year-on-year increase and 95% CUR (2021). After buying the refinery from Petrom (51.1% of shares since 2004), OMV has invested over 2 billion euros in modernization projects at its Petrobrazii location in South Romania. Planned shutdowns for routine maintenance and equipment upgrades of the country's oldest operating refinery (2021, 2023) and the first pandemic year (2020) caused its CUR indicator to fall below 90%. Its 2023 planned shutdown (for environmental compliance) reduced its output by 0.72 million tonnes (16% compared to 2022) and CUR to only 80%.

In 2019, Lukoil boasted sourcing Romania's highest gasoline quality at its Petrotel refinery, according to some independent reports. Lukoil owns 99.76% of Petrotel's shares (since 1998) through Litasco, its own oil trading company based in Switzerland, and invested almost 500 million euros in modernization projects at its Ploiesti refining

unit (2004-2015). With a share of only 20% of the country's refined petroleum products and a focus on competitive pricing, its investments slowly decreased, as did its CUR metric (around 70% in 2019 and 2020). EU sanctions against imports from the Russian Federation after invading Ukraine, combined with reduced alternative crude oil sourcing options further curtailed its refining output to around 50% of its total capacity (yield of 1.37 million tonnes), in 2023. Tax evasion lawsuits and undisclosed operations (crude oil sourcing, processed amount and downstream distribution) lead to Lukoil losing its trading license (temporarily), shortly after Petrotel suspended its operations for overhaul (November 2023).

Last year, KMG's (KazMunayGas) Petromidia refinery (owned with 54.63% of shares), produced a record 5.01 million tonnes of petroleum products (Rompertol fuel brands), an 11% year-on-year increase and 49% overall domestic refining share (2023) supported by its 97% CUR. KMG also owns Vega, a smaller refinery (capacity: 0.4 million tonnes/year and 98.5% CUR/year) in Ploiesti, that operates in an integrated system with its main refinery from Navodari. A fire interrupted activity for 83 days (July 2021) before repairs allowed regular operations to restart in September 2021, whereas refining capacity was fully restored only in April 2022. Petromidia produces Romania's highest amount of petroleum products, with one exception: 2 months of general overhaul reduced its CUR (88%) and output (4.51 million tonnes) in 2022, being overtaken by its competitor, OMV Petrom (4.56 million tonnes).

In 2019 Romania's refineries had a combined output of 12 million tonnes of petrol and diesel fuels, before 2 consecutive years of dwindle caused by the Covid-19 pandemic. Yearly volumes of petrol and diesel fuels declined by 1.78 million tonnes (15% in 2020) and a further 0.3 million tonnes (3% in 2021), before recovering in 2022 (11.29 million tonnes). 2023 marked another 10% descend (to 10.22 million tonnes), mostly due to Petrotel's ownership issues: 0.85 million tonnes less yield (39% year-on-year decrease) and OMV Petrom's planned general turnaround: 0.72 million tonnes less refined products (16% year-on-year decrease). Around 60% of refined products (petrol and diesel fuel) are sourced locally by the refineries: Petrobrazii (66%), Petromidia (58%) and Lukoil (50%). The rest is mainly imported from Hungary (MOL), Bulgaria (Lukoil), Serbia (NIS) and offshore facilities through tankers via the port of Constanta and then distributed locally (storage facilities and gas stations). Crude oil and refined petroleum products (petrol and diesel fuel) are also sourced from Asia (30%) and the United States (20%), but 80% of imports from the Caspian Sea (KMG and SOCAR) pass through Russia's pipelines before being shipped to Romania's Black Sea gateway. Imported oil products come at higher prices due to extra shipping and insurance costs and arrive at Constanta Port's Oil Terminal. MOL (Hungary) is the most important import brand, but its main refinery near Budapest is also an important supplier for the West of Romania, in addition to its controlled facilities in Slovakia (Slovnaft) and Croatia (INA).

A map of Romania's refineries, their domestic fuel depots and main alternative sourcing (import) locations is presented in Fig. 1.



Figure 1 Romania – domestic refineries' production units, storage and alternative sourcing locations (2024)

MOL's Danube refinery in Szazhalombatta (Hungary) is one of the 10 largest in Central and Eastern Europe (CEE) with its 165,000 barrels/day and/or 8.1 million tonnes/year refining capacity. MOL also owns neighboring Slovnaft (Bratislava, 6.1 million tonnes/year) and INA (Rijeka, 4.5 million tonnes/year) refineries. In the CEE region, Austria (Schwechat, OMV), Poland (Plock and Gdansk, Orlen) and Germany (7 refineries) have higher capacity refineries. MOL fuels are distributed towards Romania via the Danube River (between Regensburg and Giurgiu ports), railway (rail tank cars) and/or road (tankers), the latter being mostly used to supply the Western region (Timisoara, Arad, Oradea).

Bulgaria's Lukoil Neftohim Burgas refinery is only 250 km away from the border with Romania, being the closest neighboring unit from the south of the country, where all Romanian refineries are located. As of 2024 the refinery has ceased receiving Russian-produced oil, tax rate increased to 60% and Lukoil's concession contract to run the Rosenets Oil Terminal near the Black Sea port of Burgas was terminated. Alternative sourcing is currently supplied to the refinery from Kazakhstan, Iraq and Tunisia and new ownership is expected.

Serbia's NIS Petrol (owned by Gazprom from Russia) operates the largest network of filling stations (324 units) in Serbia (total of 1,529 units) and runs its 2 oil refineries, with a combined capacity of 7.4 million tonnes/year in the province of Vojvodina (Novi Sad and Pancevo), both within 100 km of Romania's Western border crossings. Gazprom Neft transferred some of its shares to avoid EU sanctions (May 2022), but the supply of Russian crude oil through the Adria oil pipeline from Omisalj's Oil Terminal (Croatia) ceased 6 months later. NIS Petrol mainly sources its crude oil imports from Iraq (53%), Russian Federation (46%) and Romania (as of 2022).

The Schwechat refinery (Austria) is one of Europe's top 10 units by capacity (9.6 million tonnes/year), accounts for 52% of OMV's refined output and is located approximately 450 km from the Romanian border. The Austrian refinery near Vienna is supplied by AWP (Adria-Wien-Pipeline) from the oil terminal in Trieste (Italy), being a branch connection of the Transalpine Oil Pipeline (TAL) which connects Italy, Austria and Germany. Despite important ties in the energy sector, OMV decided to cease all further investments in Russia (March 2022) and currently sources its crude oil needs

from the nearby CEE area, North Sea (mainly Norway) and North Africa regions (mainly through Algeria's port of Bejaia). Between 10-15 tankers (600,000 barrels/tanker) arrive every month at the port of Trieste, after a journey of 4-5 days, and are unloaded in 24 hours before being sent to Schwechat through the AWP and TAL pipeline systems (throughout time: 48-72 hours).

Romp petrol is KMG's fuel brand in Romania, but its distribution extends to its neighboring Black Sea countries (Bulgaria, Moldova, Turkey, Georgia) and Western Europe (France, Spain) under the same or different brands (dyneff). Since KMG took over Petromidia (2007), in addition to its local sourcing, up to 60-75% of total processed crude oil was imported from Kazakhstan via its offshore terminal.

Table 2 Romania – Overview of fuel distribution sourcing and storage units (2024)

Brand(s)	Main refinery Location	Domestic capacity	Distribution stations	Fuel depots
OMV (Petrom)	Petrobrazi (Brazi, RO)	4.8/18.2 (RO) (26.37%)	586 units (25.57%)	6 units (13.95%)
Lukoil	Petrotel (Ploiesti, RO)	2.7/64.5 (RO) (4.18%)	319 units (13.92%)	7 units (16.27%)
KazMunayGas (Romp petrol)	Petromidia (Navodari, RO)	5.1/10.6 (RO) (48.11%)	401 units (17.50%)	6 units (13.95%)
MOL (Slovnaft) (INA)	Szazhalombatta (Hungary)	0/18.7 (HU) (imported)	238 units (10.38%)	2 units (4.65%)
SOCAR	Aliaga (Turkey)	0/16.4 (AZ) (imported)	75 units (3.27%)	4 units (9.30%)
Gazprom (NIS Petrol)	Pancevo (Serbia)	0/7.4 (RS) (imported)	19 units (0.82%)	N/A
independent gas stations	petroleum wholesalers	0/12.6 (RO) (sourced)	653 units (28.50%)	18 units (41.86%)
7 brands and independents	3 domestic and 5 adjacent (A, HU, RS, BG)	12.6/135.8 (RO) (9.27%)	2291 stations (100%)	43 units (100%)

Over 70% of Romania's gas stations (2,291 total units, as of 2024) are owned by the brands listed in Tab. 2. More than half (57%) are owned by the 3 brands operating refineries in the South-East of the country (1,306 units) while the number of independent gas stations (653 units) is less than a third (28%). An outline of oil companies' and independent brands' distribution networks (sourcing, storage and stations) is presented in Tab. 2, whilst brand ownership of all 43 fuel depots in Romania is mapped out in Fig. 3.

Petrobrazi accounts for over a quarter (26.37%) of OMV's total refining capacities and supplies petroleum products for both its fuel brands (Petrom, OMV) in Romania. OMV also covers more than a quarter (25.25%) of all Romanian filling stations with its 586 units, around 30% being high-end ones (OMV brand).

Lukoil's refining capacity at Petrotel represents less than 5% of the Russian group's output volumes, whilst its focus on competitive pricing and recent EU sanctions will further add challenges to its operations in Ploiesti. Despite having the highest number of fuel depots (7) across the country, Lukoil's network of petrol stations has been steadily

decreasing since 2019 and now only adds up to 319 units (13.92% of total).

The Petromidia refinery in Navodari has an almost 50% share of KazMunayGas' overall refined output capacity and distributes its fuels in Romania under the Rompetrol brand. Rompetrol's network of filling stations (401 units) is the second largest and its fuel depot locations are the most balanced, being distributed among all regions in Romania.

MOL is Romania's third most important fuel brand (after OMV and Rompetrol), despite it having to import its petrol and diesel fuels, with the fourth largest domestic distribution network (238 filling stations and/or 10% of total) and only 2 fuel depots.

Other brands that import their petroleum products are SOCAR with 4 fuel depots and 75 filling stations (less than 5% of total) and Gazprom with less than 1% of the total fuel distribution network (only 19 stations and no fuel depots).

Independent gas stations are supplied by the petroleum wholesalers' distribution network, have seen their overall share being reduced in the last decade, but still account for almost 30% of all filling stations (653 units) and 40% of fuel depots (18 units) in Romania, as of 2023. OSCAR Downstream is the largest independent Romanian fuel distribution company and ranked in the top 5 players on the downstream market (after OMV, KMG, MOL and Lukoil). 7 fuel depots (38% share of all independent units) and 115 stations (17% share of all independent filling stations) are part of OSCAR's distribution network.

Smart Diesel (SD) is another important independent fuel company in Romania and has partnered with DKV Mobility (since 2021). The SD Group operates 36 own gas stations in Romania (5% of all independents), its own SD card (more than 3,000 active customer companies) is accepted in around 2,000 partner gas stations abroad (2023), offering bulk fuel, toll payment solutions, VAT recovery services, telematics and roadside assistance for trucks to over 13,000 B2B customers.

The Black Sea handles oil imports from the Caspian Sea (Azerbaijan and Kazakhstan) through its dedicated terminals (Oil Terminal, Midia Marine Terminal) and port pipe system (50 km) in Constanta, as shown in Fig. 2.

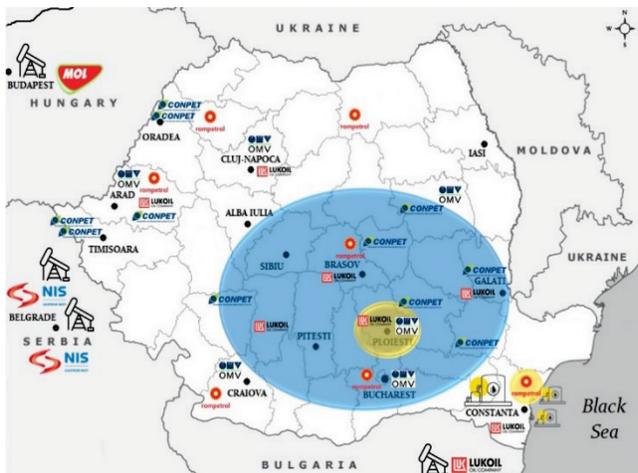


Figure 2 Romania – map of national pipeline coverage, Black Sea terminals (2024)

All Black Sea oil terminals (Oil, JT and MMT) are connected to the 3,809 km of national pipeline network, operated by Conpet across 24/41 counties, including the 3 Romanian refineries. Oil tankers dock at the specialized terminals where they unload the crude oil into storage tanks through large pipes. Constanta's Oil Terminal handles import of crude oil and gasoline, as well as export of refined and/or unrefined petroleum products and other liquid chemical products. Crude oil is pumped from storage tanks through underground pipelines to the 3 refineries in Romania or to one of 24 other export terminals (of which 11 are Conpet's own fuel depots) across the country. Almost 70% of all fuel depots (30/43) are owned by Conpet or the 3 domestic refineries. Petroleum products refined from crude oil are distributed in Romania through gas station networks and other industrial users.

Conpet's allocated storage capacity for domestic crude oil is 3 times higher than for imported crude oil. The crude oil national transport pipeline system's throughput is 200 million barrels/year (or 27.43 million tonnes/year), around 75% is used for imported oil, but is currently only used at a third of its capacity, 37% as of 2023 (in 1990 there were 10 operational refineries). Some of the refined petroleum products are also exported from Romania through pipelines, tankers and other means of transport.

Oil Terminal is the largest operator of petroleum products in the Port of Constanta, where it owns a developed network of high-capacity pipelines and tanks (storage capacity: 10 million barrels). Its terminal in the port operates 7 berths where it also stores over 70% of its capacities, the rest is deposited in an underground central storage facility in the center of Constanta. Oil Terminal inaugurated a new storage tank for crude oil and/or diesel (capacity: 55,000 cbm) in its terminal at the Constanta Port (January 2024), the largest built after 1989.

Midia Marine Terminal (MMT) is an offshore marine terminal supplying a tank farm (capacity: 390,000 cbm) and the Petromidia refinery through pipelines (onshore and offshore), located 8.6 km East off the Black Sea. MMT is operated by Rompetrol Logistics (KMG International Group), with an unloading rate of more than 600,000 barrels/day, and was built to reduce distance (by 30 km) travelled by tankers (60,000-165,000 DWT) to unload oil at Constanta Port. 32 million barrels (or 4.37 million tonnes) of crude oil, mainly from Kazakhstan, were unloaded at the MMT offshore terminal from 53 vessels (2023). Rompetrol also operates 7/14 berths at Midia's onshore oil and gas terminals, but this port's depth only allows smaller ships (10,000-20,000 DWT) to enter. Larger vessels (165,000 DWT tankers) could have been unloaded only in Constanta Port, an alternative meaning oil reached the refinery through long pipelines and with various losses along the way. On average, 2 million tonnes of additional oil products are imported and exported by 550 KMG tankers in Midia Port every year.

The newest and most modern terminal for liquid petroleum products (duty-free diesel fuel) in the Port of Constanta is operated by JT Terminal. JT Terminal offers storage and handling spaces for specific products (diesel,

biodiesel, bitumen and other liquid products) prior to their distribution on the market (domestic or export). JT Grup Oil is also a regional fuel distributor for the South-East of Romania with its own fleet (11 tankers with a 40,000 liter/tanker capacity) and more than 250 storage tanks across its distribution network. The terminal has 8 storage tanks (total capacity: 33,000 tonnes) and operates barges with different capacities (from 3,000-9,000 tonnes), trains (up to 16 wagons at the same time on 2 parallel railway lines) and tankers (up to 20 at a time in its dedicated loading/unloading area). JT Terminal's handling capacity is 700-1,000 tonnes/hour (for ships), 300-700 tonnes/hour (for trains) and 40-80 tonnes/hour (for tankers).



Figure 3 Romania – map of oil companies' storage and distribution centers (2024)

An overview of Romania's 43 fuel depots' regional distribution (per brand) is highlighted in Fig. 3, with additional details (exact locations and delivery alternatives) being provided in Tab. 3.

Conpet's national pipeline system connects to 11 fuel depots (25% of all Romanian fuel depots), half in the West and the other half around the refineries in the South-East, with free access being leased to all domestic-operating oil companies. Lukoil and Oscar Downstream each own 7 of Romania's 43 fuel depots (and 32% combined), with an almost identical spread across the country, but different fuel sourcing: Lukoil's throughput is mostly domestic (Petrotel refinery), whilst Oscar Downstream only handles imported fuels. OMV and KMG own over 25% of fuel depots (6 units each) with similar regional distribution, while Rompetrol's fuel depot in Vatra Dornei is the only one in the Northern area. Since SOCAR imports its petroleum products from Azerbaijan through the Black Sea, its fuel depot network (4 units) is scattered in the South-East, with Hungarian MOL locating its units in the West (by rail and/or road) and South (by Danube River, through the Giurgiu port).

Romania's precise fuel storage locations, their regional coverage and network access is provided in Tab. 3.

Romania's branded filling stations (multinational oil companies' petrol and diesel fuels) are supplied by a total of 18 main locations: 3 domestic refineries (16%), 8 nearby onshore (44%) and 7 offshore locations (38%). Petrobrazi is the only domestic refinery supplying 2 different brands

(standard Petrom and premium OMV fuels). Petrotel has the highest sourcing choice (6 sites, 33% of total): in addition to its domestic Ploiesti refinery it can also import fuels from neighboring onshore Burgas refinery (Lukoil, Bulgaria) or its 4 locations in Russia, currently subject to EU sanctions. Hungarian MOL is Romania's third most important brand, its 3 onshore refineries providing fuels for its neighboring CEE markets (Slovakia, Slovenia, Croatia, Serbia). The 4 nearest onshore locations are Lukoil (Bulgaria), NIS Petrol (Pancevo and Novi Sad, Serbia) and MOL (Hungary), whilst the other 4 locations (OMV and MOL refineries) are further away (Austria, Slovakia, Germany, Croatia). Only 3 of the 7 offshore locations are currently able to supply crude oil to Romania (KMG from Kazakhstan and SOCAR's 2 sites in Turkey and Azerbaijan), as Lukoil's offshore sites are all in Russia. OSCAR Downstream operates the highest number of domestic own fuel depots (7 units), despite only importing its fuels from international traders via offshore locations. All brands operating in Romania have access to Conpet's national pipeline system (3,809 km) and extensive fuel depot network (11 units, 25% of total).

Most fuel depots (30%) are located in the South (13 units), mainly around the 3 refineries, due to the proximity to the Romanian capital, Bucharest, and the manufacturing plants of Dacia and Ford, with OMV having a 23% share with its 3 units (Brazi, Jilava and Isalnita). Conpet holds the highest concentration of storage facilities both in the West (50%) and East (30%). Automotive industry is the main economic growth driver in the Western region (Timisoara and Arad), whereas the Black Sea's Port of Constanta shipping activity fuels the business stimulus in the East. Lukoil, KMG and Oscar Downstream share two thirds of fuel depots (2 fuel depots each) in the Centre, Romania's Transylvania region. The region is fuelled by Cluj's competitive IT sector and automotive industry suppliers (Mercedes Benz). Northern Romania only has 1 fuel depot (KMG), mostly due to the lack of supporting infrastructure, a local economy focused on agriculture and the shutdown of one of Romania's main nearby refineries (RAFO Onesti).

