

Ingeniería e Investigación
Journal

Abbreviated Journal Title: **Ing. Investig.**

Editor-in-chief

Sonia C. Mangones, PhD

Associate Editor

Andrés Pavas, PhD, MSc

Technical Editor

Lenin Alexander Bulla Cruz, PhD, MSc

Editorial Assistants

Julian Arcila-Forero, MSc, BSc

Ingri Gisela Camacho, BSc

Editorial Board

Paulo César Narváez Rincón, Ph.D.

Universidad Nacional de Colombia - Bogotá

Cynthia Anbuselvi Ph.D.

SEA College of Engineering and Technology, Indian

Fabio González, Ph.D.

Universidad Nacional de Colombia - Bogotá

Weronika Kruszelnicka Ph.D.

Bydgoszcz University of Science and Technology, Poland

Jayant Rajgopal, Ph.D.

University of Pittsburgh, United States

Óscar Fernando Castellanos Domínguez Ph.D.

Universidad Nacional de Colombia, Colombia

Hilda Elizabeth Reynel-Ávila Ph.D.

Instituto Tecnológico de Aguascalientes, Mexico

Gerasimos G. Rigatos Ph.D.

Unit Industrial Automation, Industrial Systems Institute, Greece

Paulina Jaramillo Ph.D.

Carnegie Mellon University, United States

Fausto Andrés Molina Gómez Ph.D.

Universidad Politécnica de Madrid, Spain

Aquiles Enrique Darghan Contreras, Ph.D.

Universidad Nacional de Colombia, Colombia

Frequency

Continuous periodicity (three issues per year)

Cover Layout and Community Manager

Ian Felipe Guarnizo Martínez

Proofreader

José Daniel Gutiérrez-Mendoza

Layout Artist

David Mauricio Valero

For additional information, please contact:

revii_bog@unal.edu.co

Bogotá - Colombia

2024

Table of Contents

Editorial

Reflecting on Two Years of Progress and Anticipating Future Challenges for
Ingeniería e Investigación

Reflexión sobre dos años de avances y anticipación de desafíos futuros para
Ingeniería e Investigación
Sonia C. Mangones M.

Agricultural Engineering

Application of Photosynthesis Models to Determine Light Requirements in Three
Cacao Clones

Aplicación de los modelos de fotosíntesis para determinar el requerimiento de luz
en tres clones de cacao

Edna I. Leiva-Rojas, Claider Hernández-Taborda, Ramiro Ramirez

A Semi-Supervised Deep Learning Model for Defective lime Classification
Modelo semisupervisado de aprendizaje profundo para la clasificación de limones

*Angel-Moisés Hernández-Ponce, Francisco-Javier Ornelas-Rodríguez, Juan-
Bautista Hurtado-Ramos, Pedro-Alfonso Ramírez-Pedraza, José-Joel González-
Barbosa*

Chemical, Food, and Environmental Engineering

Neural Networks and Fuzzy Logic-Based Approaches for Precipitation Estimation:
A Systematic Review

Enfoques basados en redes neuronales y lógica difusa para la estimación de la
precipitación: una revisión sistemática

*Andres Felipe Ruiz Hurtado, Viviana Vargas-Franco, Luis Octavio González-
Salcedo*

Civil and Sanitary Engineering

Inter-Laboratory Testing Program for the Physical Characterization of Guamo Sand
Programa inter-laboratorio de ensayos para la caracterización física de la arena del
Guamo

*Juan Carlos Ruge, Fausto Molina-Gómez, María C. Olarte, Javier Camacho-
Tauta, Óscar Reyes- Ortiz, Joan M. Larrahondo, Hermes A. Vacca, Luis F.
Prada, Alfonso Ramos-Cañón, Yezid A. Alvarado, Fernando J. Reyes, Miguel
A. Cabrera, Bernardo Caicedo-Hormaza, José S. Naranjo, Iván F. Otálvaro,
Alejandra Gómez-Jiménez, Mayra A. Galvis, July E. Carmona, Cesar A. García,
Allex E. Álvarez, Edgardo J. Díaz, Julio E. Colmenares, Carlos R. Reina, Crithian
C. Mendoza, Diego F. Gil, Laura M. Espinosa, Eliana Martínez-Rojas, Juan G.
Bastidas, Jhan P. Rojas*

Statistical Study on Domestic Gas Boiler Failures Using Various Software Platforms
Estudio estadístico de fallos en calderas de gas domésticas utilizando diversas
plataformas de software

Pavel Shcherban, Reda Abu-Khamdi

Electrical, Electronic and Telecommunications Engineering

Heuristics-Based Energy Demand Forecasting with Scarce Data in the Department of
Huila, Colombia

Pronóstico de la Demanda de Energía Basado en Heurística con Datos Escasos en el
Departamento del Huila, Colombia

Juan Cuenca, Diego Palacios-Castro, Rodolfo García

Genetic Algorithm-Based Optimization of Solar Photovoltaic Integration and
Demand Response for CO₂ Reduction in Indian Coal Power

Optimización basada en algoritmos genéticos de integración de energía solar
fotovoltaica y respuesta a la demanda para la reducción de CO₂ en la energía de
carbón de la India

Vivek Saxena, Saurabh Kumar Rajput

**Facultad de Ingeniería
Universidad Nacional de Colombia**

Sonia Esperanza Monroy Varela
Dean

Mario Enrique Velásquez Lozano
Vice Dean of Research and Extension

Iván Darío Gil Chaves

Vice Dean of Academic Affairs

Gloria Inés Beltrán Calvo

Director of the Students Welfare Service

Papers published in *Ingeniería e Investigación* are abstracted/indexed in

- Science Citation Index Expanded
- (SciSearch®), Clarivate Analytics
- Scopus - Elsevier
- Scientific Electronic Library Online - SciELO, Colombia
- Chemical Abstract
- Índice de Revistas Latinoamericanas
- en Ciencias Periódica
- Dialnet
- Sistema Regional de Información en Línea para
- Revistas Científicas de América Latina, el Caribe, España y Portugal - Latindex
- Ebsco Publishing
- DOAJ - Directory of Open Access Journals
- Redib - Red Iberoamericana de Innovación y Conocimiento Científico

The journal *Ingeniería e Investigación* was created in 1981. It is an entity in charge of spreading the teaching, Scientific, and technical research conducted at Universidad Nacional de Colombia's Department of Engineering and other national and international institutions. *Ingeniería e Investigación* deals with original, unedited scientific research and technological developments in the various disciplines related to engineering. *Ingeniería e Investigación* contributes the development of knowledge, generating a global impact on academia, industry, and society at large through an exchange of knowledge and ideas while maintaining a set of serious and recognized quality standards.

The content of the articles published in this journal does not necessarily reflect the opinions of the Editorial Team. These texts can be totally or partially reproduced provided a correct citation of the source.

Ingeniería e Investigación publications are developed for the academic community who is interested in research and engineering knowledge development. We invite readers to be part of this Journal and participate either as authors, peer reviewers, or subscribers.

For additional information, please contact:

www.revistas.unal.edu.co/index.php/ingevinv

E-mail: revii_bog@unal.edu.co

Tel: 57(1) 3 16 5000 Ext. 13674

Mechanical Engineering, Mechatronics, and Materials Science

Effect of Glycerin and Urea on the Synthesis of Potato and Cassava Starch-Based Biopolymers: Hardness, Micrography, and Thermogravimetric Analyses
Efecto de la glicerina y la urea en la síntesis de biopolímeros a base de almidón de papa y yuca: análisis de dureza, micrografía y termogravimetría
Daniel Nicolas Quintana Mariño, Diana P. Sanabria Chaparro, Hugo Felipe Salazar, Hugo Fernando Castro Silva, Ricardo Alfonso Paredes Roa

Investigating the Effects of Cutting Methods for Aluminum Metallic Foams
Investigación de los efectos de los métodos de corte para espumas metálicas de aluminio

Altug BAKIRCI, Tayfun SIGIRTMAC, Mustafa Cemal CAKIR, Agah UGUZ

Industrial Engineering

Non-linear Dynamic Behavior Identification of a Quadcopter F450 Using an Artificial Neural Network-Based NARX Model
Identificación del comportamiento dinámico no lineal de un quadcopter F450 utilizando un modelo NARX basado en redes neuronales artificiales
Howard E. Sifuentes, Carlos A. Rocha, Edgar A. Manzano

Studying the Mechanical Behavior and Strengthening of RCSACC after Exposure to Elevated Temperatures
Estudio del comportamiento mecánico y refuerzo del RCSACC tras su exposición a temperaturas elevadas
Jean Jacques Kouadjo Tchekwagep, Yiping Qui, Shifeng Huang, Shoude Wang, Xin Cheng

Electromechanical Impedance-Based Damage Detection Using Machine Learning Approaches
Detección de daños basada en impedancia electromecánica mediante métodos de aprendizaje automático
Paulo Elias Carneiro Pereira, Stanley Washington Ferreira de Rezende, Bruno Pereira Barella, José dos Reis Vieira de Moura Junior, Roberto Mendes Finzi Neto

Systems and Computer Engineering

An Intelligent System-Based Strategic Plan for a Humanoid Robot Playing the Game of Dominoes
Plan estratégico basado en sistemas inteligentes para el juego de dominó por parte de un robot humanoide
Álex Medina, Daniela Charris, Mauricio Pardo, Christian G. Quintero M.

Conecta2: An Android App for Communication between Disaster Survivors through an ad hoc Network
Conecta2: una aplicación en Android para la comunicación entre sobrevivientes de desastres a través de una red ad hoc
María del Pilar Salamanca, Elio H. Cables, Juan Camilo Ramírez

Education in Engineering

Digital Transformation: Perspectives and Post-Pandemic Adaptation in the Education Sector – A Case Study
Transformación digital: percepciones y adaptación post pandemia en el sector educativo – un estudio de caso
Jhon Alexander García Camargo, Jesús David Niño Torres, Julieth Andrea López-Castiblanco, Liz Karen Herrera-Quintero, Jorge Iván Bula Escobar

Reflecting on Two Years of Progress and Anticipating Future Challenges for *Ingeniería e Investigación*

Reflexión sobre dos años de avances y anticipación de desafíos futuros para *Ingeniería e Investigación*

Sonia C. Mangones M. ¹

In this document, as Editor-in-Chief of *Ingeniería e Investigación* for the past two years, I will reflect on the substantial advancements we have made while also acknowledging the challenges that lie ahead. With the unwavering support of the Faculty of Engineering of Universidad Nacional de Colombia (UN, Bogotá Campus) and our dedicated editorial team, we have worked to strengthen the Journal's position as a leading platform for the dissemination of high-quality, open-access scientific research in the Global South. This editorial reviews the period spanning 2023 and 2024, highlighting our accomplishments and outlining the Journal's future direction.

One of the most significant developments during this period was the expansion and professionalization of our editorial team. The appointment of Dr. Lenin A. Bulla-Cruz as technical editor has been pivotal in improving the quality and efficiency of our editorial workflows. Similarly, the invaluable contributions of our assistant editors Gisela Camacho Triana and Julián Arcila-Forero have been instrumental in enhancing the speed and quality of manuscript evaluations. Additionally, the revision of our editorial policies {emphasizing ethical rigor and transparency} has provided clear guidelines for both authors and reviewers. These efforts reflect our broader commitment to upholding the Journal's integrity and ensuring clarity in our publishing processes. For up-to-date information about the editorial team, scientific committee, and submission guidelines, we encourage stakeholders to visit our official OJS page: <https://revistas.unal.edu.co/index.php/ingeeinv>

In line with our mission to maintain a modern and accessible journal, we undertook a comprehensive redesign of our visual identity. This update was complemented by the establishment of an integrated digital marketing strategy, which we see as essential in today's rapidly evolving academic landscape. By embracing digital marketing, we aim to increase the journal's visibility, engage with a broader and more diverse audience, and adapt to the changing needs of modern researchers. As academic publishing increasingly shifts to online platforms, maintaining a strong digital presence is vital for improving the discoverability of our work and ensuring that the research we publish reaches its intended audience at both the regional and global levels. Through active engagement on platforms such as X (formerly Twitter, @IelJournalUNAL) and LinkedIn (Ingeniería e Investigación Journal), we are fostering meaningful discussions within our growing scholarly

community. While we are still in the early stages of building our digital presence, we are learning valuable lessons and remain committed to expanding our reach and deepening engagement in the years ahead.

Additionally, I am pleased to report that *Ingeniería e Investigación* has continued to strengthen its position in prestigious scientific journal ranking systems. In the Scimago Journal Rank (SJR), the Journal has maintained its position in the third quartile (Q3) since 2020, with an *h*-index of 18 and a steadily improving SJR value, which rose from 0.190 in 2021 to 0.238 in 2023. The SJR measures a journal's scientific influence by considering both the quantity and quality of citations, reflecting how central the average article is to global scientific discussions. Meanwhile, in the Journal Citation Reports (JCR), we achieved a journal impact factor (JIF) of 1.0 in 2022. Although the JIF adjusted to 0.6 in 2023, we successfully maintained our position in the third quartile (Q3). This represents a challenge to further strengthen the quality and impact of the research we publish. We are committed to meeting this challenge by continuing to enhance the visibility and academic influence of *Ingeniería e Investigación*. For more details in this regard, please visit Scimago's website: <https://www.scimagojr.com>

A key focus of our efforts has been reducing the time from submission to publication, thereby improving the overall experience for both authors and readers. Through strategic investments in editorial management and the optimization of our review processes, we have successfully reduced the average time from submission to initial decision by 30%. This achievement was made possible through streamlined coordination within our editorial team and the implementation of improved workflow systems.

While we take pride in our accomplishments, we remain mindful of the challenges ahead. One of our primary objectives in the coming years is to expand our national and international community. Achieving this goal will involve strengthening collaboration with research institutions, enhancing our indexing and visibility, and continuing to attract high-quality submissions that meaningfully contribute to the field of engineering.

¹Editor-in-Chief at *Ingeniería e Investigación* Journal. Associate Professor at Universidad Nacional de Colombia. E-mail: scmangonesm@unal.edu.co.



Attribution 4.0 International (CC BY 4.0) Share - Adapt

Furthermore, we remain steadfast in prioritizing the ethical integrity of our publishing practices. As part of this effort, we are dedicated to combating predatory publishing and adhering to the highest standards of transparency and accountability.

Looking forward, we are also preparing to introduce a series of specialized workshops and resources aimed at helping authors to deepen their understanding of open science, master digital tools such as LaTeX and OJS, and adopt best practices in academic publishing. These initiatives will reinforce our commitment to engaging with the scholarly community and strengthening *Ingeniería e Investigación*'s role in global scientific dialogue.

As we move forward, *Ingeniería e Investigación* remains dedicated to fostering a rigorous and inclusive scholarly environment. The past two years have been marked by significant strides in editorial excellence, digital engagement, and global recognition. However, our journey is far from over. With the continued support of our editorial team, authors, and UN's Faculty of Engineering, we look forward to building upon these successes and addressing future challenges, ensuring that *Ingeniería e Investigación* continues to thrive as a platform for high-impact research in the field of engineering.

Non-linear Dynamic Behavior Identification of a Quadcopter F450 Using an Artificial Neural Network-Based NARX Model

Identificación del comportamiento dinámico no lineal de un *quadcopter* F450 utilizando un modelo NARX basado en redes neuronales artificiales

Howard E. Sifuentes¹, Carlos A. Rocha², and Edgar A. Manzano³

ABSTRACT

A quadcopter drone is an extremely complex, multi-variable, highly nonlinear, and underactuated system characterized by its six degrees of freedom controlled by only four actuators as inputs. This highlights the importance of employing advanced algorithms for its identification. Therefore, this research aimed to use a nonlinear neural network model to identify the dynamic behavior of a quadcopter based on the commercially available F450 frame. Data acquisition involved four experiments in a controlled environment for both roll and pitch angles, recording the signal duty cycles and the quadcopter's attitude. Then, the selected non-linear autoregressive neural network model with exogenous inputs (N-NARX) model was trained using the acquired data along with the Levenberg-Marquardt algorithm. Afterwards, the response of the quadcopter's actual attitude angles from the validation dataset was analyzed against the predicted values generated by the neural model, obtaining an 89.44% fit with an RMSE of 2.25% for the roll angle and an 89.29% fit with an RMSE of 2.20% for the pitch angle. Both attitude angles were subjected to a statistical cross-correlation validation to assess their relationship at different time lags, observing a solid settling within the confidence bands at a 95% level. It was concluded the proposed neural network model can effectively capture the quadcopter's nonlinear dynamics.

Keywords: drone, system identification, underactuation, neural network, roll, pitch

RESUMEN

El dron *quadcopter* es un sistema extremadamente complejo, multivariable, altamente no lineal y subactuado, caracterizado por sus seis grados de libertad controlados por solo cuatro actuadores como entradas. Esto resalta la importancia de emplear algoritmos avanzados para su identificación. Por lo tanto, esta investigación tuvo como objetivo utilizar un modelo de red neuronal no lineal para identificar el comportamiento dinámico de un *quadcopter* basado en el *frame* comercial F450. La adquisición de datos involucró cuatro experimentos en un entorno controlado para los ángulos de alabeo y cabeceo, registrando los ciclos de trabajo de la señal y la actitud del *quadcopter*. Luego, se entrenó el modelo seleccionado de red neuronal autoregresiva no lineal con entradas exógenas (N-NARX), utilizando los datos adquiridos junto con el algoritmo de Levenberg-Marquardt. Posteriormente, se analizó la respuesta de los ángulos de actitud reales del *quadcopter* en el conjunto de datos de validación frente a los valores predichos generados por el modelo neuronal, obteniendo un ajuste del 89.44 % con un RMSE del 2.25 % para el ángulo de alabeo y un ajuste del 89.29 % con un RMSE del 2.20 % para el ángulo de cabeceo. Ambos ángulos de actitud fueron sometidos a una validación estadística de correlación cruzada para evaluar su relación en diferentes desfases temporales, observándose una sólida estabilización dentro de las bandas de confianza al nivel del 95 %. Se concluyó que el modelo de red neuronal propuesto puede capturar de manera efectiva las dinámicas no lineales del *quadcopter*.

Palabras clave: dron, identificación de sistema, subactuado, red neuronal, alabeo, cabeceo

Received: June 24th 2023

Accepted: September 27th 2024

Introduction

During the last decade, there has been an increase in the implementation of new technologies for unmanned aerial vehicles (UAVs) in multiple fields, e.g., logistics, surveillance and monitoring, public security, among others (Macrina *et al.*, 2020). Many cases can be highlighted, such as the design and construction of the S4 Éhecatl UAV by Hydra Technologies in Mexico, intended for the surveillance and monitoring of hazardous terrain, or the use of a UAV by the Geophysical Institute of Peru in a volcano monitoring project, which flew over the Ubinas volcano in Moquegua over 6000 m above sea level (Saito, 2019).

A quadcopter is a type of UAV that lifts and propels itself with four motors, which control its stability and mobility by varying their rotational speeds (Ahmad *et al.*, 2020).

The most commonly used controller in quadcopters is the classic PID (proportional-integral-derivative) control system, given its ease of implementation and parameter adjustment (Cedro and Wieczorkowski, 2019). Nevertheless, quadcopters pose a significant challenge in terms of control due to their complex nature, as their dynamic model is nonlinear, multivariable, strongly coupled, and underactuated, a result of the six degrees of freedom being

¹ Mechatronics Engineering, Universidad Nacional de Trujillo, Perú. E-mail: howardsg.unt@gmail.com

² Mechatronics Engineering, Universidad Nacional de Trujillo, Perú. E-mail: crocha@unitru.edu.pe

³ Electronics Engineering, Universidad Nacional del Altiplano, Perú. MSc Mechatronics Engineering, Pontificia Universidad Católica del Perú, Perú. Affiliation: Assistant professor, Departamento de Ingeniería Mecatrónica, Universidad Nacional de Trujillo, Perú. E-mail: emanzano@unitru.edu.pe



Attribution 4.0 International (CC BY 4.0) Share - Adapt

controlled by only four actuators (Zhang *et al.*, 2014). Therefore, it would be convenient to implement nonlinear control algorithms to ensure flight stability and robustness against uncertainties such as external airflows or model errors (Santoso *et al.*, 2018).

An appropriate quadcopter controller has to adapt to changes in dynamics and handle uncertainty; it is necessary to identify an accurate dynamic model, as its performance depends on it. To obtain this model, two methods can be used. The first involves modeling the system from a physical analysis using Newton's laws and Euler angles. The second method is called *system identification*, and it involves analyzing input and output data from experimental flights (Pairan and Shamsudin, 2017).

The identification method based on experimental data can be parametric or non-parametric; in the former, a reduced number of parameters are fitted to a predefined mathematical model, whereas, in the latter, more flexible techniques such as neural networks are used to fit the data without imposing a prior mathematical structure (Atteia *et al.*, 2021). Neural network-based identification methods allow obtaining an appropriate dynamic model for quadcopters since they can adapt to the complexity of the system, adjust to any variation, and obtain more precise results in real-time control (Rosales *et al.*, 2019).

System identification can be achieved through linear approaches such as Autoregressive eXogenous models (ARX), Autoregressive Moving Average models with eXogenous variables (ARMAX), and Output Error models (OE). Belge *et al.* (2020) found the OE method to be the linear model with the highest estimation accuracy for a fixed-wing UAV under different noise variations. In a comparative analysis between a linear and a nonlinear Model Predictive Controller (MPC) for trajectory tracking in a Micro Air Vehicle (MAV), the latter showed better performance regarding disturbance rejection capabilities, step response, tracking performance, and computational effort (Kamel *et al.*, 2017).

Recent studies have devised methods to effectively quantify uncertainties in identifying nonlinear systems, with a focus on cascade or block-oriented approaches such as Volterra series and Wiener and Hammerstein models (Xavier *et al.*, 2021). Similarly, there are black-box models, which have been primarily used in the analysis of electrical and control systems. These models do not incorporate prior knowledge, but instead rely on a structure flexible enough to capture all relevant physics in the measured data. Typical approaches include the Nonlinear Autoregressive eXogenous model (NARX), the Nonlinear Autoregressive Moving Average model with eXogenous inputs (NARMAX), artificial neural networks, fuzzy networks, statistical learning theory, and kernel methods (Noel and Kerschen, 2017).

In Altan *et al.* (2018), the trajectory tracking of a hexacopter was studied with a NARX neural network controller aimed at the transportation of physical cargo packages. Moreover, this approach was compared against a PID controller using performance indicators such as the Mean Squared Error (MSE) and the Root Mean Squared Error (RMSE), with the NARX controller yielding much more robust and stable results.

In addition, the implementation of a tensor network B-splines algorithm with the NARX model for nonlinear system identification demonstrated high efficiency and accuracy through numerical experiments in Single Input and Single Output nonlinear systems (SISO), surpassing the RMSE of other nonlinear methods during both prediction and simulation processes (Karagoz and Batselier, 2020).

In a detailed performance comparison between a PID and an advanced controller utilizing NARX neural networks, various indicators such as the MSE and RMSE were utilized (Karakaya and Goren, 2022). The findings revealed a similar performance for both controllers in yaw tests. However, the advanced controller yielded superior results in altitude, roll, and pitch tests, exhibiting an enhanced performance.

In this vein, Ozbek *et al.* (2015) identified the issue of modeling uncertainties in UAVs during a survey of quadrotor aerial robot control, and, in their quadcopter modeling review, Pairan *et al.* (2020) concluded that neural network-based identification is a more suitable option due to its high accuracy, cost, and resource availability. They proposed future research using Radial Basis Functions (RBF) for activation.

The most recent proposals in this field include the works by Ullah *et al.* (2020, 2021, 2022), who presented two sliding mode control designs for UAVs. These proposals were evaluated using mathematical models and simulations, with promising results. However, the progress of these works could be accelerated through more accurate models with the help of system identification.

Consequently, the design of a NARX neural network identification algorithm could be considered in order to obtain a robust model of the nonlinear systems featured by quadcopters. This would ensure proper functioning while avoiding total equipment losses associated with flight control failures.

Nonetheless, the way to ensure system identification is to consider data representing the system's non-linear behavior. Thus, input/output quadcopter data that genuinely exhibit a nonlinear behavior must be verified for use in the training and validation of quadcopter system identification models.

In light of the above, the objective of this research was to identify the non-parametric model that captures the dynamic behavior of a F450 quadcopter, with the purpose of representing the inherent nonlinearities of the system using NARX neural networks. The model was trained with experimental data acquired from the quadcopter's sensors, and a nonlinear data validation was conducted to ensure suitable data for system identification.

System identification methods

ARX

The ARX model, a linear input/output model, is commonly used due to its computationally efficient parameter estimation (Galrinho *et al.*, 2017). This is accomplished directly through the use of a least squares algorithm, which handles linear relationships between parameters and prediction errors (Tu *et al.*, 2020). However, if the actual noise in the process differs from the assumed conditions,

the parameters may deviate from their optimal values, and consistency issues may occur (Nelles, 2020).

As described by Equation (1), the ARX model establishes a relationship between the future value of the dependent output signal $y(k)$, the preceding values of the output signal, and an independent or exogenous input signal $u(k)$. In this context, na and nb represent the model orders, whereas $e(k)$ corresponds to the error of the system, indicating the discrepancy between the predicted and actual output values (Billings, 2013).

$$y(k) = [b_1 u(k-1) + \dots + b_{nb} u(k-nb)] - [a_1 y(k-1) + \dots + a_{na} y(k-na)] + e(k) \quad (1)$$

NARX

The NARX model represents a nonlinear generalization of the ARX model that substitutes the linear relationship described in Equation (1) with an unknown nonlinear function $F(\cdot)$, as outlined in Equation (2), where n_u and n_y represent the input and output memory orders, respectively (Nelles, 2020).

$$y(k) = F[y(k-1), y(k-2), \dots, y(k-n_y), u(k-d), u(k-d-1), \dots, u(k-d-n_u)] + e(k) \quad (2)$$

N-NARX

The neural network NARX representation is referred to as N -NARX. It is a dynamic neural network with a closed-loop architecture that consists of an input layer, a hidden layer, and an output layer. In addition, it incorporates lagged inputs and outputs through explicit integration or a recurrent procedure, as depicted in Figure 1.

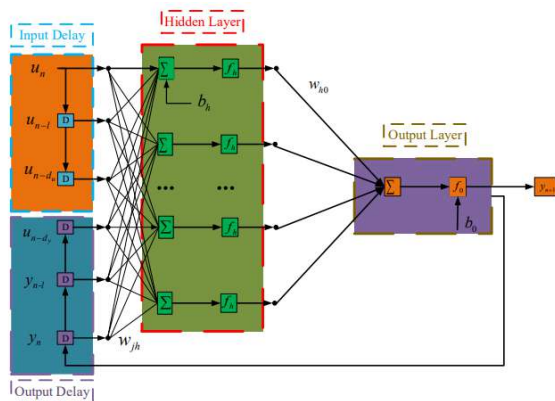


Figure 1. A recurrent single-hidden-layer NARX network
Source: Wei et al. (2020)

Materials and methods

This work was conducted while following the methodology depicted in Figure 2, which involved a systematic review of the available scientific literature to obtain relevant and up-to-date information.

After conducting the literature review, the focus shifted towards implementing the workspace, specifically a test bench, and building a prototype F450 quadcopter. With this

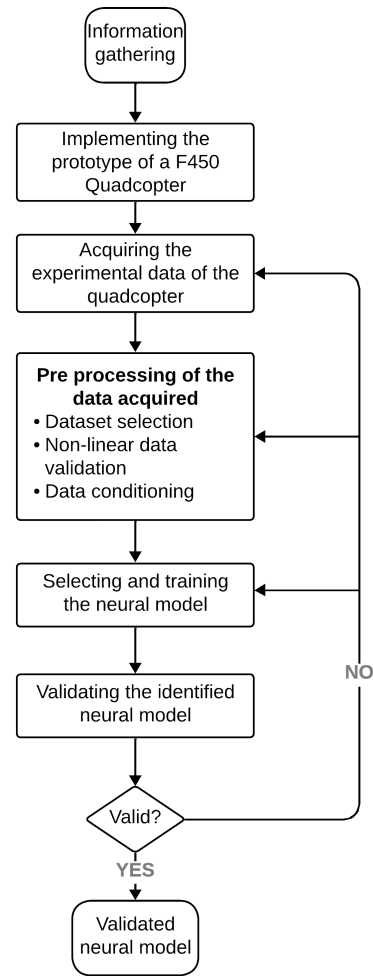


Figure 2. Methodology flowchart
Source: Authors

test bench, experimental data on the roll and pitch angle of the quadcopter were acquired. The input and output data obtained from the quadcopter were pre-processed to reduce noise. Afterwards, a neural network model was selected, its parameters were defined, and it was trained using the conditioned data. Finally, the trained model was validated using performance indicators such as the MSE and the RMSE, as well as through cross-correlation with data that were not used during training.

Implementation

A quadcopter was developed using a commercial F450 frame. The prototype was equipped with four brushless motors, four Electronic Speed Controllers (ESCs), four propellers, and a Lithium Polymer (LiPo) battery. These components were connected to an Arduino UNO hardware development board, along with an MPU9250 IMU sensor, in order to enable precise control of the system's orientation (Figure 3).

The data acquisition process involved the use of specialized equipment, i.e., a test bench specifically designed to collect data on the quadcopter's dynamic behavior. Figure 4 shows the CAD assembly of the test bench with a single degree of freedom (or one axis), which was used to analyze two of the three attitude angles of the F450 quadcopter: roll and pitch.