Arad is the county with the highest number of fuel depots (5 units or 11% of total), mostly due to its geographic position (border with Hungary and gateway to Western Europe) and excellent high speed road (A1) and railway (M200) connections. Bucharest has 4 nearby depots (Jilava, Mogosoaia, Fundulea and Giurgiu) and connects to all major economic drivers: Constanta Port and Petromidia refinery via A2 (road) and M800 (rail), carmakers Dacia and Ford via A1 (road) and DEX12 expressway (road) and Prahova refineries (Petrobrazi and Petrotel) via A3 (road). Over 40% of storage units (3 fuel depots each) are located in the counties of the following 6 big cities: Oradea (logistics hub for the Western Area, connecting Cluj to the Hungarian border), Brasov (Selgros grocery wholesaler's headquarters, automotive electronics suppliers, TotalEnergies lubricants factory) in the Centre, Bacau (connecting Iasi, the Dornesti railway and transshipment terminals (DRT, DTT) to the sea and river port of Galati), in addition to Constanta (port and Petromidia refinery), Craiova (Ford, Pirelli and power electronics

manufacturers) and Ploiesti (Petrobrazi, Petrotel and Vega refineries, main grocery retailers' logistics hub).

Table 3 Romania – Storage locations for fuel distribution (2024)

Brand(s)	Main fuel sourcing locations	Regional coverage (units)	Fuel depot locations	Intermodal connections
OMV (Petrom)	Petrobrazi (Brazi, RO) Schwechat (Austria) Burghausen (Germany)	West (1) Center (1) North (0) East (1) South (3)	Arad Cluj Brazi Isalnita Jilava Bacau	A3 (road) M200 (rail) M300 (rail) M500 (rail)
Lukoil	Petrotel (Ploiesti, RO) Burgas (Bulgaria) Kstovo, Perm, Ukhta, Volgograd (Russia)	West (1) Center (2) North (0) East (2) South (2)	Arad Cluj Rm. Valcea Brasov Ploiesti Galati Constanta	A3 (road) M200 (rail) M300 (rail) M500 (rail) ROCND (sea)
MOL (Slovnaft) (INA)	Szazhalombatta (Hungary) Bratislava (Slovakia) Rijeka (Croatia)	West (1) Center (0) North (0) East (0) South (1)	Tileagd Giurgiu	M43-A1 (road) 120-M200 (rail) RODTS (river)
KMG (KazMunayGas) (Rompetrol)	Petromidia (Navodari, RO) Atyrau (Kazakhstan)	West (1) Center (2) North (1) East (0) South (2)	Arad Craiova Mogosoia Simleu Silvaniei Vatra Dornei Zarnesti	A4, A2 (road) M800 (rail) ROCND (sea)
SOCAR	Aliaga (Turkey) Baku (Azerbaijan)	West (0) Center (1) North (0) East (2) South (1)	Teius Ploiesti Onesti Constanta	ROCND (sea) ROGAL (river)
Gazprom Neft (NIS Petrol)	Pancevo (Serbia) Novi Sad (Serbia)	West (0) Center (0) North (0) East (0) South (0)	N/A	A1-E70 (road) RODTS (river)
Conpet national pipeline system (mainly OMV Petrom, Lukoil, KMG Rompetrol)	Petrobrazi (Brazi, RO) Petrotel (Ploiesti, RO) Petromidia (Navodari, RO) KMG Atyrau (Kazakhstan) SOCAR Aliaga (Turkey)	West (5) Center (1) North (0) East (3) South (2)	Biled Valcani Pecica Marghita Suplacu de Barcau Barbatesti Imeci Berca Moinesti Independenta Ciresu	ROCND (sea) Conpet network (refineries, fuel depots) A4, A2, A3 (road) M800 (rail) M200 (rail) M300 (rail) M500 (rail)
Oscar Downstream	Szazhalombatta (Hungary) Pancevo (Serbia) Novi Sad (Serbia) other (import)	West (1) Center (2) North (0) East (2) South (2)	Fundulea Sercaia Craiova Zadareni Ocna Mures Roman Constanta	ROCND (sea) A4, A2, A1 (road) M800 (rail) M200 (rail) M500 (rail)
7 major brands 1 main wholesaler and national pipeline system	3 domestic refineries 8 onshore and 7 offshore locations	West (10) Center (9) North (1) East (10) South (13)		West (23%) Center (21%) North (2%) East (23%) South (30%)

Fig. 4 highlights fuel distribution alternatives to the 5 regions across Romania. Refineries use a mix of pipelines and tank wagons (rail) to supply their specific regional fuel depots for longer domestic hauls, tankers for imported petroleum products via Black Sea and Danube River ports

(Constanta, Galati and Giurgiu) and fuel trucks (road) for supplying their network of filling stations.

Over 60% of Petrobrazi's refined products (3.84 million tonnes, 2023) were diesel fuels. Petrom and OMV-branded fuels are distributed downstream through Conpet pipelines,

tank wagons and fuel trucks. The Oltenia region is mainly supplied through pipelines (Conpet) leading to its fuel depot located in Isalnita. West and Center fuel depots (Arad, Cluj) usually receive their products by rail (M200 and M300): 2-3 trains/day with 36-38 wagons (over 50 tonnes of fuel each), on average. Most of the total length (50 km) of the 4 operational railway lines within the Petrobrazi refinery was modernized in the past decade (around 75%). Over 50,000 railway cars loaded with petroleum products (Petrom and OMV brands) are dispatched annually from the Petrobrazi refinery to supply OMV's domestic fuel depot network. The North of Romania is mainly supplied by fuel trucks (up to 28 million liters of fuel/truck) that can carry 4 different fuel types (standard and/or premium petrol and/or diesel) on a single trip (each divided into 4,000, 8,000, 10,000 and 14,000 liters per compartment) to the fuel depot in Bacau. The Republic of Moldova (a network of 69 Petrom filling stations) is also supplied directly from the Petrobrazi refinery with fuel trucks through its own Petrom storage facility in Chisinau, 450 km away. Up to 990 fuel trucks/week transit Romania's roads with fuels supplying Petrom and OMV's 586 filling stations. OMV uses tankers from Galati, Constanta (Black Sea shipments) and Giurgiu port (via the Danube River) to supply its stations in Bulgaria (92 units) and Serbia (64 units), in addition to regular fuel truck deliveries.



Figure 4 Romania – oil companies' intermodal transportation alternatives (2024)

In addition to domestic sources, Petrotel refines Russian crude oil at its Ploiesti site, before having to find alternative suppliers after EU sanctions (2022) impeded the inflow of Lukoil's regular sources. Gradual restrictions allowed for a surge of consecutive monthly imports of Russian crude oil (up to 3-5 times higher than regular deliveries) through Constanta's Oil Terminal and hedge its storage in Lukoil's own fuel depots, as well as additional storage capacities rented from Oil Terminal and Conpet, before the embargo. Lukoil is a major fuel supplier for the Romanian public sector, despite low transparency of its operations' data, continuous net operating loss and tax evasion suspicions. Landlocked CEE countries (Czech Republic, Slovakia and Hungary) were exempt and continue to receive deliveries of (cheaper) Russian oil. Since its regular EU markets ceased

imports, Russia is now shipping (more) oil to new markets (China, India, Turkey, North Africa, South America), that now act as an intermediary in petroleum products sale. India is clearly benefiting from selling refined products to the EU (20/27 countries), at full price, but from cut-price Russian-originated oil imports. India's top 3 markets (2023) are the Netherlands (24%), France (23%) and Romania (12%). As of 2024 the Czech Republic has also fully complied to EU sanctions, having found alternative options, but Hungary's MOL (who also owns Slovnaft from Slovakia) continues to favor cheaper Russian imports, despite having alternative sourcing nearby. The Adria pipeline from Omisalj (Croatia) considered more expensive, has higher technical risks, logistics costs and Russian imports would still be necessary to supplement oil deliveries from Croatia (up to 80% of total refining needs), according to Hungary (MOL). After Ukraine blocked all Lukoil transit shipments (via the Druzhba pipeline), Hungary's MOL found Tatneft as an alternative provider. Russia also uses a shadow fleet (600-2,000 rented tankers, some being in an advanced state of degradation) for 80% of its crude oil shipments to Kaliningrad and/or other destinations through the Baltic Sea, bypassing the Swedish island of Gotland.

Every year more than two thirds of Rompetrol's refined products (3.5 million tonnes of fuel products & LPG) are shipped from Petromidia Navodari (9 intermodal terminals) and Vega Ploiesti (6 intermodal terminals) with an average of 60,000 tank wagons (rail) and 35,000 fuel trucks (road), in addition to pipeline transportation (Conpet's network). Petromidia uses up to 10,000 tank wagons/month (75% of total Rompetrol rail shipments, M800) to transport petrol and diesel fuels from Navodari to its 6 storage depots across Romania. Mogosoaia is Rompetrol's largest fuel depot (capacity: 13,000 tonnes), around 120 trucks supplying it daily with premium products. Fuel trucks then handle downstream deliveries to its network of filling stations (regular and new premium-branded "hei" concept stations). MMT enables both import (crude oil from Kazakhstan) and export (refined products mainly towards subsidiaries in Moldova, Bulgaria, Georgia) through its onshore and offshore terminals at the Black Sea port. With its 3 terminals, berth 9 is the most modern and allows transiting of 350,000 tonnes/month finished products (fuel products, ethylene and LPG), in addition to the other 4 berths (for crude oil) at Midia Port.

MOL uses trains for deliveries to the intermodal hub at Curtici Cargo Center (West) and ships on the Danube River to connect to Giurgiu port (South) which act as DCs for final deliveries, mostly by fuel trucks, to its network of 238 filling stations. KMG uses 2 routes: from Atyrau through the CPC pipeline (Tengiz-Novorossiysk Black Sea Marine Terminal) for regular shipments and through the BTC pipeline (Aktau-Baku-Ceyhan route) for annual volumes of up to 1.5 million tonnes (5-year agreement with SOCAR, since 2022). The port of Ceyhan (Turkey) then directs KMG shipments towards Romania's Black Sea port with unloading in the Petromidia refinery from its offshore marine terminal near Constanta.

OSCAR Downstream's business model relies on imported petrol and diesel fuels, their transport and storage within its network of fuel depots and distribution with the OSCAR tankers fleet. By sourcing its petroleum products from fuel traders, OSCAR is closely linked to the global fuel market, making it more vulnerable to international market trends and events. Price volatility increased as a result of a series of recent crises (Covid-19 pandemic, energy crisis, Russia's invasion of Ukraine, Houthi attacks in the Red Sea and geopolitical tensions in the Middle East) and exacerbated supply chain challenges and disruptions for major distributors.

OSCAR Downstream's 7 fuel depots have a 38% share of all independent units with a combined storage capacity of over 75,000 cbm, being located within 250-300 km of each other in all major economic regions. It owns 115 stations: 85 RO franchise stations (for B2C customers, 74%) and 31 OSCAR Drive (for B2B customers, 26%). With a network of over 2,500 DIESELpoint Access stations, the company also provides B2B customers with the ability to refuel at their own facilities (in-house). This implies the consignment of a station (5,000, 9,000 or 20,000 liters) provided on a rental basis (including installation, technical support and maintenance), regular fuel supply handled by OSCAR's fleet of 82 tanker trucks and a fuelling management application. Over 5,000 companies from the agriculture, transportation, industry and construction sector have partnered with Oscar Downstream for DIESELpoint Access stations (as of 2024).

Conpet (a 58.71% publicly-owned company) manages the 3,809 km network of Romanian pipelines, an interconnected system providing delivery terminals (receiving and storage tanks) for upstream crude oil, delivered by oil producers and/or importers (from own domestic extraction sites and/or imported), and their transfer/routing from these terminals/stations (through pumps) towards oil processing units (refineries). The total throughput of the pipeline system is 27.43 million tonnes/year, 73% being used by imported oil (20.2 million tonnes/year), while over 73% of total storage capacity is reserved for domestic oil (126,000 cbm). Most of Conpet's 15 loading/unloading ramps (86%) are used for loading crude oil, with 13 locomotives and 69 tank wagons also being part of their railway fleet. Publicly-owned (87.75%) Oil Terminal is located in the Port of Constanta and is the largest Romanian terminal for the import and/or export of oil products (crude oil, petroleum products, petrochemicals, etc.). Oil Terminal's services (receiving, storage, conditioning and shipping) cover import, export and transit of all types of petroleum products. Petrom (OMV) and Lukoil (Litasco) use Oil Terminal for their import of crude oil, petroleum products and/or diesel (including storage), as well as for the export of refined fuels.

GFR has the highest share of cargo volumes (22.45%) among all rail freight operators in Romania and is market leader in transporting petroleum products (crude oil and refined products) since 2013. With its 400 locomotives and 16,000 wagons, GFR is the main LSP for petroleum products for all 3 refineries, as well as for Conpet and Constanta Port's Oil Terminal. Depending on the specific route (loading

and/or unloading station, distance travelled), GFR hauls fuels with 15-50 tanker wagons/freight train of petroleum products.

5 DISCUSSION

This research paper's main objective is to underline the relevance of infrastructure (intermodal and pipelines) in connecting upstream oil refining production capacities with downstream distribution networks by using Romania as a case study. The 3 analyzed refineries (Petrobrazi, Petrotel and Petromidia) have very good sourcing alternatives, both East (Port of Constanta) via sea shipping and pipeline transportation networks and West (Curtici Cargo Center) via intermodal transportation (road-rail) options.

Most previous studies conduct research on correlations of oil refining with macroeconomic data [26], downstream optimization of fuel distribution networks [29] or use the particular case of a given filling station [32]. While they are a source of new insight, these approaches are either too broad [25] or only focus on partial oil supply chains [28] and thus have inconsistent outcomes. More targeted research endeavors tend to either favor obvious conclusions [27], due to a homogenous sample [30], or be too complex and require additional parameterization [31]. This research paper, on the other hand, provides an in depth analysis of the Romanian oil refining supply chain, integrating all major multinational and independent brands, as well as the national pipeline infrastructure and its storage depots. It is, at the time of writing, the first comprehensive study of a country's oil refining industry with high practical insights as its major strength. Due to the lengthy data collection process and differences in data range disclosure (NDAs) research design is mostly descriptive and explanatory as correlations (analytical or predictive models) would have implied using confidential datasets. Dataset restrictions are thus a drawback for a more proper scientific research design and corresponding conclusions, as well as a main limitation of this study.

The purpose of this article was to study the impact EU sanctions on Russia have on Romania's Petrotel refinery (Lukoil) and how its oil supply chain managed operations with alternative suppliers. Its geostrategic location, concentration of oil refineries in neighboring countries and intermodal infrastructure interconnectivity facilitated marginal disruptions in the overall fuel supply. Potential future research topics could cover the role of oil refining in local economies (jobs, infrastructure development, GDP growth contribution), the feasibility of shifting away from Russian oil sourcing alternatives or increasing digitalization in oil refining supply chains (EDI, digital twins, machine learning applications).

6 CONCLUSION

More than 80% of Romania's total output of petroleum products is currently provided by OMV's Petrobrazi and KMG's Petromidia refineries. Up to two thirds of their crude oil is sourced locally, whilst they own over 43% of all filling

stations and almost a third (28%) of all fuel depots. Lack of recent investments, tax evasion lawsuits and EU sanctions on Russian oil has scaled back production output at Lukoil's Petrotel refinery to under 20% of domestic share (2023). In addition to its refineries, Romania's filling stations are replenished from 8 other onshore and 7 offshore locations, 4 closely located alternatives being available in its neighboring countries (Bulgaria, Serbia and Hungary).

Offshore oil imports are handled through the Black Sea via the private Midia Marine Terminal (KMG), Constanta Port's dedicated terminals (Oil Terminal, new JT Terminal) and its interconnected downstream pipeline network (Conpet). Conpet owns 25% of domestic fuel depots (11/43 units), its pipelines connect 24/41 counties (58%) across Romania and are linked with the 3 domestic refineries.

Over 70% of Romania's filling stations (2,291 units) are owned by multinational oil companies, whilst more than half (1,306 units) are owned by the domestic refinery-operating brands: Petrom, OMV, Rompetrol and Lukoil. A mix of pipelines, rail freight and trucks is used to distribute petrol and diesel fuels to their network of filling stations. Almost all fuel depots (42/43 units) are located in regions with both economic development (oil refining, carmakers and/or automotive suppliers) and infrastructure connectivity (road, rail and/or port).

This case study research highlights the underlying role of infrastructure connectivity in economic development. Oil industry in Romania leverages its logistics to source, refine and distribute petroleum products across its network of fuel depots and filling stations. Insights reveal that shifting away from Russian imports and finding alternative oil sources is less of a challenge in case of multiple onshore and offshore backup options.

The paper's added value resides in its high degree of carried out on-site fieldwork, offering relevant practical insights sourced from the actual oil industry environment. Parts of collected data are not subject to disclosure (NDAs) and willingness of oil industry and/or terminal operators to grant access to internal data is very limited. The extent to which such data is processed, findings expressed within the paper's content and certain wording choices are subject to written approval, affecting scientific research approach and study replicability. These trade-offs counterbalance the empirical value of the research paper and are its major source of limitations.

7 REFERENCES

- [1] Sapnken, F. E., & Tamba, J. G. (2022). Petroleum products consumption forecasting based on a new structural auto-adaptive intelligent grey prediction model. *Expert Systems with Applications*, 203, 117579. <https://doi.org/10.1016/j.eswa.2022.117579>
- [2] Hao, X. (2023). Import competition and pressure in the international crude oil trade: A network analysis. *Resources Policy*, 82, 103468. <https://doi.org/10.1016/j.resourpol.2023.103468>
- [3] Kuby, M., Capar, I., & Kim, J.-G. (2017). Efficient and equitable transnational infrastructure planning for natural gas trucking in the European Union. *European Journal of Operational Research*, 257(3), 979-991. <https://doi.org/10.1016/j.ejor.2016.08.017>
- [4] Brkic, D. (2024). Serbian Energy Sector in a Gap between East and West. *Energy Exploration & Exploitation*, 42(1), 330-340. <https://doi.org/10.1177/01445987231215445>
- [5] Khor, C. S., & Varvarezos, D. (2017). Petroleum refinery optimization. *Optimization and Engineering*, 18(4), 943-989. <https://doi.org/10.1007/s11081-016-9338-x>
- [6] Halvorsen-Weare, E. E., & Fagerholt, K. (2017). Optimization in offshore supply vessel planning. *Optimization and Engineering*, 18(1), 317-341. <https://doi.org/10.1007/s11081-016-9315-4>
- [7] Borthen, T., Loennechen, H., Fagerholt, K., Wang, X. & Vidal, T. (2019). Bi-objective offshore supply vessel planning with costs and persistence objectives. *Computers & Operations Research*, 111, 285-296. <https://doi.org/10.1016/j.cor.2019.06.014>
- [8] Kisialiou, Y., Gribkovskaia, I., & Laporte, G. (2018). The periodic supply vessel planning problem with flexible departure times and coupled vessels. *Computers & Operations Research*, 94, 52-64. <https://doi.org/10.1016/j.cor.2018.02.008>
- [9] dos Santos Diz, G. S., Hamacher, S., & Oliveira, F. (2019). A robust optimization model for the maritime inventory routing problem. *Flexible Services and Manufacturing Journal*, 31(3), 675-701. <https://doi.org/10.1007/s10696-018-9327-9>
- [10] Cruz, R., Mendes, A. B., Bahiense, L., & Wu, Y. (2019). Integrating berth allocation decisions in a fleet composition and periodic routing problem of platform supply vessels. *European Journal of Operational Research*, 275(1), 334-346. <https://doi.org/10.1016/j.ejor.2018.11.030>
- [11] McWilliams, B., Sgaravatti, G., Tagliapietra, S., & Zachmann, G. (2023). How would the European Union fare without Russian energy? *Energy Policy*, 174, 113413. <https://doi.org/10.1016/j.enpol.2022.113413>
- [12] Johnson, S., Rachel, L., & Wolfram, C. (2023). Design and implementation of the price cap on Russian oil exports. *Journal of Comparative Economics*, 51(4), 1244-1252. <https://doi.org/10.1016/j.jce.2023.06.001>
- [13] Bento, F., Garotti, L., & Mercado, M. P. (2021). Organizational resilience in the oil and gas industry: A scoping review. *Safety Science*, 133, 105036. <https://doi.org/10.1016/j.ssci.2020.105036>
- [14] Regli, F., & Nomikos, N. K. (2019). The eye in the sky - Freight rate effects of tanker supply. *Transportation Research Part E-Logistics and Transportation Review*, 125, 402-424. <https://doi.org/10.1016/j.tre.2019.03.015>
- [15] Atmayudha, A., Syauqi, A., & Purwanto, W. W. (2021). Green logistics of crude oil transportation: A multi-objective optimization approach. *Cleaner Logistics and Supply Chain*, 1, 100002. <https://doi.org/10.1016/j.clscn.2021.100002>
- [16] Meira, W. H. T., Magatao, L., Relvas, S., Barbosa-Povoa, A. P., Neves, F., & Arruda, L. V. R. (2018). A matheuristic decomposition approach for the scheduling of a single-source and multiple destinations pipeline system. *European Journal of Operational Research*, 268(2), 665-687. <https://doi.org/10.1016/j.ejor.2018.01.032>
- [17] Meira, W. H. T., Magatao, L., Neves, F., Arruda, L. V. R., Vaqueiro, J. P., Relvas, S. & Barbosa-Povoa, A. P. (2021). A solution framework for the long-term scheduling and inventory management of straight pipeline systems with multiple-sources. *Computers & Operations Research*, 127, 105143. <https://doi.org/10.1016/j.cor.2020.105143>
- [18] Quinteros, M., Guignard, M., Weintraub, A., Llambias, M., & Tapia, C. (2019). Optimizing the pipeline planning system at the national oil company. *European Journal of Operational Research*, 277(2), 727-739.

- <https://doi.org/10.1016/j.ejor.2019.03.007>
- [19] Cafaro, D. C., & Cerda, J. (2017). Short-term operational planning of refined products pipelines. *Optimization and Engineering*, 18(1), 241-268. <https://doi.org/10.1007/s11081-016-9330-5>
- [20] Pudasaini, P. (2021). Integrated planning of downstream petroleum supply chain: a multi-objective stochastic approach. *Operations Research Perspectives*, 8, 100189. <https://doi.org/10.1016/j.orp.2021.100189>
- [21] Hsu, Y. C., Walteros, J. L., & Batta, R. (2020). Solving the petroleum replenishment and routing problem with variable demands and time windows. *Annals of Operations Research*, 294(1), 9-46. <https://doi.org/10.1007/s10479-018-3042-5>
- [22] Schroeder, T., Lauen, L.-P., & Geldermann, J. (2018). Improving biorefinery planning: Integration of spatial data using exact optimization nested in an evolutionary strategy. *European Journal of Operational Research*, 264(3), 1005-1019. <https://doi.org/10.1016/j.ejor.2017.01.016>
- [23] Khezeli, M., Najafi, E., Molana, M. H., & Seidi, M. (2023). A sustainable and resilient supply chain (RS-SCM) by using synchronisation and load-sharing approach: application in the oil and gas refinery. *International Journal of Systems Science: Operations & Logistics*, 10(1), 2198055. <https://doi.org/10.1080/23302674.2023.2198055>
- [24] Lima, C., Relvas, S., & Barbosa-Povoa, A. P. (2022). A graph modeling framework to design and plan the downstream oil supply chain. *International Transactions in Operational Research*, 29(3), 1502-1519. <https://doi.org/10.1111/itor.12969>
- [25] Al-Mahish, M., Alzahrani, F., Elzaki, R., & Ben Slama, M. (2024). Economies of scale in Saudi Arabia's refining sector: An application of modern econometric models. *Heliyon*, 10(9), e30150. <https://doi.org/10.1016/j.heliyon.2024.e30150>
- [26] Mukhamediyev, B., Temerbulatova, Z., & Mukhamediyeva, A. (2023). The significance of world oil prices and macroeconomic factors for the global competitiveness of countries. *Energy Economics*, 128, 107186. <https://doi.org/10.1016/j.eneco.2023.107186>
- [27] Guo, S., Li, H., An, H., Ma, N., Sun, Q., Feng, S., Sun, G., & Liu, Y. (2024). Detecting the horizontal/vertical price relationship patterns in the global oil industry chain through network analysis. *Energy*, 296(6), 131054. <https://doi.org/10.1016/j.energy.2024.131054>
- [28] Garcia-Verdier, T. G., Gutierrez, G., Mendez, C. A., Palacin, C. G., & de Prada, C. (2024). Optimization of crude oil operations scheduling by applying a two-stage stochastic programming approach with risk management. *Journal of Process Control*, 133, 103142. <https://doi.org/10.1016/j.jprocont.2023.103142>
- [29] Khezeli, M., Najafi, E., Molana, M. H., & Seidi, M. (2023). A sustainable and resilient supply chain (RS-SCM) by using synchronisation and load-sharing approach: application in the oil and gas refinery. *International Journal of Systems Science: Operations & Logistics*, 10(1), 2198055. <https://doi.org/10.1080/23302674.2023.2198055>
- [30] Andrade, G. B., Andrade, L. B., Conde, E., Grecco, C., Krykhtine, F. L. P., & Cosenza, C. A. (2023). A novel approach for complex and subjective variable valuation in the COPPE-Cosenza fuzzy model applied to crude oil trading. *Decision Analytics Journal*, 7, 100209. <https://doi.org/10.1016/j.dajour.2023.100209>
- [31] Freen, G., Kousar, S., Kausar, N., Pamucar, D., & Oros, G. I. (2023). Multimodal Fuzzy Downstream Petroleum Supply Chain: A Novel Pentagonal Fuzzy Optimization. *Computers, Materials & Continua*, 74(3), 4861-4879. <https://doi.org/10.32604/cmc.2023.032985>
- [32] Sihlali, B., & Telukdarie, A. (2022). Integrated Digital System in Planning, Scheduling and Logistics Processes at a Petroleum Company. *Proceedings of the IEEE 28th International Conference on Engineering, Technology and Innovation (ICE/ITMC) & 31st International Association for Management of Technology (IAMOT) Joint Conference*, 10033320. <https://doi.org/10.1109/ICE/ITMC-IAMOT55089.2022.10033320>

Author's contacts:

Attila Turi, PhD
 Politehnica University Timisoara,
 Remus 14, 300191 Timisoara, Romania
 +40.256.404.307, attila.turi@upt.ro

Application of Photogrammetry in Agriculture with an Emphasis on Close-Range

Domagoj Zimmer*, Mladen Jurišić, Hrvoje Glavaš, Mirko Karakašić, Toni Beljo

Abstract: In contemporary agricultural practices and technological advancements, the paramount focus is on enhancing productivity and operational efficiency. This entails the effective management of digital resources like obtained from photos while minimizing costs. Through the application of close-range photogrammetry and various specialized software (Agisoft Metashape, CloudCompare, MicMac, Pix4Dmapper and OpenDroneMap), farmers can effectively and accurately monitor vegetation indices and generate three-dimensional models, thereby enhancing the optimization of their agricultural practices. Photogrammetric softwares enable creating 3D models, their interaction with the processing procedure, production of point clouds, image mergin from drone photos and an automatic extraction of DEMs. The emphasis of the paper is on the great potential of photogrammetry with the main goal to improve modern agriculture such as cost reduction, precision and accuracy, data collection and similar to. Paper is based on examples of practice and field work in the agricultural sector around the world.

Keywords: Agisoft Metashape; agriculture; close-range photogrammetry; CloudCompare; MicMac

1 INTRODUCTION

The non-contact acquisition of data on the characteristics of studied surfaces and objects and their interpretation can be done using photogrammetry [1] and, in general, the technique of measuring the shape, size or position of a selected object through the interpretation of a series of photographs [2]. According to [3, 4], photogrammetry can be described as an art, science and technology to obtain good quantitative information about various objects and the surrounding environment through the process of measuring, recording and deciphering photographic images. The most economical and fastest way to collect geospatial data is by using digital photogrammetry methods [5]. In the discussed field, remote sensing generally offers two functional options, namely, satellite imagery and unmanned aerial vehicle photogrammetry [6, 7]. According to [8] the monitoring of a selected corn hybrid within pre-defined growth intervals is possible with photogrammetry images. Authors [9] used digital photogrammetry to quantify erosion on agricultural land. Photogrammetry is divided into two main types, namely aerial photogrammetry and terrestrial photogrammetry [10]. At the same time, there are two further subdivisions in relation to the position of the camera, namely horizontal and extra-terrestrial photogrammetry [11]. In close-range photogrammetry, the objects studied are examined using photographs or digital images taken with a camera at close range. The model in close-range photogrammetry is created on the basis of the central perspective projection [12]. It is mainly based on digital technology to store and manipulate data (images) in the applied area to create digital models that correspond to the real captured objects. This opens up the possibility of modeling standard concepts such as length, area and size as well as spatial coordinates [13]. The authors [14] note that with the increasing resolution of digital cameras and advances in calibration, close-range photogrammetry is increasingly being used to create three-dimensional models of objects with high precision. Model construction with close-range photogrammetry consists of a series of steps that begin with taking photographs and orienting them relative to the object [15].

2 PHOTOGRAMMETRIC SOFTWARES

2.1 Agisoft Metashape

Agisoft Metashape is used for 3D modeling, which creates high-quality 3D spatial data by processing digital photos [16]. On figure 1 is presented the process of importing and aligning photos from plough share. Agisoft Metashape is a program that performs photogrammetric processing of digital photos and generates three-dimensional spatial information used in geographic information systems and visual effects production, as well as for indirect measurements of various objects. The authors [17] claim that it is an improved photo-based three-dimensional modeling that can be used to generate professional-quality three-dimensional content from still photographs. This program is one of the best photogrammetry tools. Author [2] states that the mentioned application is used to process 3D data on the wear and tear of parts of agricultural machinery. The program also makes it possible to work with laser scanning and combine data from LiDAR (Light Detection and Ranging) and cameras in the same project. LiDAR is a widely used method for reconstructing three-dimensional images for various applications and points to reconstruct three-dimensional maps of the investigated environment [18].

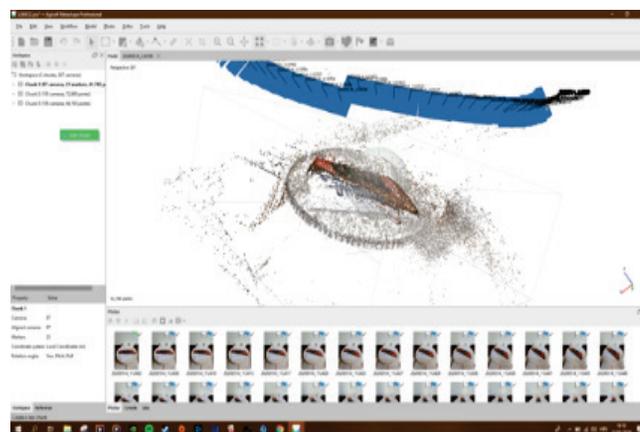


Figure 1 Agisoft Metashape – importing and aligning photos from plough share [2]

2.2 MicMac

MicMac (Fig. 2) is a freeware program for photogrammetry [19]. It is used by the French National Geographic Institute for its research, which includes algorithms for three-dimensional image processing for a wider audience, universities, public institutions and companies [20]. The MicMac processing method is fully customizable by the user, allowing for fine-tuning [21]. As a photogrammetric program, it is characterized by precision and reliability and offers capabilities not available in most alternative programs [14]. For example, according to the authors [20], the program allows access to intermediate results in open data formats, which enables interaction with the processing procedure at each stage as well as qualitative evaluation of the results via quality indicators. In addition, different models for camera calibration are available, e.g. models configured for cameras with basic functionality, large-format aerial cameras, cameras with very long focal lengths, fisheye cameras and cameras that can capture a 360-degree panorama [20]. Authors [22] state that the MicMac was successfully used for the purposes of determining key agricultural strategic factors.

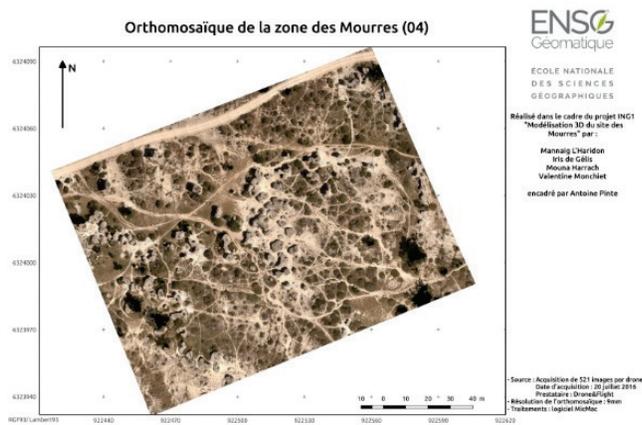


Figure 2 MicMac ortho-mosaic [23]

2.3 Pix4Dmapper

Pix4D (Fig. 3) is an advanced program for creating geocoded maps and models from aerial imagery. Sophisticated photogrammetry software is used to create professional orthomosaics, point clouds and models from images captured by drones [24]. According to the official manufacturer Pix4D, the main process consists of 7 steps: Capture, Digitize, Control, Measure, Review, Collaborate and Share. With the mentioned program, different performances of final products can be obtained, such as point clouds in full color, orthomosaics, three-dimensional textured meshes, a series of maps of general interest and temperature mapping with corresponding data processing that includes a report on the credibility of the data [25]. According to [26] the aforementioned application was used for software for processing photogrammetric data, more precisely for creating maps for precision fertilization.



Figure 3 Plant health in Pix4DMapper [27]

2.4 OpenDroneMap

OpenDroneMap (Fig. 4) contains a set of open source software packages that provide a complete photogrammetric solution for small drones [28]. The software works as an application for processing non-referenced images with a high degree of overlap, including color point clouds, digital surface models, textured digital surface models, and orthophotos [29]. After the process of image merging, the software has a visualization option that uses Meshlab, Python, and the WebOdm web application to integrate features such as user authentication, map visualization, and three-dimensional displays [30]. According to [31] it was possible to create a weed spread map with the help of OpenDroneMap.



Figure 4 Plant health in OpenDroneMap [32]



Figure 5 Laser Scanner PlanTeye builds 3D point clouds [34]

Laser scanners are often used for non-contact measurement (Fig. 5) of roughness and provide information on accuracy in the range of 0.1 – 0.5 mm and 0.1 – 2 mm for vertical and horizontal measurements respectively.

Numerous studies of this type have been carried out under controlled indoor conditions, as certain devices must be used in field research due to their complexity, which has a direct impact on variable costs, technical maintenance and ultimately low resistance to meteorological conditions and dust, as well as processing time [33].

3 APPLICATION OF PHOTOGRAMMETRY IN AGRICULTURE

3.1 Vegetation Indices

Vegetation indices (Tab. 1) provide a wealth of data for various practical applications in precision agriculture and offer quantitative information on plant growth and health [35]. Vegetation indices (VARI, NGRDI, GLI and ExG) are the result of very simple and effective algorithms for quantitative and qualitative prediction of vegetation cover, lushness and growth rate, which are used in environmental monitoring, biodiversity conservation, agriculture, forestry and other similar fields [36]. Vegetation indices are information obtained by calculating from numerous channels of multispectral images based on the absorption, transmission and reflectance of plants in different spectral channels. Vegetation indices are used to graphically predict vegetation activity on the production area under investigation. Vegetation indices in production areas with vegetation have higher pixel values than in areas without vegetation [3].

4 CLOSE-RANGE PHOTOGRAMMETRY IN AGRICULTURE

Close-range photogrammetry usually captures an object within a few metres from different angles. Used in agriculture, it can generate 3D models using various algorithms.

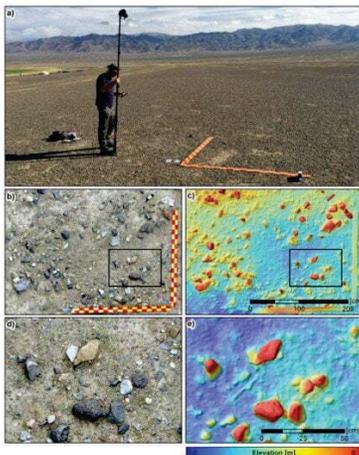


Figure 6 Measurement of soil surface roughness by close-range photogrammetry method processed by Agisoft Photoscan [41]

4.1 Improving the Accuracy of Soil Roughness Measurements

The quantification of soil roughness is actually the extent of irregularity of the production surface due to soil texture, aggregates, rock fragments and the way the production surface is worked. It has a significant impact on surface water

storage, infiltration, overland flow and ultimately sediment deposition and erosion [37]. The roughness of the production surface can be determined by contact techniques such as roller chain calculations [38], profile measuring devices or using a needle [39] and automated relief measuring devices [27]. Automated close-range photogrammetry is increasingly used today (Fig. 6) [40]. Stereophotogrammetry, which enables very fast information acquisition and automatic extraction of DEMs, is used to obtain micro-reliefs of the production surface with acceptable accuracy (vertical 0.2–6 mm, horizontal 1–5 mm).

4.2 3D Surface Modeling of a Tomato Plant and Tractor Tire Footprints

The creation of a 3D model is considered a complete process, the first step of which is the collection of information, while the last step is the creation of a virtual 3D model that can be displayed on a computer multimedia [42]. Close-range photogrammetry enables very high-resolution point clouds (Fig. 7). Efficient methods for 3D reconstruction from images are widely available on web-based online processing services and allow the export of high-quality textured 3D models (Fig. 8) [43].