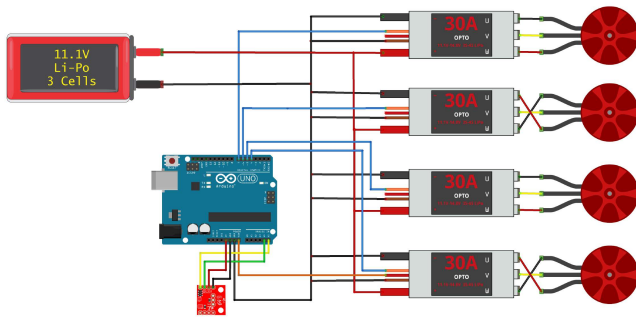


Figure 3. Electronic diagram of the drone
Source: Authors



Figure 4. CAD assembly of a one-degree-of-freedom test bench
Source: Authors

Data acquisition

The data acquisition architecture relied on the test bench, which combined the use of an Arduino board and a computer running the MatLab software. Furthermore, to ensure ease of resetting in case of failure, the Arduino board was located outside the quadcopter.

Figure 5 presents the equipment used for data acquisition alongside a block diagram illustrating the control system of the quadcopter's initial programming, which includes a PID control algorithm with deliberately unsynchronized gains, causing the quadcopter to exhibit an oscillatory motion. The goal of this was to gather crucial insights into the quadcopter's dynamic performance. This motion was critical in capturing the necessary data for subsequent identification processes. Additionally, an algorithm called *Motor Mixer* was implemented to distribute the output of the PID controller into four Pulse-Width Modulation (PWM) signals that served as inputs for the quadcopter's four Electric Propulsion Units (EPUs).

Subsequently, four experiments were conducted for each angle, with each lasting 40 seconds and featuring a sampling time of 10 ms, in order to gather a comprehensive range of data on the quadcopter's dynamic behavior.

The input and output data obtained from Experiment 1 with the roll angle are presented in Figures 6 and 7, respectively.

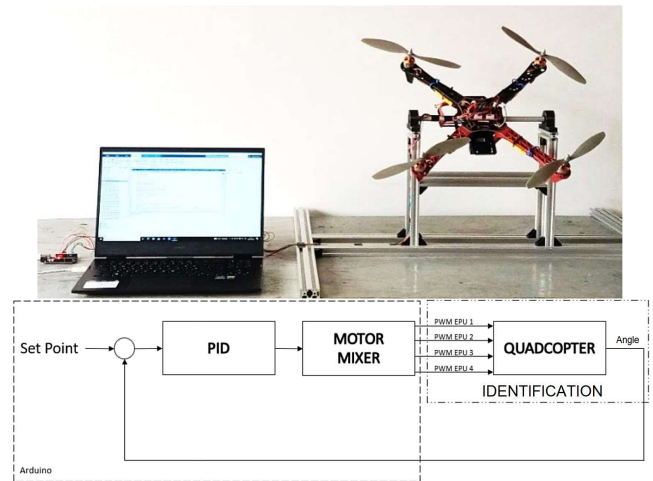


Figure 5. Equipment for data acquisition and block diagram of the control system
Source: Authors

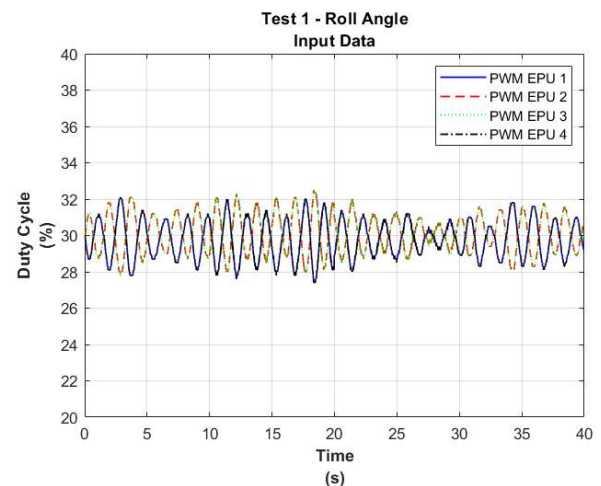


Figure 6. Input data for Experiment 1 – Roll angle
Source: Authors

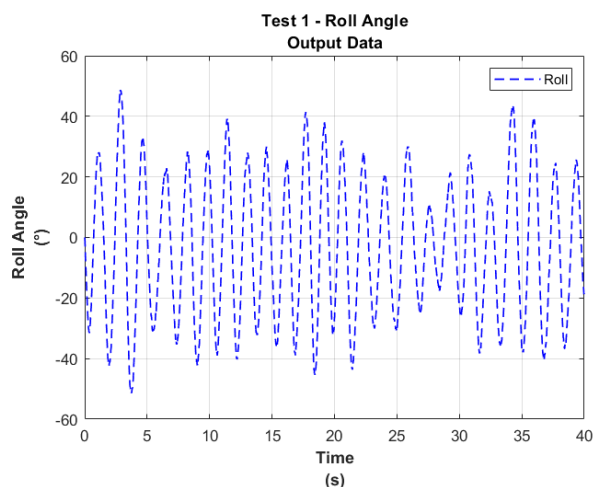


Figure 7. Output data from Experiment 1 – Roll angle
Source: Authors

Data pre-processing

Dataset selection:

The selection of the dataset considered the type of model to be used for the identification process, i.e., a neural network. To train the model, it was important for the data to contain the maximum possible information on the system's dynamics. To ensure this, the dominant frequency from the dynamic system's output data was taken into account alongside its maximum amplitude range.

From the datasets collected during the four experiments conducted for each angle, only two were selected: one for the identification process and the other for cross-validation. Dataset selection for the training process was based on the experiment whose output data had a higher dominant frequency and maximum amplitude range. For the cross-validation process, however, the experiment with lower dominant frequency and a higher maximum amplitude range in its output data would be selected.

The output data for the roll angle data from Experiment 1 are presented in Figure 8, showing a maximum amplitude range of -51.42 to 48.77° .

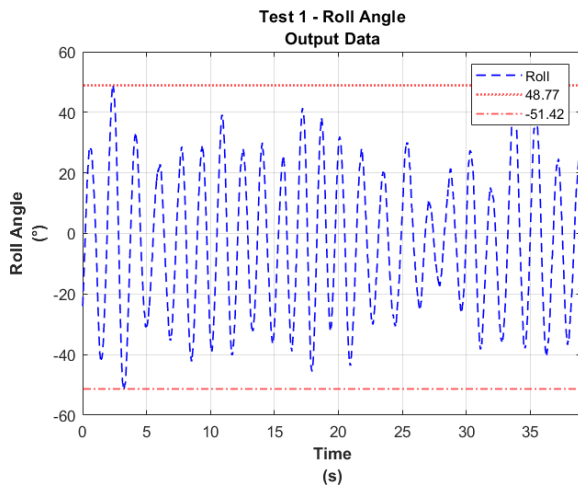


Figure 8. Amplitude of output data from Experiment 1 – Roll angle
Source: Authors

The output pitch angle data from Experiment 1 are shown in Figure 9, indicating a maximum amplitude range of -50.22 to 44.19° .

The Fourier transform was applied to calculate the frequency spectrum of the signal and analyze its dominant frequency. The entire frequency domain of the data was limited to the interval from 0 Hz up to twice the dominant frequency.

The frequency spectrum of the roll angle for Experiment 1 is depicted in Figure 10, indicating a dominant frequency of 0.61523 Hz. The frequency spectrum of the pitch angle from Experiment 1 is presented in Figure 11, revealing a dominant frequency of 0.6665 Hz.

Therefore, based on the criterion of higher dominant frequency, the output data from the first experiments involving the roll and pitch angles were selected as the training dataset for the corresponding neural network model. Similarly, following the criterion of lower dominant frequency and lower amplitude, the data from Experiment 3 were selected for the cross-validation of both angles.

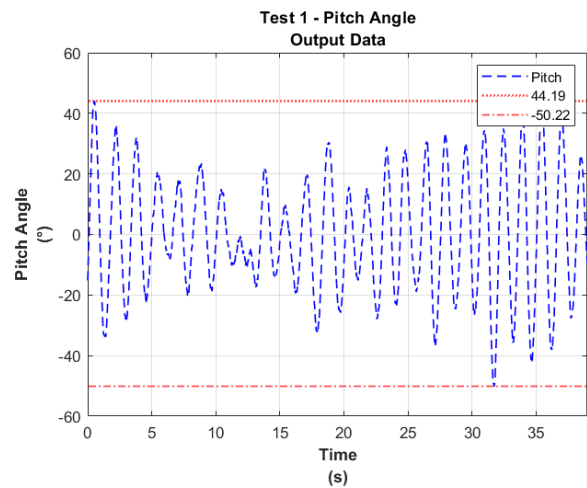


Figure 9. Amplitude of output data from Experiment 1 – Pitch angle
Source: Authors

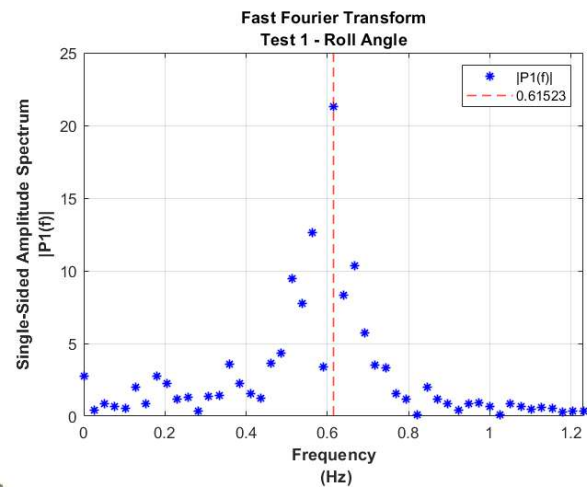


Figure 10. Frequency spectrum of the output data from Experiment 1 – Roll angle
Source: Authors

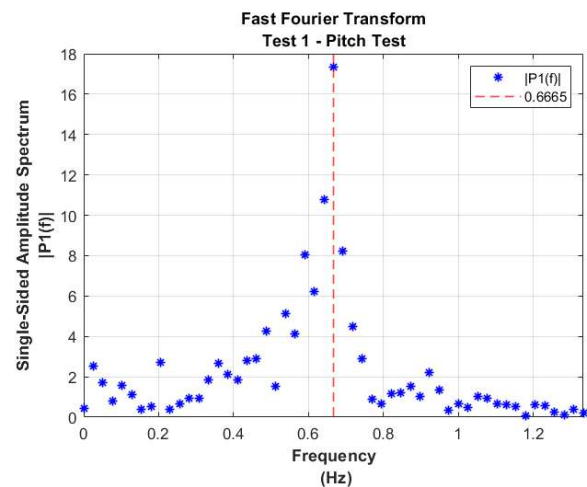


Figure 11. Frequency spectrum of the output data from Experiment 1 – Pitch angle
Source: Authors

Nonlinear data validation:

In order to validate the nonlinear relationship between input and output data (ensuring that the nonlinear behavior of the system would be obtained), two evaluation criteria were considered:

1. The quotient between the RMS value of each wave of the output's oscillatory signal with respect to the RMS value of each wave of the oscillatory signal of the pitch or roll torque.
2. A higher-order correlation function of the system's output signal, as long as the input signal satisfied the nonlinear conditions established by (Billings, 2013). This is described in Equation 3.

$$\begin{cases} \phi_{y'y^2}(\tau) = 0, (\tau = 0, 1, \dots) & \Leftrightarrow \text{linear} \\ \phi_{y'y^2}(\tau) \neq 0, (\tau = 0, 1, \dots) & \Leftrightarrow \text{non-linear} \end{cases} \quad (3)$$

where $\phi_{y'y^2}$ is the cross-correlation between the average level output and its power squared, and τ denotes the torque.

Data conditioning:

Data conditioning was performed for both the input and output of the system. In the case of the input, an internal numerical rounding took place in the Arduino micro-controller during the execution of the Motor Mixer algorithm. On the other hand, the output data exhibited measurement variations as a consequence of the quadcopter's vibrations and the presence of white noise in the signal. These inherent factors contributed to the fluctuations and noise observed in the recorded data.

Figure 12 shows a preliminary demonstration, prior to interpolation, of the one-second sampled stepped input data from the first roll angle experiments.

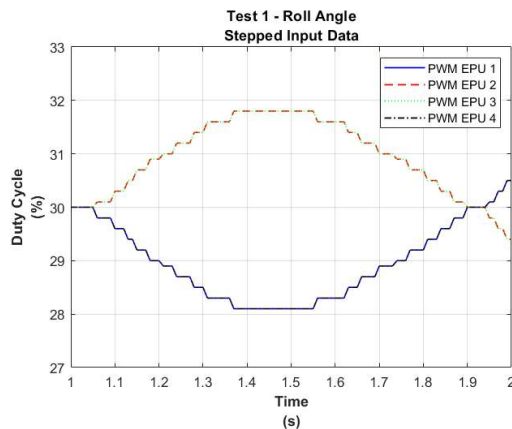


Figure 12. Stepped input data from Experiment 1 – Roll angle
Source: Authors

Afterwards, the acquired stepped input data underwent a smoothing procedure using linear interpolation. This process aimed to obtain a refined dataset that would be more suitable for the subsequent identification.

The interpolated input data from the first set of experiments, with a sample interval of one second, is illustrated in Figure 13 for the roll angle and in Figure 14 for the pitch angle.

It is important to note that this interpolation technique was consistently applied to all the input data gathered throughout the experiments.

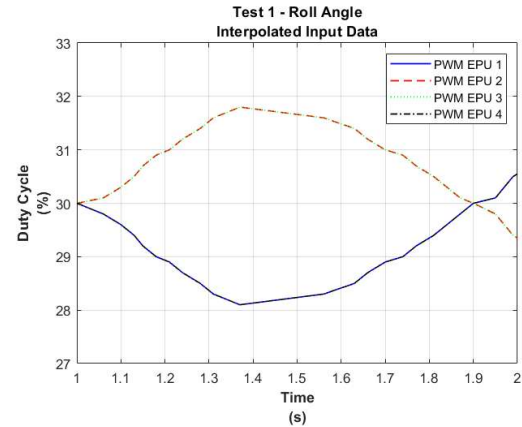


Figure 13. Interpolated input data from Experiment 1 – Roll angle
Source: Authors

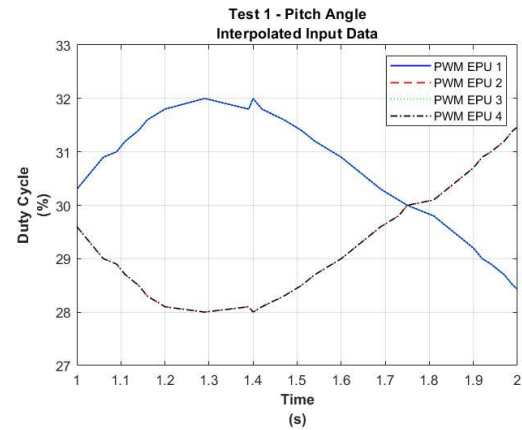


Figure 14. Interpolated input data from Experiment 1 – Pitch Angle.
Source: Authors

The output signals correspond to the roll and pitch tilt angles, which are considered to be the attitude angles of an aircraft with respect to the horizontal plane. Their processing is based on the implementation of the Kalman filter, a robust estimation algorithm that effectively removes inconsistencies and variations in the measurements, thereby ensuring more precise, reliable, and consistent results.

Figure 15 shows a demonstration, prior to filtering, of a one-second sample interval of the output data for the first experiment with the roll angle.

The one-second sample interval of the output data from the first experiments with the roll and pitch angles, as filtered using the Kalman filter, is shown in Figures 16 and 17.

Training the neural network model

Based on Al-Mahasneh et al. (2017), a neural model was chosen for the quadcopter identification process. Comparisons between different neural network approaches were made to determine their usefulness and versatility in this task.

The N-NARX model was selected based on El Dakrory and Tawfik (2016). This model consistently demonstrated its

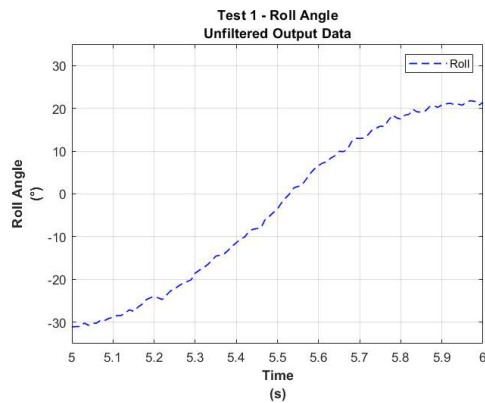


Figure 15. Unfiltered output data from Experiment 1 – Roll Angle
Source: Authors

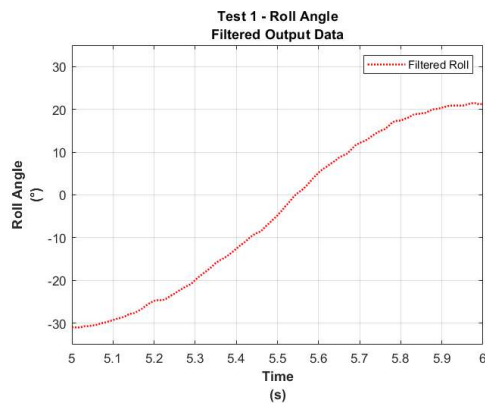


Figure 16. Filtered output data from Experiment 1 – Roll angle
Source: Authors

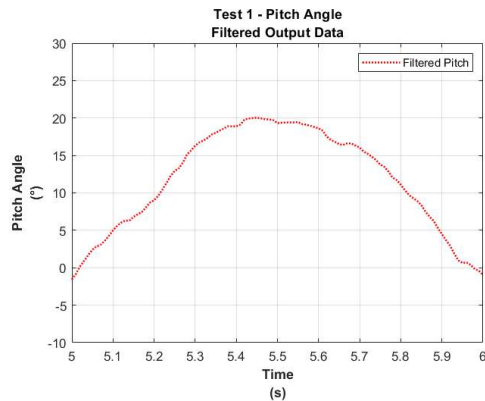


Figure 17. Filtered output data from Experiment 1 – Pitch angle
Source: Authors

promising performance over other methods, surpassing the capabilities of its linear counterpart, the ARX model.

The system architecture (i.e., the number of layers and neurons) followed the mathematical model obtained by Oktay and Kose (2019), with RBF activation functions stemming from prior work by Pairan and Shamsudin (2017). This was used as a starting point for the training and weight adjustment process. The backpropagation training method and the Levenberg-Marquardt algorithm were also employed to minimize the MSE.

Table 1 presents an overview of the features considered while designing the architecture for the selected neural

network. This includes the network type, the number of layers, and the neuron activation function used.

Table 1. Architecture of the selected neural network

Neural architecture features	Description
Model	Neural network
Neural network type	NARX structure
Network layers	One input layer, one hidden layer with five neurons, one output layer
Neuron activation function	RBF

Source: Authors

In the N-NARX model, one lag was considered for the input and three for the output. This decision was based on the literature review, as illustrated in Figure 18 (Pairan and Shamsudin, 2017). As previously mentioned, an RBF was considered.

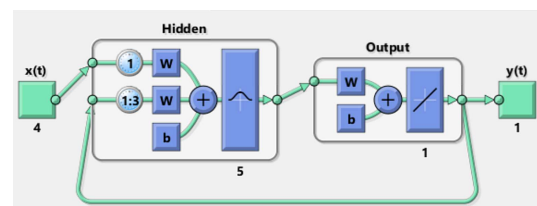


Figure 18. N-NARX model architecture

Source: Pairan and Shamsudin (2017)

After identifying the parameters of the neural network model, the processed roll and pitch angle datasets were trained. Firstly, the roll angle information from Experiment 1 was used, consisting of 3901 input and output data. The N-NARX model's training stopped at 19.48 seconds, achieving a minimized MSE of 4.59.

Meanwhile, for the pitch angle dataset from Experiment 1, also consisting of 3901 input and output data, the training stopped at 2.16 seconds, with a minimized MSE of 1.44.

Results and discussion

Results

Based on the previously mentioned indicators, this study compared the temporal response of the non-parametric quadcopter system's actual attitude angles (from Experiment 1) and those predicted by the N-NARX model.

Two validation steps were implemented:

- Graphical and quantitative validation, showing how closely the neural model's results approximated the real system data, with quantitative metrics such as the MSE, the RMSE, and the fit percentage.
- Cross-validation, using experimental data not employed in training the N-NARX model. This procedure is detailed below.

For the roll angle (Figure 19), the following values were obtained: an MSE of 4.59, an RMSE of 2.14, an RMSE percentage of 2.14%, and a fit percentage of 90.87%.

As for the pitch angle, the comparison is shown in Figure 20, with the following results: a MSE of 1.44, a RMSE of 1.20, a RMSE percentage of 1.28%, and a fit percentage of 93.80%.

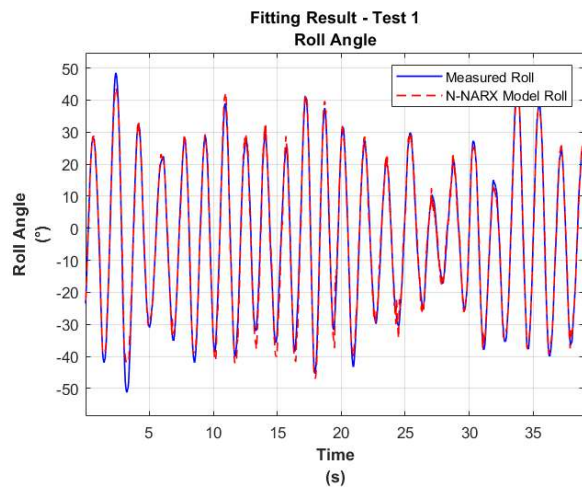


Figure 19. Approximation of the N-NARX model using data from Experiment 1 – Roll angle
Source: Authors

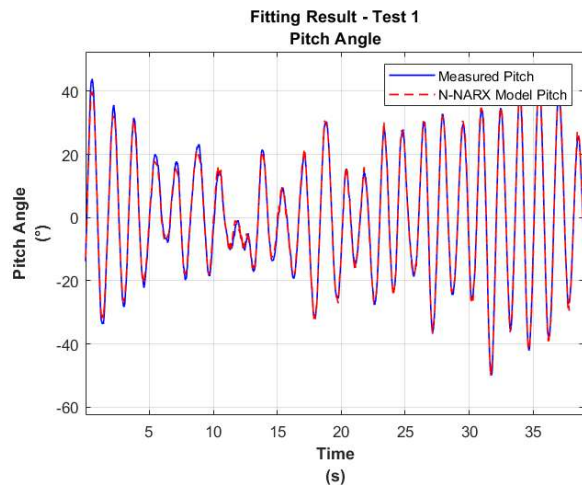


Figure 20. Approximation of the N-NARX model using data from Experiment 1 – Pitch angle
Source: Authors

Afterwards, the selected neural model underwent validation on a separate dataset, specifically that of Experiment 3 (*i.e.*, cross-validation).

For the roll angle (Figure 21), the following values were obtained: a MSE of 5, a RMSE of 2.23, a RMSE percentage of 2.25%, and a fit percentage of 89.44%.

For the pitch angle, as depicted in Figure 22, the performance evaluation yielded the following results: a MSE of 2.70, a RMSE of 1.64, a RMSE percentage of 2.20%, and a fit percentage of 89.29%.

Table 2 presents a quantitative comparison between the performance indicators obtained from Experiment 1, which were used in the training dataset, and those from Experiment 3, *i.e.*, the dataset for cross-validation. This comparison enables an assessment of the model's performance across different datasets.

It is worth mentioning that the results obtained from Experiment 1 evinced a higher level of accuracy and a lower error in predicting both angle values when compared to

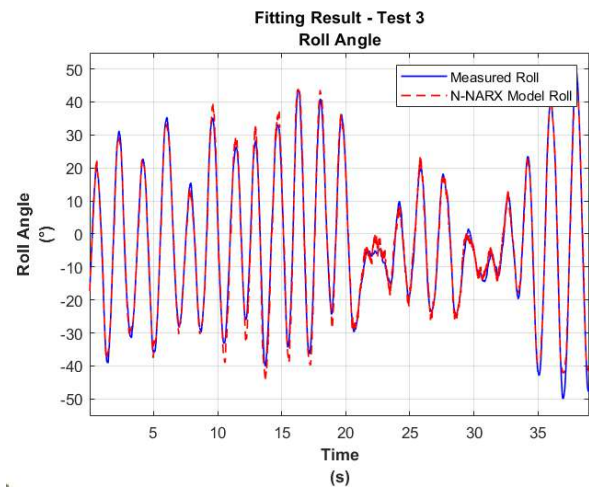


Figure 21. Approximation of the N-NARX model using data from Experiment 3 – Roll angle
Source: Authors

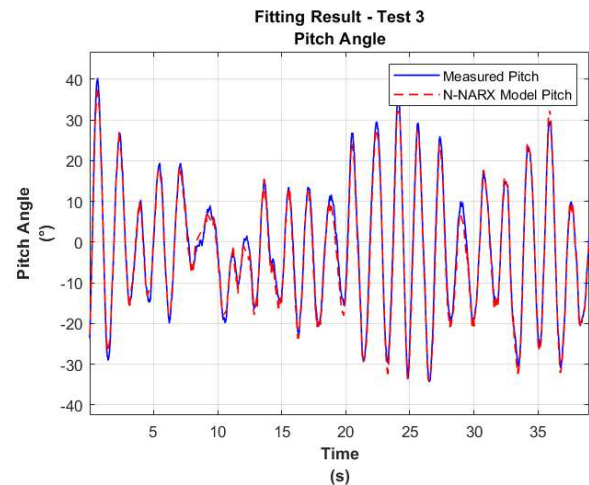


Figure 22. Approximation of the N-NARX model using data from Experiment 3 – Pitch angle
Source: Authors

Table 2. Performance indicators of experiments 1 and 3 from roll and pitch angles

	MSE	RMSE	RMSE(%)	Fit(%)
Roll 1	4.59	2.14	2.14	90.87
Roll 3	5.00	2.23	2.25	89.44
Pitch 1	1.44	1.20	1.28	93.80
Pitch 3	2.70	1.64	2.20	89.29

Source: Authors

the actual data, as the model was trained on the former. However, it is important to note that the N-NARX model showed a remarkable prediction performance even when evaluated on the unseen data from Experiment 3.

A statistical analysis was conducted to assess the behavior of the trained model's predictions and the actual values obtained from Experiment 3 for the roll angle (Figure 23) and the pitch angle (Figure 24).

The results confirm that the simple correlation consistently falls within the established confidence bands, at a 95% level for different time lags. This indicates a strong and

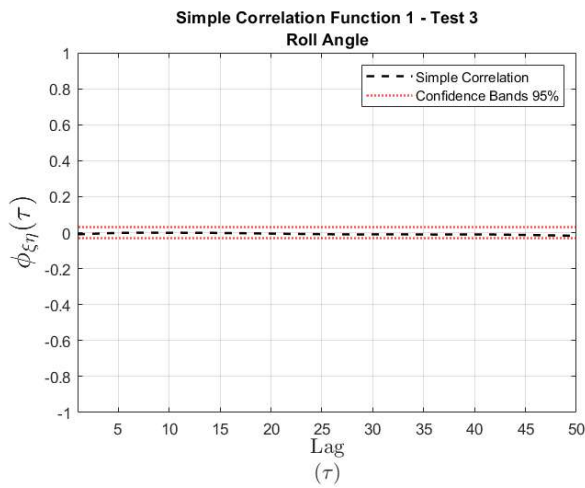


Figure 23. Correlation for Experiment 3 – Roll angle
Source: Authors

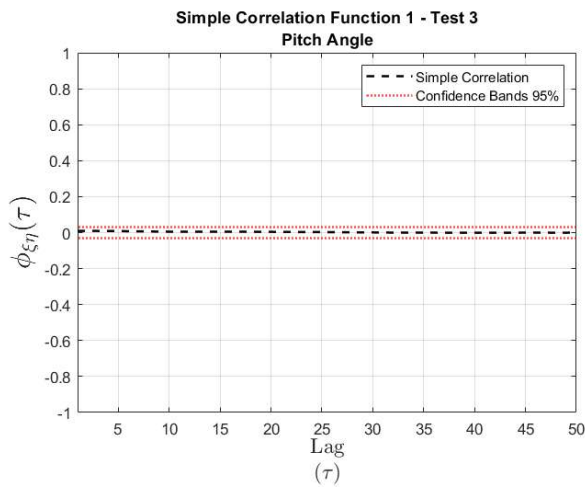


Figure 24. Correlation for Experiment 3 – Pitch Angle
Source: Authors

consistent association between the predicted and actual values, supporting the model's ability to capture the patterns and behaviors of the analyzed nonlinear system.

Discussion

The proposed artificial neural network method showed a high capability to predict attitude angles in a complex nonlinear system model such as a quadcopter. This observation aligns with the results of Muresan *et al.* (2013) and Zhang *et al.* (2014), who reported the model's high performance and accuracy in predicting both roll and pitch angles with a trained nonlinear model, outperforming the capacities of linear counterparts.

In light of the above, a comparison was conducted (Table 3) to assess the RMSE percentages of our method and the 4-neuron Minimal Resource Allocating Network (MRAN) linear model proposed by Pairan and Shamsudin (2017), whose work provided the essential basis for parameter selection during the training of the designed neural network. Note that the trained N-NARX model exhibits a lower RMSE percentage than the MRAN model in both roll and pitch angles.

Table 3. Comparison of the neural models' RMSE percentage

RMSE percentage	N-NARXI (our method)	MRAN (Pairan and Shamsudin 2017)
Roll	2.25 %	10.99 %
Pitch	2.20 %	13.53 %

Source: Authors

Similarly, Table 4 presents a comparison of the fit percentage between our method and the 4-input, 1-output Multiple Input and Multiple Output (MIMO) ARX linear model presented by Salameh *et al.* (2015). Note that the trained N-NARX model significantly outperforms the MIMO ARX model regarding the fit percentage for both roll and pitch angles.