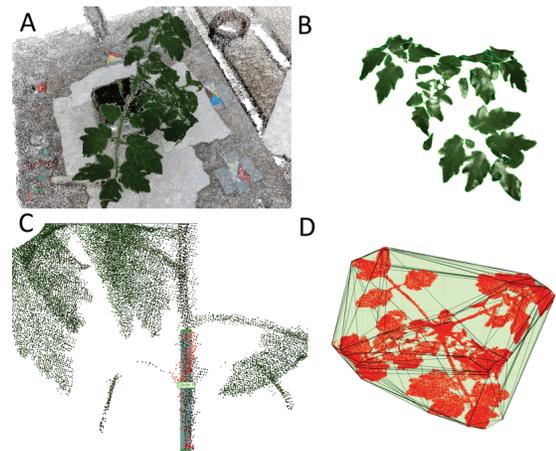


Figure 7 Three-dimensional mesh cloud of a tomato vegetable [5]

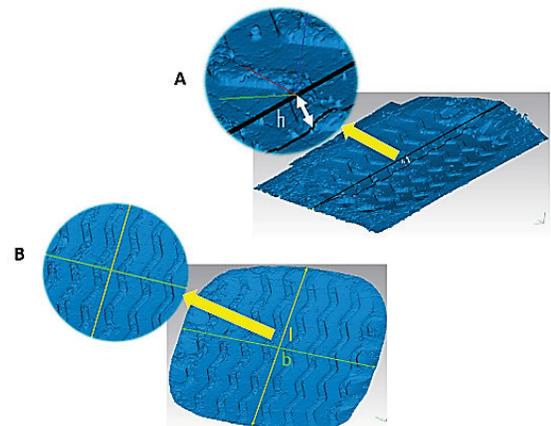


Figure 8 3D mesh of triangles of the tire imprint on the ground: A - with intact ground (h - depth of the imprint); B - determined imprint (l - length of the imprint and b - width of the imprint) in Smarttech3Dmeasure [44]

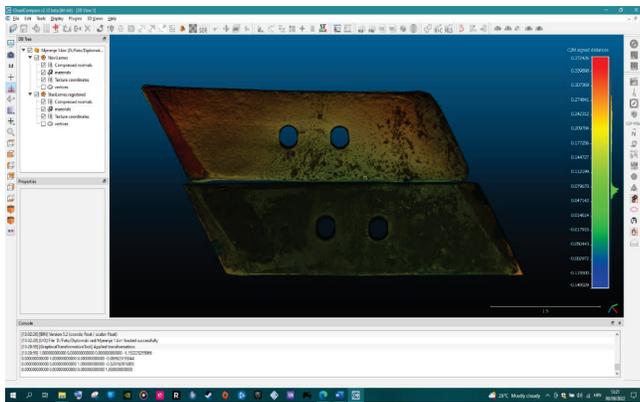


Figure 9 CloudCompare Surface differences on plough share's [2]

5 CLOUDCOMPARE

CloudCompare (Fig. 9) is a software for processing 3D point clouds and triangle meshes. Also, CloudCompare can show differences in surfaces, measure surface and measure volume. Very important for this kind of application is that it can shown and export differences by diagrams. In addition, it includes advanced data processing algorithms to perform actions such as displaying images in multiple projections (rectangular, cylindrical, conical), registering point clouds, calculating distances (between two point clouds, between a cloud and a triangle mesh, between two individual points), test statistics, and evaluating geometric properties [45]. It can be easily used for the visualization of large data sets.

6 CONCLUSIONS

Tribology is the science that studies phenomena on the surfaces of two bodies in contact or relative motion (triboelements), primarily friction, material wear, and the effects of lubrication. At the beginning tribology was based on measuring the mass of objects before and after using, in this way, an insight was gained into the state of the materials, i.e. today's 3D models, which can be created much faster thanks to the mentioned programs. The use of photogrammetry in agriculture represents a significant and promising approach for the further development of the agricultural sector. This paper has also emphasized the potential of photogrammetry, i.e., short-range photogrammetry, in various segments or aspects of agriculture, ranging from crop monitoring using vegetation indices to the accuracy of soil surface roughness measurements and the creation of 3D plant models through short-range photogrammetry. The use of photogrammetric technologies and the processing of images with photogrammetric software enable farmers and experts to optimize the use of resources and improve productivity and work efficiency. Close-range photogrammetry in agriculture offers great opportunities for precision agriculture. Agisoft Metashape is one of the best known photogrammetry programs and is characterized by its ability to create very high quality 3D models. It is mainly intended for professionals. MicMac and OpenDroneMap are open source programs that provide tools for image processing and 3D

model creation. Their best feature is their cost-effectiveness. Pix4DMapper is a program, which uses photos to create high quality orthomosaics and is often used in agriculture. CloudCompare, while not photogrammetry software, is used for processing and analyzing 3D point cloud data.

7 REFERENCES

- [1] Schenk, T. (2005). *Introduction to Photogrammetry*. Department of Civil and Environmental Engineering and Geodetic Science the Ohio State University. Retrieved from https://www.google.com/url?sa=t&source=web&rct=j&opi=89978449&url=https://www.mat.uc.pt/~gil/downloads/IntroPhoto.pdf&ved=2ahUKewix1amT_KaLaxVR8LsIHUVvHfEQFnoECBgQAQ&usq=AOvVaw0KUMtgzelidB0HYfdFyhUf
- [2] Beljo, T. (2020). Application of close – range Photogrammetry in Agriculture. *Undergraduate thesis*. Josip Juraj Strossmayer University of Osijek, Faculty of Agrobiotechnical Sciences Osijek. (in Croatian)
- [3] Sigurnjak, E. (2020). Monitoring of more important phenological and agrotechnical soybean parameters using an unmanned aircraft and vegetation indices. *Master thesis*. Josip Juraj Strossmayer University of Osijek, Faculty of Agrobiotechnical Sciences Osijek. (in Croatian)
- [4] Kovačić, B. (2021). Creation of 3D models using the photogrammetric method. *Undergraduate thesis*. University of Zagreb, Faculty of Mining, Geology and Petroleum Engineering. (in Croatian)
- [5] Marčeta, M. (2014). Collection of information for creation of GIS with automatic digital Photogrammetry methods. *Synthesis 2024 - International Scientific Conference of Singidunum University*, Serbia. Retrieved from <https://portal.sinteza.singidunum.ac.rs/>
- [6] Ferrer-González, E., Agüera-Vega, F., Carvajal-Ramírez, F., Martínez-Carricondo, P. (2020). UAV Photogrammetry Accuracy Assessment for Corridor Mapping Based on the Number and Distribution of Ground Control Points. *Remote Sensing*, 12, 2447. <https://doi.org/10.3390/rs12152447>
- [7] Martínez-Carricondo, P., Agüera-Vega, F., & Carvajal-Ramírez, F. (2020). Use of UAV-Photogrammetry for Quasi-Vertical Wall Surveying. *Remote Sensing*, 12, 2221. <https://doi.org/10.3390/rs12142221>
- [8] Janoušek, J., Jambor, V., Marchoň, P., Dohnal, P., Synková, H., & Fiala, P. (2021). Using UAV-based photogrammetry to obtain correlation between the vegetation indices and chemical analysis of agricultural crops. *Remote Sensing*, 13(10), 1878. <https://doi.org/10.3390/rs13101878>
- [9] Arriola-Valverde, S., Villalobos-Avellán, L. C., Villagra-Mendoza, K., & Rimolo-Donadio, R. (2020). Erosion quantification in runoff agriculture plots by multitemporal high-resolution UAS digital photogrammetry. *Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 13, 6326-6336. <https://doi.org/10.1109/JSTARS.2020.3027880>
- [10] Kalin, J. (2021). Using photogrammetry for purposes of cultural heritage preservation. *Master thesis*. University in Ljubljana, Faculty of Computing and Informatics. (in Slovenian)
- [11] Pandur, A. (2019). Creation of 3D models using visualfsm. *Undergraduate thesis*. Medimurje University of Applied Sciences in Čakovec. (in Croatian)
- [12] Yilmaz, H. M., Yakar, M., & Yildiz, F. (2008). Documentation of historical caravansaries by digital close range photogrammetry. *Automation in Construction*, 17(4), 489-498.

- <https://doi.org/10.1016/j.autcon.2007.09.003>
- [13] Fawzy, El-Din H. (2019). Study the accuracy of digital close range photogrammetry technique software as a measuring tool. *Alexandria Engineering Journal*, 58(1), 171-179. <https://doi.org/10.1016/j.aej.2018.04.004>
- [14] Firdaus, M., & Rau, J.-Y. (2017). Comparisons of the Three-Dimensional Model Reconstructed Using MicMac, Pix4DMapper and PhotoscanPro. <https://doi.org/10.13140/RG.2.2.33613.74720>
- [15] Samaan, M., Raphaële, H., & Deseilligny, P. M. (2013). Close-range photogrammetric tools for small 3D archeological objects. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XL-5(W2). <https://doi.org/10.5194/isprsarchives-XL-5-W2-549-2013>
- [16] Marić, I., Šiljeg, A., & Domazetović, F. (2019). Geospatial technologies in 3D documentation and promotion of cultural heritage - example of Fortica fortress on the island of Pag. *Geodetic Gazette*, 50, 19-44. <https://doi.org/10.58817/2233-1786.2019.53.50.19>
- [17] Regmi, Y., & Manandhar, S. G. (2020). Applications of UAVs and digital photogrammetry. Tribhuvan University, Institute of engineering. Pashchimanchal Campus, Department of Civil and Geomatics Engineering.
- [18] Villa, F., Severini, F., Madonini, F., & Zappa, F. (2021). SPADs and SiPMs Arrays for Long-Range High-Speed Light Detection and Ranging (LiDAR). *Sensors*, 21, 3839. <https://doi.org/10.3390/s21113839>
- [19] Guimarães, N., Pádua, L., Adão, T., Hruška, J., Peres, E., & Sousa, J. J. (2020). VisWebDrone: A Web Application for UAV Photogrammetry Based on Open - Source Software. *International Journal of Geo-Information*, 9, 679. <https://doi.org/10.3390/ijgi9110679>
- [20] Rupnik, E., Daakir, M., & Pierrot Deseilligny, M. (2017). MicMac – a free, open-source solution for photogrammetry. *Open geospatial data, software and Standards*, 2(14). <https://doi.org/10.1186/s40965-017-0027-2>
- [21] Cutugno, M., Robustelli, U., & Pugliano, G. (2022). Structure-from-Motion 3D Reconstruction of the Historical Overpass Ponte della Cerra: A Comparison between MicMac® Open Source Software and Metashape®. *Drones*, 6(9), 242. <https://doi.org/10.3390/drones6090242>
- [22] Barati, A. A., Azadi, H., Dehghani Pour, M., Lebailly, P., & Qafari, M. (2019). Determining Key Agricultural Strategic Factors Using AHP-MICMAC. *Sustainability*, 11(14), 3947. <https://doi.org/10.3390/su11143947>
- [23] Wiki dedicated to MicMac. *MicMac*. <https://micmac.ensg.eu/index.php/Fichier:MourresOrtho>
- [24] Hinge, L., Gundorph, J., Ujang, U., Azri, S., Anton, F., & Abdul Rahman, A. (2019). Comparative Analysis of 3D photogrammetry modeling software packages for drones survey. *The International Archives of the Photogrammetry Remote Sensing and Spatial Information Sciences*, XLII(4/W12), 95-100. <https://doi.org/10.5194/isprs-archives-XLII-4-W12-95-2019>
- [25] Lau, B. S., Mohd Ariff, M. F., & Razali, A. F. (2022). Comparison between AliceVision Meshroom and Pix4Dmapper Software in Generating Three-Dimensional (3D) Photogrammetry. *Journal of Advanced Geospatial Science and Technology*, 2(2), 117-138.
- [26] Plaščak, I., Jurišić, M., Radočaj, D., Rapčan, I., & Galić Subašić, D. (2024). Application of an unmanned aerial vehicle for the creation of a precise fertilization map. *Zbornik sažetaka 13. međunarodnog simpozija Kopački rit jučer, danas, sutra*, 194-197.
- [27] Feature Request – Plant health monitoring. *PIX4D*. Retrieved from <https://community.pix4d.com/t/feature-request-plant-health-monitoring/8781/2>
- [28] Psirofonia, P., Samaritakis, V., Eliopoulos, P., & Potamitis, I. (2017). Use of unmanned aerial vehicles for agricultural applications with emphasis on crop protection: three novel case-studies. *Journal of Agricultural Science and Technology*, 5(1), 30-39. <https://doi.org/10.12783/ijast.2017.0501.03>
- [29] Park, J. W., Jeong, H. H., Kim, J. S., & Choi, C. U. (2016). Development of Open source-based automatic shooting and processing UAV imagery for Orthoimage Using Smart Camera UAV. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLI-B7, 941-944. <https://doi.org/10.5194/isprs-archives-XLI-B7-941-2016>
- [30] Delgado-Vera, C., Aguirre-Munizaga, M., Jiménez-Icaza, M., Manobanda-Herrera, N., & Rodríguez-Méndez, A. (2017). A Photogrammetry Software as a Tool for Precision Agriculture: A Case Study. In: Valencia-García, R., Lagos-Ortiz, K., Alcaraz-Mármol, G., Del Cioppo, J., Vera-Lucio, N., & Bucaram-Leverone, M. (Eds) *Technologies and Innovation. CITI 2017. Communications in Computer and Information Science*, vol 749. Springer, Cham. https://doi.org/10.1007/978-3-319-67283-0_21
- [31] Mattivi, P., Pappalardo, S. E., Nikolić, N., Mandolesi, L., Persichetti, A., De Marchi, M., & Masin, R. (2021). Can commercial low-cost drones and open-source GIS technologies be suitable for semi-automatic weed mapping for smart farming? A case study in NE Italy. *Remote sensing*, 13(10), 1869. <https://doi.org/10.3390/rs13101869>
- [32] OpenDroneMap Update: Cloud Optimized GeoTIFFs, Plant Health, Histograms and More. *OpenDroneMap*. Retrieved from <https://www.opendronemap.org/2019/12/opendronemap-update-cloud-optimized-geotiffs-plant-health-histograms-and-more/>
- [33] Aguilar M. A., Aguilar F. J., & Negreiros, J. (2009). Off – the - shelf laser scanning and close-range digital photogrammetry for measuring agricultural soils microrelief. *Biosystems Engineering*, 103(4), 504-517. <https://doi.org/10.1016/j.biosystemseng.2009.02.010>
- [34] Will 3D Laser Scanning Services End World Hunger? Retrieved from <https://arrival3d.com/will-3d-laser-scanning-end-world-hunger/>
- [35] Radočaj, D., Šiljeg, A., Marinović, R., & Jurišić, M. (2023). State of Major Vegetation Indices in Precision Agriculture Studies Indexed in Web of Science: A Review. *Agriculture*, 13(707). <https://doi.org/10.3390/agriculture13030707>
- [36] Xue, J., & Baofeng, S. (2017). Significant Remote Sensing Vegetation Indices: A Review of Developments and Applications. *Journal of Sensors*, 2017(1), 1-17. <https://doi.org/10.1155/2017/1353691>
- [37] Thomsen, L. M., Baartman, J. E. M., Barneveld, R. J., Starkloff, T., & Stolte, J. (2015). Soil surface roughness: comparing old and new measuring methods and application in a soil erosion model. *SOIL Discussions*, 1(1), 399-410. <https://doi.org/10.5194/soil-1-399-2015>
- [38] Merrill, D. (1998). Comments on the chain method for measuring soil surface roughness: use of the chain set. *Soil Science Society American Journal*, 62, 1147-1149. <https://doi.org/10.2136/sssaj1998.03615995006200040040x>
- [39] Moreno, G. R., Álvarez, M. C., Tarquis, A., Barrington, S., & Saa Requejo, A. (2008). Tillage and soil type effects on soil roughness at semiarid climatic conditions. *Soil and Tillage Research*, 98, 35-44. <https://doi.org/10.1016/j.still.2007.10.006>

- [40] Rieke-Zapp, D. H., & Nearing, M. A. (2005). Digital close range photogrammetry for measurement of soil erosion. *The Photogrammetric Record*, 20(109), 69-87. <https://doi.org/10.1111/j.1477-9730.2005.00305.x>
- [41] Ullmann, T., & Stauch, G. (2020). Surface roughness estimation in the orog nuur basin (Southern Mongolia) using sentinel-1 SAR time series and ground-based photogrammetry. *Remote sensing*, 12(19), 3200. <https://doi.org/10.3390/rs12193200>
- [42] Bitelli, G., Girelli, V., Remondino, F., & Vittuari, L. (2006). Surface modelling of complex archaeological structures by digital close-range photogrammetry. *Bar International Series*, 1568(321), 80.
- [43] Pierzchala, M., Talbot, B., & Astrup, R. (2016). Measuring wheel ruts with close-range photogrammetry. *Forestry: An International Journal of Forest Research*, 89(4), 383-391. <https://doi.org/10.1093/forestry/cpw009>
- [44] Tomatoes 3D Modelh. *Free3D*. Retrieved from <https://free3d.com/3d-model/tomatoes-4820.html>
- [45] How to compare two 3D entities. *CloudCompare*. https://www.cloudcompare.org/doc/wiki/index.php/How_to_compare_two_3D_entities

Authors' contacts:

Domagoj Zimmer, PhD, Assistant Professor
(Corresponding author)
Faculty of Agrobiotechnical Sciences Osijek,
Vladimira Preloga 1, 31000 Osijek, Croatia
dzimmer@fazos.hr

Mladen Jurišić, PhD, Full Professor
Faculty of Agrobiotechnical Sciences Osijek,
Vladimira Preloga 1, 31000 Osijek, Croatia
mjurisic@fazos.hr

Hrvoje Glavaš, PhD, Full professor
Faculty of Electrical Engineering, Computer Science and Information Technology
Osijek, University of Osijek,
Kneza Trpimira 2B, 31000 Osijek, Croatia
hrvoje.glavas@ferit.hr

Mirko Karakašić, PhD, Full professor
University of Slavonski Brod, Mechanical Engineering Faculty in Slavonski Brod,
Trg Ivane Brlić-Mažuranić 2, 35000 Slavonski Brod, Croatia
mirko.karakasic@unisb.hr

Toni Beljo, MSc
Faculty of Agrobiotechnical Sciences Osijek,
Vladimira Preloga 1, 31000 Osijek, Croatia
tbeljo@fazos.hr

The Crucial Role of Graphic Design in Shaping User Experience: A Study Case

Amanda Wiana, Menul Teguh Riyanti, Virginia Suryani Setiadi, Bambang Tri Wardoyo, Muhammad Hafiz Hassan, Wegig Murwonugroho*

Abstract: This study explores the integral role of graphic design in shaping user experiences on digital platforms, particularly on Instagram. Using Peirce's semiotic method, the research combines analyses of icons, indexes, and signs within the Instagram interface on user engagement, usability, and emotional connection. Findings indicate that the intentional design of these elements significantly influences users' initial engagement and overall satisfaction. However, frequent and disruptive changes to the graphic design, especially layout modifications, can negatively impact the user experience by disorienting users, reducing efficiency, and creating frustration. The study underscores the importance of thoughtful and consistent graphic design in creating engaging and user-friendly digital platforms. By understanding the semiotic significance of design elements, designers and developers can enhance user experiences and improve the overall platform performance.

Keywords: digital platform; graphic design; study case; user experience; user interface

1 INTRODUCTION

In the ever-evolving digital era, the role of graphic design in shaping user experience has become a critical aspect that influences interactions across various platforms. Graphic design not only provides aesthetic elements but also significantly impacts how users interact with digital media. Graphic design includes the use of visual elements, such as color, typography, and images, which play an important role in conveying messages and creating a brand identity [1]. The definition of graphic design in this context is the art and practice of combining visual elements to create effective communication. With a focus on digital platforms, the media studied spans a wide range of forms, from websites to mobile applications, where graphic design serves as the primary means of conveying information and guiding users through their digital experience [2]. The media characteristics studied include various aspects, such as page layout, navigation, and design elements related to brand identity. Additionally, the design should be simple and easy to understand, with optimal readability and clear typography [3].

With the rapid development of digital technology, the use of various digital platforms is becoming increasingly widespread [4]. One of the most significant impacts of digital platforms has been on communication. Social media provide a convenient and accessible means of sharing information, expressing thoughts and opinions, and building and maintaining relationships. One of the most popular social media platforms is Instagram. It is a social media platform that primarily focuses on photo and video sharing. It has rapidly gained popularity due to its visually appealing interface and user-friendly design. Its ability to capture and share moments in a visually striking manner has resonated with millions of users worldwide. The platform's focus on high-quality images and videos has attracted a large number of users who appreciate the aesthetic appeal of visual storytelling. The ability to apply filters and editing tools has empowered users to create visually stunning content that stands out from the crowd. This has led to the emergence of a thriving community of photographers, artists, and influencers who use Instagram to showcase their work and connect with their audience. Instagram's combination of

visual appeal, user-friendly design, and social features has made it a popular and influential social media platform.

Amidst these dynamics, the role of graphic design in creating engaging and effective user experiences is increasingly important. However, graphic design is often overlooked in the digital product development process, even though well-designed visual elements can increase user engagement and increase interaction efficiency [5]. In analyzing the influence of graphic design on user experience, this research will answer the following main questions:

- 1) What is the crucial role of graphic design on Instagram in improving user experience in the digital world?
- 2) What errors or shortcomings in graphic design on Instagram can negatively impact the user experience?

Accordingly, this research aims to reveal and analyze the impact of graphic design on Instagram for improving user experience in digital environments. The influence of graphic design on user experience will be studied through an in-depth analysis of user interactions with these design elements to understand more deeply how graphic design contributes to user satisfaction, engagement, and efficiency in the context of digital media [6]. Through an in-depth understanding of the crucial role of graphic design in various forms of digital media, this research objective is to identify how these visual elements contribute to the level of user satisfaction, engagement, and efficiency. By drawing attention to the interactions between users and graphic design elements, this research is expected to provide more comprehensive insight into how graphic design influences the overall user experience in the digital realm. Consequently, it can help developers, designers, and other stakeholders to optimize graphic designs to achieve more positive and satisfying results in the use of digital media.

2 LITERATURE REVIEW

In an increasingly digitally connected world, graphic design is not just an aesthetic addition but the foundation that shapes the entire user experience [7]. Graphic design as a discipline involves the use of visual elements, including color, typography, and images, to create communication that

is not only effective but also engaging [8]. Graphic design theory provides an important foundation for practitioners and researchers to understand the basic principles that guide the process of creating attractive and functional interfaces. One of the central concepts in graphic design theory is visual hierarchy, which highlights the importance of arranging visual elements in such a way that the user's eyes naturally fall on the most relevant and significant information. In the context of user experience, visual hierarchy is key to guiding users through an interface in an intuitive way, thereby increasing understanding and engagement [9].

Moreover, the concept of usability underlines the importance of designs that can be used easily by users without significant problems. An easily accessible interface, with clear navigation and intuitive understanding, is key to creating a positive and efficient user experience. By detailing these principles, graphic design theory provides an in-depth look at how design can influence user interactions [10]. This combination of visual elements creates a form of communication that goes beyond words, creating a visual narrative that is evocative and empowering [11]. In the digital era, graphic designers and UX (User Experience) professionals have increasingly greater responsibilities. They not only create attractive interfaces but also consider the overall user experience. Users are taken through a visual journey, from color selection to carefully designed page layouts, shaping their perception of the digital platform or product.

The visual design of a digital platform or product plays a crucial role in shaping user perception and experience. From the initial color palette to the carefully crafted page layouts, every visual element contributes to the overall aesthetic and functionality of the interface. Color, as a key element in graphic design, plays a central role in shaping user experience in various digital environments [12]. The implementation of color psychology in digital settings includes the use of colors to achieve specific goals [13]. The appropriate implementation of color psychology can significantly increase user satisfaction and engagement in digital settings [14]. Choosing the right color can create an atmosphere that supports the intended use, increases visual appeal, and builds an emotional connection with users [15].

Using fonts that suit the context and purpose of the communication is also vital. The choice of font, size, and text arrangement is not just a technical aspect of graphic design; it is an art that plays an important role in shaping the user experience [16]. Typography is not just a tool for conveying information; it also creates a visual identity. It is a medium where messages are conveyed, accessibility is enhanced, and typographic aesthetics create an unforgettable brand identity [17]. Carefully chosen typography can improve accessibility, ensuring that information is accessible and understandable to a wide range of users [18]. Text size also plays an important role. In responsive design, text size must be adjusted to remain consistent and easy to read across devices. Choosing contrast between text and background colors also contributes to readability.

Furthermore, page layout is another essential aspect of visual design. The arrangement of elements on a page can significantly impact user experience [19]. A well-designed layout should be easy to navigate, with a clear visual

hierarchy and a logical flow of information. Designers should use white space effectively to create a sense of balance and avoid clutter. Visual design also plays a crucial role in user engagement. The use of high-quality images, illustrations, and animations can enhance the visual appeal of the interface and make it more engaging. However, it is important to ensure that visual elements are not overwhelming and do not detract from the core functionality of the platform [20].

In addition to aesthetics, visual design can also improve usability. Utilizing clear and consistent visual cues enables designers to improve users' understanding of how to engage with the interface effectively [21]. For example, the use of buttons, icons, and other visual elements can provide clear indications of functionality. A well-designed interface should be intuitive and easy to use, even for users who are not tech-savvy. Knowing content media enhanced the users' fantasies and reinstruction with pleasant feelings [22].

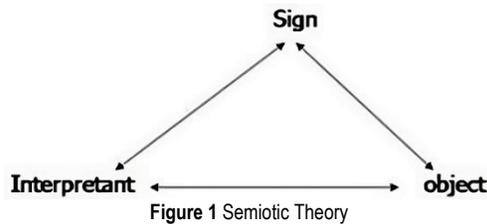
In this overall context, it is important to realize that graphic design is not just about aesthetics. Graphic design is a visual language that can strengthen messages, shape brand identity, and build emotional connections with users [23]. It is a tool that allows the transmission of information, ideas, and emotions through globally accessible media. A carefully planned combination of visual elements creates more than just an image but a visual narrative that speaks to the user through each carefully selected element. Therefore, an in-depth exploration of the impact of graphic design on user experience is not only relevant but also urgent to advance our understanding of the ever-evolving digital world.

3 METHODOLOGY

By using Peirce's semiotic theory, this research becomes very relevant when it is associated with the influence of graphic design on user experience. In this context, Pierce's semiotic approach allows research to describe in detail the graphic design elements involved, analyze the interactions between these elements, and interpret their impact on user experience [24]. This method allows researchers to detail every aspect of graphic design, such as the use of color, layout, typography, and other visual elements. An in-depth analysis of these factors provides a better understanding of how graphic design can influence user perception and response to a product or service. This research can also identify correlations between certain design choices and levels of user engagement, satisfaction, or even potential confusion. The analysis process provides deeper insight into how graphic design can contribute to creating a positive user experience or vice versa [25].

In addition, the interpretation stage in Pierce's semiotic theory allows researchers to interpret findings by assembling a broader context, including how graphic design can influence brand perception, company image, or even user attitudes toward a product or service. By connecting this descriptive approach to the impact of graphic design on user experience, this research provides a powerful tool for uncovering the complexity of interactions between design elements and user responses, providing a deeper understanding of how graphic design can be improved. The use of Pierce's semiotic theory in research regarding "the influence of graphic design on user experience" offers a

structured approach and allows systematic measurement of the variables involved [26].



The use of sign semiotics, according to Charles Sanders Peirce's theory in research regarding "the influence of graphic design on user experience", can provide in-depth insight into how visual elements in graphic design form meaning and influence user perceptions. The following is the implementation of sign semiotics in this context:

- a) **Icon.** An icon is a sign that has a physical or visual similarity to the object it represents. In the UI design, the camera button has a camera-shaped icon that depicts the physical object, then that is an example of an icon. These icons allow users to recognize the button's function quickly. For example, the use of a lock icon represents a security page in a banking application design. Users can quickly identify that the lock icon indicates a security-related area.
- b) **Symbol.** A symbol is a sign that has a conventional relationship or social agreement with the object it represents. Users generally identify the arrow symbol pointing to the right as a button to move to the next page. Although there is no physical resemblance to the concept of "next", social conventions make the symbol easy to understand. For instance, heart symbols indicate "favorites" or "likes" in social media platform designs. The heart symbol is universally recognized as a positive expression, creating a user's emotional connection to the content.
- c) **Index.** An index is a sign that has a cause-and-effect relationship or is directly related to the object it represents. For illustration, an index can be seen in the network signal indicator on the UI design of a mobile device. If it is in a full mark, it indicates a strong signal; meanwhile, if it is blank, it indicates a weak signal. A direct relationship between a sign and a network condition is an example of an index. Another example is a display of delivery status indicators (e.g., a checklist icon or loading animation) when a user sends a message, which provides real-time feedback regarding the status of the message delivery process.

4 RESULTS AND DISCUSSION

In this research, the Instagram application is used as an object of analysis. The analysis of UI/UX on Instagram according to Peirce's Semiotic theory is presented as follows:

4.1 Instagram's Icons

In Peirce's semiotics theory, an icon is a sign that physically resembles the object it represents. Within the

context of the Instagram app, icons play a pivotal role in guiding user interaction and facilitating seamless navigation. These visual signs include navigation and action icons, which assist in providing clear visual signs and streamlining the user experience.

a) Navigation Icons

Navigation icons are essential components of user interface design that guide users through digital environments. They play a crucial role in helping users navigate the app's various features and content. These icons are strategically placed throughout the Instagram interface, providing clear and intuitive pathways for users to explore and interact with the platform.

▪ Top navigation bar

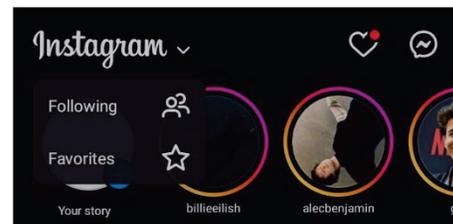


Figure 2 Icons in the top navigation bar

There are several icons at the top navigation bar. First, there is an arrow icon next to the Instagram logo. The arrow icon serves as a visual cue that guides users toward additional options or settings within the Instagram interface. In the context of the top navigation bar, this icon represents the action of switching between the "Following" and "Favorites" feeds. The arrow icon's simple and universally recognizable design contributes to its effectiveness. Its clear and concise representation of the underlying action ensures that users can quickly understand its purpose and utilize it without confusion. This icon suggests a sense of depth or expansion, implying that more options are available beyond the immediately visible content. The meaning assigned to this icon is intuitive; thus, it is easily understandable to the users. This, in turn, can lead to increased engagement and satisfaction with the platform [27].

Second, there is a heart icon. The heart icon is a universal symbol on social media platforms like Instagram. The heart icon in the top navigation bar functions as a notification feature. It notifies users when their posts receive additional likes, comments, or mentions. Its ability to convey positive emotions, foster social interaction, and provide notifications makes it an essential tool for connecting users and building a sense of community. Additionally, the design of the heart icon can evoke specific emotions and associations, influencing users' perceptions of the platform and its content [28]. Hence, the heart icon's design and functionality contribute to a more engaging and rewarding user experience on Instagram.

Third, there is a messenger icon. The messenger icon, a speech bubble with a horizontal lightning bolt inside, is a visual symbol that represents the platform's direct messaging feature. The speech bubble, a universally recognized symbol for communication, is an iconic representation that

immediately conveys the idea of messaging or conversation. On the other hand, the lightning bolt within the speech bubble adds a symbolic dimension that suggests speed, efficiency, and real-time communication. It allows users to share thoughts, ideas, and experiences in a private and intimate setting than in the platform's public sphere. This simple yet effective symbol represents the concept of communication and evokes associations with interpersonal interactions in the real world. This can lead to increased engagement with friends and followers, thereby fostering stronger social connections and enhancing the overall user experience.

- *Bottom navigation bar*



Figure 3 Icons in the bottom navigation bar

The bottom navigation bar has five icons. First, there is a home icon. This familiar symbol allows users to quickly identify and access the home page, where they can discover new content, interact with their feed, and explore the platform's various features. The feed is personalized to each user as it displays posts from accounts they follow. This allows users to stay updated on the latest activities of their chosen influencers and communities easily. Accordingly, it encourages users to engage with posts, leave comments, and share their own content, resulting in a more vibrant and dynamic social media environment. Therefore, the home icon contributes to a more intuitive and enjoyable user experience by providing a reliable and familiar point of reference within the app's interface.

Second, there is a magnifying glass icon. It is a universally recognized symbol of search and discovery. This icon represents the user's ability to actively seek out new content and trends beyond their personalized feed. The magnifying glass icon empowers users to expand their horizons and discover content they might not have encountered otherwise. Beyond its role in content discovery, the magnifying glass icon also functions as a search tool. Users can quickly find relevant posts, profiles, or reels by typing keywords or hashtags into the search bar. So, this streamlined search functionality enhances user experience by facilitating content discovery and fostering community engagement.

Third, there is a plus icon. The plus icon serves as Instagram's creative center, where users can create and share their content. By sharing their own photos, videos, and thoughts, users can connect with others, build their personal brand, and become part of an online community. It empowers users to share their unique perspectives, connect with like-minded individuals, and foster a sense of belonging. Therefore, the plus icon on the bottom navigation bar of Instagram significantly improves user experience by providing easy access to content creation tools, empowering users to share their content, and fostering community engagement.

Fourth, there is a reel video icon. The reel feature is a short vertical video of around 15 to 90 seconds within the app. The icon's design, a stylized film reel symbol, is easily recognized and understood by users. Through a reel video icon, users can express themselves in the form of video and engage with the platform's short-form video feature. The platform's algorithm would suggest reels based on the user's viewing history and preferences so each user would have a personalized and enjoyable experience [29]. Hence, the reel video icon plays a crucial role in the user experience by giving users a way to create and engage with interesting short-form video content.

Fifth, there is a profile icon. The profile picture icon serves as a visual basis for users' personal identities within the platform. Its strategic placement at the bottom right corner of the app interface provides easy access to a user's profile page, where they can view and manage their personal information, posts, and followers. They can showcase their content, interact with followers, and manage their account settings. Additionally, since users can have multiple Instagram accounts, this feature is particularly useful as it provides a convenient way to switch between different profiles. In conclusion, the profile picture icon enhances the user experience by providing a visual representation of the user's identity, facilitating easy navigation, and fostering a sense of personalization and ownership.

b) Action Icons

Action icons are small, symbolic images that represent specific actions or functions within the Instagram app. They are typically found under individual posts or user profiles, which allow users to interact with content directly.



Figure 4 Action icons below the Instagram post

There are four action icons at the bottom of Instagram's posts. First, there is a heart icon. Not only does it serve as a notification icon, but the heart icon also appears below the post as a means of appreciation or liking. Its stylized shape represents the concept of affection, care, and appreciation. This association with positive emotions makes the heart icon a natural choice for indicating the action of liking a post. The act of liking a post can be seen as a form of digital affirmation, a way of acknowledging the content and its creator. This can lead to a sense of belonging and connection, as users feel valued and appreciated by their peers.

Second, there is a comment icon. The comment icon serves as a gateway to social interaction on Instagram. This icon represents the concept of conversation, evoking associations with interpersonal communication in the real world. Users can instinctively understand that it is the means through which they can engage with the content and its creator by leaving a written comment. It allows users to share their thoughts, opinions, and reactions to posts, creating a space for dialogue and discussion. This can create a more interactive and educational experience for users, as they can learn from each other and expand their knowledge. Hence, it

provides users with a means of engaging in conversation and fostering community.

Third, there is a share icon. The share icon, a stylized paper airplane, serves as a symbol of dissemination and distribution on Instagram. This iconic image represents the user's ability to spread content beyond their immediate circle of followers. Users can use the icon to share the post with their friends, followers, or specific groups, effectively extending its reach and visibility. It empowers users to become active participants in the platform's content network. Therefore, the share icon plays a pivotal role in shaping the user experience by fostering community engagement and content discovery.

Fourth, there is a bookmark icon. The bookmark icon is a familiar symbol associated with saving and organizing information. This icon represents the user's ability to store content for future reference, ensuring that they do not miss out on interesting or important posts. This allows them to easily revisit the content at a later time without having to search for it again. Users can also create personalized collections that reflect their interests and preferences. This can help users stay organized, avoid information overload, and easily access relevant content when they need it. In conclusion, the bookmark icon is a valuable tool for Instagram users, providing a way to save and organize content for future reference.

4.2 Instagram's Index

Index in Peirce's semiotics theory refers to the direct or causal relationship between a sign and the object it represents. On Instagram, several elements can be considered indexical signs. First, there is a notification.



Figure 5 Index in the form of notification

Notifications on Instagram are a clear example of an index. They are directly linked to specific user actions or events, such as new posts, likes, comments, or direct messages. It helps users stay connected to their network and participate in conversations. Timely and relevant notifications encourage users to return to the platform, fostering a sense of community and belonging. They keep users informed about the latest activities of their friends and followers, enabling them to stay up to date with the platform's content. Additionally, Instagram's algorithm tailors notifications to individual users' interests and preferences, enhancing the relevance and value of the information presented [29]. Thus, Instagram notifications are critical for maintaining user engagement and connection.

Second, there is a location tag. This feature allows users to add a specific location to a post. This tag typically appears between the username and the post content and serves as a

direct connection between the content and a specific geographical location.



Figure 6 Index in the form of location tag

The location tag allows users to quickly identify where the post was taken and potentially explore other content from that location. This feature is particularly useful for travelers, locals seeking recommendations, or those interested in connecting with people in their community. Hence, location tags have become integral to the Instagram users' experience, significantly impacting user engagement and interaction.

Next, there is a timestamp. This time-based label is usually located below the post content and indicates when the post was published.