Table 4. Comparison of the neural models' fit percentage

Fit percentage	N-NARX model (our method)	MIMO ARX (Salameh <i>et al.</i> , 2015)
Roll	89.44 %	26.28 %
Pitch	89.29 %	26.28 %

Source: Authors

The results shown in Tables 3 and 4, which compare the designed model with those proposed by other authors, clearly demonstrate the system's nonlinear behavior. Moreover, these findings highlight our model's robustness and its exceptional accuracy in delivering precise predictions, further validating its effectiveness and reliability in real-world applications.

Conclusions

Quadcopter drones are extremely complex and highly nonlinear systems, underscoring the importance of employing advanced algorithms for their identification. Thereupon, in this study, the roll and pitch angles of an F450 prototype were captured using a neural network model. To facilitate experimentation, a test bench was employed, enabling data collection from four tests conducted for each angle. Simultaneously, a real-time signal monitoring system was set up by connecting the Arduino to a computer, enabling the tracking of inputs consisting of four PWM signals sent to the quadcopter's EPU's, while the MPU9250 IMU sensor determined the device's orientation and produced the corresponding output signals.

These experiments revealed inherent variations, and data quality was assessed through amplitude and frequency spectrum analyses, leading to the selection of Experiments 1 and 3 as databases for each angle. For enhanced modeling, the input data were pre-processed using linear interpolation, while the output data underwent Kalman filtering, resulting in smoother signals and improved dynamic capture. Next, the NARX neural network, with a hidden layer comprising five neurons and utilizing RBF activation, was employed. The model was then trained using the dataset derived from Experiment 1, using the Levenberg-Marquardt algorithm. This yielded optimal outcomes.

The neural network model effectively captured the quadcopter's nonlinear dynamics during both the training and validation processes, resulting in an 89.44% fit with an

RMSE of 2.25% for the roll angle, and an 89.29% fit with an RMSE of 2.20% for the pitch angle. Furthermore, the designed neural network model outperformed the proposals of other authors, clearly showcasing its ability to accurately represent the nonlinear behavior of the system.

Future work and scenarios

Considering the work by Ucgun *et al.* (2022), future research could also build on both the NARX model and the Vertical Take-Off and Landing (VTOL) UAV testbed by evaluating free-flight drones. This would allow validating neural network-based control algorithms and their integration with the robust sliding mode controller developed for the testbed. The goal would be to enhance drone stability and performance, comparing the combined approach against traditional control methods and demonstrating improved robustness and tracking in real-world scenarios. This synergy between the NARX model and the testbed could lead to more reliable UAV control systems.

Building on previous research using a NARX neural network model to identify nonlinear quadcopter dynamics, future work could be extended to free-flying drones, capturing a fully dynamic behavior that includes yaw motion. By moving from controlled environments to free flight, a neural network-based system could identify and stabilize the drone in real time, adapting to conditions like wind disturbances or payload changes. Combining the NARX model with traditional methods like PID or model predictive control could help to create a hybrid system, enhancing performance in complex scenarios. Real-time experiments would validate the approach for more robust flight control in unpredictable environments.

In addition, according to Mechali *et al.* (2021), the NARX neural model for UAVs can be extended by integrating it with control strategies such as Continuous Non-singular Terminal Sliding Mode Control (CNTSMC) and Disturbance Observer-Based Control (DOBC). In future research, the use of NARX models to accurately predict and identify the nonlinear dynamics of a free-flying UAV could be explored. Then, the CNTSMC scheme could be applied for robust attitude and position control. By combining the predictive capabilities of the NARX model with CNTSMC's robustness, disturbance rejection (DOBC) and Fixed-time eXtended State Observer (FXESO), a highly accurate and resilient control system for UAVs could be developed which performs well even under uncertainties, nonlinearities, and external disturbances. This approach would allow enhancing trajectory tracking and system stability in real-time scenarios, validating the combined method through both simulations and real-world tests.

Conflicts of interest

All authors declare no conflicts of interest.

CRedit author statement

Howard E. Sifuentes: conceptualization, data curation, formal analysis, funding acquisition, investigation, methodology, software, resources, validation, visualization, writing - original draft.

Carlos A. Rocha: investigation, methodology, writing - original draft, writing (review and editing).

Edgar A. Manzano: project administration, supervision, writing (review and editing).

References

- Ahmad, F., Kumar, P., Bhandari, A., and Patil, P. (2020). Simulation of the quadcopter dynamics with LQR based control. *Materials Today: Proceedings*, 24, 326–332. <https://doi.org/10.1016/j.matpr.2020.04.282>
- Al-Mahasneh, A. J., Anavatu, S. G., and Garratt, M. (2017). *Nonlinear multi-input multi-output system identification using neuro-evolutionary methods for a quadcopter* [Conference paper]. 2017 Ninth International Conference on Advanced Computational Intelligence. <https://doi.org/10.1109/icaci.2017.7974512>
- Altan, A., Aslan, O., and Hacıoglu, R. (2018). *Real-time control based on NARX neural network of hexarotor UAV with load transporting system for path tracking* [Conference paper]. International Conference on Control, Engineering & Information Technology. <https://doi.org/10.1109/ceit.2018.8751829>
- Atteia, G., Mengash, H. A., and Samee, N. A. (2021). Evaluation of using Parametric and Non-parametric Machine Learning Algorithms for Covid-19 Forecasting. *International Journal of Advanced Computer Science and Applications*, 12(10), 0121071. <https://doi.org/10.14569/ijacsa.2021.0121071>
- Belge, E., Kaba, H., Parlak, A., Altan, A., and Hacıoğlu, R. (2020). Estimation of small unmanned aerial vehicle lateral dynamic model with system identification approaches. *Balkan Journal of Electrical & Computer Engineering*, 8(2), 121–126. <https://doi.org/10.17694/bajece.654499>
- Billings, S. (2013). *Nonlinear system identification: NARMAX methods in the time, frequency, and spatio-temporal domains*. John Wiley & Sons.
- Cedro, L., and Wiczorkowski, K. (2019). Optimizing PID controller gains to model the performance of a quadcopter. *Transportation Research Procedia*, 40, 156–169. <https://doi.org/10.1016/j.trpro.2019.07.026>
- El Dakrory, A. M., and Tawfik, M. (2016). *Identifying the attitude of dynamic systems using neural network* [Conference paper]. 2016 International Workshop on Recent Advances in Robotics and Sensor Technology for Humanitarian Demining and Counter-IEDs (RST). <https://doi.org/10.1109/RST.2016.7869856>
- Galrinho, M., Everitt, N., and Hjalmarsson, H. (2017). ARX modeling of unstable linear systems. *Automatica*, 75, 167–171. <https://doi.org/10.1016/j.automatica.2016.09.041>
- Kamel, M., Burri, M., and Siegwart, R. (2017). Linear vs nonlinear MPC for trajectory tracking applied to rotary wing micro aerial vehicles. *IFAC-PapersOnLine*, 50(1), 3463–3469. <https://doi.org/10.1016/j.ifacol.2017.08.849>
- Karagoz, R., and Batselier, K. (2020). Nonlinear system identification with regularized Tensor Network B-splines. *Automatica*, 122, 109300. <https://doi.org/10.1016/j.automatica.2020.109300>

- Karakaya, Ş., and Goren, A. (2022). Performance comparison of PID and NARX neural network for attitude control of a quadcopter UAV. *Journal of Materials and Mechatronics: A*, 3(1), 1–19. <https://doi.org/10.55546/jmm.1010919>
- Macrina, G., Pugliese, L., Guerriero, F., and Laporte, G. (2020). Drone-aided routing: A literature review. *Transportation Research Part C-emerging Technologies*, 120, 102762. <https://doi.org/10.1016/j.trc.2020.102762>
- Mechali, O., Xu, L., Huang, Y., Shi, M., and Xie, X. (2021). Observer-based fixed-time continuous nonsingular terminal sliding mode control of quadrotor aircraft under uncertainties and disturbances for robust trajectory tracking: Theory and experiment. *Control Engineering Practice*, 111, 104806. <https://doi.org/10.1016/j.conengprac.2021.104806>
- Muresan, B., Folea, S., Nascu, I., Ionescu, C. M., and De Keyser, R. (2013). Identification and modeling of the three rotational movements of a miniature coaxial helicopter. *Simulation*, 89(12), 1490–1504. <https://doi.org/10.1177/0037549713504788>
- Nelles, O. (2020). *Nonlinear system identification: From classical approaches to neural networks, fuzzy models, and Gaussian processes* (2nd ed.). Springer Nature.
- Noel, J., and Kerschen, G. (2017). Nonlinear system identification in structural dynamics: 10 more years of progress. *Mechanical Systems and Signal Processing*, 83, 2–35. <https://doi.org/10.1016/j.ymssp.2016.07.020>
- Ozbek, N. S., Onkol, M., and Efe, M. O. (2015). Feedback control strategies for quadrotor-type aerial robots: A survey. *Transactions of the Institute of Measurement and Control*, 38(5), 529–554. <https://doi.org/10.1177/0142331215608427>
- Oktay, T., and Kose, O. (2019). Farklı uçuş durumları için quadcopter dinamik modeli ve simülasyonu. *Eur. J. Sci. Technol.*, 15(1), 132–142. doi: <https://doi.org/10.31590/ejosat.507222>
- Pairan, M., and Shamsudin, S. (2017). *System identification of an unmanned quadcopter system using MRAN neural* [Preprint]. IOP Conference Series. <https://doi.org/10.1088/1757-899x/270/1/012019>
- Pairan, M. F., Shamsudin, S. S., and Zulkafli, M. F. (2020). Neural network-based system identification for quadcopter dynamic modeling: A review *Journal of Advanced Mechanical Engineering Applications*, 2(1), 20–33. <https://doi.org/10.30880/ijie.2020.02.01.003>
- Rosales, C., Soria, C., and Rossomando, F. G. (2019). Identification and adaptive PID Control of a hexacopter UAV based on neural networks. *International Journal of Adaptive Control and Signal Processing*, 33(1), 74–91. <https://doi.org/10.1002/acs.2955>
- Saito, C. (2019). *Modelamiento dinámico de los parametros de control de vuelo de una aeronave del tipo ala volante utilizando redes neuronales artificiales* [Master's thesis, Pontificia Universidad Católica del Perú]. <https://tesis.pucp.edu.pe/repositorio/handle/20.500.12404/13308>
- Salameh, I., Ammar, E. M., and Tutunji, T. A. (2015). *Identification of quadcopter hovering using experimental data* [Conference paper]. 2015 IEEE Jordan Conference on Applied Electrical Engineering and Computing Technologies. <https://doi.org/10.1109/aeect.2015.7360559>
- Santoso, F., Garratt, M., Anavatti S., and Petersen, I. (2018). Robust hybrid nonlinear control systems for the dynamics of a quadcopter drone. *IEEE Transactions on Systems, Man, and Cybernetics*, 50(8), 3059–3071. <https://doi.org/10.1109/tsmc.2018.2836922>
- Tu, Q., Rong, Y., and Chen, J. (2020). Parameter identification of ARX models based on modified momentum gradient descent algorithm. *Complexity*, 2020, 1–11. <https://doi.org/10.1155/2020/9537075>
- Ucgun, H., Okten, I., Yuzgec, U., and Kesler, M. (2022). Test platform and graphical user interface design for vertical take-off and landing drones. *Romanian Journal of Information Science and Technology*, 25(3), 350–367. <https://www.romjist.ro/abstract-727.html>
- Ullah, S., Mehmood, A., Khan, Q., Rehman, S., and Iqbal, J. (2020). Robust integral sliding mode control design for stability enhancement of under-actuated quadcopter. *International Journal of Control, Automation and Systems*, 18(7), 1671–1678. <https://doi.org/10.1007/s12555-019-0302-3>
- Ullah, S., Mehmood, A., Ali, K., Javaid, U., Hafeez, G., and Ahmad, E. (2021). *Dynamic modeling and stabilization of surveillance quadcopter in space based on integral super twisting sliding mode control strategy* [Conference paper]. 2021 International Conference on Artificial Intelligence (ICAI). <https://doi.org/10.1109/icai52203.2021.9445268>
- Ullah, S., Khan, Q., and Mehmood, A. (2022). Neuro-adaptive fixed-time non-singular fast terminal sliding mode control design for a class of under-actuated nonlinear systems. *International Journal of Control*, 96(6), 1529–1542. <https://doi.org/10.1080/00207179.2022.2056514>
- Wei, M., Ye, M., Li, J. B., Wang, Q., and Xu, X. (2020). State of charge estimation of lithium-ion batteries using LSTM and NARX neural networks. *IEEE Access*, 8, 189236–189245. <https://doi.org/10.1109/ACCESS.2020.3031340>
- Xavier, J., Patnaik, S., and Panda, R.C. (2021). Process modeling, identification methods, and control schemes for nonlinear physical systems { A comprehensive review. *ChemBioEng Reviews*, 8(4), 392–412. <https://doi.org/10.1002/cben.202000017>
- Zhang, X., Li, X., Wang, K., and Lu, Y. (2014). A survey of modelling and identification of quadrotor robot. *Abstract and Applied Analysis*, 2014, 320526. <https://doi.org/10.1155/2014/320526>

A Semi-Supervised Deep Learning Model for Defective lime Classification

Modelo semisupervisado de aprendizaje profundo para la clasificación de limones

Angel-Moisés Hernández-Ponce¹, Francisco-Javier Ornelas-Rodríguez¹, Juan-Bautista Hurtado-Ramos¹, Alfonso Ramírez-Pedraza^{1,2}, and José-Joel González-Barbosa¹

ABSTRACT

For consumers, the predominant fruit selection criterion is visual quality, a factor that classification models emulate when employing images as input data. Most classification paradigms presuppose a balance across classes. In the field of defective fruit detection, databases commonly exhibit a pronounced imbalance between healthy and defective fruit counts. Such disparity can compromise the robustness of classification models or introduce biases stemming from insufficient data. This study introduces a semi-supervised classification framework based on anomaly detection to identify defective lime fruits (*Citrus aurantifolia*). The framework employs the reconstruction error obtained from an autoencoder neural network and a calculated anomaly probability to locate samples within a two-dimensional space designed for such purpose. Based on the defined parameter ranges, the limes are categorized as either healthy or defective. The proposed classification model underwent training utilizing the publicly accessible Fruits360 database and was tested with a set of 118 new and unlabeled lime images. The classification model attained a precision of 94%, a recall of 0.88, and an F1-score of 0.91 across the test set. These results corroborate that models based on anomaly detection constitute a promising solution to the inherent challenges of unbalanced classification tasks. They offer the advantage of requiring minimal training data and reduced training times while maintaining efficacy, even when the evaluation dataset diverges substantially from the training set. Thus, the proposed model can serve as a decision support tool for farmers, producers, and consumers.

Keywords: semi-supervised learning, citrus fruit classification, anomaly detection, precision agriculture applications

RESUMEN

Para los consumidores, el criterio predominante en la selección de frutas es la calidad visual, un factor que los modelos de clasificación emulan cuando emplean imágenes como datos de entrada. La mayoría de los paradigmas de clasificación presuponen un equilibrio entre las clases. En el ámbito de la detección de frutas defectuosas, las bases de datos suelen presentar un desequilibrio pronunciado entre el recuento de frutas sanas y defectuosas. Esta disparidad puede comprometer la solidez de los modelos de clasificación o introducir sesgos derivados de la insuficiencia de datos. En este estudio se introduce un marco de clasificación semisupervisada basado en la detección de anomalías para identificar frutos defectuosos de limón (*Citrus aurantifolia*). El modelo emplea el error de reconstrucción de una red neuronal *autoencoder* y una probabilidad de anomalía calculada para localizar muestras dentro de un espacio bidimensional diseñado para tal propósito. A partir de los rangos de parámetros definidos, los limones se clasifican como sanos o defectuosos. El modelo de clasificación propuesto fue entrenado mediante la base de datos de acceso público Fruits360 y evaluado con un conjunto de 118 imágenes de limones nuevas y sin etiquetar. El modelo de clasificación obtuvo una precisión del 94 %, una recuperación del 0,88 y un valor F1 0,91 en el conjunto de pruebas. Estos resultados corroboran que los modelos basados en la detección de anomalías constituyen una solución prometedora a los retos inherentes de las tareas de clasificación no equilibradas; ofrecen la ventaja de requerir datos de entrenamiento mínimos y tiempos de entrenamiento reducido, manteniendo la eficacia incluso cuando el conjunto de datos de evaluación diverge sustancialmente del conjunto de entrenamiento. Así, el modelo propuesto puede servir como herramienta de apoyo en las decisiones de agricultores, productores y consumidores.

Palabras clave: aprendizaje semisupervisado, clasificación de frutos cítricos, detección de anomalías, aplicaciones para agricultura de precisión

Received: February 7th 2024

Accepted: October 2nd 2024

Introduction

Citrus fruits are among the most consumed and produced in the world. However, it is estimated that their production will decrease within the next 10 years due to factors such as climate change, land use issues, and diseases.

In recent years, precision agriculture has helped farmers and producers (Ayoub Shaikh *et al.*, 2022; Nowak, 2021; D. Li *et al.*, 2021; González-Barbosa *et al.*, 2022) to create optimized systems for crop production, crop estimation, weather monitoring, disease detection, and defect identification (Wang *et al.*, 2022; Mzoughi and Yahiaoui, 2023; E. Li *et al.*, 2023; Gokulnath and Usha Devi, 2021; Chaturvedi

et al., 2023; Naranjo-Torres *et al.*, 2021; Aparicio Pico *et al.*, 2022; Puerto Cuadros, 2024). Other factors can affect crop production, such as diseases and pests (George *et al.*, 2022; Bao *et al.*, 2021). For a producer, the fruit must have the best possible visual appearance, as it is the main aspect that the customers look for in the market (Blasco *et al.*, 2016). Hence, it is essential to distinguish good

¹Instituto Politécnico Nacional, Cerro Blanco 141, Querétaro, México. Email: *jgonzlezba@ipn.mx

²Dirección Adjunta de Desarrollo Científico, IxM, CONAHCyT, Alvaro Obregón, Mexico City 03940, Mexico



Attribution 4.0 International (CC BY 4.0) Share - Adapt

fruit from those with defects or poor visual appearance (Ghazal *et al.*, 2021). The early detection of diseases and defects has become a critical task to ensure crop production and quality (Soltani Firouz, 2022), easier post-harvest processing (Zhang *et al.*, 2021), and longer shelf-life (Bhole, 2021). The consumers purchasing decision is heavily influenced by the quality and appearance of the fruit; a good healthy fruit without defects is the first choice (Blasco *et al.*, 2016). Therefore, having an automated and efficient system to monitor the citrus production process is essential in continuing to meet the global demand (Food and Agriculture Organization, 2022; Lozano and Archibald, 2022; Peng *et al.*, 2023).

Despite the fact that some systems have been proposed and developed to improve citrus production, most of them focus on the study of diseases and defects in tree leaves (Arnal Barbedo, 2019; Khanramaki *et al.*, 2021; Ümit Atila *et al.*, 2021). The evaluation of leaves is one of the fastest available methods, but it can lead to biased results since the decision is based on limited samples rather than on the entire tree or an inspection of the fruit produced. Furthermore, these systems often use private and customized databases (Lu and Young, 2020) and provide limited results. In addition, the fact that a leaf is defective does not mean that the fruit produced is. Therefore, fruit evaluation arises as an alternative method to leaf assessment.

Classifying healthy fruits from defective ones is the final step in the post-harvest process. At this stage, the fruit is prepared for storage, and the selection is based on its visual quality, so both producers and consumers look for the fruit with the best appearance.

The use of images and their processing is an alternative to detect and identify diseases and defects in tree leaves and fruits. Machine vision systems allow for precise defect localization using RGB (Cubero *et al.*, 2014; Tan *et al.*, 2021), X-ray, multi-spectral, satellite (Cándido-Mireles *et al.*, 2023; Toosi *et al.*, 2022), and aerial images (Futerman *et al.*, 2023; Istiak *et al.*, 2023). More recently, the use of machine and deep learning models has helped to make machine vision systems more automated, accurate, and efficient (Palei *et al.*, 2023). Deep learning models based on convolutional neural networks (CNNs) can handle large amounts of information and make decisions based on key features such as color, shape, and texture. However, to obtain a good result using a CNN, it is necessary to have a large and diverse database, a balance of classes, and labeled samples (Gron, 2017; Goodfellow *et al.*, 2016; Ibrahim and Kuban, 2023). Additionally, it takes a considerable amount of time to train these networks, with the risk of overfitting or bias when making decisions. One paradigm that addresses some of these challenges is semi-supervised learning (Thoidis *et al.*, 2021; Gao *et al.*, 2022; Memarzadeh *et al.*, 2022).

An alternative to CNNs are autoencoders (Bank *et al.*, 2021). These types of neural networks adopt a different approach by using a latent space that retains the most important information of the input data. They also have the advantage of being able to work with unlabeled or barely labeled data, so they can work as unsupervised or semi-supervised classification models and can be robust to input noise and anomalous data (Cazzonelli and Kulbach, 2023). Given these advantages, an autoencoder can be used as a classification model based on anomaly detection, e.g., in a

dataset of samples that do not match the learned features and are considered anomalous. This approach is useful when dealing with imbalanced datasets. In some real-world scenarios, the data available for different classes can be highly imbalanced. In contrast, traditional machine learning models assume a class balance, so they can be prone to overfitting or to bias.

In the field of precision agriculture, anomaly detection can be applied to detect defective or diseased fruit. In a lime dataset, anomalous data correspond to damaged, injured, or sick fruits, or to any other fruit with a compromised visual appearance. Fruit diseases are a serious threat to lime harvesting, yet defective limes can only be observed during the post-harvest process or in the delivery stages. Therefore, the citrus industry and the horticultural industry in general necessitate rapid and automated disease detection tools throughout the post-harvest period.

The objective of this work is to develop a semi-supervised classification system to separate healthy limes from defective ones. The paradigm of autoencoders, combined with the advantages of semi-supervised learning, constitutes a versatile yet robust combination for such tasks. In this case, defective limes are difficult to obtain due to current high-quality standards. By using anomaly detection, defective limes can be identified with relatively little effort by teaching the model the characteristics of a healthy lime. Consequently, if a lime does not meet these characteristics, it is considered defective. The anomaly detection approach addresses the issues of data scarcity, overfitting, and heightened training times commonly associated with traditional classification models. This article proposes a classification system based on anomaly detection using autoencoders to obtain the reconstruction error of the input data (images) as well as the value of a kernel density estimation function.

The main contributions of this work are presented below.

- The anomaly detection model was trained exclusively with images of healthy limes.
- Predictions were made with the probability of an image and the reconstruction error value.
- The encoder can be used as a feature extractor for further classification models.

Related work

A novel deep learning model used to detect defected regions in citrus fruits is proposed in (Dhiman *et al.*, 2022). The model uses the Felzenszwalb algorithm as the main approach to measure pixel intensity in the images. The fruits in this database can have one of three damage severity levels: low, medium, and high. Additionally, there is a fourth class for healthy fruits. The model achieved an accuracy value of 99%, 97%, and 96% for low-, medium-, and high-severity levels. For healthy fruit, the model showed 96% accuracy.

The study by (Fan *et al.*, 2020) presents a deep learning architecture based on neural networks with a low-cost vision system to detect defective apples in a fruit sorting machine. The proposed architecture used images of both classes for training and performance validation, achieving an accuracy of 96.5%. The trained model was validated by loading it onto

an independent sorting machine using 200 apples, obtaining an accuracy of 92%.

Another approach for fruit quality classification is presented in (Hanh and Bao, 2022). This article proposes a machine vision system to classify limes in the Vietnam region into three quality groups: bad, regular, and best. For classification, the authors use a P-score, which counts the number of pixels in the image that are considered healthy within the fruit peel. The quality of limes is poor if the P-score is less than 35. The lime quality is considered good if P is between 35 and 85. Lastly, if the P-score is greater than 85, the quality is the best.

An orange categorization system using different features such as color, fruit size, shape, and surface defects was reported in (Hasan et al., 2021). In their system, oranges were classified using three CNN models, with the third model achieving an accuracy of 92.25%. In parallel, a classification model using a conditional generative adversarial network (CGAN) and the Lemon Quality Dataset was developed in (Bird et al., 2022), aiming to address the shortage of defective lime images with the CGAN. Although the reconstructions lacked consistency regarding the shape, color, and texture of the fruit, the authors managed to achieve an accuracy value of 88.75% in their experiments.

A computer vision system was reported in (Chen et al., 2018) for the automatic detection and classification of oranges based on their external features. Utilizing a combination of color- and texture-based characteristics, this system classifies oranges into four categories, each corresponding to a different level of ripeness. The model's performance was evaluated by the authors, with the system achieving an impressive accuracy of 97.5%. This led to the conclusion that the proposed system presents significant opportunities for enhancing both efficiency and accuracy in industrial applications.

A spectroscopy system developed to evaluate the phytopathological condition of mango fruit is presented in (Cabrera Ardila et al., 2020b), given mangoes' susceptibility to anthracnose infections during the harvesting phase. The authors monitored the progression of the pathogen, categorizing the fruits into three stages: healthy, asymptomatic, and diseased. Various classification algorithms, such as linear discriminant analysis (LDA), random forests (RF) and support vector machines (SVM), were used to assess the data. The best accuracy value was given by LDA, ranging from 91-100% across the three stages.

A novel approach for identifying surface defects in oranges using computer vision techniques was proposed in (Rong et al., 2017). A sliding window algorithm was applied by the authors to segment the images, which allowed detecting defects of varying sizes and shapes. The proposed method was tested on a custom dataset and achieved an accuracy rate of 91.5%.

A robust and generalized CNN model for detecting black spot disease and ripeness levels in orange fruit was introduced in (Momeny et al., 2022) by fine-tuning pre-trained models. Data-augmentation techniques were used by the authors to increase the dataset size and enhance their model's performance. The learning-to-augment strategy employed is a technique that generates new training data by adding noise to existing images. This helped to improve

the performance of the deep neural network model, making it more robust and better able to generalize to new unseen data with an accuracy of 99.5%.

The state of the art shows deep learning to be the most marked trend in recent years. However, one of the most important challenges is the absence of a standard or universal database to establish a common frame of reference to evaluate the work done. Each proposal tends to create its own database, adapted to specific needs and particular applications, hindering direct comparison between methods and results.

Classification was identified as the main task within the studies found, with a focus on assessing the health of fruits or categorizing them by type. Despite advances in deep learning techniques, the amount of data used in these studies is often limited, which restricts the generalization capacity of the models developed. This data limitation not only affects the accuracy and robustness of the models, but also poses interesting challenges for their application in broader and more varied scenarios and tasks. In summary, although deep learning shows potential in citrus fruit classification, the absence of standard databases and the current limitations regarding the amount of available data are critical challenges that need to be addressed in order to improve overall performance and results comparability in this field.

Methodology

Model description

The proposed model is based on the autoencoder algorithm. An autoencoder is a type of algorithm with the primary purpose of learning an abstract representation of the data by learning to reconstruct a set of input observations (Bank et al., 2021). This representation can then be used for different applications. In recent years, convolutions have been added to autoencoders to reduce dimensions. The variant studied herein is known as *convolutional autoencoders* (CAEs). CAEs are a type of CNN that use convolutional, pooling, and deconvolutional layers to create and refine feature maps from input images, ultimately reconstructing the original input through unsupervised learning. By employing the convolution operator, CAEs filter input signals to extract significant content, encoding inputs into fundamental signals and reconstructing them to minimize error. Unlike traditional CNNs, which are primarily used for supervised classification tasks, CAEs focus on learning optimal filters for feature extraction and input reconstruction. This approach modifies the standard autoencoder by incorporating convolutional layers in the encoder and transposing convolutional layers in the decoder, effectively capturing spatial statistics in image data (Maheshwari et al., 2022). CAEs are better suited for image processing tasks since they capture spatial patterns and relationships between pixels, resulting in a more accurate image reconstruction (Michelucci, 2022). Figure 1 shows the workflow followed to carry out the classification tasks.

The model takes a color image of any size as input. Then, the input image passes from the input layer to a preprocessing layer, where its size changes to 96×96 . The image is also standardized in this layer. After the pre-processing layer, there is an image augmentation layer. Here, the image is transformed through various adjustments,

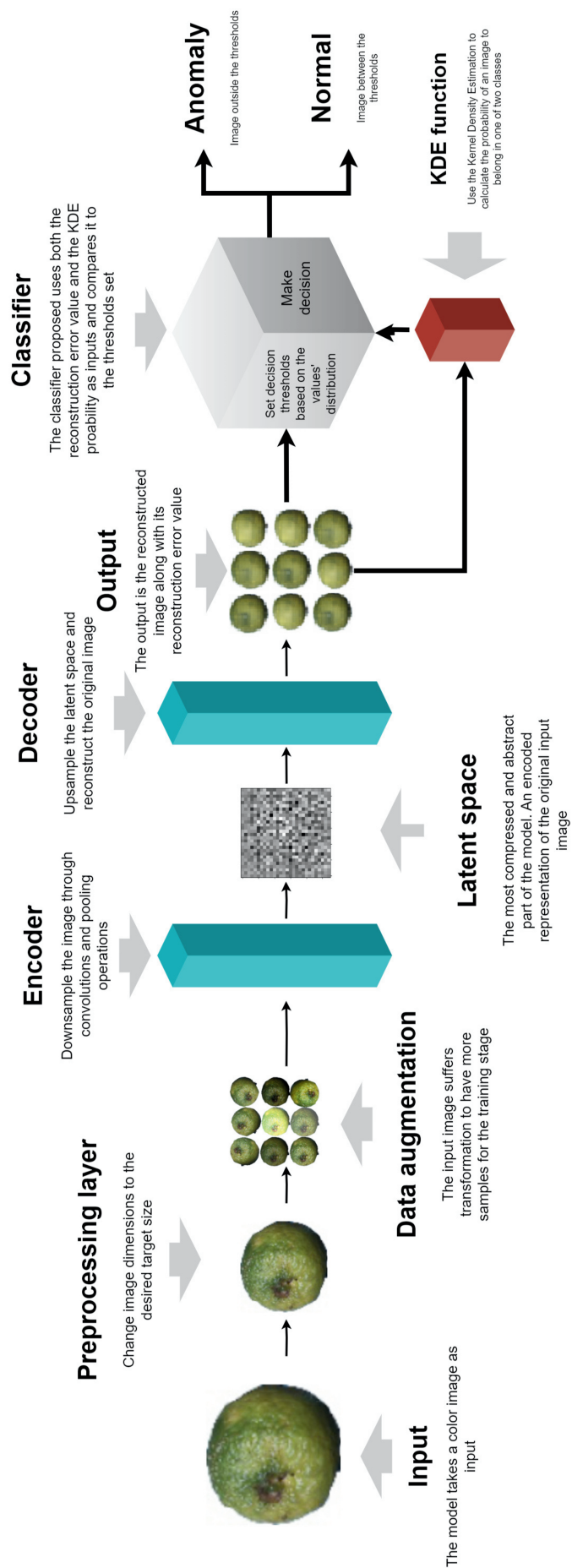


Figure 1. Graphical representation of the workflow proposed for the classification of the lime images
Source: Authors

including rotation, brightness, contrast, and changes in pose. These changes provide the autoencoder with more data in the training stage (dos Santos Tanaka and Aranha, 2019; Shijie et al., 2017). The augmentation stage helps the autoencoder learn from new situations or instances in which the lime fruit may appear.

The autoencoder processes the input image into a condensed representation known as the *latent space*. In this part of the network, the input data become a lower-dimensional representation of the input image, capturing its most significant features in a compact, encoded format. Because the latent space is smaller than the input image, it effectively reduces data complexity. Exploring this space enables the generation of data points that mirror the original, allowing for advanced operations like clustering, classification, and anomaly detection.

The encoder

The next segment of the model is the encoder, which converts input data into the latent space. This process aims to retain as much of the initial information as possible. The proposed encoder design incorporates three down-sampling stages. In each stage, the dimensions of the image are reduced, while its complexity is increased. This is achieved through a sequence of convolutional layers paired with max-pooling layers. Consistency is maintained throughout the process by using identical kernel sizes and stride values for the convolutional and max-pooling layers, while the rectified linear unit (ReLU) serves as the activation function for all layers. Variations arise only in the number of filters within the convolutional layers and in the pool sizes for the max-pooling layers. Initially, the convolutional layers start with 16 filters. This number then decreases to 8 and subsequently to 3 for the second and third layers, respectively. In tandem, the pool size expands progressively as the model delves deeper. Upon completion of the encoding process, the transformed data emerge with the dimensions $3 \times 3 \times 3$, which constitutes the latent space of the proposed model. This compressed output then proceeds to the next phase: the decoder.

The decoder

The decoder is responsible for reconstructing the input data from the encoder's compressed latent representation, producing an output that closely resembles the original input. This step is crucial in an autoencoder's architecture, as the quality of the decoder directly influences the fidelity of the reconstruction (Gron, 2017). The decoder reverses the encoder's process by up-sampling the compressed data to its original dimensionality. A well-crafted decoder ensures high-quality reconstructions, whereas a sub-optimal one may result in poor quality. Reflecting the autoencoder's symmetrical design, the decoder of the proposed model mirrors the encoder's structure with reversed operations.

Training the model

The next phase in the workflow involves training the model. Google Colaboratory's services and TensorFlow were utilized for coding and training. The basic GPU environment of Google Colaboratory, equipped with an NVIDIA K80 with 12GB of VRAM and an Intel Xeon CPU

Table 1. Parameters used for the training stage

Parameter	Value
Training set size	656
Validation set size	166
Training time	500 epochs (2 s/epoch)
Loss function	Mean squared error
Optimizer	Adam
Learning rate	0.001
Batch size	32

Source: Authors

with 13GB of RAM, was selected for the training process. The Fruits360 dataset (Murean and Oltean, 2018), a publicly accessible collection of 90 483 images across 120 fruit and vegetable categories, including apples, bananas, oranges, limes, and tomatoes, was employed for both training and validation. The dataset was built for object recognition, classification, and detection tasks. It includes color images 96×96 pixels in size, captured from various angles, under different lighting conditions, and against contrasting poses and backgrounds. Every image has an assigned label (organized in folders) that serves to identify the type of fruit or vegetable. Furthermore, each class is divided into training and validation sets.

The Fruits360 dataset is used as a valuable benchmark for the development of fruit classification or detection models (Siddiqi, 2020; Latif et al., 2023; Rathnayake et al., 2022; Dandekar et al., 2020). Within the dataset, there is a subset of lime images (*Citrus aurantifolia*). The lime images were chosen for their close resemblance to the lime varieties commonly found in Mexico regarding both color and visual features. The lime subset comprises a total of 822 lime images, which were partitioned into 656 for training and 166 for validation, i.e., approximately an 80-20 split. This distribution was adopted due to the fixed number of training and validation images in the dataset.

Table 1 presents the data distribution and the hyperparameters selected for model training. The training process spanned 500 epochs and incorporated an early-stopping mechanism to mitigate the risk of overfitting.

Subsequent to training, the model became adept at recognizing the characteristics of a healthy lime. This allowed applying the trained model for predictions on both familiar and novel lime images. A lime was classified as defective if it lacked the characteristics that the model was trained to recognize as indicative of a healthy specimen. Conversely, a lime that exhibited these learned traits was deemed healthy.

Making predictions for new images

The model returns a loss value after the image goes through all the convolution layers. This value is called the *reconstruction error* (RE). For this case, the RE is a measure that quantifies how similar the decoded output of the autoencoder is to the original input data. This RE

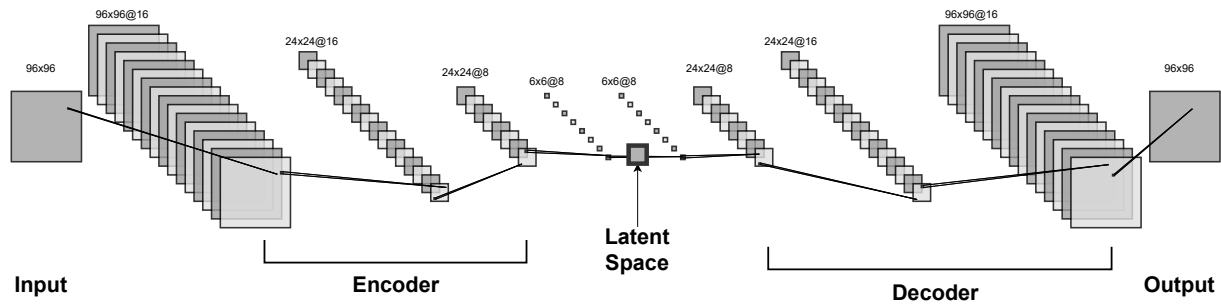


Figure 2. Architecture of the convolutional autoencoder used in this work. It consists of an encoder, a latent space, and a decoder. The input is a 96×96 color image that is progressively compressed by convolution and max-pooling layers to a latent representation of size $3 \times 3 \times 3$. The decoder then up-samples the latent representation to reconstruct the original 96×96 image via up-sampling and convolutional layers. The network's total parameter count is 3738, with 16 convolutional filters in the initial and final layers, ensuring efficient feature extraction and image reconstruction.
Source: Authors

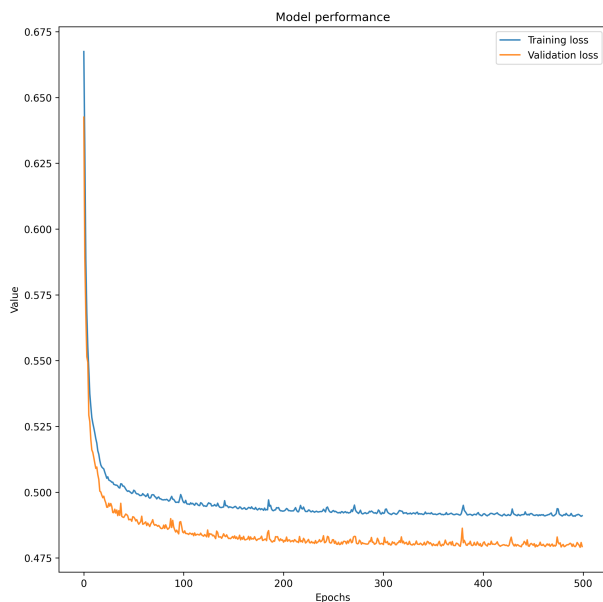


Figure 3. Loss curve of the model during the training stage. The blue line denotes the loss value for the training set, and the orange one denotes the loss value for the validation set.

Source: Authors

serves as the main feedback during the training phase of an autoencoder; by optimizing this error, it should learn to capture the important features and patterns of the input data. Ideally, the RE value would be 0, meaning that the model can recreate the input image without errors. In practice, this does not occur often.

The RE value is used to evaluate the quality of a reconstructed image. One way to ensure that an autoencoder is trained correctly is to compare its inputs and outputs; the differences should be insignificant (Gron, 2017). Since the model is trained to identify healthy lime images, when a faulty image is introduced, the RE should be higher. However, the RE value should not be the only metric to evaluate the performance of a model (Beggel et al., 2020).

A second reference value was calculated to this effect. This value, derived from the latent space, was utilized to calculate a probability density using the kernel density

estimation (KDE) technique (Hu et al., 2020). KDE is a non-parametric technique for estimating the probability density function (PDF) of a random variable. It can estimate the probability distribution of a set of data and constitutes a helpful tool for data analysis, as it allows the user to visualize the data distribution more effectively than other visual representations. Unlike parametric methods, KDE does not rely on assuming a specific parametric form of the density function; instead, KDE learns the shape of the density from the data itself. This kind of flexibility makes this method popular as a tool for analyzing data from complex distributions (Lang et al., 2022; Cao et al., 2016). By combining the autoencoder's RE and the data's KDE value, anomalies (defective limes in this case) can be detected in an unknown dataset. This approach enables the automatic detection of anomalies without prior knowledge of the expected data distribution. Anomalies are data points with a high RE value and a low probability density, as estimated by the KDE function.

Results

Another dataset of 118 lime images was constructed to evaluate the model's performance on new, previously unseen samples. These images differed significantly in size (1080×1080 pixels), color, texture, and geometry from those in the Fruits360 dataset. Moreover, the background of the images was removed. This approach allowed assessing the classification model's ability to generalize its learning of the normal class. A sample of these new images is shown in Figure 4.

The RE was calculated for the new test set by passing the each new image through the model. Then, this value, as given by the trained model, was used to calculate the KDE probability for all the datasets. Figure 5 shows the distribution of the REs and the values of the KDE function for all the three datasets. By illustrating these values in a scatter plot, we can evaluate whether the model can detect an anomaly (and locate it if present) in an unknown dataset. Figure 6 illustrates the data point distribution using both values.

We propose two ways to evaluate the performance of the classification model. The first approach is to look at where



Figure 4. Some of the limes used to test the model's performance. These images had the background segmented. The limes were harvested from orchards located in the town of Tecolapa, in the state of Colima, Mexico. As can be seen, all the fruits exhibit at least one type of defect that affects their visual quality.

Source: Authors

the value of a new image would lie in the plot after the model reconstructs it and calculates the value of the KDE function. However, this strictly necessitates visualization in the scatter plot in order to provide an accurate result. The second approach is to establish two thresholds, as shown in Equation (1), where *healthy* (normal class) and *anomaly* (anomalous class) are the possible output classes; *error* refers to the RE value; *density* represents the KDE function value; e_{min} and e_{max} denote the lower and upper threshold values for the RE, respectively; and d_{min} and d_{max} represent the lower and upper threshold values for the KDE function. These values are based on the minimum and maximum values observed in both histograms in Figure 5.

$$label = \begin{cases} \text{healthy if } e_{min} < error < e_{max} \ \& \ d_{min} < density < d_{max} \\ \text{anomaly otherwise} \end{cases} \quad (1)$$

These thresholds are used to automate the classification of the faulty limes. Therefore, if a new lime image has a value within both thresholds, it will be considered a healthy lime; otherwise, it will be deemed defective. The threshold values determine the cutoff points for distinguishing between anomalies and normal instances.

In an anomaly detection application, the objective of the model is to find all possible anomalies in a given dataset. In this work, a confusion matrix (Table 2) was elaborated under the following considerations: true positives (TP) were the anomalies (defective limes) that lay outside the proposed thresholds; false positives (FP) corresponded to all the limes that were anomalies but fell between the established thresholds; true negatives (TN) were all healthy limes (if any) that were correctly classified as such; and false negatives (FN) are the healthy ones misclassified as defective.

Table 3 presents the classification reports, as calculated using the SkLearn framework (Pedregosa et al., 2011), the k-nearest neighbors algorithm (kNN), and principal components analysis (PCA).

The value of the classification model's metrics has different interpretations. The recall value (0.88) is higher for the anomalous class than for the normal class (0.80). For the normal class, the precision is 0.65, meaning that the number of instances detected is lower than that of the anomalous class (0.94). To summarize, the classification model performs better at identifying anomalies (unhealthy lime fruits) than in normal instances. This behavior suggests that the model can identify instances of the anomalous class, i.e., it has a strong ability to detect the TPs of said class in comparison with the normal one. The high recall values for both classes are also desirable under these circumstances,

as they minimize the FN rate while increasing the detection rate of TPs without causing misclassification. Given that the value is almost identical, it can be assumed that the model captures the same proportion of positive instances (either normal or anomalous) across different scenarios and circumstances in different databases, meaning that the model is consistent when making predictions with unknown data.

A comparison with a more recent approach was carried out, i.e., an autoencoder using the VGG16 architecture. This architecture is among the most widely used for classification and feature extraction tasks. Additionally, it benefits from being pre-trained on the ImageNet database. By utilizing the TensorFlow framework, it was possible to work with a pre-trained version of VGG16 and employ transfer learning, followed by fine-tuning with images from the Fruits360 dataset. The VGG16 architecture is a deep CNN that consists of 16 layers, primarily composed of convolutional and fully connected layers. The architecture includes five convolutional blocks, each followed by max-pooling layers, and three fully connected layers at the end, making it highly effective for image classification tasks. The decoder part of VGG16 was implemented using TensorFlow and Python, following the same structure as the network, albeit inversely. Since the decoder's task is merely to reconstruct the input data, fine-tuning this section of the autoencoder was not necessary. The number of epochs, data distribution, and hyperparameters were consistent with those used in our autoencoder model. Moreover, as with our proposal, the RE and KDE values were calculated. Figure 7 shows their distribution.

An analysis of the density plots and the distribution of the points revealed distinct patterns between the training and anomaly datasets. Upon evaluating the overlap between these distributions, a small number of anomaly points (orange crosses) were found to lie within the high-density region of the training data (blue stars), signifying incorrect classification.

This approach demonstrates a similar precision value to our proposed method (around 95%). Notably, the number of anomalous points that lie within the normal data is approximately 20, which is higher than that of our method. Nonetheless, a lower proportion of the normal points are within the region of anomalous data.

Discussion

Since this work aims to detect as many anomalies as possible, the precision metric was prioritized. For example, a FN rate is more critical in a medical diagnosis scenario than a FP. In this case, classifying a healthy lime as faulty is less critical than misclassifying a faulty one as healthy. As shown in Table 2, the model has fewer FNs than FPs, indicating its effectiveness for the given task.

Table 4 lists studies that solve a task similar to that proposed in our study, and it offers a broader perspective on our proposal and some related works found in the literature. Most of these studies use some kind of machine learning (ML) algorithm for classification, with CNNs being among the most popular approaches and the ones that report better accuracy values. As for the feature extraction process, the methods differ significantly according to the various targets

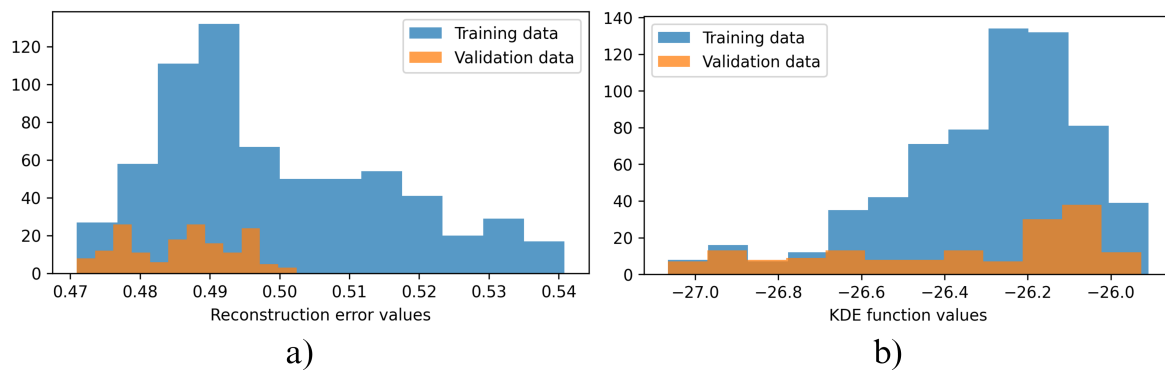


Figure 5. Distribution of values for the training and validation datasets. a) Histogram of RE values for both datasets; b) histogram showing the KDE function values for both datasets. The blue bars represent training data, and the orange bars represent the validation data. The distributions reveal that the RE values for both datasets lie within the expected ranges, highlighting the effectiveness of the proposed method in identifying regions corresponding to healthy lime fruit.

Source: Authors

Table 2. Confusion matrix built with the predictions

		Predicted		Total
		Anomaly	Normal	
Actual	Anomaly	82 (TP)	11 (FN)	93
	Normal	5 (FP)	20 (TN)	25
Total		87	31	118

Source: Authors

of the studies. The accuracy value for the listed works ranges from 88 to 98%.

In comparison with an autoencoder built upon a well-established architecture such as VGG16, our method demonstrates a slightly better performance despite being trained entirely from scratch, while the VGG16 autoencoder utilized a pre-trained version. Although the precision values are comparable, the key differences lie in training times and computational efficiency. Our autoencoder required approximately 45 minutes for training, whereas the VGG16 model took nearly 120 minutes. Furthermore, despite the shallower depth of our autoencoder's architecture compared to VGG16, it achieved comparable results, highlighting that effective performance can be obtained by training a new model from scratch. Nevertheless, one advantage of using pre-trained models is the ability to leverage pre-existing weights, which significantly aids in feature extraction and the identification of relevant information.

The anomalous dataset reveals similarities with the training images in terms of color, shape, and texture. Nonetheless, certain anomalies were mistakenly classified as healthy by the model. This misclassification could be attributed to outliers within the anomalous set, indicating more severe damage or advanced stages of disease. The RE values for the anomalous set range from 0.2 to 0.9, suggesting that the autoencoder models reconstructions are generally reliable. However, a significant limitation of this semi-supervised approach is its reliance on determining thresholds for the KDE density function and RE values in the two-dimensional space. These thresholds are based on the source data, which may introduce inherent biases. Furthermore, the choice of

autoencoder architecture and data pre-processing methods can impact the RE and the KDE density function values, potentially contributing to biases in the classification system.

Additionally, most of the works found in the literature use their own datasets, which were created for their specific purposes. As a result, it is not possible to make direct comparisons between works, and there is no common benchmark to compare the results. This is important because it limits the solutions to local and specific contexts. The solutions proposed so far work with regional varieties of lime fruit. Using these varieties to build datasets can introduce biases that hinder the construction of a general solution for this type of precision agriculture application. It should also be considered that the types of defects change depending on the harvest location. This phenomenon is multi-factorial due to the different soil types, climatic conditions, and harvest times present. Although some of the studies found show interesting solutions, they are applied in local contexts, either due to a lack of data or to solve a particular need. However, there is a trend to develop more general solutions that can be applied and scaled to broader and more general contexts (Verma and Verma, 2022; Bhardwaj et al., 2022; Shahi et al., 2022).

The experimental results confirm that the proposed classification model is sufficiently robust, as it accurately classifies new lime images, even when they differ significantly from the training set in key features such as color, hue, shape, size, and lighting conditions.

Another advantage of the classification model is that it uses an innovative and robust approach to solve a problem that

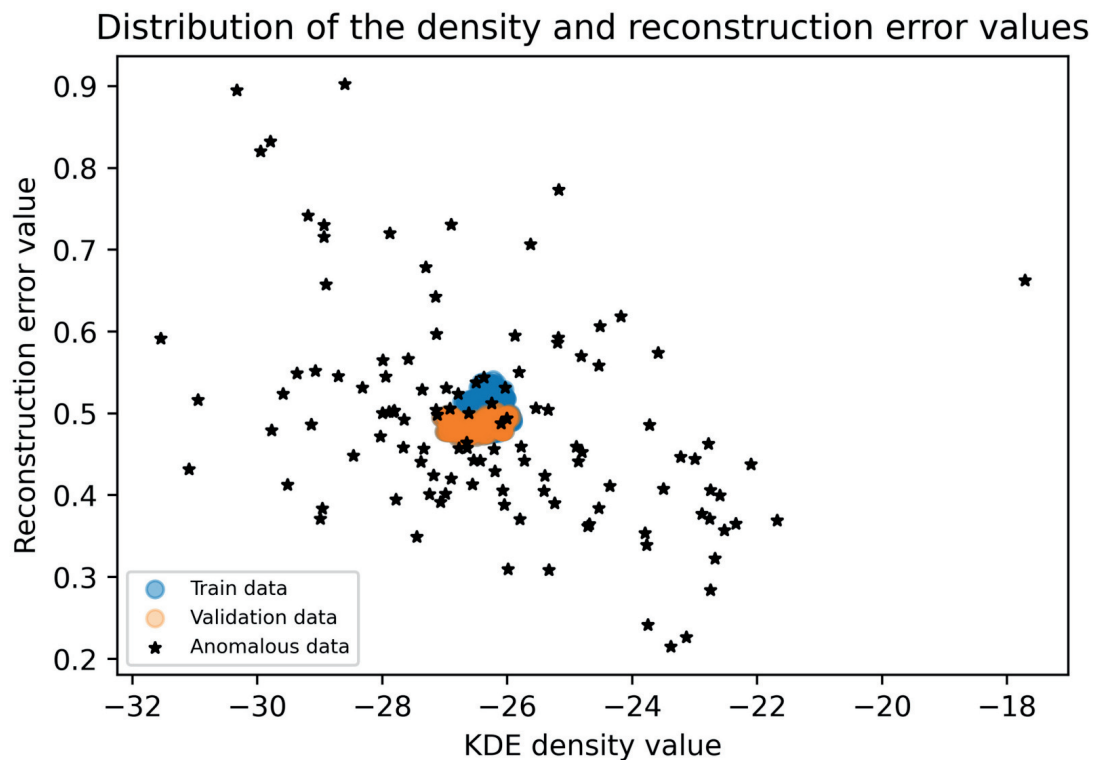


Figure 6. Scatter plot made using the KDE probability value and the RE. The red and blue regions indicate the training and validation datasets, respectively. The black stars are the distribution of the data points for the test set. The majority of the test data points lie outside the red and blue regions. However, some data points fell into them, implying that there are some limes that the model missed to detect as anomalies.

Source: Authors

Table 3. Classification report provided by Sklearn with the metrics of the proposed model, the kNN algorithm, and the PCA

	Our method			KNN			PCA		
	Precision	Recall	F1 score	Precision	Recall	F1 score	Precision	Recall	F1 score
Normal	0.65	0.80	0.71	0.64	0.95	0.70	0.76	0.73	0.76
Anomalies	0.94	0.88	0.91	0.97	0.84	0.91	0.92	0.93	0.93

Source: Authors

has been present for many years. The anomaly detection method, along with the proposed architecture, showed that it is possible to achieve similar results without large amounts of data and long training times. Additionally, this method can detect unknown anomalies that are not present in the training data, whereas a balanced classification approach assumes that all classes are known and equally present in the training set.

Although the limes used for validation have at least one kind of visual defect, the model misclassified some of them as healthy. This phenomenon is understandable since some of the defects are not as evident as in the training images or have not reached a critical state. For example, color change is a known defect in citrus fruit, but the transition from green to yellow is so subtle that is difficult to identify. Another example is when limes have tiny black spots on their peel. According to the experts, the lime is defective (since it implies the presence of a plague or disease), but, for the classification model, it is interpreted as healthy. However, including images with this kind of characteristic may significantly improve the performance.

The proposed classification model uses color as a critical feature for lime classification. This behavior can be noticed in the training and validation data distribution. The images used for this purpose have a similar green hue. Meanwhile, the test images have a brighter green hue, and, in some cases, the fruit color is yellow instead of green.

Another factor that could compromise the results is the type of defect: if the defect is more evident, i.e., brown spots or peel injuries, the model will more easily identify the fruit as defective. On the other hand, less obvious defects such as plagues or diseases could produce FNs during classification. This behavior could suggest that the human factor is still needed. Looking at the scatter plot is useful and can serve as a quick way to decide, but looking at the lime image can ensure that the decision is correct.

The study most similar to ours is the one by (Yilmaz et al., 2023), as their methodology, fruit studied, and network design are based on a variant of autoencoders: stacked autoencoders (SAE). Their SAE used different color and morphological features as input to the network. These features were extracted using techniques involving the

Table 4. Comparison of the works found in the literature

Work	Metric used	Value (%)	Approach	Application
Proposed method	Precision	94%	Semi-supervised anomaly detection	Lime fruit classification
(Hernández et al., 2021)	Accuracy	92%	Convolutional neural networks	Lime quality classification
(Schor et al., 2016)	Accuracy	95%	PCA-based classification	Tomato virus detection
(Cabrera Ardila et al., 2020a)	Accuracy	91%	Spectroscopy and Linear discriminant analysis	Anthraxose detection in mangoes
(Rong et al., 2017)	Accuracy	97%	Window local segmentation algorithm	Defective orange peel detection
(Hanh and Bao, 2022)	Accuracy	95%	Convolutional neural networks	Lemon quality classification
(Bird et al., 2022)	Accuracy	88.75%	Conditional GAN network	Lemon quality and defect classification
(Hasan et al., 2021)	Accuracy	94.11%	Convolutional neural networks	Citrus fruit categorization
(Chen et al., 2018)	Accuracy	97.5%	Neural networks	Orange detection and classification
(Yilmaz et al., 2023)	Accuracy	98.96%	Stacked autoencoders	Lemon quality classification
(Roy et al., 2021)	Accuracy	97.54%	Semantic segmentation	Rotten and fresh fruit detection

Source: Authors



Figure 7. Scatter plot obtained from the KDE and RE values of the VGG16 autoencoder. The orange crosses are the anomalous points (defective limes), while the blue stars correspond to the healthy fruit. Source: Authors

gray-level co-occurrence matrix (GLCM), color space, and morphological methods. The authors used labeled data with two classes (healthy and defective limes) from a public database. This model achieved an accuracy of 98.96% while using only 32 of the previously mentioned features. Furthermore, the authors mention that their metrics are appropriate due to the size and quality of the database images, which allowed them to discard redundant features during training. This study presents interesting advancements in lime classification. However, one of its constraints lies in the methodology for extracting features from the images; as the number of features increases, the computational cost also does, which may be problematic for real-time applications or *in situ* implementations. However, the utilization of hybrid models suggests a new way to approach classification tasks.

Conclusions

The method proposed in this paper used a semi-supervised anomaly detection approach for lime fruit classification, and it achieved an accuracy of 94% for the anomalous data. This

means that the model is competitive with other approaches found in the literature in terms of performance.

As future work, we will test our model with a larger dataset and redesign it for a more robust classification. Although our dataset was large enough for our specific purpose, the color of the limes in the Fruits360 dataset is darker than that of the subspecies in Mexico. Recognizing this issue, we found that having a dataset with Mexican limes is necessary; in this way, the predictions will enrich model performance, benefiting local producers.

We will explore some variants of the autoencoders, such as variational autoencoders, or even other deep learning approaches like reinforcement learning. Recent advances in these artificial intelligence paradigms show promising results and could tackle the problem in different ways, leading to a more efficient or accurate classification method.

Acknowledgements

The authors would like to thank the lime producers from Tecolapa, Colima, Mexico, for the fruit provided, their insightful point of view, and for their knowledge. In addition, the authors would like to thank José J. Rico-Jiménez, who reviewed the overall grammar and style of the manuscript. This work was supported by Secretaría de Investigación y Posgrado of the Instituto Politécnico Nacional, under grants 20241653 and 20240650.

CRedit author statement

Angel-Moisés Hernández-Ponce: conceptualization, methodology, software, writing (original draft) Francisco-Javier Ornelas-Rodríguez: writing (review and editing), supervision Juan-Bautista Hurtado-Ramos: writing (review and editing) Jose-Joel Gonzalez-Barbosa: writing (review and editing), software, supervision Alfonso Ramírez-Pedraza: writing (review and editing)

Conflicts of interest

The authors declare no conflict of interest.

Data availability

The dataset used in the training stage was obtained from <https://www.kaggle.com/datasets/moltean/fruits>.

The dataset used in the experimental stage can be made available upon request.