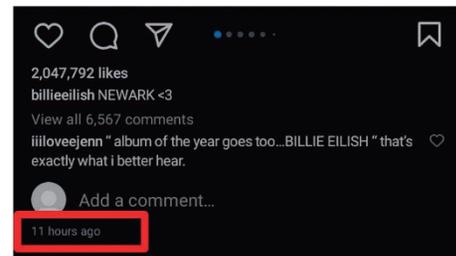


Figure 7 Index in the form of a timestamp

Timestamps provide a temporal context that helps users navigate the platform more effectively and engage with content in a meaningful way. It encourages engagement with the most recent content, fostering a sense of community and immediacy. Moreover, timestamps help users understand the relevance of content in relation to their own timeline. This also can build trust in the platform and its users, as it demonstrates a commitment to transparency and accuracy [30]. Thus, this personalized approach enhances the overall user experience by tailoring recommendations to individual preferences.

4.3 Instagram's Symbol

A symbol is a sign that has no direct relationship to the object it represents but derives its meaning from convention or agreement. This distinguishes it from an icon, which often has a direct visual resemblance to the thing it represents. Instagram hashtags are prime examples of symbols. They do not have any intrinsic connection to the content they tag but rather acquire meaning through their widespread use and understanding by the platform's users. It is symbolized by a

pound sign (#) that allows users to categorize and find content related to a specific topic. It can be found in several places within the Instagram app, such as post captions, comments, bios, and stories.

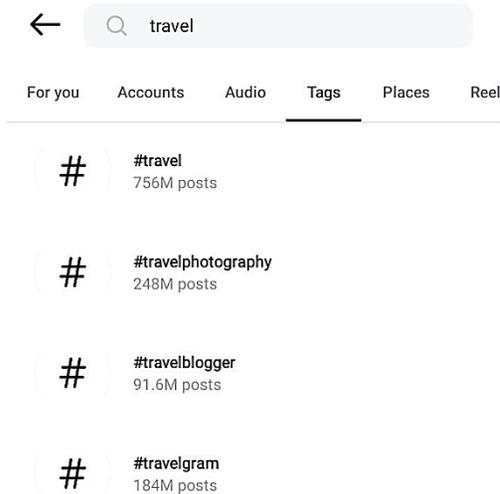


Figure 8 Hashtag symbol

Hashtags have a profound impact on Instagram's user experience. Hashtags on Instagram serve as powerful tools for discovery, categorization, and engagement. They allow users to tag their posts with relevant keywords, making them easily searchable by others who are interested in similar topics. This functionality enhances discoverability, as users can find content such as posts, reels, or stories that align with their interests without having to follow specific accounts. This feature empowers users to participate in conversations and connect with others who are interested in the same things. Therefore, Instagram hashtags serve as a powerful example of how symbols can be used to enhance user experience. Their ability to facilitate content discovery, categorization, and community building makes them an essential feature of the platform.

Hence, the Instagram UI design, characterized by its intuitive navigation and icons, significantly contributes to a positive user experience. The top and bottom navigation bars provide easy access to core features, while the index elements like the notification bar, location tag, and timestamps enhance interaction and context. Additionally, the hashtag symbol serves as a powerful tool for discovery and engagement. By carefully considering these elements, Instagram has created a user-friendly interface that facilitates seamless content exploration and interaction within the platform.

However, Instagram is also known for frequently updating its layout design, which makes many users complain about the frequent adaptation. It has undergone several UI design changes over the years, receiving mixed reactions from users. Some of these changes may result in some users finding the new interface difficult or uncomfortable [31]. Here are some examples and considerations:

1) **Additional Icon.** Instagram's continuous evolution involves introducing new features, such as Reels and shopping features. These changes result in modifications to the UI design. In this case, the reels and shopping buttons on Instagram were added to the bottom navigation bar.



Figure 9 Reels and shopping button

Placing the Reels and Shopping buttons prominently in the bottom navigation bar increases their visibility and accessibility, making them easier for users to discover and use. Nevertheless, adding these two new buttons to the bottom navigation bar may make the interface feel cluttered and less intuitive for some users, especially if they are accustomed to the previous navigation layout [32]. Consequently, the introduction of new features or buttons could disrupt the established workflows and potentially reduce the overall user experience.



Figure 10 Exchange of the DM button with the explore/ search button

2) **Layout Changes.** Apart from adding new icons, Instagram has also changed the location of the icons/buttons several times. For instance, the exchange of the DM button with the explore/search button.

The layout change may confuse users who are used to the previous design, leading to frustration and a less positive overall experience. It can also disrupt the established user flow since users develop mental models of how to navigate the platform, and when these models are disrupted, it can create cognitive dissonance and hinder efficient interaction.

Thus, Instagram's frequent UI design changes can lead to mixed reactions from users. While these updates aim to

improve the user experience, they can also disrupt established workflows and potentially cause confusion or frustration for those accustomed to the previous design [33]. As Instagram continues to evolve, it is essential to carefully consider the potential impact of UI changes on user satisfaction and ensure that updates are implemented in a way that minimizes disruption and enhances the overall user experience. Designers should carefully consider the potential impact of changes on existing users and strive to minimize disruption. Clear communication about updates and gradual changes can help mitigate user resistance and ensure a smooth transition to new designs. Regular user testing and continuous improvement through feedback help improve the quality of the experience [34].

Therefore, Instagram's strategic use of icons, indexes, and symbols significantly enhances user experience by fostering intuitive navigation and understanding. These visual cues streamline interactions, reducing cognitive load and allowing users to engage with the platform efficiently. However, while aiming to improve the user experience, the platform's frequent updates can sometimes disrupt the established familiarity, potentially leading to temporary confusion as users adapt to the new interface.

5 CONCLUSION

Graphic design is not just about visual aesthetics but also about how messages are conveyed and how users interact with information. The quality of user experience is greatly influenced by choosing the right colors, structured layout, and clear navigation. Design responsiveness to multiple devices is becoming increasingly important, given the variations in digital device usage. Additionally, designs that consider inclusivity and accessibility show concern for user diversity. The interactive responses and feedback provided to users help create an informative experience and guide users through the interaction process. Through regular user testing and continuous improvement, designers can continually improve their designs according to user feedback. The task-based approach also emphasizes the importance of design, which supports users in achieving their goals efficiently. Through this research, it can be concluded that there are six principles of UX design with a focus on users, namely understanding the target user, simplicity in layout, good readability, visual consistency, and responsiveness to various devices. By integrating these principles, designers can create graphic designs that are not only aesthetically attractive but also prioritize user comfort and usability in the context of digital technology. In an increasingly digitally connected world, understanding and applying these principles is key to creating satisfying user experiences and meeting the needs of users with diverse backgrounds and preferences.

6 REFERENCES

- [1] Zhang, L., Lin, D., Sun, X., Curth, U., Drost, C., Sauerhering, L., Becker, S., Rox, K., & Hilgenfeld, R. (2020). Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved α -ketoamide inhibitors. *Science*, 368(6489), 409-412. <https://doi.org/10.1126/science.abb3405>
- [2] Tian, Z. (2020). Dynamic Visual Communication Image Framing of Graphic Design in a Virtual Reality Environment. *IEEE Access*, 8, 211091-211103. <https://doi.org/10.1109/ACCESS.2020.3022644>
- [3] Murwonugroho, W. (2019). Mediating role of social media in the memorability of street sculpture art: Jogja Street sculpture Project 2017 as case study. *Wacana Seni Journal of Arts Discourse*, 18, 95-124. <https://doi.org/10.21315/ws2019.18.5>
- [4] Gawer, A. (2022). Digital platforms and ecosystems: remarks on the dominant organizational forms of the digital age. *Innovation*, 24(1), 110-124. <https://doi.org/10.1080/14479338.2021.1965888>
- [5] Salehudin, M., Hamid, A., Zakaria, Z., Rorimpandey, W. H. F., & Yunus, M. (2020). Instagram User Experience in Learning Graphic Design. *International Journal of Interactive Mobile Technologies (IJIM)*, 14(11), 183. <https://doi.org/10.3991/ijim.v14i11.13453>
- [6] Suh, A., & Prophet, J. (2018). The state of immersive technology research: A literature analysis. *Computers in Human Behavior*, 86, 77-90. <https://doi.org/10.1016/j.chb.2018.04.019>
- [7] Yuan, F., Li, S., Yang, B., & Liu, J. (2024). Research on Graphic Visual Communication Design for User Perception Experience. *Applied Mathematics and Nonlinear Sciences*, 9(1). <https://doi.org/10.2478/amns-2024-1347>
- [8] Murchie, K. J. & Diomedede, D. (2020). Fundamentals of graphic design—essential tools for effective visual science communication. *FACETS*, 5(1), 409-422. <https://doi.org/10.1139/facets-2018-0049>
- [9] Johnson, J. (2020). *Designing with the mind in mind: simple guide to understanding user interface design guidelines*. Morgan Kaufmann. <https://doi.org/10.1145/3411763.3444997>
- [10] Huang, S. (2020). The influence of digital multimedia communication forms on graphic design. *Computer Systems Science and Engineering*, 35(3), 215-222. <https://doi.org/https://doi.org/10.32604/csse.2020.35.215>
- [11] ul Haq, A. (2020). *Graphic Acts: Narrative Desire and Design in Comics and Architecture. Doctoral Dissertation*. Northeastern University.
- [12] Suhendra, F. M., Wragg, N., & Barnes, C. (2020). Social Media Research and the Impact of Graphic Design: a case study examining an Indonesian political campaign. In Boess, S., Cheung, M. & Cain R. (Eds.), *Synergy - DRS International Conference 2020*. Held online. <https://doi.org/10.21606/drs.2020.265>
- [13] Khattak, S. R., Ali, H., Khan, Y., & Shah, M. (2021). Color Psychology in Marketing. *Journal of Business & Tourism*, 4(1), 183-190. <https://doi.org/10.34260/jbt.v4i1.99>
- [14] Yu, J., & Egger, R. (2021). Color and engagement in touristic Instagram pictures: A machine learning approach. *Annals of Tourism Research*, 89, 103204. <https://doi.org/10.1016/j.annals.2021.103204>
- [15] Bhandari, U., Chang, K., & Neben, T. (2019). Understanding the impact of perceived visual aesthetics on user evaluations: An emotional perspective. *Information & Management*, 56(1), 85-93. <https://doi.org/10.1016/j.im.2018.07.003>
- [16] Zheng, X., Qiao, X., Cao, Y., & Lau, R. W. H. (2019). Content-aware generative modeling of graphic design layouts. *ACM Transactions on Graphics*, 38(4), 1-15. <https://doi.org/10.1145/3306346.3322971>
- [17] Budelmann, K., & Kim, Y. (2019). *Brand Identity Essentials, Revised and Expanded: 100 Principles for Building Brands*. Rockport Publishers.
- [18] Fisk, A. D., Czaja, S. J., Rogers, W. A., Charness, N., Czaja, S. J., & Sharit, J. (2018). *Designing for Older Adults*. CRC Press.

- <https://doi.org/10.1201/9781420080681>
- [19] O'Hara, J., & Fleger, S. (2020). *Human-System Interface Design Review Guidelines*. <https://doi.org/10.2172/1644018>
- [20] Turja, I. (2020). The effects of visual aesthetics and functionality in user interface design and perception. *Bachelor's thesis*. South-Eastern Finland University of Applied Sciences.
- [21] Miraz, M. H., Ali, M., & Excell, P. S. (2021). Adaptive user interfaces and universal usability through plasticity of user interface design. *Computer Science Review*, 40, 100363. <https://doi.org/10.1016/j.cosrev.2021.100363>
- [22] Murwonugroho, W., & Ardianto, D. T. (2019). Visual fantasy in children's learning through virtual & augmented reality. *International Journal of Scientific & Technology Research*, 8, 12.
- [23] Pan, Z., Pan, H., & Zhang, J. (2024). The application of graphic language personalized emotion in graphic design. *Heliyon*, 10(9), e30180. <https://doi.org/10.1016/j.heliyon.2024.e30180>
- [24] Riyanti, M. T. (2018). Development of Learning Devices Commercial Graphic Based Planning Project. *International Journal of Education, Training and Learning*, 2(1), 1-6. <https://doi.org/10.33094/6.2017.2018.21.1.6>
- [25] Murwonugroho, W., & Yudarwati, G. A. (2020). Exposure to Unconventional Outdoor Media Advertising. *Pertanika Journal of Social Sciences and Humanities*, 28(4). <https://doi.org/10.47836/pjssh.28.4.50>
- [26] Lewis, J. R., & Sauro, J. (2021). *Usability and User Experience: Design and Evaluation*. In *Handbook of Human Factors and Ergonomics*. Wiley. 972-1015. <https://doi.org/10.1002/9781119636113.ch38>
- [27] Jylhä, H., & Hamari, J. (2020). Development of measurement instrument for visual qualities of graphical user interface elements (VISQUAL): a test in the context of mobile game icons. *User Modeling and User-Adapted Interaction*, 30(5), 949-982. <https://doi.org/10.1007/s11257-020-09263-7>
- [28] Bai, H.-Y., Bao, Y.-W., & Kim, J.-H. (2024). Effects of app icon familiarity and authority on app icon image fit and usage intention: focusing on the moderating effects of app icon features and app type. *Asia Pacific Journal of Marketing and Logistics*. <https://doi.org/10.1108/APJML-09-2023-0930>
- [29] Shin, D., Zhong, B., & Biocca, F. A. (2020). Beyond user experience: What constitutes algorithmic experiences? *International Journal of Information Management*, 52, 102061. <https://doi.org/10.1016/j.ijinfomgt.2019.102061>
- [30] Abrahão, S., Insfran, E., Sluÿters, A. & Vanderdonckt, J. (2021). Model-based intelligent user interface adaptation: challenges and future directions. *Software and Systems Modeling*, 20(5), 1335-1349. <https://doi.org/10.1007/s10270-021-00909-7>
- [31] Liu, Y. C. (2024). A New Framework for Evaluating Social Transparency Factors and Personal Brands in Social Networks. *Social Indicators Research*, 171(2), 701-728. <https://doi.org/10.1007/s11205-023-03289-1>
- [32] Buley, L., & Natoli, J. (2024). The user experience team of one: A research and design survival guide. Rosenfeld Media.
- [33] Pham, K. (2024). Optimizing User Experience-A Design-Driven Approach to Enhancing Web Application Interfaces. *Master's Thesis*. University of Gothenburg.
- [34] Chun Tie, Y., Birks, M., & Francis, K. (2019). Grounded theory research: A design framework for novice researchers. *SAGE Open Medicine*, 7, 205031211882292. <https://doi.org/10.1177/2050312118822927>

Authors' contacts:

Amanda Wiana, Student
Universitas Trisakti, Faculty of Art and Design,
JL. Kyai Tapa No. 1 Grogol Petamburan, Jakarta 11440, Indonesia
E-mail: 091302100008@std.trisakti.ac.id

Menul Teguh Riyanti, Lecturer in Visual Communication Design
Universitas Trisakti, Faculty of Art and Design,
JL. Kyai Tapa No. 1 Grogol Petamburan, Jakarta 11440, Indonesia
E-mail: menulteguh@trisakti.ac.id

Virginia Suryani Setiadi, Lecturer in Visual Communication Design
Universitas Trisakti, Faculty of Art and Design,
JL. Kyai Tapa No. 1 Grogol Petamburan, Jakarta 11440, Indonesia
E-mail: virginia@trisakti.ac.id

Bambang Tri Wardoyo, Lecturer in Visual Communication Design
Universitas Trisakti, Faculty of Art and Design,
JL. Kyai Tapa No. 1 Grogol Petamburan, Jakarta 11440, Indonesia
E-mail: bambangtri@trisakti.ac.id

Muhamad Hafiz Hassan, Graphic Design Lecturer
Universiti Teknologi MARA Sarawak Branch, College of Creative Arts,
Persiaran Selangor, Damansara Seksyen 16, 40200 Shah Alam, Selangor,
Malaysia
E-mail: hafizhassan@uitm.my

Wegig Murwonugroho, Professor of Art and Design
(Corresponding author)
Universitas Trisakti, Faculty of Art and Design,
JL. Kyai Tapa No. 1 Grogol Petamburan, Jakarta 11440, Indonesia
E-mail: wegig@trisakti.ac.id

Development of Solar-Powered Herbicide Spraying Robot Controlled Using a Mobile Application

Samuel O. Owoeye*, Folasade O. Durodola, Babatunde M. Bisiriyu, Kehinde M. Adegbenro, Taofeek A. Adetoro

Abstract: The evolution of digital agriculture is significantly influenced by agricultural robots, which contribute uniquely and present numerous benefits to the efficiency of farming operations. The trend of population explosion that the world is experiencing mandates the need to provide an adequate food supply. Hence, a conscious effort must be taken to eradicate weeds on the farmland. This research explored the use of technology to implement a weed elimination system by combining IOT and wireless navigation while also implementing solar energy which encourages green energy initiatives. The research encompasses an unmanned ground vehicle driven by four wheels and controlled by a mobile application that establishes a Bluetooth connection with the system. The ground vehicle was tested to establish the solar charging efficiency, the mobile application connection and operation and the ground vehicle spraying action. The result obtained at the end of the research showed that remote controlled operation of agricultural robots can help improve production efficiency while also reducing environmental pollution with solar charging systems.

Keywords: agriculture; bluetooth; herbicide; mobile application; solar; weed

1 INTRODUCTION

Agriculture is a scientific discipline that employs contemporary technologies grounded in scientific methodologies and principles to enhance yield and production efficiency [1]. These scientific principles include crop breeding, crop protection, crop improvement and veterinary procedures for livestock and animals [2]. The natural rise in global population leads to heightened food demands, and projections indicate that by 2050, the world population may reach nine billion, necessitating a 70% increase in agricultural output [3]. Weeds pose a significant challenge in agriculture due to their unpredictable emergence and competition with crops for essential resources, ultimately leading to reduced crop yields [4]. Weeds emerge sporadically throughout the field, competing with crops for essential resources such as water, nutrients, and sunlight, potentially leading to negative effects on both crop yields and quality if not effectively managed [4-5]. Weed management involves various methods and strategies to control and reduce weed populations [6].

The demand for food is rising continuously, while conventional agricultural practices are proving inadequate. Additionally, the imperative of sustainable farm solutions has become critical in light of significant global issues like energy sustainability and migration crises [7]. The conventional approach to weed control in agriculture relies on manual labour and widespread herbicide application confronts significant issues, such as resource inefficiency and environmental degradation. Indiscriminate herbicide spraying not only escalates operational expenses but also raises ecological concerns, jeopardizing biodiversity [8]. Some farmers still use shoulder-mounted pesticide sprayers, which can release harmful pesticide residues into the air, posing health risks and potential fatalities [9].

The use of herbicides is common in agriculture that depends on agrochemicals, especially due to the promotion of genetically modified crops that can withstand herbicides [10]. The effectiveness of this application method is a major doubt due to how expansive some farmlands can be. Hence,

an automated means of application is required. Solar power is incorporated into the system to explore the effective combination of green energy with robotics to solve the problem of herbicide application in agriculture. This integrated system offers more than efficiency; it harnesses solar energy, enabling independent operation in remote agricultural areas without traditional power sources. Powering onboard sensors and devices with solar energy allows this system to run continuously without frequent recharging, enhancing its scalability and adaptability in diverse agricultural settings. Solar energy stands out as a renewable energy source that is not constrained by limitations in agricultural production. Its applicability transcends factors such as property size, terrain configuration, and specific agricultural sectors [19]. Solar energy has the potential to meet global energy demands through environmentally friendly methods and at a reduced cost [20]. Many countries have not fully explored the potentials of solar energy [22].