References

- Aparicio Pico, L. E., Devia Lozano, P., and Amaya Marroquin, O. J. (2022). Aplicación de deep learning para la identificación de defectos superficiales utilizados en control de calidad de manufactura y producción industrial: Una revisión de la literatura. *Ingeniería*, 28(1), e18934. <https://dx.doi.org/10.14483/23448393.18934>.
- Arnal Barbedo, J. G. (2019). Plant disease identification from individual lesions and spots using deep learning. *Biosystems Engineering*, 180, 96-107. <https://dx.doi.org/10.1016/j.biosystemseng.2019.02.002>.
- Ayoub Shaikh, T., Rasool, T., and Rasheed Lone, F. (2022). Towards leveraging the role of machine learning and artificial intelligence in precision agriculture and smart farming. *Computers and Electronics in Agriculture*, 198, 107119. <https://dx.doi.org/10.1016/j.compag.2022.107119>.
- Bank, D., Koenigstein, N., and Giryas, R. (2021). Autoencoders. *arXiv preprint*. <https://doi.org/10.48550/arXiv.2003.05991>.
- Bao, M., Zheng, Z., Chen, J., and Deng, X. (2021). Investigation of citrus hlb symptom variations associated with candidatus liberibacter asiaticus strains harboring different phages in southern china. *Agronomy*, 11(11), 2262. <https://dx.doi.org/10.3390/agronomy11112262>.
- Beggel, L., Pfeiffer, M., and Bischl, B. (2020). Robust anomaly detection in images using adversarial autoencoders. In U. Brefeld, E. Fromont, A. Hotho, A. Knobbe, M. Maathuis, and C. Robardet (Eds.), *Machine Learning and Knowledge Discovery in Databases*. Springer International Publishing. https://doi.org/10.1007/978-3-030-46150-8_13, pages="206--222".
- Bhardwaj, A., Hasteer, N., Kumar, Y., and Yogesh. (2022). Deep learning based fruit defect detection system. In S. Mekhilef, R. N. Shaw, and P. Siano (Eds.), *Innovations in Electrical and Electronic Engineering* (pp. 306-313. https://doi.org/10.1007/978-981-19-1677-9_28). Springer Singapore.
- Bhole, A., Varsha Kumar. (2021). A transfer learning-based approach to predict the shelf life of fruit. *Inteligencia Artificial*, 24(67), 102120. <https://doi.org/10.4114/intartif.vol24iss67pp102-120>.
- Bird, J. J., Barnes, C. M., Manso, L. J., Ekárt, A., and Faria, D. R. (2022). Fruit quality and defect image classification with conditional gan data augmentation. *Scientia Horticulturae*, 293, 110684. <https://doi.org/10.1016/j.scienta.2021.110684>.
- Blasco, J., Cubero, S., and Moltó, E. (2016, 4). Quality evaluation of citrus fruits. In *Computer Vision Technology for Food Quality Evaluation: Second Edition* (pp. 305-325). Elsevier. <https://dx.doi.org/10.1016/B978-0-12-802232-0.00012-8>.
- Cabrera Ardila, C. E., Alberto Ramirez, L., and Prieto Ortiz, F. A. (2020a). Spectral analysis for the early detection of anthracnose in fruits of sugar mango (*mangifera indica*). *Computers and Electronics in Agriculture*, 173, 105357. <https://doi.org/10.1016/j.compag.2020.105357>.
- Cabrera Ardila, C. E., Alberto Ramirez, L., and Prieto Ortiz, F. A. (2020b). Spectral analysis for the early detection of anthracnose in fruits of sugar mango (*Mangifera indica*). *Computers and Electronics in Agriculture*, 173, 105357. <https://doi.org/10.1016/j.compag.2020.105357>.
- Cao, V. L., Nicolau, M., and McDermott, J. (2016). A hybrid autoencoder and density estimation model for anomaly detection. In J. Handl, E. Hart, P. R. Lewis, M. López-Ibáñez, G. Ochoa, and B. Paechter (Eds.), *Parallel problem solving from nature – ppsn xiv* (pp. 717-726). Springer International Publishing. https://doi.org/10.1007/978-3-319-45823-6_67.
- Cazzonelli, L., and Kulbach, C. (2023). Detecting anomalies with autoencoders on data streams. In M.-R. Amini, S. Canu, A. Fischer, T. Guns, P. Kralj Novak, and G. Tsoumakas (Eds.), *Machine learning and knowledge discovery in databases* (pp. 258-274). Springer International Publishing. https://doi.org/10.1007/978-3-031-26387-3_16.
- Chaturvedi, A., Sharma, S., and Janghel, R. R. (2023). Detection of external defects in tomatoes using deep learning. *Journal of Ambient Intelligence and Humanized Computing*, 14(3), 2709-2721. <https://doi.org/10.1007/s12652-023-04514-y>.
- Chen, Y., Wu, J., and Cui, M. (2018). Automatic classification and detection of oranges based on computer vision. In *2018 IEEE 4th International Conference on Computer and Communications (ICCC)* (p. 1551-1556. <https://doi.org/10.1109/CompComm.2018.8780680>).
- Cubero, S., Aleixos, N., Albert, F., Torregrosa, A., Ortiz, C., García-Navarrete, O., and Blasco, J. (2014). Optimised computer vision system for automatic pre-grading of citrus fruit in the field using a mobile platform. *Precision Agriculture*, 15(1), 80-94. <https://doi.org/10.1007/s11119-013-9324-7>.
- Cándido-Mireles, M., Hernández-Gama, R., and Salas, J. (2023). Detecting vineyard plants stress in situ using deep learning. *Computers and Electronics in Agriculture*, 210, 107837. <https://doi.org/10.1016/j.compag.2023.107837>.
- Dandekar, M., Pun, N. S., Sonbhadra, S. K., and Agarwal, S. (2020). Fruit classification using deep feature maps in the presence of deceptive similar classes. *CoRR*. <https://arxiv.org/abs/2007.05942>.
- Dhiman, P., Kukreja, V., Manoharan, P., Kaur, A., Kamruzzaman, M. M., Dhaou, I. B., and Iwendi, C. (2022). A novel deep learning model for detection of severity level of the disease in citrus fruits. *Electronics*, 11(3), 495. <https://doi.org/10.3390/electronics11030495>.
- dos Santos Tanaka, F. H. K., and Aranha, C. (2019). Data augmentation using gans. *arXiv preprint*. <https://doi.org/10.48550/ARXIV.1904.09135>.

- Fan, S., Li, J., Zhang, Y., Tian, X., Wang, Q., He, X., ... Huang, W. (2020). On line detection of defective apples using computer vision system combined with deep learning methods. *Journal of Food Engineering*, 286, 110102. <https://doi.org/10.1016/j.jfoodeng.2020.110102>.
- Food and Agriculture Organization. (2022). *World food and agriculture statistical yearbook 2022* (Tech. Rep.). Food and Agriculture Organization of the United Nations. <https://openknowledge.fao.org/handle/20.500.14283/cc2211en>.
- Futerman, S. I., Laor, Y., Eshel, G., and Cohen, Y. (2023). The potential of remote sensing of cover crops to benefit sustainable and precision fertilization. *Science of The Total Environment*, 891, 164630. <https://dx.doi.org/10.1016/j.scitotenv.2023.164630>.
- Gao, X., Huang, C., Teng, S., and Chen, G. (2022). A deep-convolutional-neural-network-based semi-supervised learning method for anomaly crack detection. *Applied Sciences*, 12(18), 9244. <https://dx.doi.org/10.3390/app12189244>.
- George, A., Rao, C. N., and Mani, M. (2022). Pests of citrus and their management. In M. Mani (Ed.), *Trends in Horticultural Entomology* (pp. 551–575). Springer Nature Singapore. https://doi.org/10.1007/978-981-19-0343-4_17.
- Ghazal, S., Qureshi, W. S., Khan, U. S., Iqbal, J., Rashid, N., and Tiwana, M. I. (2021). Analysis of visual features and classifiers for fruit classification problem. *Computers and Electronics in Agriculture*, 187, 106267. <https://doi.org/10.1016/j.compag.2021.106267>.
- Gokulnath, B. V., and Usha Devi, G. (2021). Identifying and classifying plant disease using resilient LF-CNN. *Ecological Informatics*, 63, 101283. <https://doi.org/10.1016/j.ecoinf.2021.101283>.
- González-Barbosa, J.-J., Ramírez-Pedraza, A., Ornelas-Rodríguez, F.-J., Cordova-Esparza, D.-M., and González-Barbosa, E.-A. (2022). Dynamic measurement of portos tomato seedling growth using the kinect 2.0 sensor. *Agriculture*, 12(4), 449. <https://doi.org/10.3390/agriculture12040449>.
- Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep learning*. MIT Press. <http://www.deeplearningbook.org>.
- Gron, A. (2017). *Hands-on machine learning with scikit-learn and tensorflow: Concepts, tools, and techniques to build intelligent systems* (1st ed.). O'Reilly Media.
- Hanh, L. D., and Bao, D. N. T. (2022, May 29). Autonomous lemon grading system by using machine learning and traditional image processing. *International Journal on Interactive Design and Manufacturing (IJIDeM)*, 17, 445-452. <https://doi.org/10.1007/s12008-022-00926-w>.
- Hasan, M. M., Salehin, I., Moon, N. N., Kamruzzaman, T. M., Baki-Ul-Islam, and Hasan, M. (2021). A computer vision system for the categorization of citrus fruits using convolutional neural network. In IEEE (Ed.), *2021 International Symposium on Electronics and Smart Devices (ISESD)* (p. 1-6). <https://doi.org/10.1109/ISESD53023.2021.9501790>. IEEE.
- Hernández, A., Ornelas-Rodríguez, F. J., Hurtado-Ramos, J. B., and González-Barbosa, J. J. (2021). Accuracy comparison between deep learning models for mexican lemon classification. In M. F. Mata-Rivera and R. Zagal-Flores (Eds.), *Telematics and Computing* (pp. 62–73). Springer International Publishing.
- Hu, W., Gao, J., Li, B., Wu, O., Du, J., and Maybank, S. (2020). Anomaly detection using local kernel density estimation and context-based regression. *IEEE Transactions on Knowledge and Data Engineering*, 32(2), 218-233. <https://doi.org/10.1109/TKDE.2018.2882404>.
- Ibrahim, M., and Kuban, K. (2023). Key generation method from fingerprint image based on deep convolutional neural network model. *Nexo Revista Científica*, 36, 906-925. <https://dx.doi.org/10.5377/nexo.v36i06.17447>.
- Istiaq, M. A., Syeed, M. M., Hossain, M. S., Uddin, M. F., Hasan, M., Khan, R. H., and Azad, N. S. (2023). Adoption of Unmanned Aerial Vehicle (UAV) imagery in agricultural management: A systematic literature review. *Ecological Informatics*, 78, 102305. <https://doi.org/10.1016/j.ecoinf.2023.102305>.
- Khanramaki, M., Askari Asli-Ardeh, E., and Kozegar, E. (2021). Citrus pests classification using an ensemble of deep learning models. *Computers and Electronics in Agriculture*, 186, 106192. <https://doi.org/10.1016/j.compag.2021.106192>.
- Lang, C. I., Sun, F.-K., Lawler, B., Dillon, J., Dujaili, A. A., Ruth, J., ... Boning, D. S. (2022). One class process anomaly detection using kernel density estimation methods. *IEEE Transactions on Semiconductor Manufacturing*, 35(3), 457-469. <https://doi.org/10.1109/TSM.2022.3181468>.
- Latif, G., Mohammad, N., and Alghazo, J. (2023). Deepfruit: A dataset of fruit images for fruit classification and calories calculation. *Data in Brief*, 50, 109524. <https://doi.org/10.1016/j.dib.2023.109524>.
- Li, D., Song, Z., Quan, C., Xu, X., and Liu, C. (2021). Recent advances in image fusion technology in agriculture. *Computers and Electronics in Agriculture*, 191, 106491. <https://doi.org/10.1016/j.compag.2021.106491>.
- Li, E., Wang, L., Xie, Q., Gao, R., Su, Z., and Li, Y. (2023). A novel deep learning method for maize disease identification based on small sample-size and complex background datasets. *Ecological Informatics*, 75, 102011. <https://doi.org/10.1016/j.ecoinf.2023.102011>.
- Lozano, E., and Archibald, D. (2022). *Citrus annual*. United States Department of Agriculture. <https://www.fas.usda.gov/data/mexico-citrus-annual-7>.

- Lu, Y., and Young, S. (2020). A survey of public datasets for computer vision tasks in precision agriculture. *Computers and Electronics in Agriculture*, 178, 105760. <https://doi.org/10.1016/j.compag.2020.105760>.
- Maheshwari, A., Mitra, P., and Sharma, B. (2022). Autoencoder: Issues, challenges and future prospect. In M. Vashista, G. Manik, O. P. Verma, and B. Bhardwaj (Eds.), *Recent Innovations in Mechanical Engineering* (pp. 257–266). https://doi.org/10.1007/978-981-16-9236-9_24. Springer Singapore.
- Memarzadeh, M., Akbari Asanjan, A., and Matthews, B. (2022). Robust and explainable semi-supervised deep learning model for anomaly detection in aviation. *Aerospace*, 9(8), 437. <https://doi.org/10.3390/aerospace9080437>.
- Michelucci, U. (2022). An introduction to autoencoders. *CoRR*. <https://arxiv.org/abs/2201.03898>.
- Momeny, M., Jahanbakhshi, A., Neshat, A. A., Hadipour-Rokni, R., Zhang, Y.-D., and Ampatzidis, Y. (2022). Detection of citrus black spot disease and ripeness level in orange fruit using learning-to-augment incorporated deep networks. *Ecological Informatics*, 71, 101829. <https://doi.org/10.1016/j.ecoinf.2022.101829>.
- Murean, H., and Oltean, M. (2018, 06). Fruit recognition from images using deep learning. *Acta Universitatis Sapientiae, Informatica*, 10, 26-42. <https://doi.org/10.2478/ausi-2018-0002>.
- Mzoughi, O., and Yahiaoui, I. (2023). Deep learning-based segmentation for disease identification. *Ecological Informatics*, 75, 102000. <https://doi.org/10.1016/j.ecoinf.2023.102000>.
- Naranjo-Torres, J., Mora, M., Fredes, C., and Valenzuela, A. (2021). Disease and defect detection system for raspberries based on convolutional neural networks. *Applied Sciences*, 11(24), 11868. <https://doi.org/10.3390/app112411868>.
- Nowak, B. (2021). Precision agriculture: Where do we stand? a review of the adoption of precision agriculture technologies on field crops farms in developed countries. *Agricultural Research*, 10(4), 515–522. <https://doi.org/10.1007/s40003-021-00539-x>.
- Palei, S., Behera, S. K., and Sathy, P. K. (2023). A systematic review of citrus disease perceptions and fruit grading using machine vision. *Procedia Computer Science*, 218, 2504-2519. <https://doi.org/10.1016/j.procs.2023.01.225>.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830. <https://dl.acm.org/doi/10.5555/1953048.2078195>.
- Peng, K., Ma, W., Lu, J., Tian, Z., and Yang, Z. (2023). Application of machine vision technology in citrus production. *Applied Sciences*, 13(16), 9334. <https://doi.org/10.3390/app13169334>.
- Puerto Cuadros, E. G. (2024, Nov.). Advanced neural model for spanish spell-checking. *Ingeniería*, 29(3), e21135. <https://dx.doi.org/10.14483/23448393.21135>.
- Rathnayake, N., Rathnayake, U., Dang, T. L., and Hoshino, Y. (2022). An efficient automatic fruit-360 image identification and recognition using a novel modified cascaded-anfis algorithm. *Sensors*, 22(12), 4401. <https://doi.org/10.3390/s22124401>.
- Rong, D., Rao, X., and Ying, Y. (2017). Computer vision detection of surface defect on oranges by means of a sliding comparison window local segmentation algorithm. *Computers and Electronics in Agriculture*, 137, 59-68. <https://doi.org/10.1016/j.compag.2017.02.027>.
- Roy, K., Chaudhuri, S. S., and Pramanik, S. (2021, September). Deep learning based real-time industrial framework for rotten and fresh fruit detection using semantic segmentation. *Microsystem Technologies*, 27(9), 3365–3375. <https://doi.org/10.1007/s00542-020-05123-x>.
- Schor, N., Bechar, A., Ignat, T., Dombrovsky, A., Elad, Y., and Berman, S. (2016). Robotic disease detection in greenhouses: Combined detection of powdery mildew and tomato spotted wilt virus. *IEEE Robotics and Automation Letters*, 1(1), 354-360. <https://doi.org/10.1109/LRA.2016.2518214>.
- Shahi, T. B., Sitaula, C., Neupane, A., and Guo, W. (2022). Fruit classification using attention-based mobilenetv2 for industrial applications. *PLOS ONE*, 17(2), 1-21. <https://doi.org/10.1371/journal.pone.0264586>.
- Shijie, J., Ping, W., Peiyi, J., and Siping, H. (2017). *Research on data augmentation for image classification based on convolution neural networks*. 2017 Chinese Automation Congress (CAC) [Conference paper]. <https://doi.org/10.1109/CAC.2017.8243510>.
- Siddiqi, R. (2020). Comparative performance of various deep learning based models in fruit image classification. In ACM (Ed.), *11th International Conference on Advances in Information Technology* (p. 14). Association for Computing Machinery. <https://doi.org/10.1145/3406601.3406619>.
- Soltani Firouz, H., Mahmoud Sardari. (2022). Defect detection in fruit and vegetables by using machine vision systems and image processing. *Food Engineering Reviews*, 14(3), 353–379. <https://doi.org/10.1007/s12393-022-09307-1>.
- Tan, A., Zhou, G., and He, M. (2021). Surface defect identification of citrus based on kf-2d-renyi and abc-svm. *Multimedia Tools and Applications*, 80(6), 9109–9136. <https://doi.org/10.1007/s11042-020-10036-y>.
- Thoidis, I., Giouvanakis, M., and Papanikolaou, G. (2021). Semi-supervised machine condition monitoring by learning deep discriminative audio features. *Electronics*, 10(20), 2471. <https://doi.org/10.3390/electronics10202471>.

- Toosi, A., Javan, F. D., Samadzadegan, F., Mehravar, S., Kurban, A., and Azadi, H. (2022). Citrus orchard mapping in juybar, iran: Analysis of ndvi time series and feature fusion of multi-source satellite imageries. *Ecological Informatics*, 70, 101733. <https://doi.org/10.1016/j.ecoinf.2022.101733>.
- Verma, R., and Verma, A. K. (2022). Fruit classification using deep convolutional neural network and transfer learning. In V. E. Balas, G. R. Sinha, B. Agarwal, T. K. Sharma, P. Dadheech, and M. Mahrishi (Eds.), *Emerging Technologies in Computer Engineering: Cognitive Computing and Intelligent IoT* (pp. 290–301). Springer International Publishing.
- Wang, T., Zhao, L., Li, B., Liu, X., Xu, W., and Li, J. (2022). Recognition and counting of typical apple pests based on deep learning. *Ecological Informatics*, 68, 101556. <https://doi.org/10.1016/j.ecoinf.2022.101556>.
- Yılmaz, E. K., Adem, K., Kılıçarslan, S., and Aydın, H. A. (2023, June). Classification of lemon quality using hybrid model based on stacked AutoEncoder and convolutional neural network. *European Food Research and Technology*, 249(6), 1655–1667. <https://doi.org/10.1007/s00217-023-04245-6>.
- Zhang, H., Chen, Y., Liu, X., Huang, Y., Zhan, B., and Luo, W. (2021). Identification of common skin defects and classification of early decayed citrus using hyperspectral imaging technique. *Food Analytical Methods*, 14(6), 1176–1193. <https://doi.org/10.1007/s12161-020-01960-8>.
- Ümit Atila, Uçar, M., Akyol, K., and Uçar, E. (2021). Plant leaf disease classification using efficientnet deep learning model. *Ecological Informatics*, 61, 101182. <https://doi.org/10.1016/j.ecoinf.2020.101182>.

Neural Networks and Fuzzy Logic-Based Approaches for Precipitation Estimation: A Systematic Review

Enfoques basados en redes neuronales y lógica difusa para la estimación de la precipitación: una revisión sistemática

Andrés F. Ruiz-Hurtado ¹, Viviana Vargas-Franco ², and Luis O. González-Salcedo ³

ABSTRACT

Precipitation estimation at the river basin level is essential for watershed management, the analysis of extreme events and weather and climate dynamics, and hydrologic modeling. In recent years, new approaches and tools such as artificial intelligence techniques have been used for precipitation estimation, offering advantages over traditional methods. Two major paradigms are artificial neural networks and fuzzy logic systems, which can be used in a wide variety of configurations, including hybrid and modular models. This work presents a literature review on hybrid metaheuristic and artificial intelligence models based on signal processes, focusing on the applications of these techniques in precipitation analysis and estimation. The selection and comparison criteria used were the model type, the input and output variables, the performance metrics, and the fields of application. An increase in the number of this type of studies was identified, mainly in applications involving neural network models, which tend to get more sophisticated according to the availability and quality of training data. On the other hand, fuzzy logic models tend to hybridize with neural models. There are still challenges related to prediction performance and spatial and temporal resolution at the basin and micro-basin levels, but, overall, these paradigms are very promising for precipitation analysis.

Keywords: precipitation, river basin, neural networks, fuzzy logic, machine learning, fuzzy inference systems

RESUMEN

La estimación de la precipitación a nivel de cuenca hidrográfica es esencial para la gestión de cuencas, el análisis de eventos extremos y dinámicas meteorológicas y climáticas, y el modelado hidrológico. En los últimos años se han empleado nuevos enfoques y herramientas como las técnicas de inteligencia artificial para estimar la precipitación, ofreciendo ventajas sobre los métodos tradicionales. Dos paradigmas principales son las redes neuronales artificiales y los sistemas de lógica difusa, que pueden utilizarse en una amplia variedad de configuraciones, incluyendo modelos híbridos y modulares. Este trabajo presenta una revisión de la literatura sobre modelos híbridos metaheurísticos y de inteligencia artificial basados en procesos de señales, centrándose en las aplicaciones de estas técnicas en el análisis y la estimación de la precipitación. Los criterios de selección y comparación utilizados fueron el tipo de modelo, las variables de entrada y salida, las métricas de desempeño y los campos de aplicación. Se identificó un aumento en el número de este tipo de estudios, principalmente en aplicaciones que involucran modelos de redes neuronales, los cuales tienden a volverse más sofisticados según la disponibilidad y calidad de los datos de entrenamiento. Por otro lado, los modelos de lógica difusa tienden a hibridarse con modelos neuronales. Aún existen desafíos relacionados con el desempeño de las predicciones y la resolución espacial y temporal a nivel de cuenca y microcuenca, pero, en general, estos paradigmas son muy prometedores para el análisis de la precipitación.

Palabras clave: precipitación, cuenca hidrográfica, redes neuronales, lógica difusa, aprendizaje automático, sistemas de inferencia difusa

Received: May 4th, 2023

Accepted: December 11th, 2024

Introduction

Precipitation is a critical component of the global water cycle, significantly influencing both climatic and hydrological dynamics [1]. Variations in precipitation intensity have diverse impacts on natural and societal systems [2]. For instance, light rainfall, which soils readily absorb, aids in drought mitigation and boosts agricultural productivity. In contrast, intense downpours frequently result in catastrophic floods and landslides. Consequently, a thorough understanding of the precipitation intensity spectrum is vital for developing specific adaptation strategies. Estimating precipitation at the watershed level is highly valuable for environmental studies, given its role as the primary input in a hydrological system,

directly contributing to the analysis of the water budget and related socio-economic and ecosystem interactions [3]. Therefore, accurately estimating precipitation is crucial for understanding meteorological and hydroclimatic processes and their impact on extreme events such as floods and

¹ Mechatronics engineer, Universidad Nacional, Colombia. Affiliation: Master's student in Environmental Engineering at Universidad Nacional, Palmira campus, Colombia. Email: anfrizhu@unal.edu.co

² PhD in Environmental Engineering. Affiliation: Universidad Nacional de Colombia, Palmira campus, Colombia. Email: vvargasf@unal.edu.co

³ PhD in Materials Engineering, Universidad del Valle, Colombia. Affiliation: Universidad Nacional de Colombia, Palmira campus, Colombia. Email: logonzalezsa@unal.edu.co



Attribution 4.0 International (CC BY 4.0) Share - Adapt

droughts [4]. Various statistical, analytical, and numerical methods are employed for precipitation estimation [5]. The main approaches involve developing models with explanatory and response variables. Diverse meteorological and physiographical parameters have been included as explanatory variables, and, in recent years, data from remote sensing systems such as satellite images and radars have been progressively incorporated. Significant models like the Global Circulation Model (GCM) and numerical weather prediction (NWP) models are particularly relevant and extensively used on the macroscale [6]. However, more robust and locally adapted models are required for regional and local scales.

Monitoring precipitation enables the acquisition of data for historical analysis, facilitating the development of estimation and prediction models. Measurements are obtained through rain gauges, weather radars, or satellite products with varying spatial and temporal resolutions [7]. The challenges in accurately estimating precipitation on the river basin scale include improving the spatial density of gauges and addressing the coarse resolution of remote sensing products [8]. Furthermore, the evident impacts of climate change in recent years, such as the progressive alteration of precipitation regimes and variations in the frequency and intensity of extreme events (including heavy rain and droughts) underscore the need for more robust and precise estimation at the regional level. Another limiting factor is coupling precipitation with the chaotic behavior of atmospheric dynamics. For example, a known issue in numerical systems corresponds to the errors and significant deviations in predictions caused by even slight changes in initial conditions [9]. Statistically, it has been also recognized that precipitation does not necessarily follow a normal distribution and can be modeled using asymmetrical distributions [10]. Consequently, more effective and powerful approaches, such as the use of artificial intelligence, are being studied to better approximate the correct behavior.

As a result of technological advances in the field of artificial intelligence and related areas such as data science, new approaches for processing and analyzing data for precipitation estimation are being employed. These include machine learning techniques like neural networks and the application of expert knowledge through fuzzy logic [11]. Such techniques provide flexibility and facilitate the development of more robust models for estimating precipitation, given their inherent ability to model complex nonlinear behaviors [12]. Few studies have been found which review artificial intelligence techniques for precipitation assessment, especially in relation to neural networks and fuzzy logic. [13] presented a review on resilient rainfall forecasting models using artificial intelligence techniques, with an emphasis on artificial neural networks (ANNs) as well as on hybrid models including neuro-fuzzy systems.

This document presents a bibliographic review of artificial intelligence techniques used for estimating precipitation.

The main objective is to compare ANNs against fuzzy logic models, focusing on the differences between machine learning and expert systems approaches. The methodology for the literature search and the criteria for selection are detailed in the next section. Afterwards, the theoretical basis for each method is explained, followed by a discussion of their main applications, and the article concludes with a succinct comparison of the two types of models.

Methodology

We conducted a systematic review to identify relevant studies on rainfall forecasting using artificial intelligence (AI), specifically fuzzy logic, neural networks, and neuro-fuzzy models. A literature search was conducted in the Scopus database, utilizing strategically selected keywords to capture a comprehensive overview of the most relevant studies. With the search criteria presented below, 134 articles were selected for analysis. Each study was systematically reviewed in a specific reading sequence: abstract, conclusion, results and discussion, methodology, and, finally, the introduction. This method facilitated the identification of potential themes and categories in the information presented by each paper.

Selection criteria

The main objective of this systematic review was to analyze the use of AI for precipitation estimation at the river basin level. The main selection criterion was a focus on precipitation analysis, with a preference for river basins and limited to neural networks and fuzzy logic approaches. A secondary objective involved determining and understanding the input and output variables, the model architecture, the performance metrics, and the scope of each case.

Search equations

The set of keywords encompassed terms like *river basin*, *precipitation*, and *artificial intelligence*, with additional specific terms for each technique: *neural networks* and *fuzzy logic*. It should be acknowledged that these terms were consulted in several permutations, including synonyms, nomenclatures, and broader keywords, in order to enhance the search breadth. The following variations were included in the search equations:

- River basin, catchment, watershed
- Precipitation, rainfall, rain estimation, rain rate, precipitation estimation
- Artificial intelligence, machine learning, soft computing
- Artificial neural networks, neural networks, deep learning, machine learning, artificial intelligence
- Fuzzy logic, fuzzy inference systems, expert systems, soft computing, artificial intelligence

Eq. (1) was used for the initial search in Scopus:

(precipitation OR rainfall OR rain OR "precipitation estimation") AND ("river basin" OR watershed OR basin OR catchment) AND ("artificial intelligence" OR "machine learning" OR "neural network" OR "fuzzy logic" OR "soft computing") (1)

Subsequently, a more specialized search was conducted, using several widespread databases (Elsevier, Springer, IEEE) as well as conventional search engines like Google or those oriented towards academic results (Google Scholar), in order to include non-indexed results (e.g., Arxiv). To correctly filter by AI paradigm, Search Eqs. (2) and (3) were employed.

(“fuzzy logic”) AND (watershed OR “river basin” OR “catchment”) AND (precipitation OR “precipitation estimation” OR “rainfall”) (2)

backpropagation neural networks (BPNNs). In several studies, ANN, NN, and BPNN are used interchangeably, but it must be clarified that are differences in the in the configuration of the network and in internal parameters like the weights, the bias matrix, the transfer function, and the optimization algorithm [18]. [19] conducted a study aimed at identifying relationships between atmospheric temperature and rainfall with ANN models.