Agrobot was designed by [11] for agricultural applications and represents a concept that, once optimized for performance and cost, is expected to enhance efficiency in agricultural spraying operations. The Agrobot is not cost effective and the concept design height is too low for real-life application [21]. Remote-controlled herbicide spraying systems are a better alternative to manual herbicide application as they factor in the health hazards associated with the manual application on labourers and farmers. The justification for this research is to develop a solar-powered herbicide spraying robot, managed through a mobile app to tackle agricultural labour shortages and supports sustainable, precision farming. This eco-friendly, cost-effective solution boosts efficiency, reduces chemical reliance, and prioritizes worker well-being.

2 MATERIALS AND METHODS

2.1 Hardware

The components required for the execution of this research are a buck converter, ESP 32 microcontroller

module, Photovoltaic cell, relay module, ultrasonic sensor, DC motor, DC pump, charge controller, battery, connection cables, metallic frame, wheels, hose with nozzles and tank.

2.1.1 ESP 32 Microcontroller

The ESP32 microcontroller (shown in Fig. 1 below) is a popular choice for Internet of Things (IoT) projects due to its low cost and power consumption, making it ideal for battery-powered devices. Its integration of Wi-Fi and Bluetooth capabilities allows for seamless connectivity to the internet and other devices, making it versatile for a wide range of applications. The dual-core Tensilica Xtensa LX6 microprocessor provides high performance and efficiency, allowing for the smooth operation of multiple tasks simultaneously.

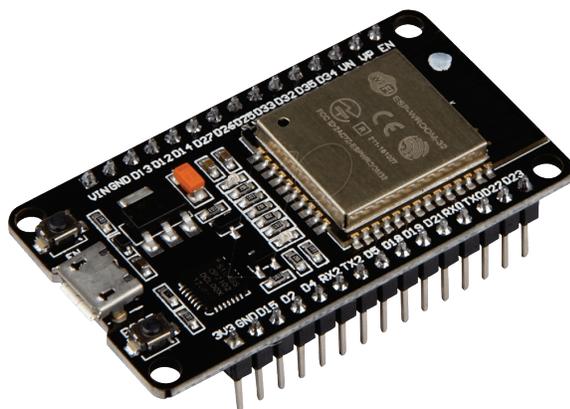


Figure 1 ESP 32 microcontroller



Figure 2 LM 2596 Buck Converter

2.1.2 LM 2596 Buck Converter

The LM2596 regulator, shown in Fig. 2, is a monolithic integrated circuit designed for easy implementation of a step-down switching regulator (buck converter). It handles loads up to 3.0 A with excellent line and load regulation, featuring internal compensation to minimize external components. Operating at a 150 kHz switching frequency, it offers higher efficiency than traditional linear regulators, especially at high input voltages, and allows for smaller filter components. Available in a 5-lead TO-220 package and D2PAK surface mount option, it has a guaranteed output voltage tolerance of

4%, a 15% tolerance on oscillator frequency, and an external shutdown feature with a standby current of 80 μ A. Protection features include cycle-by-cycle current limiting and thermal shutdown. Key specifications include an input voltage range of 3.2 V to 40 V, adjustable output voltage from 1.25 V to 35 V, maximum output current of 3A, up to 92% efficiency, output ripple below 30 mV, and an operational temperature range of -45°C to $+85^{\circ}\text{C}$, with dimensions of $43.2 \times 21.0 \times 14.0$ mm [16-17].

2.1.3 Photovoltaic Cell

Photovoltaic cells also known as solar panels (see Fig. 3) are devices that convert sunlight into electrical energy, providing an alternative power source for the system. Photovoltaic (PV) cells, also known as solar cells, are tiny power generators that convert sunlight directly into electricity. When sunlight hits the PV cell, photons (light particles) are absorbed by a semiconductor material, typically silicon. The absorbed photon energy excites electrons within the semiconductor, freeing them from their atoms. Imagine bumping an electron with enough force to break it free from its usual spot. These freed electrons don't want to stay put. Due to an internal electric field within the cell, it starts to flow through it, creating an electric current. This flow of electrons constitutes electricity that can be used to power electronic devices. This movement of electrons is what we use as electricity.



Figure 3 Photovoltaic cell



Figure 4 12 V AD20P DC Pump

2.1.4 DC Pump

The DC pump (see Fig. 4) is an electromechanical device that helps transmit fluid from one point to another by creating a differential pressure due to a voltage change. The AD20P DC pump 12 V has a brushless permanent magnetic rotor and a very efficient running life of more than 30,000 hours. It features a highly efficient ceramic shaft to reduce wear and

tear, and it employs static sealing instead of dynamic to minimize any leaking problems. The design is flexible to allow the pump to be submerged and fully water-tight while being energy-saving and low noise-emitting at less than 35 dB. It functions with a 12 V input voltage, which can be supplied either from batteries or from solar panels.

2.1.5 DC Motor

The DC electric motor (shown in Fig. 5) is an electrical machine that converts electric power into mechanical energy. The DC motor is the prime mover of the unmanned vehicle. The specification are as follow:

Voltage: 12 V

- Weight: 310 g
- Motor diameter: 37 mm
- Output shaft length: 21mm
- No load speed: 2000 rpm
- Ratio: 1:6.3.



Figure 5 DC Motor

2.1.6 Relay Module

A relay is an electromagnetic switch that opens or closes an electrical circuit when activated by an electric current. A relay module (shown in Fig. 6) is a component that includes a relay mounted on a board with other elements to provide isolation and protection. Its main function is to switch electrical devices on and off, while also isolating control circuits to ensure safety. The relay can be triggered by different voltage signals like 3.3 V or 5 V, making them versatile for use in controlling the pump and DC motors.

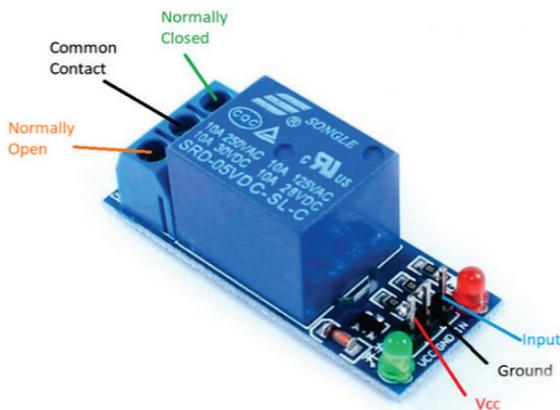


Figure 6 Relay module

2.1.7 Battery

The battery, shown in Fig. 7, converts chemical energy into electrical energy, powering the system. This system will use a rechargeable battery with a voltage of 12 V to support its various components. The other specifications of the battery are:

- Capacity: 7.5 Ah
- Weight: 0.64 kg



Figure 7 Battery

2.2 Software

2.2.1 Mobile Application

The mobile control application was created using MIT App Inventor, a user-friendly visual programming tool from the Massachusetts Institute of Technology (MIT) that allows users with little programming experience to develop Android apps [18]. It features a drag-and-drop interface for building applications with functional blocks. The platform includes a designer layout for the user interface and a block layout for backend operations with the microcontroller. It also supports non-visible components like Bluetooth and Wi-Fi for communication with the ESP32 module used in the research. The mobile application front page layout is shown in Fig. 8 below.

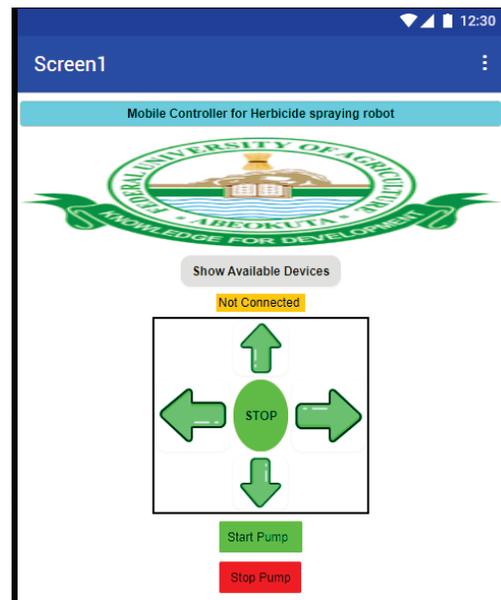


Figure 8 Mobile application layout

2.3 Construction and Assembly of Chassis Frame

The chassis was developed utilizing PTC Creo 9.0 parametric software, incorporating all essential parameters. Fig. 9 illustrates the isometric and lateral views of the CAD model. The chassis structure was constructed from L-shaped metals, which were assembled through arc welding.

2.4 Overview

The system consists of an electrical circuit that controls the robot's movement and pump function, powered by a rechargeable 12 V battery charged via a solar cell and managed by a charge controller to prevent overcharging. The controller has indicators for battery status and a switch to deactivate the circuit. A boost converter ensures a stable 5-volt output for the ESP32 control unit, which manages inputs and outputs. The motor control relays use 4 of the ESP32's 40 GPIO pins to control four motors for speed and direction.

Additionally, a relay module and ultrasonic sensors are connected for pump control and level measurement. The circuit diagram was created using Kicad software, as shown in Fig. 10.



Figure 9 3D CAD model of the chassis

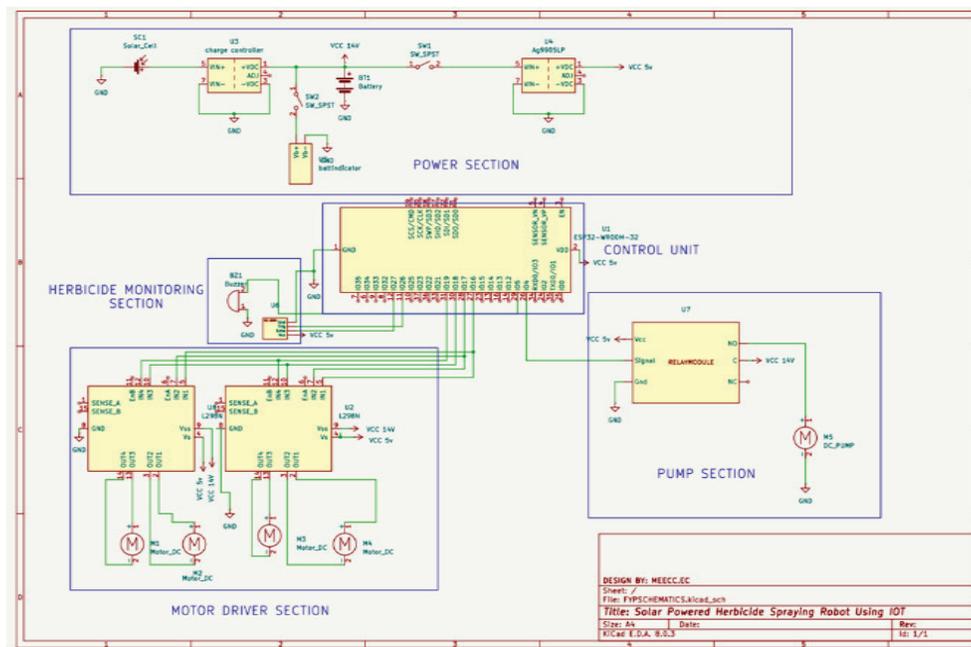


Figure 10 Circuit Diagram

2.5 Design Calculation

2.5.1 Weight Consideration

The power required to operate the system relies on the characteristics of the DC motors utilized for wheel propulsion [12, 13] stem is approximately 8 kg. Hence, selecting the right driving motor that fits the system's functionality is imperative.

$$P = I \cdot V, \quad (1)$$

$$P = m \cdot g \cdot \frac{2\pi \cdot N}{60}, \quad (2)$$

where: P – electric power (W), I – current rating of motor (A), V – voltage rating of the system (V), m – mass to be

driven by the wheel (kg), g – acceleration due to gravity (approximately 10 m/s^2), N – speed of the motor (rpm).

The weight of the motor can calculate by Eq. (3).

$$W = m \cdot g = \frac{60 \cdot P}{2\pi \cdot N} = \frac{60 \cdot I \cdot V}{2\pi \cdot N}. \quad (3)$$

Considering the specification of the motor: $V = 12 \text{ V}$, $I = 5 \text{ A}$ and $N = 75 \text{ rpm}$.

Substituting into the Eq. (3),

$$W = \frac{60 \cdot 12 \cdot 5}{2\pi \cdot 75} = 8 \text{ kg}.$$

Since two motors spraying are considered, the total weight will be 16 kg, up to the required rating.

2.5.2 Power Calculation

The power consumption of each component (shown in Tab. 1) was calculated using the power formula to determine if the battery choice would be able to power all the selected components.

Table 1 Power consumption.

Component	Voltage (V)	Maximum Current (A)	Power (W)
Motor (2 units)	9	2	36
Pump	12	0.3	3.6
ESP 32 module	3.5	0.21	0.735
Relay (3 units)	3.5	0.02	0.21
Buzzer	3.5	0.02	0.07
LCD screen	3.5	0.02	0.07
TOTAL			40.685

The total power supplied by the battery is 90 W. Hence, the battery is sufficient to power the system.

3 TESTING AND RESULT

The completed setup (as shown in Fig. 11) was put to test to determine if the aims of the research was achieved. The mobile application connection, charging via the photovoltaic cell, the vehicle motion and the spraying action were all tested.



Figure 11 Research Prototype

3.1 Mobile Application Connection

The following steps are taken to connect the mobile application to the robot after installing on the mobile device:

- 1) Use Bluetooth to pair to the ESP 32 module named "ESP32test".
- 2) Open the mobile application and select the "Connect device" button.
- 3) Navigate through the available devices option and select "ESP32test".
- 4) The ESP32 module will immediately connect to the application and the connection status on the mobile application will be changed to "connected".
- 5) The buttons on the mobile application can be clicked to navigate the robot in any direction as required. The "start pump" and "stop pump" buttons are used to control the spraying action of the herbicide.

The function of each button implemented on the mobile application is shown in Tab. 2 below.

Table 2 Table showing the functions of the buttons present on the mobile application

S/N	Button	Function
1	Forward arrow	When clicked, this actuates the motors to move forward.
2	Reverse arrow	When clicked, this causes the motors to move backwards.
3	Left arrow	When clicked, this causes the motors to move to the left.
4	Right arrow	When clicked, this causes the motors to move to the right.
5	Stop	When clicked, this stops the motors bringing the vehicle to a halt.
6	Start Pump	When clicked, this actuates the pump to spray the herbicide
7	Stop Pump	When clicked, this turns off the pump and spraying stops.

3.2 Photovoltaic Cell Charge Testing

The power rating of the selected solar panel is 8 W. The voltage output collected at 11:26 GMT under a temperature of 28°C was 16.72V. This was collected by connecting a voltmeter to measure the result. This voltage was sufficient enough to charge the battery. The effectiveness of a solar panel depends on various factors, including its orientation, angle, sunlight exposure, temperature, shading, and the electrical load it's connected to [14]. The performance ratio of the photovoltaic cell is proportional to the incoming radiation on the surface which also corresponds to the quantity of electricity it supplies [23]. The charging rate of the photovoltaic cell will vary under different environmental conditions. Tab. 3 below shows the relationship between the sunlight intensity and current. Figs. 12 and 13 show the graph of light intensity against average power output supplied by the photovoltaic cell.

Table 3 Table showing the relationship between the light intensity and current supplied by the solar cell

Light Intensity (W/m ²)	Charging Current (A)
200	0.13
400	0.27
600	0.40
800	0.53
1000	0.67

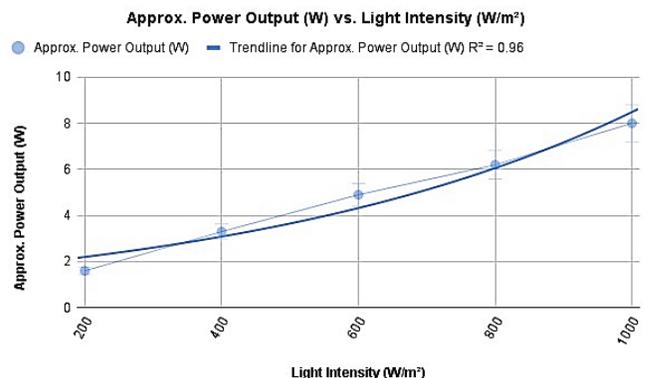


Figure 12 Graph showing the light intensity vs approximate power output

3.3 Spraying Test

The sprayer was implemented with three nozzles to spray a larger cross-section area during the spraying action. The effective flow rate of the pump is 1 L/min. Hence, the flow rate at each nozzle will be 1/3 L/min. The spraying action is immediately activated at the tap of the "start pump" button and it is deactivated at the tap of the "stop pump" button.

3.4 Motion Test

The vehicle's motion was tested on different terrain to observe its speed. The tests were carried out on three different terrains which are: smooth surface floor, rough surface and a low growing lawn. Each surface exhibits distinct properties that can affect the motion of an object, including factors like friction, resistance, and texture. The objective of comparing these different conditions was to analyze how surface characteristics influence the object's speed, travel distance, and overall dynamics of motion. Tab, 4 below shows the time taken to cover similar distance on each surface and Fig. 13 shows a bar chart of the time taken on the different test terrain for the same area.

Table 4 Table showing the time taken to cover an area of 3m² on different terrain.

S/N	Terrain	Area (m ²)	Time (s)
1	Smooth surface Floor	3.0	5.1
2	Rough Surface	3.0	7.3
3	Low Growing Lawn	3.0	13.3

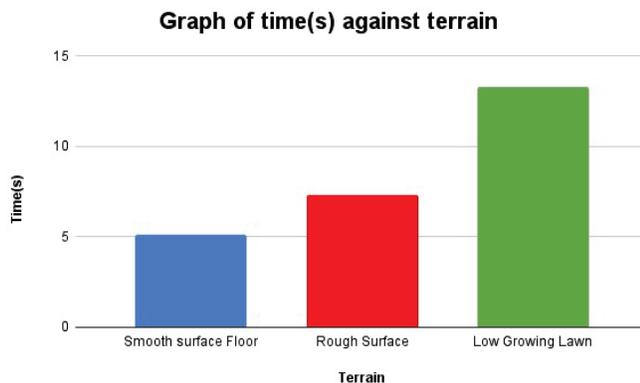


Figure 13 Bar chart showing the time taken to cover the same area on different terrains

4 CONCLUSION AND RECOMMENDATION

A comprehensive assessment of all system functions was performed to validate their alignment with the research's defined goals and objectives. The mobile application, created through the MIT Inventor online platform, effectively interfaced with the robot via a Bluetooth module connected to the ESP 32. This configuration was instrumental in navigating the research course and controlling the pump that facilitated the transfer of herbicide to the nozzles for the purpose of weed elimination in agricultural fields. To augment the system's functionality, several recommendations can be considered:

- Integrating Wi-Fi alongside Bluetooth to enhance connectivity in environments prone to interference.