Different types of AI and machine learning models can be used for precipitation prediction and forecasting applications, such as expert systems, NNs, and deep learning. In the realm of deep learning, it is possible to find models like convolutional NNs, recurrent NNs, and generative adversarial networks. ANNs have been used to complete missing data in precipitation time series [20], as well as in autoregressive models, where precipitation is modeled from historical data, as was done by [21] for 15 min precipitation, by [22] for daily precipitation, by [23] for daily precipitation with wavelets analysis, and by [24] for monthly precipitation from rain gauge data between 1961 and 2018 in the Wujiang River Basin while using an artificial bee algorithm. Moreover, [25] performed a similar study in Greece. Simpler single-layer models like the ADALINE network have been used for monthly precipitation forecasting [26].

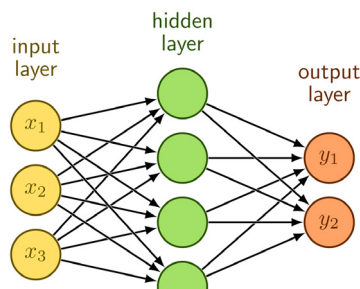


Figure 3. Neural network topology
Source: Adapted from [27]

Different input variables can be used besides precipitation. [28] included precipitable water vapor, pressure, temperature, relative humidity, cloud top temperature, cloud top pressure, and cloud top altitude to predict hourly precipitation. Other studies have used climate indices such as the southern oscillation index (SOI), the interdecadal pacific oscillation index (IPO), La Niña 3.4 [29], [30], and the standard precipitation index (SPI) [31] as input variables.

[32] used different types of NNs to estimate monthly mean precipitation and temperature based on data from 90 weather stations, with the purpose of elaborating a climatic cartography of Chile. Likewise, [33] delved into spatiotemporal predictions in Brazil. [34] used precipitation time series derived from stations monitoring data and radar and satellite images from different weather products, and [35] applied NNs to estimate precipitation using the WSR-88D radar in Oklahoma.

[36] were the first to describe the application of ANNs to satellite images in order to improve spatial precipitation estimation. Multiple products were derived from their studies, e.g., the PERSIANN system. Furthermore, with the advent of new weather products, precipitation databases, and new research, new studies have mostly taken interest in integrating data from various sources [12]. [37] used data from satellite products (ERA-5, CHIRPS, IMD, PERSIANN-CDR) to create a machine learning algorithm that combined different sources to achieve what they called *secondary precipitation estimate merging using machine learning* (SPEM2L).

Since the target variable is precipitation, most papers seek to implement regressions. However, classification processes can also be applied, as was the case with [38], who used data from the global navigation satellite system (GNSS) to identify heavy precipitation.

Recurrent neural networks

Recurrent NNs are a special type of network whose neurons include an additional connection to themselves that works as a buffer or memory element (Fig. 4). This configuration is particularly useful to approximate relations depending on previous data such as time series [39]. There are different types: the basic recurrent neural network (RNN), the gated recurrent unit (GRU), and long short-term memory (LSTM).

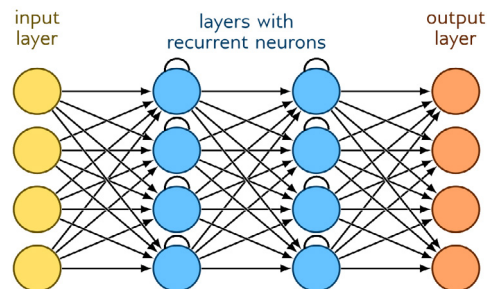


Figure 4. Recurrent neural network topology
Source: [40]

Deep and convolutional neural networks

Deep neural networks (DNNs) are a relatively new concept that involves ANNs containing many neurons, hidden layers, and training data. This kind of architecture has shown very good results in practice, especially with large amounts of quality data and high computational power available for training and validation [41]. The field of deep learning has gained ground for its great performance, to the point that the term *neural networks* is now directly associated with deep learning. [42] integrated data from different sources to predict precipitation using a deep network. Meanwhile, [43] implemented a classification model to identify heavy rain events, and [44] used bio-spectral images to predict precipitation.

The progress of DNNs also allowed developing new configurations like convolutional neural networks (CNNs). These networks use specialized nodes (Fig. 5) that work as sliding filters (i.e., they convolute) on the input data to identify the particular characteristics that activate them [45]. This behavior is suitable for image analysis aimed at object detection, instance segmentation, and image and pixel classification [46]. In precipitation analysis, this can be applied in the detection of clouds, weather fronts, and atmospheric dynamics in radar products, etc.

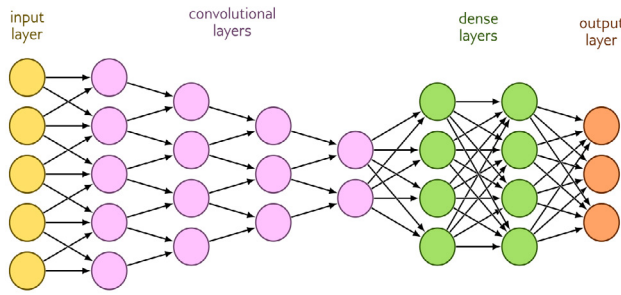


Figure 5. Convolutional neural network topology
Source: Adapted from [40]

[47] merged data from rain gauges, radar satellite images, and digital elevation models for precipitation estimation. They used CNNs and an additional post-processing step related to precipitation probability and intensity. The integration of radar data was improved, and the station bias was reduced in subsequent research [48]. Precipitation dynamics were analyzed in another study using both DNNs and CNNs applied to images obtained from terrestrial radars [49]. A CNN-based deep learning method was used to improve rainfall-runoff modeling in the Mekon River Basin [50]. One study explored the application of a CNN-based architecture for detecting and estimating near real-time precipitation in the USA [51].

In recent years, methods based on deep CNNs have achieved significant success, and their performance continues to improve [52]. [53] set about correcting the bias of daily satellite precipitation in tropical regions using a DNN. Most deep and convolutional models use non structured data as input (e.g., images). A specific study on precipitation presented a nowcasting method based on sparse correspondence and a DNN [49]. The necessary data can be obtained directly from remote sensing products, generated from curated data provided by multiple sources, or generated from statistical or numerical models. For example, [54] used data from the ERA5 numerical and reanalysis model and the E-OBS database to apply a U-net (deep and convolutional network). The input data included weather and physical variable maps considering temperature, wind speed, water vapor, and geopotential altitude to generate the output, in the form of an hourly precipitation map.

Another project, focused on short-term weather forecasting (i.e., nowcasting), mainly used CNNs or variants with recurrent components. Here, [55] used precipitation data from radar and satellite images provided by the Geostationary

Satellite Server (GOES), together with physical weather models commonly used in meteorology. They obtained good results for 12 h forecasts. [56] applied a hybrid MLP and CNN model to predict extreme regional precipitation in central-eastern China. Similarly, [57] conducted a quantitative precipitation forecasting study for China with a multi-stream CNN. On the other hand, [58] applied CNNs in the United Kingdom. They added a generative component, wherein two modules (the generator and the discriminator) compete to generate an optimal output.

Optical flow can also be used on radar images [59] and in direct processing and detection from satellite images [60], [61]. Due to the sequential nature of precipitation data, it is possible to merge image and temporal analysis [62] using models that integrate convolutions and LSTM [63]. [64] proposed a transformer-enhanced spatiotemporal neural network called *TransLSTMUNet* for the post-processing of precipitation forecasts, and, using a DNN, [65] developed a forecasting model based on the global normalized difference vegetation index (NDVI), air temperature, soil moisture, and precipitation.

Thanks to the availability of precipitation data from satellite images, videos, and climate reanalysis products, a whole new wave of studies using computer vision has emerged. For instance, [66] compared several convolutional models (LSTM and U-Nets) for precipitation nowcasting within a 15 min temporal scale. Notably, a large volume of precipitation images was required.

Downscaling methods

The downscaling and regionalization of data allows improving the spatial scale of weather data or radar and optical images obtained via remote sensing in order to produce information that better captures the study area [11]. [67] applied downscaling with different machine learning models for precipitation estimation, using data from the Coupled Model Intercomparison Project Phase 5 (CMIP5). [68] and [69] used CNNs for the micro-regional monitoring of precipitation, while [70] analyzed the probability of extreme events through downscaling. [71] applied radial-basis NNs based on downscaling, integrating data from precipitation time series, global circulation models, and different climate change scenarios as inputs. Downscaling can be applied by means of different models (e.g., statistical methods) or through classical NNs [72], CNNs, and U-nets [73]. Depending on the data available, this can be done on different temporal scales (annual, monthly, or daily) [74].

Fuzzy logic for precipitation estimation

Fuzzy inference systems

Fuzzy logic is based on the concept of *fuzzy sets*. A fuzzy set is a set with no crisp or clear boundary. Unlike two-valued Boolean logic, fuzzy logic is multi-valued, and it deals with degrees of membership and truth. Fuzzy logic uses

any logical value from the set of real numbers between 0 (completely false) and 1 (completely true). This is known as the *membership value*, and the function that represents such value is called a *membership function* [75]. Fuzzy logic takes advantage of expert knowledge and the flexibility of fuzzy sets to model complex systems [18]. It allows representing numerical variables as identifiable linguistic values through membership functions (facilitating the representation of uncertainty and vagueness) (Fig. 6). Moreover, interpretable logic rules can be applied to these linguistic variables in the inference process. The fuzzy inference system (FIS) is the common configuration, comprising three main steps: fuzzification, inference, and defuzzification (Fig. 7).

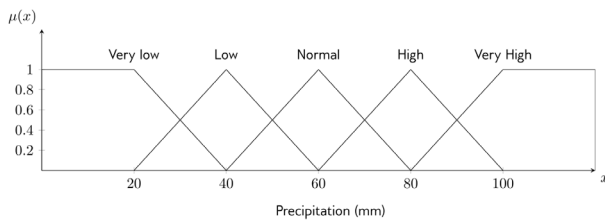


Figure 6. Example of a membership function
Source: Adapted from [76]

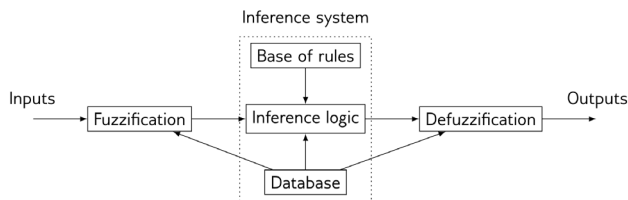


Figure 7. Fuzzy inference system
Source: Adapted from [77]

A special instance of this approach is the *Mamdani fuzzy inference system* (MFIS), which is widely accepted among the scientific community due to its interpretability. Here, the consequent of the implication rules is a single value. On the other hand, the *Sugeno fuzzy inference system* (SFIS) has a consequent with an arbitrary fuzzy function that considers all the variables in the antecedent [16]. The behavior of a FIS can be visualized, for two inputs and a single output, as a three-dimensional surface indicating the non-linear relation between the variables (Fig. 8) – when more variables are added, it generates an n-dimensional hyperplane [78].

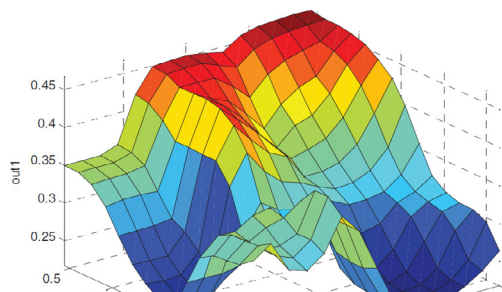


Figure 8. FIS output surface example for precipitation estimation from time series data
Source: [78]

[79] applied triangular membership functions to a FIS for precipitation data imputation. Precipitation prediction from other weather variables is also possible: [80] implemented a FIS using maximum, minimum, and mean values for wind speed, precipitation, and temperature as input in a model with 23 inference rules. [76] only used wind speed and air temperature. [81] applied fuzzy logic to a set of geographical variables including altitude, distance to the coastline, and slope – in addition to rain gauge data – to improve precipitation maps from meteorological radars.

[82] incorporated atmospheric pressure, humidity, dew point, temperature, and wind speed as input variables. The membership functions for each variable were triangular, with simple linguistic categories ranging from *very low* to *very high* in a MFIS. Furthermore, [83] added a temporal variable to differentiate the current day from the day before in their accumulated daily precipitation analysis. It is also possible to use preprocessed data such as those from the meteorological aerodrome report (METAR), a very common source in aerospace applications and weather analysis for air bases [84]; or those from the National Oceanic and Atmospheric Administration (NOAA) which offers data on different weather variables [85]. The main objective of the study by [84] was to predict rainfall events using a rule-based FIS that incorporated five parameters: relative humidity, total cloud cover, wind direction, temperature, and surface pressure. Similarly, [86] analyzed the uncertainties associated with extreme rainfall in terms of return levels. They also quantified the potential risk of these events in the coastal wetlands of India using fuzzy logic. [87] worked with fuzzy rainfall-runoff models to generate predictions for claypan catchments with conservation buffers in northeastern Missouri. Finally, [88] studied the climate sensitivity of mountainous regions to natural hazards through a fuzzy logic approach, identifying alterations in the level, intensity, or type of precipitation as the main drivers, together with glacier melting and permafrost thawing.

Fuzzy clustering and interpolation

Fuzzy systems can be implemented to improve the spatial interpolation of precipitation. [89] applied fuzzy logic to inverse distance weighting (IDW) for the spatial interpolation of precipitation, aiming to reduce the estimation error at river basin level. There are similar methods exclusively based on spatial interpolation [90] or classification, as is the case of [91], who used fuzzy logic to zone monthly precipitation and improve decision-making for cacao cultivation.

On the other hand, fuzzy clustering, or fuzzy C-means (FCM), is the use of membership functions to cluster, group, or categorize elements according to a similarity criterion. For example, [92] implemented this method to estimate precipitation and generate flood maps, and [90] applied it to validate spatial precipitation estimation. Fuzzy clustering can also be applied for downscaling precipitation data [93].

Fuzzy time series

Although FIS are mainly used for a system of inputs and outputs where the temporal component is not clearly incorporated, fuzzy logic can also be used for time series analysis. In this case, the time series should be interpreted as a fuzzy set. For example, [94] used fuzzy time series and NNs to predict rainfall, and, in complementary work, [78] focused exclusively on precipitation time series.

Within a purely autoregressive approach, membership functions are created by temporally dividing the precipitation time series [96]. In said cases, the membership functions split the data according to their temporal scale, i.e., the linguistic variable can be the month of the year, and, after the fuzzification of the inputs, the inference rules can directly reference the known experimental behavior of the precipitation in certain months (Fig. 9).

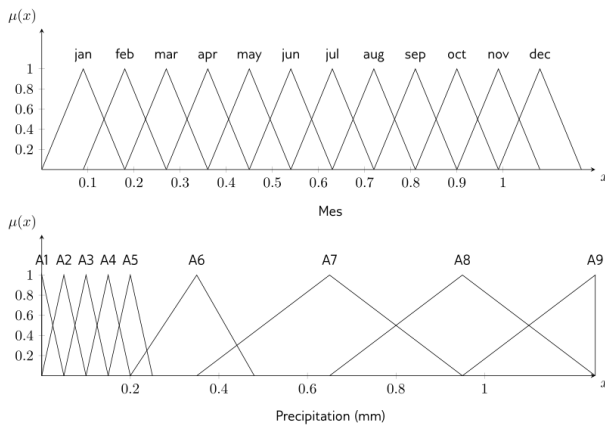


Figure 9. Membership functions for fuzzy time series
Source: Adapted from [97]

Hybrid models: neuro-fuzzy systems

Hybrid models refers to instances that integrate machine learning components to complement FIS, e.g., NNs and genetic algorithms. Given the high effectiveness recently shown by machine learning applied to big data applications, it is increasingly common to include it as an additional step for expert systems. For example, NNs can be used to automatically generate membership functions for FIS, or even to generate inference rules [98]. [99] used NNs to generate inference rules within a so-called *neuro-fuzzy system* (NFS), using coordinates and their corresponding precipitation values, in a study similar to that by [100]. [101] merged data from stations, radar, and satellite images using a neuro-fuzzy network.

Another very common architecture in the literature corresponds to the adaptive neuro-fuzzy inference system (ANFIS) (Fig. 10). Neuro-fuzzy hybridization results in a hybrid intelligent system that synergizes ANNs and fuzzy logic by combining the human-inspired reasoning of fuzzy systems with the learning and connectionist structure of NNs [75]. [102] applied ANFIS to estimate precipitation from several rain gauge stations in Serbia, reporting improved

reliability against uncertainty. Using ANFIS, [103] managed to identify the most relevant meteorological variables and their influence on precipitation estimation. They included data on vapor pressure, air temperature, the monthly frequency of wet days and the percent monthly cloud cover. Meanwhile, [104] used this approach to improve precipitation estimation from radar data. Some comparative studies have implemented the ANFIS method [23], as well as others focused on predicting precipitation-related climatic indices [105] or on using historical precipitation series.

Several models can also be merged into this approach, wherein the fuzzy logic component serves as a module integrator [106]. [107] presented a self-identification neuro-fuzzy inference model (SINFIM) for modeling the relationship between rainfall and runoff on a Chilean watershed. Another work studied the trends and patterns of rainfall to conduct an analysis of the city of Mumbai via the rainfall regionalization approach coupled with fuzzy logic and clustering [108]. [109] applied an ANFIS to evaluate rainfall-runoff modeling in a sub-catchment of the Kranji Basin in Singapore, and another study used NNs and fuzzy logic in statistical downscaling to support daily precipitation forecasting [110].

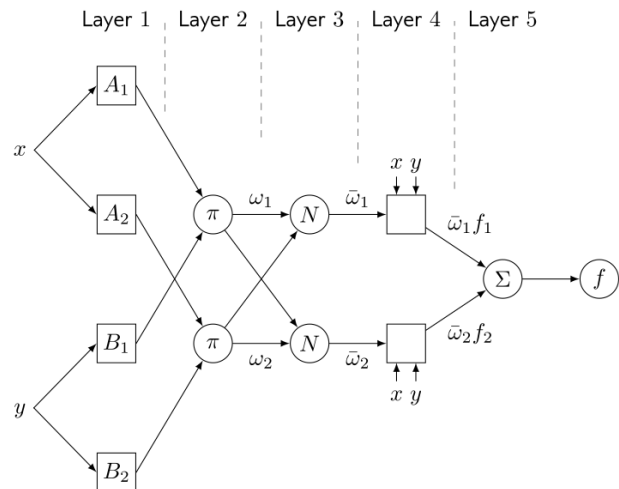


Figure 10. Topology of an ANFIS model
Source: Adapted from [103]

Hybrid metaheuristic algorithms

Hybrid metaheuristic algorithms are advanced tools in the field of AI [111]. These techniques can solve problems via prediction errors, hyperparameter determination, and feature selection using machine learning algorithms [112], which is why they are gaining popularity and are being used for the development of hybrid models for hydrological research [113], including those dealing with the prediction of reference evapotranspiration (ET_o), a very important parameter for determining the availability of water resources and in hydrological studies. However, they are mainly used to predict ET_o, as stated by [114]. To this effect, they studied and compared the prediction capabilities of two support vector regression (SVR) models along with three metaheuristic

algorithms, *i.e.*, particle swarm optimization (PSO), gray wolf optimization (GWO), and the gravitational search algorithm (GSA), using meteorological variables in monthly ETO prediction used meteorological variables as input.

Hybrid metaheuristic algorithms have also been used to elaborate flood susceptibility maps, and the optimization capabilities offered by different machine learning algorithms has been leveraged by means of metaheuristic algorithms [111]. In the Haraz Basin, Iran, [115] employed an ANFIS coupled with the cropping (CA), bee (BA), and invasive weed optimization (IWO) algorithms. [116] used a combination of ANFIS, the genetic algorithm (GA), ant colony optimization (ACO), and PSO to generate a flood susceptibility map for the municipality of Jahrom, Iran. [24] performed ANFIS optimization with biogeography-based optimization (BBO) and the imperialist competitive algorithm (ICA). [117] used differential evolution (DE), the GA, and PSO along with an ANFIS to elaborate a flood susceptibility map for the Ganges Plain in India. [118] also used a combination of SVR, the GWO, and the bat optimizer (Bat) to generate this type of map. [119] used GWO and the whale optimization algorithm (WOA) to optimize SVR and create a flood susceptibility map for the Ardabil province in Iran. [120] combined SVR, PSO, and the grasshopper optimization algorithm (GOA) to develop a flood susceptibility map. [121] used the group method of data management (GMDH), DE, and the GA to generate a flood susceptibility map for the Haraz-Neka Basin, Iran. Moreover, [122] conducted GMDH optimization with the help of GWO in flood modeling.

[123] explored the accurate prediction of daily rainfall via AI methods. These methods were grounded in an ANFIS. Some metaheuristic optimization algorithms were also employed: the artificial bee colony algorithm (ABC), the GA, and simulated annealing (SA). [124] presented a method for providing explainability in the integration of inductive rules, combined with fuzzy logic and data mining techniques, when dealing with meteorological predictions.

Machine learning

Machine learning (ML) is a field of AI that deals with the development and study of statistical algorithms capable of learning from data and generalizing to unseen data, allowing them to perform tasks without explicit instructions.

In this vein, there are some studies related to precipitation forecasting and ML. [125] developed a conceptual metaheuristics-based framework for improving runoff time series simulation in glacierized catchments, combining hydrological model with a series predictor model and the optimization-driven parameter tuning of the firefly algorithm. Furthermore, [126] used a MLP network – optimized via the GA, PSO, the firefly algorithm, and teleconnection pattern indices – for rainfall modeling in the Mediterranean Basin. In addition, nested hybrid rainfall-runoff modelling has been performed via embedding ML techniques [127]. [128] used a combination of approaches, *i.e.*, statistical, ML, deep learning

(DL), and hybrid algorithms, in order to build a precipitation forecasting system. In addition, [129] proposed a new rainfall prediction model that employs different techniques as well as indicator features like average directional movement (ADX), moving average convergence divergence (MACD), and Welles Wilder's smoothing average (WWS). [130] developed a metaheuristic evolutionary DL model based on a temporal convolutional network for rainfall-runoff simulation and multi-step runoff prediction. [131] assessed some rainfall prediction models to explore the advantages of ML and remote sensing approaches. Furthermore, an assessment of hybrid ML algorithms using TRMM rainfall data for daily inflow forecasting was carried out in eastern Brazil [132]. In China, a study on automated ML for rainfall-induced landslide hazard mapping was conducted [133]. [134] performed a comparative assessment of rainfall-based water level prediction methods using ML. [135] evaluated traditional and ML approaches to rainfall prediction, and [136] examined a combination of the ERA5 dataset and ML. Long-term rainfall prediction was performed by [137], using atmospheric synoptic patterns in semiarid climates with statistical and ML methods. [138] studied ML-based rainfall models for accurate flood mapping in Pakistan. [139] conducted specific studies on short-term rainfall forecasting using cumulative precipitation fields from station data with a probabilistic ML approach.

Comparative analysis

Input variables

Both NNs and fuzzy logic models depend on the available input variables. An initial knowledge of the objective function and the possible relationships between the explanatory and response variables is assumed in order to build the model. Fig. 11 shows the common input variables for the studied field. In general, these parameters can be classified as meteorological, physiographic, or hydrological variables; climatic indices; data derived from physical or numerical models; satellite or radar products; or other derived databases.

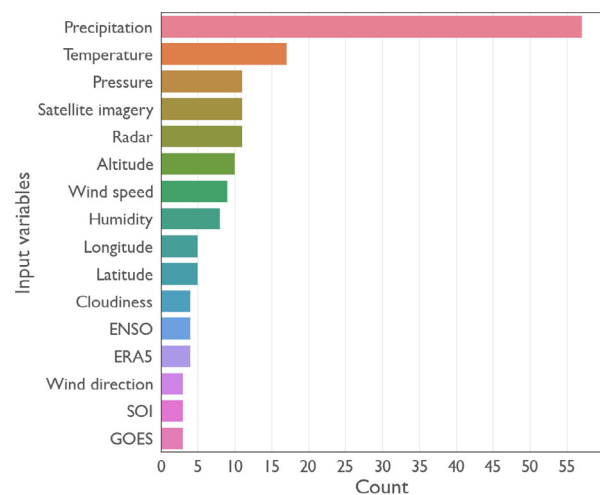


Figure 11. Input variables used in the references

Source: Authors

Output variables

As the purpose of these models is estimating it, precipitation should be the output or target variable in most cases. However, this variable can be expressed in diverse temporal scales, units, or configurations, as shown in Fig. 12. In some studies, both precipitation and temperature are included as output variables [32]. The most widely used output is monthly precipitation, mainly in the fields of weather forecasting and climate analysis. These variables also allow evaluating extreme events and return periods.

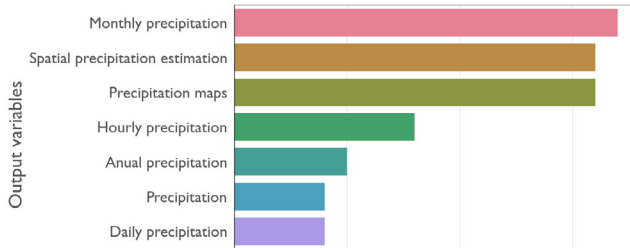


Figure 12. Output variables used in the references
Source: Authors

Model architectures

Fig. 13 shows the common NN architectures for precipitation estimation. ANNs and BPNN are differentiated as in the referenced literature. Although DNNs, CNNs, and convolutional-recurrent networks (CRNs) are shown separately, they could be grouped into a single category (i.e., deep networks) that is representative of the selected references.

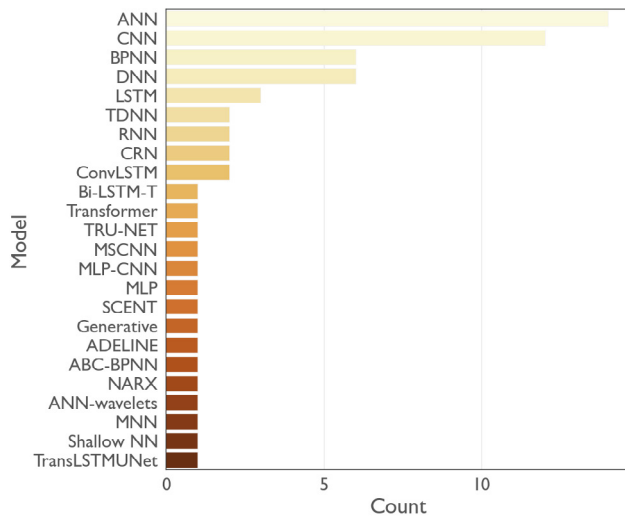


Figure 13. Neural network models found in the references
Source: Authors

Fig. 14 shows fuzzy logic models for precipitation estimation. ANFIS, FIS, and MFIS are the most commonly used. It could be said that MFIS are just a special case of FIS. On the other hand, both ANFIS and NFS integrate neural elements, so they represent the hybrid models in the references.

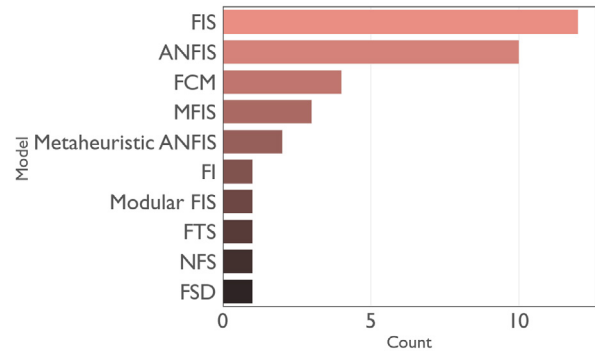


Figure 14. Fuzzy logic models found in the references
Source: Authors

Performance metrics

To validate the techniques discussed herein, it is necessary to use certain performance metrics or criteria in order to compare actual values to those generated by the models. Some of the most common metrics include the root mean square error (RMSE), the correlation coefficient (R), the determination coefficient (R^2), and the mean absolute error (MAE), which are mainly applied to regression models. In the case of classification models, performance evaluation should be mixed; for example, a confusion matrix can be used, as well as the F1 score or accuracy values. Among the performance metrics used in the referenced literature (Fig. 15), there are specific indicators for the field of hydrology, such as the average flood exposure risk (AFER), a specialized metric for flood analysis; Nash-Sutcliffe efficiency (NSE), widely employed in model assessment; the fractions skill score (FSS) for forecasting; and the skill score (SS) denominations, which are employed in quantitative precipitation forecasting (QPF). Apart from these, the RMSE, MAE, and R stand out as the most common parameters in fuzzy logic models implementing regression approaches.

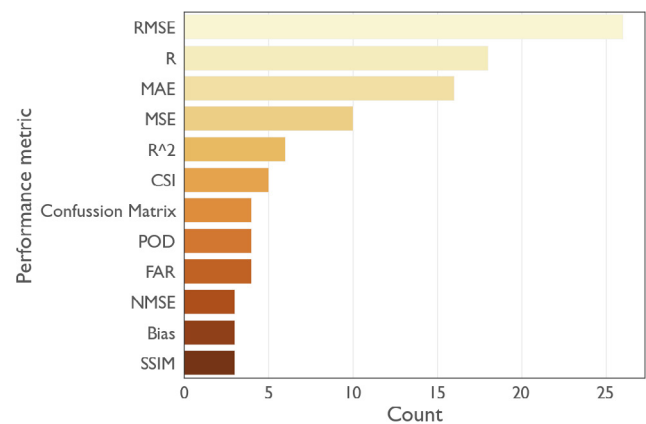


Figure 15. Neural network performance metrics
Source: Authors

Software tools and implementation

AI models can be implemented using different software tools and programming languages. A few references clearly describe the software used for implementation, but most of

them do not provide clear information in this regard. The MATLAB software is notably used to implement of both NNs and fuzzy logic models [80]. In addition, the R language is applied for statistical analysis and downscaling [71], and some Python libraries are used for DL models [62]. It is important to highlight that the use of statistical software and geographical information systems is essential in this field.

Applications

For the general applications mentioned in the literature on precipitation estimation, some categories are identified. Firstly, as expected, precipitation forecasting on different temporal scales tends to be the main objective of several papers. For long-term temporal scales (30 years or more), the objective is climate analysis. Some papers emphasize the usefulness of AI techniques for issuing extreme event early warnings and in watershed management [21], [24], [78], [92].

Advantages and challenges of AI methodologies

AI methods exhibit both limitations and advantages. The fundamental aspects of the main methods are outlined below.

ML offers significant advantages regarding automation, accuracy, and scalability, but it poses challenges related to data dependence, model complexity, resource requirements, and ethical considerations. Fuzzy logic is quite advantageous in handling uncertainty, providing intuitive solutions and adaptability across various domains. However, its limitations are related to precision, rule design complexity, computational effort, and the lack of self-learning capabilities. Moreover, NNs are powerful and versatile tools capable of learning complex patterns from large datasets with good learning performance, adaptability, versatility, and the possibility of continuous improvement. However, they come with significant challenges related to data dependence, computational load, interpretability, and overfitting, all of which need to be carefully managed to ensure an effective and ethical use. NFS offer a powerful combination of the learning capabilities of NNs and the interpretability and uncertainty management of fuzzy logic. These hybrid models are particularly valuable in applications that require both adaptive learning and human-like reasoning. However, they pose challenges pertaining to complexity, computational effort, overfitting, and data quality dependence. Careful design and implementation are required to fully realize these techniques' potential while managing their limitations.

Conclusions

This work presents the results of a thorough review of the literature on prediction precipitation using AI techniques. Our findings provide academia and society in general with perspectives for future research in the field. There are various

approaches for precipitation estimation using AI, even when limiting the search to two specific paradigms such as NNs and fuzzy logic. Model selection widely depends on the type, quantity, and quality of the available data, and there is no single configuration that guarantees the best results. The integration of multiple data sources holds great potential for performing regression in future studies.

Although the number of studies involving fuzzy logic has decreased, these models remain a relevant option due to their interpretability. Access to large amounts of data could benefit fuzzy logic, as achieved through the inclusion of ML components to create hybrid models, allowing for scaling while maintaining interpretability.

NN research applied to precipitation estimation has grown in recent years, with more sophisticated models like deep, recurrent, and convolutional networks being incorporated and showing significantly better results. However, among their limitations is the availability of and access to large amounts of data or high computational power, as well as the lack of interpretability and implementation issues.

There are still many challenges for precipitation estimation at the river basin level. Advances in the field of AI and access to new data sources, models, and software tools have yielded very promising results for the study of precipitation at different levels, from mere forecasting to extreme events forecasting and hydrological and environmental modeling.

CRedit author statement

All authors: conceptualization, methodology, software, validation, formal analysis, investigation, writing (original draft, review and editing), data curation, supervision. All authors contributed to the writing of the manuscript and approved its definitive version for publication.

Conflicts of interest

The authors declare no conflict of interest.

References

- [1] UNESCO, Eds., *Informe mundial de las Naciones Unidas sobre el desarrollo de los recursos hídricos 2020: agua y cambio climático*. Paris, France: UNESCO, 2020. https://unesdoc.unesco.org/ark:/48223/pf0000372985_spa
- [2] J. Wang, P. Zhai, and C. Li, "Non-uniform changes of daily precipitation in China: Observations and simulations," *Weather Clim. Ext.*, vol. 44, Article 100665, 2024. <https://doi.org/10.1016/j.wace.2024.100665>
- [3] MAVDT, "Política nacional para la gestión integral del recurso hídrico," 2010. [Online]. Available: https://www.minambiente.gov.co/wp-content/uploads/2021/10/Politica-nacional-Gestion-integral-de-recurso-Hidrico-web.pdf?utm_source=chatgpt.com

- [4] S. Zhu, Z. Xu, X. Luo, C. Wang, and J. Wu, "Assessing coincidence probability for extreme precipitation events in the Jinsha River basin," *Theor. Appl. Climatol.*, vol. 139, no. 1, pp. 825–835, Jan. 2020. <https://doi.org/10.1007/s00704-019-03009-1>
- [5] R. Prudden et al., "A review of radar-based nowcasting of precipitation and applicable machine learning techniques," 2020. [Online]. Available: <https://arxiv.org/abs/2005.04988v1>
- [6] IPCC, Eds., *Climate change 2007: The physical science basis: Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge, UK; New York, NY, USA: Cambridge University Press, 2007. <https://www.ipcc.ch/report/ar4/wg1/>
- [7] T. Davie, *Fundamentals of hydrology*, 3rd ed. London, UK; New York, NY, USA: Routledge, Taylor & Francis Group, 2019.
- [8] B. Hingray, C. Picouet, A. Musy, and B. Hingray, *Hydrology: A science for engineers*. Boca Raton, FL, USA: CRC Press, Taylor & Francis Group, 2014.
- [9] R. Buizza, "Chaos and weather prediction," 2002. [Online]. Available: <https://www.ecmwf.int/node/16927>
- [10] D. S. Wilks, *Statistical methods in the atmospheric sciences*, 4th ed. Cambridge, UK: Elsevier, 2019.
- [11] Q. Yuan et al., "Deep learning in environmental remote sensing: Achievements and challenges," *Rem. Sens. Environ.*, vol. 241, art. 111716, May 2020. <https://doi.org/10.1016/j.rse.2020.111716>
- [12] R. S. Khan and M. A. E. Bhuiyan, "Artificial intelligence-based techniques for rainfall estimation integrating multisource precipitation datasets," *Atmosphere*, vol. 12, no. 10, art. 10, Oct. 2021. <https://doi.org/10.3390/atmos12101239>
- [13] Md. A. Saleh, H. M. Rasel, and B. Ray, "A comprehensive review towards resilient rainfall forecasting models using artificial intelligence techniques," *Green Tech. Sust.*, vol. 2, no. 3, art. 100104, Sep. 2024. <https://doi.org/10.1016/j.grets.2024.100104>
- [14] A. Perianes-Rodriguez, L. Waltman, and N. J. van Eck, "Constructing bibliometric networks: A comparison between full and fractional counting," *J. Informetrics*, vol. 10, no. 4, pp. 1178–1195, Nov. 2016. <https://doi.org/10.1016/j.joi.2016.10.006>
- [15] N. J. van Eck and L. Waltman, "VOSviewer Manual," 2021. [Online]. Available: <https://www.vosviewer.com/getting-started>
- [16] K.-L. Du and M. N. S. Swamy, "Neural networks and statistical learning," 2019. [Online]. Available: <https://doi.org/10.1007/978-1-4471-7452-3>
- [17] M. T. Hagan, H. B. Demuth, M. Hudson, and O. Jesús, "Neural network design," 2014. [Online]. Available: <https://pdfs.semanticscholar.org/ee5d/5f138cb3220fbc642ec099d1bab9b41e410.pdf>
- [18] S. Russell and P. Norvig, *Artificial intelligence: A modern approach*, 3rd ed. Upper Saddle River, NJ, USA: Pearson, 2009.
- [19] U. R. Anushka-Perera, "Rainfall and atmospheric temperature against the other climatic factors: A case study from Colombo, Sri Lanka," *Math. Prob. Eng.*, vol. 2019, art. 5692753, 2019. <https://doi.org/10.1155/2019/5692753>
- [20] J. Kajornrit, K. W. Wong, and C. C. Fung, "Estimation of missing precipitation records using modular artificial neural networks," in *Int. Conf. Neural Info. Proc.*, 2012, pp. 52–59. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-642-34478-7_7
- [21] K. C. Luk, J. E. Ball, and A. Sharma, "An application of artificial neural networks for rainfall forecasting," *Math. Comp. Model.*, vol. 33, no. 6, pp. 683–693, Mar. 2001. [https://doi.org/10.1016/S0895-7177\(00\)00272-7](https://doi.org/10.1016/S0895-7177(00)00272-7)
- [22] W. Hong, "Rainfall forecasting by technological machine learning models," *App. Math. Comp.*, vol. 200, no. 1, pp. 41–57, Jun. 2008. <https://doi.org/10.1016/j.amc.2007.10.046>
- [23] A. Solgi, V. Nourani, and A. Pourhaghi, "Forecasting daily precipitation using hybrid model of wavelet-artificial neural network and comparison with adaptive neuro-fuzzy inference system (case study: Verayneh Station, Nahavand)," *Adv. Civil Eng.*, vol. 2014, pp. 1–12, 2014. <https://doi.org/10.1155/2014/279368>
- [24] Y. Wang, J. Liu, R. Li, X. Suo, and E. Lu, "Precipitation forecast of the Wujiang River Basin based on artificial bee colony algorithm and backpropagation neural network," *Alexandria Eng. J.*, vol. 59, no. 3, pp. 1473–1483, Jun. 2020. <https://doi.org/10.1016/j.aej.2020.04.035>
- [25] K. P. Moustris, I. K. Larissi, P. T. Nastos, and A. G. Paliatatos, "Precipitation forecast using artificial neural networks in specific regions of Greece," *Water Res. Manag.*, vol. 25, no. 8, pp. 1979–1993, Jun. 2011. <https://doi.org/10.1007/s11269-011-9790-5>
- [26] I. Sutawinaya, I. N. G. Astawa, and N. K. Hariyanti, "Comparison of adaline and multiple linear regression methods for rainfall forecasting," *J. Phys. Conf. Ser.*, vol. 953, pp. 1–8, Feb. 2018. <https://doi.org/10.1088/1742-6596/953/1/012046>
- [27] S. Araghinejad, *Data-driven modeling: Using MATLAB® in water resources and environmental engineering*. Dordrecht, Netherlands: Springer, 2014. <https://doi.org/10.1007/978-94-007-7506-0>
- [28] P. Benevides, J. Catalao, and G. Nico, "Neural network approach to forecast hourly intense rainfall using GNSS precipitable water vapor and meteorological sensors," *Rem. Sens.*, vol. 11, no. 8, art. 8, Jan. 2019. <https://doi.org/10.3390/rs11080966>
- [29] J. Abbot and J. Marohasy, "Input selection and optimisation for monthly rainfall forecasting in Queensland, Australia, using artificial neural networks," *Atmos. Res.*, vol. 138, pp. 166–178, Mar. 2014. <https://doi.org/10.1016/j.atmosres.2013.11.002>
- [30] J. Abbot and J. Marohasy, "The application of artificial intelligence for monthly rainfall forecasting in the Brisbane Catchment, Queensland, Australia," *WIT Tran. Ecol. Environ.*, vol. 172, pp. 125–135, May 2013. <https://doi.org/10.2495/RBM130111>

- [31] S. Azimi and M. Azhdary-Moghaddam, "Modeling short term rainfall forecast using neural networks, and Gaussian process classification based on the SPI drought index," *Water Res. Manag.*, vol. 34, no. 4, pp. 1369–1405, Mar. 2020. <https://doi.org/10.1007/s11269-020-02507-6>
- [32] F. Neira, "Elaboración de la cartografía climática de temperaturas y precipitación mediante redes neuronales artificiales: caso de estudio en la Región del Libertador Bernardo O'Higgins," 2010, [Online]. Available: <http://repositorio.uchile.cl/handle/2250/112354>
- [33] P. S. Lucio, F. C. Conde, I. F. A. Cavalcanti, A. I. Serrano, A. M. Ramos, and A. O. Cardoso, "Spatiotemporal monthly rainfall reconstruction via artificial neural network -case study: South of Brazil," *Adv. Geosci.*, vol. 10, pp. 67–76, 2007. <https://doi.org/10.5194/adgeo-10-67-2007>
- [34] Y. Seo, S. Kim, and V. P. Singh, "Estimating spatial precipitation using regression kriging and artificial neural network residual kriging (RKNRK) hybrid approach," *Water Res. Manag.*, vol. 29, no. 7, pp. 2189–2204, May 2015. <https://doi.org/10.1007/s11269-015-0935-9>
- [35] T. B. Trafalis, M. B. Richman, A. White, and B. Santosa, "Data mining techniques for improved WSR-88D rainfall estimation," *Comp. Ind. Eng.*, vol. 43, no. 4, pp. 775–786, Sep. 2002. [https://doi.org/10.1016/S0360-8352\(02\)00139-0](https://doi.org/10.1016/S0360-8352(02)00139-0)
- [36] K. L. Hsu, X. Gao, S. Sorooshian, and H. V. Gupta, "Precipitation estimation from remotely sensed information using artificial neural networks," *J. Appl. Meteorol.*, vol. 36, no. 9, pp. 1176–1190, Sep. 1997. [https://doi.org/10.1175/1520-0450\(1997\)036<1176:PEFRSI>2.0.CO;2](https://doi.org/10.1175/1520-0450(1997)036<1176:PEFRSI>2.0.CO;2)
- [37] V. Kolluru, S. Kolluru, N. Wagle, and T. D. Acharya, "Secondary precipitation estimate merging using machine learning: Development and evaluation over Krishna River Basin, India," *Rem. Sens.*, vol. 12, no. 18, art. 18, Jan. 2020. <https://doi.org/10.3390/rs12183013>
- [38] H. Li *et al.*, "A neural network-based approach for the detection of heavy precipitation using GNSS observations and surface meteorological data," *J. Atmos. Solar-Terres. Phys.*, vol. 225, art. 105763, Nov. 2021. <https://doi.org/10.1016/j.jastp.2021.105763>
- [39] D. Z. Haq *et al.*, "Long short-term memory algorithm for rainfall prediction based on El-Nino and IOD data," *Proc. Comp. Sci.*, vol. 179, pp. 829–837, 2021. <https://doi.org/10.1016/j.procs.2021.01.071>
- [40] A. Tch, "The mostly complete chart of neural networks, explained," Medium. [Online]. Available: <https://towardsdatascience.com/the-mostly-complete-chart-of-neural-networks-explained-3fb6f2367464>
- [41] Y. Tao, X. Gao, A. Ihler, K. Hsu, and S. Sorooshian, "Deep neural networks for precipitation estimation from remotely sensed information," in *2016 IEEE Cong. Evol. Comp. (CEC)*, 2016, pp. 1349–1355. [Online]. Available: <http://ieeexplore.ieee.org/abstract/document/7743945/>
- [42] H. G. Damavandi and R. Shah, "A learning framework for an accurate prediction of rainfall rates," 2019. [Online]. Available: <http://arxiv.org/abs/1901.05885>
- [43] M. Sangiorgio *et al.*, "A comparative study on machine learning techniques for intense convective rainfall events forecasting," in *Theory and Applications of Time Series Analysis*, O. Valenzuela, F. Rojas, L. J. Herrera, H. Pomares, and I. Rojas, Eds. Cham, Germany: Springer International Publishing, 2020, pp. 305–317.
- [44] Y. Tao, X. Gao, A. Ihler, S. Sorooshian, and K. Hsu, "Precipitation identification with bispectral satellite information using deep learning approaches," *J. Hydrometeorol.*, vol. 18, no. 5, pp. 1271–1283, May 2017. <https://doi.org/10.1175/JHM-D-16-0176.1>
- [45] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, art. 7553, May 2015. <https://doi.org/10.1038/nature14539>
- [46] X. X. Zhu *et al.*, "Deep learning in remote sensing: A comprehensive review and list of resources," *IEEE Geosci. Rem. Sens. Mag.*, vol. 5, no. 4, pp. 8–36, Dec. 2017. <https://doi.org/10.1109/MGRS.2017.2762307>
- [47] A. Moraux, S. Dewitte, B. Cornelis, and A. Munteanu, "Deep learning for precipitation estimation from satellite and rain gauges measurements," *Rem. Sens.*, vol. 11, no. 21, art. 21, Jan. 2019. <https://doi.org/10.3390/rs11212463>
- [48] A. Moraux, S. Dewitte, B. Cornelis, and A. Munteanu, "A deep learning multimodal method for precipitation estimation," *Rem. Sens.*, vol. 13, no. 16, art. 16, Jan. 2021. <https://doi.org/10.3390/rs13163278>
- [49] W. Fang, F. Zhang, V. S. Sheng, and Y. Ding, "SCENT: A new precipitation nowcasting method based on sparse correspondence and deep neural network," *Neurocomputing*, art. S0925231221003283, Mar. 2021. <https://doi.org/10.1016/j.neucom.2021.02.072>
- [50] X.-H. Le, Y. Kim, D.-V. Binh, S. Jung, and D.-H. Nguyen, "Improving rainfall-runoff modeling in the Mekon river basin using bias-correct satellite precipitation products by convolutional neural networks," *J. Hydrol.*, vol. 630, art. 130762, 2024. <https://doi.org/10.1016/j.jhydrol.2024.130762>
- [51] M. Sadeghi, P. Nguyen, K. Hsu, and S. Sorooshian, "Improving near real-time precipitation estimation using a U-Net convolutional neural network and geographical information," *Environ. Model. Soft.*, vol. 134, art. 104856, Dec. 2020. <https://doi.org/10.1016/j.envsoft.2020.104856>
- [52] J. Wang, L. Lin, Z. Zhang, S. Gao, and H. Yu, "Deep neural network based on dynamic attention and layer attention for meteorological data downscaling," *ISPRS J. Photogramm. Rem. Sens.*, vol. 215, pp. 157–176, Sep. 2024. <https://doi.org/10.1016/j.isprsjprs.2024.06.020>
- [53] X. Yang *et al.*, "Correcting the bias of daily satellite precipitation estimates in tropical regions using deep neural network," *J. Hydrol.*, art. 608 127656, 2022. <https://doi.org/10.1016/j.jhydrol.2022.127656>
- [54] R. Adewoyin, P. Dueben, P. A. G. Watson, Y. He, and R. Dutta, "TRU-NET: A deep learning approach to high resolution prediction of rainfall," 2020. [Online]. Available: <https://research-information.bris.ac.uk/en/publications/tru-net-a-deep-learning-approach-to-high-resolution-prediction-of>
- [55] L. Espeholt *et al.*, "Skillful twelve hour precipitation forecasts using large context neural networks," 2021. [Online]. Available: <http://arxiv.org/abs/2111.07470>

- [56] Q. Jiang, F. Cioffi, W. Li, J. Tan, X. Pan, and X. Li, "Hybrid multilayer perceptron and convolutional neural network model to predict extreme regional precipitation dominated by the large-scale atmospheric circulation," *Atmos. Res.*, vol. 304, art. 107362, 2024. <https://doi.org/10.1016/j.atmosres.2024.107362>
- [57] Y. Tian, Y. Ji, X. Gao, X. Yuan, and X. Zhi, "Post-processing of short-term quantitative precipitation forecast with the multi-stream convolutional neural network," *Atmos. Res.*, vol. 309, art. 107584, Oct. 2024. <https://doi.org/10.1016/j.atmosres.2024.107584>
- [58] S. Ravuri et al., "Skilful precipitation nowcasting using deep generative models of radar," *Nature*, vol. 597, no. 7878, pp. 672–677, Sep. 2021. <https://doi.org/10.1038/s41586-021-03854-z>
- [59] S. Agrawal, L. Barrington, C. Bromberg, J. Burge, C. Gazen, and J. Hickey, "Machine learning for precipitation nowcasting from radar images," 2019. [Online]. Available: <https://doi.org/10.48550/arXiv.1912.12132>
- [60] V. A. Gorooh, A. A. Asanjan, P. Nguyen, K. Hsu, and S. Sorooshian, "Deep neural network high spatiotemporal resolution precipitation estimation (Deep-STEP) using passive microwave and infrared data," *J. Hydrometeorol.*, vol. 23, no. 4, pp. 597–617, Apr. 2022. <https://doi.org/10.1175/JHM-D-21-0194.1>
- [61] W. Li, B. Pan, J. Xia, and Q. Duan, "Convolutional neural network-based statistical post-processing of ensemble precipitation forecasts," *J. Hydrol.*, vol. 605, art. 127301, Feb. 2022. <https://doi.org/10.1016/j.jhydrol.2021.127301>
- [62] X. Shi et al., "Deep learning for precipitation nowcasting: A benchmark and a new model," *NeurIPS*, vol. 30, art. 11, 2017. <https://doi.org/10.48550/arXiv.1706.03458>
- [63] C. K. Søndersby et al., "MetNet: A neural weather model for precipitation forecasting," 2020. [Online]. Available: <https://arxiv.org/abs/2003.12140>
- [64] M. Jiang, B. Weng, J. Chen, T. Huang, F. Ye, and L. You, "Transformer-enhanced spatiotemporal neural network for post-processing of precipitation forecasts," *J. Hydrol.*, vol. 630, art. 130720, 2024. <https://doi.org/10.1016/j.jhydrol.2024.130720>
- [65] L. Fathollahi, F. Wu, R. Melaki, P. Jamshidi, and S. Sarwar, "Global normalized difference vegetation index forecasting from air temperature, soil moisture and precipitation using a deep neural network," *App. Comp. Geosci.*, vol. 23, art. 100174, 2024. <https://doi.org/10.1016/j.acags.2024.100174>
- [66] M. C. Bakkay et al., "Precipitation nowcasting using deep neural network," 2022. [Online]. Available: <http://arxiv.org/abs/2203.13263>
- [67] R. Xu, N. Chen, Y. Chen, and Z. Chen, "Downscaling and projection of multi-CMIP5 precipitation using machine learning methods in the Upper Han River Basin," *Adv. Meteorol.*, vol. 2020, pp. 1–17, Mar. 2020. <https://doi.org/10.1155/2020/8680436>
- [68] X. Shi, "Enabling smart dynamical downscaling of extreme precipitation events with machine learning," *Geophys. Res. Lett.*, vol. 47, no. 19, Oct. 2020. <https://doi.org/10.1029/2020GL090309>
- [69] B. Kumar, R. Chattopadhyay, M. Singh, N. Chaudhari, K. Kodari, and A. Barve, "Deep-learning based down-scaling of summer monsoon rainfall data over Indian region," 2020. [Online]. Available: <http://arxiv.org/abs/2011.11313>
- [70] H. Hu and B. M. Ayyub, "Machine learning for projecting extreme precipitation intensity for short durations in a changing climate," *Geosciences*, vol. 9, no. 5, art. 209, May 2019. <https://doi.org/10.3390/geosciences9050209>
- [71] D. D. Montenegro Murillo, M. A. Pérez Ortiz, and V. Vargas Franco, "Predicción de precipitación mensual mediante redes neuronales artificiales para la cuenca del río Cali, Colombia," *DYNA*, vol. 86, no. 211, pp. 122–130, Oct. 2019. <https://doi.org/10.15446/dyna.v86n211.76079>
- [72] T. Trinh, N. Do, V. T. Nguyen, and K. Carr, "Modeling high-resolution precipitation by coupling a regional climate model with a machine learning model: an application to Sai Gon–Dong Nai Rivers Basin in Vietnam," *Clim. Dyn.*, vol. 57, no. 9, pp. 2713–2735, Nov. 2021. <https://doi.org/10.1007/s00382-021-05833-6>
- [73] A. Y. Sun and G. Tang, "Downscaling satellite and reanalysis precipitation products using attention-based deep convolutional neural nets," *Front. Water*, vol. 2, art. 536743, 2020. [Online]. Available: <https://www.frontiersin.org/article/10.3389/frwa.2020.536743>
- [74] B. Pan, K. Hsu, A. AghaKouchak, and S. Sorooshian, "Improving precipitation estimation using convolutional neural network," *Water Resour. Res.*, vol. 55, no. 3, pp. 2301–2321, Mar. 2019. <https://doi.org/10.1029/2018WR024090>
- [75] A. Talei, L. H. C. Chua, and C. Quek, "A novel application of a neuro-fuzzy computational technique in event-based rain-fall-runoff modeling," *Exp. Sys. App.*, vol. 37, pp. 7456–7468, 2010. <https://doi.org/10.1016/j.eswa.2010.04.015>
- [76] M. A. Rahman, "Improvement of rainfall prediction model by using fuzzy logic," *Am. J. Clim. Change*, vol. 9, no. 4, art. 4, Nov. 2020. <https://doi.org/10.4236/ajcc.2020.94024>
- [77] A. Sözen, M. Kurt, M. A. Akçayol, and M. Özalp, "Performance prediction of a solar driven ejector-absorption cycle using fuzzy logic," *Renew. Energy*, vol. 29, no. 1, pp. 53–71, Jan. 2004. [https://doi.org/10.1016/S0960-1481\(03\)00172-1](https://doi.org/10.1016/S0960-1481(03)00172-1)
- [78] J. Kaiornrit, K. W. Wong, and C. C. Fung, "A modular technique for monthly rainfall time series prediction," in *2013 IEEE Symp. Comp. Intel. Dyn. Uncert. Environ. (CIDUE)*, 2013, pp. 76–83. <https://doi.org/10.1109/CIDUE.2013.6595775>
- [79] C. Tzimopoulos, L. Mpallas, and C. Evangelide, "Fuzzy model comparison to extrapolate rainfall data," *J. Environ. Sci. Tech.*, vol. 1, no. 4, pp. 214–224, Sep. 2008. <https://doi.org/10.3923/jest.2008.214.224>
- [80] R. Janarthanan, R. Balamurali, A. Annappoorani, and V. Vimala, "Prediction of rainfall using fuzzy logic," *Mater. Today Proc.*, vol. 37, pp. 959–963, 2021. <https://doi.org/10.1016/j.matpr.2020.06.179>

- [81] M. Silver, T. Svoray, A. Karnieli, and E. Fred, "Improving weather radar precipitation maps: A fuzzy logic approach," *Atmos. Res.*, vol. 234, art. 104710, Oct. 2019. <https://doi.org/10.1016/j.atmosres.2019.104710>
- [82] A. Helen, A. Gabriel, A. E., and B. Alese, "Development of a fuzzy logic based rainfall prediction model," *Int. J. Eng. Tech.*, vol. 3, pp. 427–435, Jan. 2013. https://www.researchgate.net/publication/285799840_Development_of_a_fuzzy_logic_based_rainfall_prediction_model
- [83] M. Hasan, T. Tsegaye, X. Shi, G. Schaefer, and G. Taylor, "Model for predicting rainfall by fuzzy set theory using USDA scan data," *Agri. Water Manag.*, vol. 95, no. 12, pp. 1350–1360, Dec. 2008. <https://doi.org/10.1016/j.agwat.2008.07.015>
- [84] S. Askani, K. Elhelou, I. Youssef, and M. El-Wahab, "Rainfall events prediction using rule-based fuzzy inference system," *Atmos. Res.*, vol. 101, pp. 228–236, Jul. 2011. <https://doi.org/10.1016/j.atmosres.2011.02.015>
- [85] G. Abbas Fall, M. Mousavi-Ba, and M. Habibi Nok, "Annual rainfall forecasting by using mamdani fuzzy inference system," *Res. J. Environ. Sci.*, vol. 3, no. 4, pp. 400–413, Apr. 2009. <https://doi.org/10.3923/rjes.2009.400.413>
- [86] S. Rakkasagi, M. K. Goyal, and S. Jha, "Evaluating the future risk of coastal Ramsar wetlands in India to extreme rainfalls using fuzzy logic," *J. Hydrol.*, vol. 632, art. 130869, 2024. <https://doi.org/10.1016/j.jhydrol.2024.130869>
- [87] G. M. M. A. Senaviratne, R. P. Udawatta, S. H. Anderson, C. Baf-faut, and A. Thompson, "Use of fuzzy rainfall-runoff predictions for claypan watersheds with conservation buffers in Northeast Missouri," *J. Hydrol.*, vol. 517, pp. 1008–1018, 2014. <https://doi.org/10.1016/j.jhydrol.2014.06.023>
- [88] P. Mani, S. Allen, S. Kotlarski, and M. Stoffel, "Climate sensitivity of natural hazards processes in mountain regions: A fuzzy logic approach," *Geomorphology*, vol. 461, art. 109329, 2024. <https://doi.org/10.1016/j.geomorph.2024.109329>
- [89] C. L. Chang, S. L. Lo, and S. L. Yu, "Applying fuzzy theory and genetic algorithm to interpolate precipitation," *J. Hydrol.*, vol. 314, no. 1, pp. 92–104, Nov. 2005. <https://doi.org/10.1016/j.jhydrol.2005.03.034>
- [90] J. Kajornrit and K. W. Wong, "Cluster validation methods for localization of spatial rainfall data in the northeast region of Thailand," in *2013 Int. Conf. Mach. Learn. Cyber.*, 2013, pp. 1637–1642. <https://doi.org/10.1109/IC-MLC.2013.6890861>
- [91] L. B. Franco, C. D. G. C. de Almeida, M. M. Freire, G. B. Franco, and S. de A. Silva, "Rainfall zoning for cocoa growing in Bahia state (Brazil) using fuzzy logic," *Eng. Agric.*, vol. 39, special no., pp. 48–55, Sep. 2019. <https://doi.org/10.1590/1809-4430-eng.agric.v39nep48-55/2019>
- [92] M. Zare, G. J.-P. Schumann, F. N. Teferle, and R. Mansorian, "Generating flood hazard maps based on an innovative spatial interpolation methodology for precipitation," *Atmosphere*, vol. 12, no. 10, art. 10, Oct. 2021. <https://doi.org/10.3390/atmos12101336>
- [93] P. Sinha *et al.*, "Downscaled rainfall projections in south Florida using self-organizing maps," *Sci. Total Environ.*, vol. 635, pp. 1110–1123, Sep. 2018. <https://doi.org/10.1016/j.scitotenv.2018.04.144>
- [94] P. Singh, "Rainfall and financial forecasting using fuzzy time series and neural networks based model," *Int. J. Mach. Learn. Cyber.*, vol. 9, no. 3, pp. 491–506, Mar. 2018. <https://doi.org/10.1007/s13042-016-0548-5>
- [95] J. Kajornrit, K. W. Wong, and C. C. Fung, "Rainfall prediction in the northeast region of Thailand using Modular Fuzzy Inference System," in *2012 IEEE Int. Conf. Fuzzy Syst.*, 2012, pp. 1–6. <https://doi.org/10.1109/FUZZ-IEEE.2012.6250785>
- [96] S. Dani and S. Sharma, "Forecasting rainfall of a region by using fuzzy time series," *Asian J. Math. App.*, vol. 2013, 2013. <https://scienceasia.asia/files/65.pdf>
- [97] J. Kajornrit, "Interpretable fuzzy systems for monthly rainfall spatial interpolation and time series prediction," PhD dissertation, Murdoch Univ., Perth, Australia, 2014. [Online]. Available: <https://researchrepository.murdoch.edu.au/id/eprint/26220/>
- [98] J.-S. R. Jang, "ANFIS: adaptive-network-based fuzzy inference system," *IEEE Trans. Syst. Man. Cyber.*, no. 3, art. 665, 1993. <https://doi.org/10.1109/21.256541>
- [99] K. W. Wong, P. M. Wong, T. D. Gedeon, and C. C. Fung, "Rainfall prediction model using soft computing technique," *Soft Comp.*, vol. 7, no. 6, pp. 434–438, May 2003. <https://doi.org/10.1007/s00500-002-0232-4>
- [100] P. V. de Campos Souza, L. Batista de Oliveira, and L. A. Ferreira do Nascimento, "Fuzzy rules to help predict rains and temperatures in a Brazilian capital state based on data collected from satellites," *App. Sci.*, vol. 9, no. 24, art. 5476, Dec. 2019. <https://doi.org/10.3390/app9245476>
- [101] F. Chang, Y.-M. Chiang, M.-J. Tsai, M.-C. Shieh, K.-L. Hsu, and S. Sorooshian, "Watershed rainfall forecasting using neuro-fuzzy networks with the assimilation of multi-sensor information," *J. Hydrol.*, vol. 508, pp. 374–384, Jan. 2014. <https://doi.org/10.1016/j.jhydrol.2013.11.011>
- [102] D. Petković, M. Gocić, and S. Shamshirband, "Adaptive neuro-fuzzy computing technique for precipitation estimation," *Facta Univ. Ser. Mech. Eng.*, vol. 14, no. 2, pp. 209–218, 2016. <https://doi.org/10.22190/FU-ME1602209P>
- [103] R. Hashim *et al.*, "Selection of meteorological parameters affecting rainfall estimation using neuro-fuzzy computing methodology," *Atmos. Res.*, vol. 171, supp. C, pp. 21–30, May 2016. <https://doi.org/10.1016/j.atmosres.2015.12.002>
- [104] M. Hessami, F. Anctil, and A. A. Viau, "An adaptive neuro-fuzzy inference system for the post-calibration of weather radar rainfall estimation," *J. Hydroinfo.*, vol. 5, no. 1, pp. 63–70, Jan. 2003. <https://doi.org/10.2166/hydro.2003.0005>
- [105] S. Sharma, P. Srivastava, X. Fang, and L. Kalin, "Hydrologic simulation approach for El Niño Southern Oscillation (ENSO)-affected watershed with limited raingauge stations," *Hydrol. Sci. J.*, vol. 61, no. 6, pp. 991–1000, Apr. 2016. <https://doi.org/10.1080/02626667.2014.952640>