- Utilizing computer vision for accurate navigation, which would enable real-time trajectory modifications for optimal coverage.

5 REFERENCES

- [1] Pretty, J. (2008). Agricultural sustainability: concepts, principles and evidence. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 363(1491), 447-465. <https://doi.org/10.1098/rstb.2007.2163>
- [2] Palmer, P. G., Curtis, R. A., Ekstrom, S. E., Campbell, G. F., Campbell, J. R., & Thomas, S. B. (2023, November 16). Agricultural sciences. *Encyclopedia Britannica*. <https://www.britannica.com/science/agricultural-sciences>
- [3] Sylvester, G. (2018). E-Agriculture in Action: Drones for Agriculture. *Food and Agriculture Organization of the United Nations and International Telecommunication Union*. Bangkok, Thailand, 11-22.
- [4] Hamuda, E., Glavin, M., & Jones, E. (2016). A survey of image processing techniques for plant extraction and segmentation in the field. *Computers and electronics in agriculture*, 125, 184-199. <https://doi.org/10.1016/j.compag.2016.04.024>
- [5] Berge, T. W., Aastveit, A. H., & Fykse, H. (2008). Evaluation of an algorithm for automatic detection of broad-leaved weeds in spring cereals. *Precision Agriculture*, 9, 391-405. <https://doi.org/10.1007/s11119-008-9083-z>
- [6] Debasish, B. (2023). Different methods of weed management. *Agroscience today*, 4, 660-662.
- [7] Kiani, F., Randazzo, G., Yelmen, I., Seyyedabbasi, A., Nematzadeh, S., Anka, F. A., Erenel, F., Zontul, M., Lanza, S., & Muzirafuti, A. (2022). A Smart and Mechanized Agricultural Application: From Cultivation to Harvest. *Applied Sciences*, 12(12), 6021. <https://doi.org/10.3390/app12126021>
- [8] Otokpa, O. J. (2017). Health impact of the indiscriminate use of herbicides in Nigeria. *Texila International Journal of Public Health*, 5(1), 1-9. <https://doi.org/10.21522/TIJPH.2013.05.01.Art004>
- [9] Dange, K. M., Bodile, R. M., & Srinivasa Varma, B. (2023). A Comprehensive Review on Agriculture-Based Pesticide Spraying Robot. In: Pandit, M., Gaur, M. K., & Kumar, S. (Eds) *Artificial Intelligence and Sustainable Computing (ICSISCE 2022) - Algorithms for Intelligent Systems*. Springer, Singapore. https://doi.org/10.1007/978-981-99-1431-9_28
- [10] Mesnage, R., Székács, A., & Zaller, J. G. (2021). Herbicides: Brief history, agricultural use, and potential alternatives for weed control. *Herbicides - Emerging Issues in Analytical Chemistry*, 1-20. <https://doi.org/10.1016/B978-0-12-823674-1.00002-X>
- [11] Chalwa, V. N., & Gundagi, S. S. (2014). Mechatronics-based remote-controlled agricultural robot. *International Journal of Emerging Trends in Engineering Research*, 2(7).
- [12] Neha, S. A., & Asra, S. (2018). Automated Grass Cutter Robot Based on IoT. *International Journal of Trend in Scientific Research and Development*, 2(5), 334337. <https://doi.org/10.31142/ijtsrd15824>
- [13] Oyelami, A. T., Bamgbose, O. M., & Akintunlaji, O. A. (2023). Mission-Planner Mapped Autonomous Robotic Lawn Mower. *Journal Européen des Systèmes Automatisés*, 56(2), 253-258. <https://doi.org/10.18280/jesa.560210>
- [14] Saleem, A., Mehmood, E. K., & Rashid, E. F. (2016, December). The efficiency of solar PV system. In *Proceedings*

- of 2nd International Multi-Disciplinary Conference, Vol. 19, p. 20.
- [15] LM2596 buck converter datasheet. Theengineeringprojects. <https://www.theengineeringprojects.com/2020/09/lm2596-buck-converter-datasheet-pinout-features-applications.html#:~:text=LM2596%20is%20a%20step-down%20voltage%20regulator,%20also%20known>
- [16] LM2596 Datasheet. ALLDATASHEET.COM. <https://www.alldatasheet.com/datasheet-pdf/pdf/315343/ONSEMI/LM2596.html#:~:text=The%20LM2596%20regulator%20is%20monolithic%20integrated%20circuit%20ideally>
- [17] AD20P-1230C Mini DC12V Brushless Motor Submersible water pump. Etusker.com. <https://etusker.com/product/ad20p-1230c-mini-dc12v-brushless-motor-submersible-water-pump/#:~:text=Nominal%20voltage:%2012V%20DC.%20Maximum%20current:%20400%20mA>.
- [18] About us. Appinventor.mit.edu. <https://appinventor.mit.edu/about-us>
- [19] Ašonja, A., Pekez, J., Janjić, N., & Mikić, D. (2014). The validity for the application of solar energy in irrigation of perennial plants in fruit growing in the Republic of Serbia *Appl. Eng. Lett*, 1(3), 85-90.
- [20] Erten-Ela, S., & Cagatay Cakır, A. (2015). Dye sensitized solar cells for conversion of solar energy into electricity. *Energy Sources, Part A: Recovery, Utilization, and Environmental Effects*, 37(8), 807-816. <https://doi.org/10.1080/15567036.2011.572118>
- [21] Tarun Ayyagari, S., Kizhakke Erakkat, S. K., Srikanth, T. S., & Neerati, M. (2021). Design and fabrication of solar powered remote controlled all terrain sprayer and mower robot. *arXiv e-prints*, arXiv-2106. <https://doi.org/10.48550/arXiv.2106.05236>
- [22] Ašonja, A., & Vuković, V. (2018). The potentials of solar energy in the Republic of Serbia: Current situation, possibilities and barriers. *Appl. Eng. Lett*, 3, 90-97. <https://doi.org/10.18485/aeletters.2018.3.3.2>
- [23] Aksoy, M. H., & Ispir, M. (2023). Techno-economic feasibility of different photovoltaic technologies. *Applied Engineering Letters*, 8 (1), pp 1-9. <https://doi.org/10.18485/aeletters.2023.8.1.1>

Authors' contacts:

Samuel O. Owoeye

(Corresponding Author)
Department of Mechatronics Engineering,
Federal University of Agriculture, Abeokuta, Nigeria
owoeyeso@funaab.edu.ng

Folasade O. Durodola

Department of Mechatronics Engineering,
Federal University of Agriculture, Abeokuta, Nigeria

Babatunde M. Bisiriyu

Department of Mechatronics Engineering,
Federal University of Agriculture, Abeokuta, Nigeria

Kehinde M. Adegbenro

Department of Mechatronics Engineering,
Federal University of Agriculture, Abeokuta, Nigeria

Taofeek A. Adetoro

Department of Mechatronics Engineering,
Federal University of Agriculture, Abeokuta, Nigeria

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 Narrow, 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 Narrow, 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

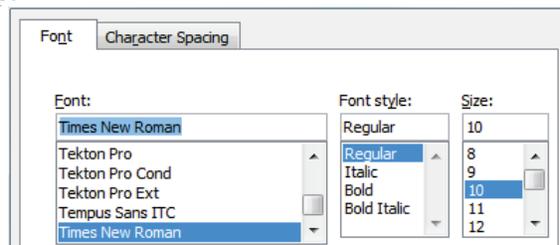


Figure 1 Text under the figure [1]
(Style: Arial Narrow, 8pt, Align Centre)

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 Narrow, 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

$$F_{\text{avg}}(t, t_0) = \frac{1}{t} \int_{t_0}^{t_0+t} F[q(\tau), p(\tau)] d\tau, \quad (1)$$

$$\cos \alpha + \cos \beta = 2 \cos \frac{\alpha + \beta}{2} \cdot \cos \frac{\alpha - \beta}{2}, \quad (2)$$

$$(AB)^T = B^T A^T, \quad (3)$$

$$AAMC = \frac{1}{n} \sum_{i=1}^n PVMC_i. \quad (4)$$

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

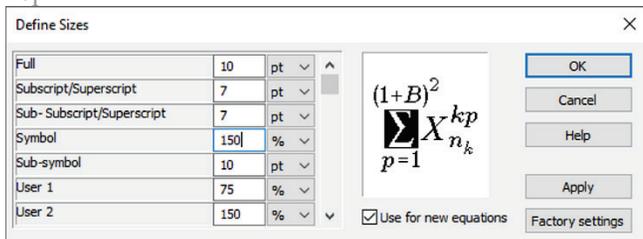


Figure 2 The texts under figures
(Style: Arial Narrow, 8pt, Align Centre)

10pt

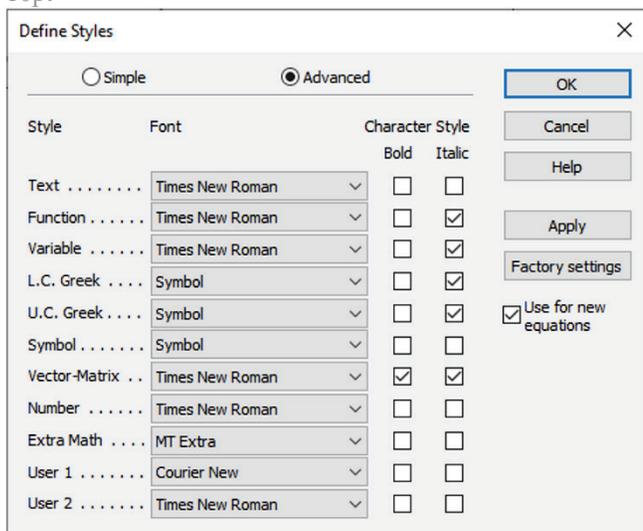


Figure 3 The texts under figures
(Style: Arial Narrow, 8pt, Align Centre)

Figures and tables that are a part of the article have to be mentioned inside the text and thus connected to the content i.e. "... as shown in Fig. 1..." or "data from Tab. 1..." and similar.

10pt

2 PRELIMINARY ANNOTATION

10pt

Article that is offered for publication cannot be published beforehand, be it in the same or similar form, and it cannot be offered at the same time to a different journal. Author or authors are solely responsible for the content of the article and the authenticity of information and statements written in the article.

Articles that are accepted for publishing are classified into four categories: original scientific papers, preliminary communications, subject reviews and professional papers.

Original scientific papers are articles that according to the reviewer and the editorial board contain original theoretical or practical results of research. These articles need to be written in such a way that based on the information given, the experiment can be repeated and the results described can be achieved together with the author's observations, theoretical statements or measurements.

Preliminary communication contains one or more pieces of new scientific information, but without details that allow recollection as in original scientific papers. Preliminary communication can give results of an experimental research, results of a shorter research or research in progress that is deemed useful for publishing.

Subject review contains a complete depiction of conditions and tendencies of a specific domain of theory, technology or application. Articles in this category have an overview character with a critical review and evaluation. Cited literature must be complete enough to allow a good insight and comprehension of the depicted domain.

Professional paper can contain a description of an original solution to a device, assembly or instrument, depiction of important practical solutions, and similar. The article need not be related to the original research, but it should contains a contribution to an application of known scientific results and their adaptation to practical needs, so it presents a contribution to spreading knowledge, etc.

Outside the mentioned categorization, the Editorial board of the journal will publish articles of interesting content in a special column. These articles provide descriptions of practical implementation and solutions from the area of production, experiences from device application, and similar.

10pt

3 WRITING AN ARTICLE

10pt

Article is written in the English language and the terminology and the measurement system should be adjusted to legal regulations, standards and the International System of Units (SI) (Quantities and Units: ISO 80 000 - from Part 1 to Part 14). The article should be written in third person.

Introduction contains the depiction of the problem and an account of important results that come from the articles that are listed in the cited literature.

Main section of the article can be divided into several parts or chapters. Mathematical statements that obstruct the reading of the article should be avoided. Mathematical statements that cannot be avoided can be written as one or more addendums, when needed. It is recommended to use an example when an experiment procedure, the use of the work in a concrete situation or an algorithm of the suggested method must be illustrated. In general, an analysis should be experimentally confirmed.

Conclusion is a part of the article where the results are being given and efficiency of the procedure used is emphasized. Possible procedure and domain constraints where the obtained results can be applied should be emphasized.

AI and AI-assisted tools do not qualify for authorship under TG/TJ's authorship policy. Authors who use AI or AI-assisted tools during the manuscript writing process are asked to disclose their use in a separate section of the manuscript. The publishing agreement process works as usual, with the authors keeping the copyright to their own work.

10pt

4 RECAPITULATION ANNOTATION

10pt

In order for the articles to be formatted in the same manner as in this template, this document is recommended for use when writing the article. Finished articles written in MS Word for Windows and formatted according to this template must be submitted using our The Paper Submission Tool (PST) (<https://tehnickiglasnik.unin.hr/authors.php>) or eventually sent to the Editorial board of the Technical Journal to the following e-mail address: tehnickiglasnik@unin.hr

The editorial board reserves the right to minor redaction corrections of the article within the framework of prepress procedures. Articles that in any way do not follow these authors' instructions will be returned to the author by the editorial board. Should any questions arise, the editorial board contacts only the first author and accepts only the reflections given by the first author.

10pt

5 REFERENCES (According to APA)

10pt

The literature is cited in the order it is used in the article. No more than 35 references are recommended. Individual references from the listed literature inside the text are addressed with the corresponding number inside square brackets i.e. "... in [7] is shown ...". If the literature references are web links, the hyperlink is to be removed as shown with the reference number 8. Also, the hyperlinks from the e-mail addresses of the authors are to be removed. In the literature list, each unit is marked with a number and listed according to the following examples (omit the subtitles over the references – they are here only to show possible types of references):

9pt

- [1] See <http://www.bibme.org/citation-guide/apa/>
- [2] See http://sites.umuc.edu/library/libhow/apa_examples.cfm
- [3] (Style: Times New Roman, 9pt, according to APA)
- [4] Amidzic, O., Riehle, H. J. & Elbert, T. (2006). Toward a psychophysiology of expertise: Focal magnetic gamma bursts

as a signature of memory chunks and the aptitude of chess players. *Journal of Psychophysiology*, 20(4), 253-258. <https://doi.org/10.1027/0269-8803.20.4.253>

- [5] Reitzes, D. C. & Mutran, E. J. (2004). The transition to retirement: Stages and factors that influence retirement adjustment. *International Journal of Aging and Human Development*, 59(1), 63-84. Retrieved from <http://www.baywood.com/journals/PreviewJournals.asp?Id=0091-4150>
- [6] Jans, N. (1993). *The last light breaking: Life among Alaska's Inupiat Eskimos*. Anchorage, AK: Alaska Northwest Books.
- [7] Miller, J. & Smith, T. (Eds.). (1996). *Cape Cod stories: Tales from Cape Cod, Nantucket, and Martha's Vineyard*. San Francisco, CA: Chronicle Books.
- [8] Chaffe-Stengel, P. & Stengel, D. (2012). *Working with sample data: Exploration and inference*. <https://doi.org/10.4128/9781606492147>
- [9] Freitas, N. (2015, January 6). People around the world are voluntarily submitting to China's Great Firewall. Why? Retrieved from http://www.slate.com/blogs/future_tense/2015/01/06/tencent_s_wechat_worldwide_internet_users_are_voluntarily_submitting_to.html
(Style: Times New Roman, 9pt, according to APA)

10pt

10pt

Authors' contacts:

8pt

Full Name, title
Institution, company
Address
Tel./Fax, e-mail

8pt

Full Name, title
Institution, company
Address
Tel./Fax, e-mail

Note: Gray text should be removed in the final version of the article because it is for guidance only.



WIT
CONFERENCES
Call for Papers

BEM/MRM

48

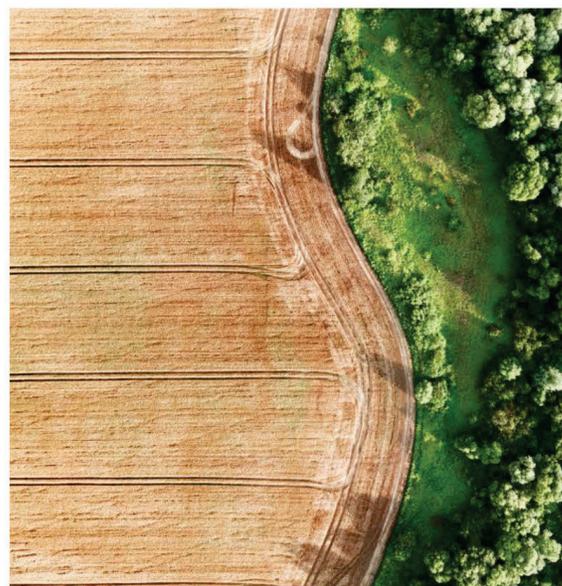
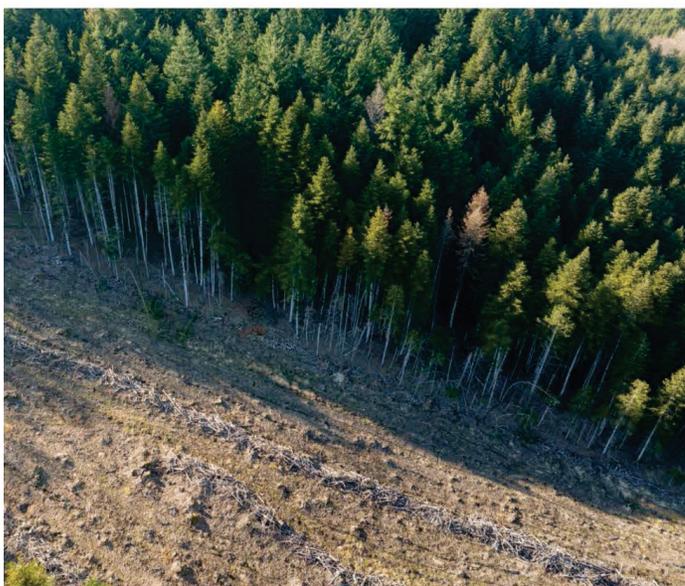
16-18 October 2026 | Dalian, China

Organised by
Dalian University of Technology, CHINA

Conference Affiliated with
Wessex Institute of Technology, UK

Sponsored by
State Key Laboratory of Structural Analysis, Optimization and CAE
Software for Industrial Equipment, CHINA

<https://BEM-MRM48.dlut.edu.cn>



www.witconferences.com/impact2026

TEHNIČKI GLASNIK / TECHNICAL JOURNAL – GODIŠTE / VOLUME 20 – BROJ / ISSUE 1

OŽUJAK 2026 / MARCH 2026 – STRANICA / PAGES 1-172



Sveučilište
Sjever

SVEUČILIŠTE SJEVER / UNIVERSITY NORTH – CROATIA – EUROPE

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

TEHNICKIGLASNIK@UNIN.HR – [HTTP://TEHNICKIGLASNIK.UNIN.HR](http://tehnickiglasnik.unin.hr)