- [106] A. Rahman *et al.*, "Rainfall prediction system using machine learning fusion for smart cities," *Sensors*, vol. 22, pp. 1–14, May 2022. <https://doi.org/10.3390/s22093504>
- [107] Y. Morales, M. Querales, H. Rosas, H. Allende-Cid, and R. Salas, "A self-identification Neuro-Fuzzy inference framework for modeling rainfall-runoff in a Chilean watershed," *J. Hydrol.*, vol. 23, art 125910, Aug. 2020. <https://doi.org/10.1016/j.jhydrol.2020.125910>
- [108] Shahfahad, M. W. Naikoo, S. Talukdar, T. Das, and A. Rahman, "Identification of homogenous rainfall regions with trend analysis using fuzzy logic and clustering approach coupled with advanced trend analysis techniques in Mumbai city," *Urb. Clim.*, vol. 46, art. 101306, Dec. 2022. <https://doi.org/10.1016/j.uclim.2022.101306>
- [109] A. Talei, L. H. C. Ghua, and T. S. W. Wong, "Evaluation of rainfall and discharge inputs used by adaptive network-based fuzzy inference systems (ANFIS) in rainfall-runoff modeling," *J. Hydrol.*, vol. 391, no. 3-4, pp. 248–262, 2010. <https://doi.org/10.1016/j.jhydrol.2010.07.023>
- [110] M. C. Valverde, E. Araujo, and H. Campos Velho, "Neural network and fuzzy logic statistical downscaling of atmospheric circulation-type specific weather pattern for rainfall forecasting," *App. Soft Comp.*, vol. 22, pp. 681–694, 2014. <https://doi.org/10.1016/j.asoc.2014.02.025>
- [111] S. Askar *et al.*, "Flood susceptibility mapping using remote sensing and integration of decision table classifier and metaheuristic algorithms," *Water*, vol. 14, no. 19, art. 3062, 2022. <https://doi.org/10.3390/w14193062>
- [112] N.-D. Hoang and X.-L. Tran, "Remote sensing-based urban green space detection using marine predators algorithm optimized machine learning approach," *Math. Probl. Eng.*, art. 5586913, 2021. <https://doi.org/10.1155/2021/5586913>
- [113] H. E. Khairan, S. L. Zubaidi, S. F. Raza, M. Hameed, N. Al-Ansari, and H. M. Ridha, "Examination of Single- and hybrid-based metaheuristic algorithms in ANN reference evapotranspiration estimating," *Sustainability*, vol. 15, art. 14222, 2023. <https://doi.org/10.3390/su151914222>
- [114] R. Adnan *et al.*, "Advanced hybrid metaheuristic machine learning models application for reference crop evapotranspiration prediction," *Agronomy*, vol. 13, art. 98, 10 3390 13010098, 2022. <https://doi.org/10.3390/agronomy13010098>
- [115] D. T. Bui *et al.*, "Novel hybrid evolutionary algorithms for spatial prediction of floods," *Sci. Rep.*, vol. 8, p. 15364, 2018. <https://doi.org/10.1038/s41598-018-33620-1>
- [116] S. V. R. Termeh, A. Kornejady, H. R. Pourghasemi, and S. Keesstra, "Flood susceptibility mapping using novel ensembles of adaptive neuro fuzzy inference system and metaheuristic algorithms," *Sci. Total Environ.*, vol. 615, pp. 438–451, 2018. <https://doi.org/10.1016/j.scitotenv.2017.09.262> Get rights and content
- [117] A. Arora *et al.*, "Optimization of state-of-the-art fuzzy-metaheuristic anfis-based machine learning models for flood susceptibility prediction mapping in the middle Ganga plain, India," *Sci. Total Environ.*, vol. 750, art. 141565, 2021. <https://doi.org/10.1016/j.scitotenv.2020.141565>
- [118] O. Rahmati *et al.*, "Development of novel hybridized models for urban flood susceptibility mapping," *Sci. Rep.*, vol. 10, art. 12937, 2020. <https://doi.org/10.1038/s41598-020-69703-7>
- [119] F. Rezaie, M. Panahi, S. M. Bateni, C. Jun, C. M. Neale, and S. Lee, "Novel hybrid models by coupling support vector regression (SVR) with meta-heuristic algorithms (WOA and GWO) for flood susceptibility mapping," *Nat. Haz.*, pp. 1–37, 2022. <https://doi.org/10.1007/s11069-022-05424-6>
- [120] M. Panahi *et al.*, "Flood spatial prediction modeling using a hybrid of meta-optimization and support vector regression modeling," *Catena*, vol. 199, art. 105114, 2021. <https://doi.org/10.1016/j.catena.2020.105114>
- [121] E. Dodangeh *et al.*, "Novel hybrid intelligence models for flood-susceptibility prediction: Meta optimization of the GMDH and SVR models with the genetic algorithm and harmony search," *J. Hydrol.*, vol. 590, art. 125423, 2020. <https://doi.org/10.1016/j.jhydrol.2020.125423>
- [122] F. Rezaie, S. M. Bateni, E. Heggy, and S. Lee, "Utilizing the SAR, GIS, and novel hybrid metaheuristic-GMDH algorithm for flood susceptibility mapping," in *Proc. 2021 IEEE Int. Geosci. Rem. Sens. Symp. IGARSS*, 2021, pp. 11–16. <https://doi.org/10.1109/IGARSS47720.2021.9553468>
- [123] B. T. Pham, K.-T. T. Bui, I. Prakash, and B.-B. Ly, "Hybrid artificial intelligence models based on adaptive neuro fuzzy inference system and metaheuristic optimization algorithms for prediction of daily rainfall," *Phys. Chem. Earth*, vol. 134, art. 103563, 2024. <https://doi.org/10.1016/j.pce.2024.103563>
- [124] C. Peláez-Rodríguez *et al.*, "A general explicable forecasting framework for weather events based on ordinal classification and inductive rules combined with fuzzy logic," *Knowledge-Based Syst.*, vol. 291, art. 111556, 2024. <https://doi.org/10.1016/j.knosys.2024.111556>
- [125] B. Mohammadi, S. Vazifehkhah, and Z. Duan, "A conceptual metaheuristic-based framework for improving runoff time series simulation in glacierized catchments," *Eng. App. Art. Intel.*, vol. 127, art. 107302, Jan. 2024. <https://doi.org/10.1016/j.engappai.2023.107302>
- [126] B. Zerouali *et al.*, "Artificial intelligent systems optimized by metaheuristic algorithms and teleconnection indices for rainfall modeling: The case of a humid region in the mediterranean basin," *Heliyon*, vol. 9, no. 4, art. e15355, Apr. 2023. <https://doi.org/10.1016/j.heliyon.2023.e15355>
- [127] U. Okkan, Z. B. Ersoy, A. A. Kumanlioglu, and O. Fistikoglu, "Embedding machine learning techniques into a conceptual model to improve monthly runoff simulation: A nested hybrid rainfall runoff modeling," *J. Hydrol.*, vol. 598, art. 126433, 2021. <https://doi.org/10.1016/j.jhydrol.2020.125423>
- [128] S. E. Priestly, K. Raimond, Y. Cohen, J. Brema, and D. J. Hemanth, "Evaluation of a novel hybrid lion swarm optimization – AdaBoostRegressor model for forecasting monthly precipitation," *Sust. Comp. Info. Syst.*, vol. 39, art. 100884, 2023. <https://doi.org/10.1016/j.suscom.2023.100884>

- [129] T. Anuradha, P. S. G. Aruna Sri Formal, and J. RamaDevi, "Hybrid model for rainfall prediction with statistical and technical indicator feature set," *Exp. Syst. App.*, vol. 249, art. 123260, Sep. 2024. <https://doi.org/10.1016/j.eswa.2024.123260>
- [130] X. Qiao *et al.*, "Metaheuristic evolutionary deep learning model based on temporal convolutional network, improved aquila optimizer and random forest for rainfall-runoff simulation and multi-step runoff prediction," *Exp. Syst. App.*, vol. 229, art. 120616, Nov. 2023. <https://doi.org/10.1016/j.eswa.2023.120616>
- [131] S. D. Latif *et al.*, "Assessing rainfall prediction models: Exploring the advantages of machine learning and remote sensing approaches," *Alexandria Eng. J.*, vol. 82, pp. 16–25, Nov. 2023. <https://doi.org/10.1016/j.aej.2023.09.060>
- [132] E. Gomaa *et al.*, "Assessment of hybrid machine learning algorithms using TRMM rainfall data for daily inflow forecasting in Três Marias Reservoir, eastern Brazil," *Heliyon*, vol. 9, no. 8, art. e18819, Aug. 2023. <https://doi.org/10.1016/j.heliyon.2023.e18819>
- [133] T. Li, C. Xie, C. Xu, W. Qi, Y. Huang, and L. Li, "Automated machine learning for rainfall-induced landslide hazard mapping in Luhe County of Guangdong Province, China," *China Geol.*, vol. 7, art. 315 329, 2024. <https://doi.org/10.31035/cg2024064>
- [134] A. I. Pathan *et al.*, "Comparative assessment of rainfall-based water level prediction using machine learning (ML) techniques," *Ain Shams Eng. J.*, vol. 15, art. 102854, 2024. <https://doi.org/10.1016/j.asej.2024.102854>
- [135] Y. Yin, J. He, J. Guo, W. Song, H. Zheng, and J. Dan, "Enhancing precipitation estimation accuracy: An evaluation of traditional and machine learning approaches in rainfall predictions," *J. Atmos. Solar-Terres. Phys.*, vol. 255, art. 106175, Feb. 2024. <https://doi.org/10.1016/j.jastp.2024.106175>
- [136] W. Dai *et al.*, "Estimation of rainfall erosivity on the Chinese Loess Plateau: A new combination of the ERA5 dataset and machine learning," *J. Hydrol.*, vol. 624, art. 129892, Sep. 2023. <https://doi.org/10.1016/j.jhydrol.2023.129892>
- [137] J. Diez-Sierra and M. Del Jesus, "Long-term rainfall prediction using atmospheric synoptic patterns in semi-arid climates with statistical and machine learning methods," *J. Hydrol.*, vol. 586, art. 124789, Jul. 2020. <https://doi.org/10.1016/j.jhydrol.2020.124789>
- [138] U. Rasool *et al.*, "Rainfall-driven machine learning models for accurate flood inundation mapping in Karachi, Pakistan," *Urb. Clim.*, vol. 49, art. 101573, May 2023. <https://doi.org/10.1016/j.uclim.2023.101573>
- [139] D. Pirone, L. Cimorelli, G. Del Giudice, and D. Pianese, "Short-term rainfall forecasting using cumulative precipitation fields from station data: A probabilistic machine learning approach," *J. Hydrol.*, vol. 617, art. 128949, Feb. 2023. <https://doi.org/10.1016/j.jhydrol.2022.128949>

Table I. Main references by year and category regarding artificial neural networks and fuzzy logic-based approaches

Source: Authors

Title	Authors	Year	Country	Category	AI Model
<i>Precipitation estimation from remotely sensed information using artificial neural networks</i>	Hsu et al.	1997	USA	Artificial Neural Networks	ANN
<i>An application of artificial neural networks for rainfall forecasting</i>	Luk et al.	2001	Australia	Artificial Neural Networks	BPNN, RNN, TDNN
<i>Data mining techniques for improved WSR-88D rainfall estimation</i>	Trafalis et al.	2002	USA	Artificial Neural Networks	ANN
<i>Precipitation estimation from remotely sensed imagery using an Artificial Neural Network Cloud Classification System</i>	Hong et al.	2004	USA	Artificial Neural Networks	ANN
<i>Spatiotemporal monthly rainfall reconstruction via artificial neural network -case study: south of Brazil</i>	Lucio et al.	2007	Brazil	Artificial Neural Networks	ANN
<i>Rainfall forecasting by technological machine learning models</i>	W. Hong	2008	Taiwan	Artificial Neural Networks	RNN
<i>Elaboración de la cartografía climática de temperaturas y precipitación mediante redes neuronales artificiales: caso de estudio en la Región del Libertador Bernardo O'Higgins</i>	Román & Andrés	2010	Chile	Artificial Neural Networks	BPNN
<i>Precipitation forecast using artificial neural networks in specific regions of Greece</i>	Moustris et al.	2011	Greece	Artificial Neural Networks	ANN
<i>Estimation of missing precipitation records using modular artificial neural networks</i>	Kajornrit et al.	2012	Thailand	Artificial Neural Networks	BPNN, MNN
<i>The application of artificial intelligence for monthly rainfall forecasting in the Brisbane Catchment, Queensland, Australia</i>	Abbot and Marohasy	2013	Australia	Artificial Neural Networks	ANN
<i>Artificial neural networks modeling for forecasting the maximum daily total precipitation at Athens, Greece</i>	Nastos et al.	2014	Greece	Artificial Neural Networks	TDNN
<i>Forecasting daily precipitation using hybrid model of wavelet-artificial neural network and comparison with adaptive neurofuzzy inference system (case study: Verayneh Station, Nahavand)</i>	Solgi	2014	Iran	Artificial Neural Networks	ANN-wavelets
<i>Input selection and optimisation for monthly rainfall forecasting in Queensland, Australia, using artificial neural networks</i>	Abbot and Marohasy	2014	Australia	Artificial Neural Networks	ANN
<i>Estimating spatial precipitation using regression kriging and artificial neural network residual kriging (RKNRK) hybrid approach</i>	Seo et al.	2015	South Korea	Artificial Neural Networks	ANN
<i>Artificial neural networks in precipitation nowcasting: An Australian case study</i>	Schroeter	2016	Australia	Artificial Neural Networks	Shallow NN
<i>Precipitation identification with bispectral satellite information using deep learning approaches</i>	Tao et al.	2017	USA	Artificial Neural Networks	DNN
<i>Comparison of adaline and multiple linear regression methods for rainfall forecasting</i>	Sutawinaya et al.	2018	Indonesia	Artificial Neural Networks	ADELIN
<i>A learning framework for an accurate prediction of rainfall rates</i>	Damavandi and Shah	2019	China, India, Pakistan	Artificial Neural Networks	BPNN


























<i>Deep learning for precipitation estimation from satellite and rain gauges measurements</i>	Morau et al.	2019	Germany, Belgium, Netherlands	Artificial Neural Networks	DNN, CNN
<i>Neural network approach to forecast hourly intense rainfall using GNSS precipitable water vapor and meteorological sensors</i>	Benevides et al.	2019	Portugal	Artificial Neural Networks	NARX
<i>Predicción de precipitación mensual mediante redes neuronales artificiales para la cuenca del río Cali, Colombia</i>	Montenegro Murillo et al.	2019	Colombia	Artificial Neural Networks	ANN
<i>Improving precipitation estimation using convolutional neural network</i>	Pan et al.	2019	USA	Artificial Neural Networks	CNN
<i>Rainfall and atmospheric temperature against the other climatic factors: A case study from Colombo, Sri Lanka</i>	Anushka-Perera	2019	Sri Lanka	Artificial Neural Networks	ANN
<i>Modeling short term rainfall forecast using neural networks, and Gaussian process classification based on the SPI drought index</i>	Azimi and Azhdary-Moghaddam	2020	Iran	Artificial Neural Networks	BPNN
<i>Precipitation forecast of the Wujiang River Basin based on artificial bee colony algorithm and backpropagation neural network</i>	Wang et al.	2020	China	Artificial Neural Networks	ABC-BPNN
<i>Secondary precipitation estimate merging using machine learning: Development and evaluation over Krishna River Basin, India</i>	Kolluru et al.	2020	India	Artificial Neural Networks	ANN
<i>A comparative study on machine learning techniques for intense convective rainfall events forecasting</i>	Sangiorgio et al.	2020	Italy	Artificial Neural Networks	DNN
<i>Deep multilayer perceptron for knowledge extraction: Understanding the Gardon de Mialet flash floods modeling</i>	Saint Fleur et al.	2020	France	Artificial Neural Networks	ANN
<i>Downscaling satellite and reanalysis precipitation products using attention-based deep convolutional neural nets</i>	Sun and Tang	2020	USA	Artificial Neural Networks	CNN
<i>MetNet: A neural weather model for precipitation forecasting</i>	Sønderby et al.	2020	USA	Artificial Neural Networks	ConvLSTM
<i>Improving near real-time precipitation estimation using a U-Net convolutional neural network and geographical information</i>	Sadegui et al.	2020	USA	Artificial Neural Networks	CNN
<i>Artificial intelligence-based techniques for rainfall estimation integrating multisource precipitation datasets</i>	Khan and Bhuiyan	2021	Ethiopia	Artificial Neural Networks	ANN
<i>Long short-term memory algorithm for rainfall prediction based on El-Niño and IOD data</i>	Haq et al.	2021	Indonesia	Artificial Neural Networks	LSTM
<i>SCENT: A new precipitation nowcasting method based on sparse correspondence and deep neural network</i>	Fang et al.	2021	China	Artificial Neural Networks	DNN,CNN
<i>Skilful precipitation nowcasting using deep generative models of radar</i>	Ravuri et al.	2021	United Kingdom	Artificial Neural Networks	Generativo
<i>Skillful twelve hour precipitation forecasts using large context neural networks</i>	Espeholt et al.	2021	USA	Artificial Neural Networks	CRN
<i>TRU-NET: A deep learning approach to high resolution prediction of rainfall</i>	Adewoyin et al.	2021	United Kingdom	Artificial Neural Networks	CRN
<i>Modeling high-resolution precipitation by coupling a regional climate model with a machine learning model: an application to Sai Gon-Dong Nai Rivers Basin in Vietnam</i>	Trinh et al.	2021	Vietnam	Artificial Neural Networks	ANN
<i>A deep learning multimodal method for precipitation estimation</i>	Morau et al.	2021	Germany, Belgium, Netherlands	Artificial Neural Networks	DNN, CNN

<i>A neural network-based approach for the detection of heavy precipitation using GNSS observations and surface meteorological data</i>	Li et al.	2021	China	Artificial Neural Networks	BPNN
<i>SCENT: A new precipitation nowcasting method based on sparse correspondence and deep neural network</i>	Fang et al.	2021	China	Artificial Neural Networks	SCENT
<i>Precipitaion nowcasting using deep neural network</i>	Bakkay et al.	2022	France	Artificial Neural Networks	CNN, LSTM
<i>Extreme precipitation prediction based on neural network model – A case study for southeastern Brazil</i>	de Sousa Araújo et al.	2022	Brazil	Artificial Neural Networks	LSTM
<i>Convolutional neural network-based statistical post-processing of ensemble precipitation forecasts</i>	Li et al.	2022	China	Artificial Neural Networks	CNN
<i>Deep neural network high spatiotemporal resolution precipitation estimation (Deep-STEP) using passive microwave and infrared data</i>	Gorooh et al.	2022	USA	Artificial Neural Networks	CNN
<i>Correcting the bias of daily satellite precipitation estimates in tropical regions using deep neural network</i>	Yang et al.	2022	China	Artificial Neural Networks	Bi-LSTM-T
<i>Hybrid multilayer perceptron and convolutional neural network model to predict extreme regional precipitation dominated by the large-scale atmospheric circulation</i>	Jiang et al.	2024	China	Artificial Neural Networks	MLP, CNN, MLP-CNN
<i>Post-processing of short-term quantitative precipitation forecast with the multi-stream convolutional neural network</i>	Tian et al.	2024	China	Artificial Neural Networks	CNN, MSCNN
<i>Transformer-enhanced spatiotemporal neural network for post-processing of precipitation forecasts</i>	Jiang et al.	2024	China	Artificial Neural Networks	Transformer, TransLSTMUNet, ConvLSTM
<i>Improving rainfall-runoff modeling in the Mekon river basin using bias-correct satellite precipitation products by convolutional neural networks</i>	Le et al.	2024	China, Burma, Laos, Thailand, Cambodia, Vietnam	Artificial Neural Networks	CNN
<i>Deep neural network based on dynamic attention and layer attention for meteorological data downscaling</i>	Wang et al.	2024	China	Artificial Neural Networks	DNN
<i>Rainfall prediction model using soft computing technique</i>	Wong et al.	2003	Italy	Fuzzy Logic	NFS
<i>An adaptive neuro-fuzzy inference system for the post-calibration of weather radar rainfall estimation</i>	Hessami et al.	2003	Canada	Fuzzy Logic	ANFIS
<i>Applying fuzzy theory and genetic algorithm to interpolate precipitation</i>	Chang et al	2005	Taiwan	Fuzzy Logic	FI
<i>Fuzzy model comparison to extrapolate rainfall data</i>	Tzimopoulou et al.	2008	Greece	Fuzzy Logic	ANFIS
<i>Model for predicting rainfall by fuzzy set theory using USDA scan data</i>	Hasan et al.	2008	USA	Fuzzy Logic	FIS
<i>Annual rainfall forecasting by using mamdani fuzzy inference system</i>	Abbas et al.	2009	Iran	Fuzzy Logic	MFIS
<i>Evaluation of rainfall and discharge inputs used by adaptive network-based fuzzy inference systems (ANFIS) in rainfall-runoff modeling</i>	Talei et al.	2010	Singapore	Fuzzy Logic	ANFIS
<i>Rainfall events prediction using rule-based fuzzy inference system</i>	Asklany et al.	2011	Egypt	Fuzzy Logic	FIS
<i>Rainfall prediction in the northeast region of Thailand using modular fuzzy inference system</i>	Kajornrit et al.	2012	Thailand	Fuzzy Logic	FIS
<i>A modular technique for monthly rainfall time series prediction</i>	Kaionrit et al.	2013	Thailand	Fuzzy Logic	MFIS
<i>Cluster validation methods for localization of spatial rainfall data in the northeast region of Thailand</i>	Kajornrit and Wong	2013	Thailand	Fuzzy Logic	FCM
<i>Development of a fuzzy logic based rainfall prediction model</i>	Helen et al.	2013	Nigeria	Fuzzy Logic	MFIS

<i>Forecasting rainfall of a region by using fuzzy time series</i>	(Dani & Sharma	2013		Fuzzy Logic	FIS
<i>An integrated intelligent technique for monthly rainfall time series prediction</i>	Kajornrit et al.	2014	Thailand	Fuzzy Logic	ANFIS
<i>Interpretable fuzzy systems for monthly rainfall spatial interpolation and time series prediction</i>	Kajornrit	2014	Thailand	Fuzzy Logic	FIS, ANFIS, Modular FIS
<i>Watershed rainfall forecasting using neuro-fuzzy networks with the assimilation of multi-sensor information</i>	F. Chang et al.	2014	Taiwan	Fuzzy Logic	ANFIS
<i>Neural network and fuzzy logic statistical downscaling of atmospheric circulation-type specific weather pattern for rainfall forecasting</i>	Valverde et al.	2014	Brazil	Fuzzy Logic	FSD
<i>Adaptive neuro-fuzzy computing technique for precipitation estimation</i>	Petković et al.	2016	Serbia	Fuzzy Logic	ANFIS
<i>Hydrologic simulation approach for El Niño Southern Oscillation (ENSO)-affected watershed with limited raingauge stations</i>	Sharma et al.	2016	USA	Fuzzy Logic	ANFIS
<i>Selection of meteorological parameters affecting rainfall estimation using neuro-fuzzy computing methodology</i>	Hashim et al.	2016	India	Fuzzy Logic	ANFIS
<i>Rainfall and financial forecasting using fuzzy time series and neural networks based model</i>	Singh	2018	India	Fuzzy Logic	FTS
<i>Downscaled rainfall projections in south Florida using self-organizing maps</i>	Sinha et al.	2018	USA	Fuzzy Logic	FCM
<i>Fuzzy rules to help predict rains and temperatures in a brazilian capital state based on data collected from satellites</i>	de Campos Souza et al.	2019	Brazil	Fuzzy Logic	ANFIS
<i>Improving weather radar precipitation maps: A fuzzy logic approach</i>	Silver et al.	2019	Israel	Fuzzy Logic	FIS
<i>Rainfall zoning for cocoa growing in Bahia state (Brazil) using fuzzy logic</i>	Franco et al.	2019	Brazil	Fuzzy Logic	FIS
<i>Improvement of rainfall prediction model by using fuzzy logic</i>	Rahman	2020	Bangladesh	Fuzzy Logic	FIS
<i>Generating flood hazard maps based on an innovative spatial interpolation methodology for precipitation</i>	Zare et al.	2021	Luxembourg, Germany, France	Fuzzy Logic	FCM
<i>Prediction of rainfall using fuzzy logic</i>	Janarthanan et al.	2021	India	Fuzzy Logic	FIS
<i>Optimization of state-of-the-art fuzzy-metaheuristic ANFIS-based machine learning models for flood susceptibility prediction mapping in the middle Ganga Plain, India</i>	Arora et al.	2021	India	Fuzzy Logic	Metaheuristic ANFIS
<i>Rainfall prediction system using machine learning fusion for smart cities</i>	Rahman et al.	2022	Pakistan	Fuzzy Logic	FIS
<i>Identification of homogenous rainfall regions with trend analysis using fuzzy logic and clustering approach coupled with advanced trend analysis techniques in Mumbai city</i>	Shahfahad et al.	2022	India	Fuzzy Logic	FCM
<i>A general explicable forecasting framework for weather events based on ordinal classification and inductive rules combined with fuzzy logic</i>	Peláez-Rodríguez et al.	2024	Spain	Fuzzy Logic	FIS
<i>Hybrid artificial intelligence models based on adaptive neuro fuzzy inference system and metaheuristic optimization algorithms for prediction of daily rainfall</i>	Pham et al.	2024	Vietnam	Fuzzy Logic	Metaheuristic ANFIS
<i>Evaluating the future risk of coastal Ramsar wetlands in India to extreme rainfalls using fuzzy logic</i>	Rakkasagi et al.	2024	India	Fuzzy Logic	FIS

Inter-Laboratory Testing Program for the Physical Characterization of Guamo Sand

Programa inter-laboratorio de ensayos para la caracterización física de la arena del Guamo

Juan Carlos Ruge ¹, Fausto Molina-Gómez ², María C. Olarte ³, Javier Camacho-Tauta ⁴,
Óscar Reyes-Ortiz ⁵, Joan M. Larrahondo ⁶, Hermes A. Vacca ⁷, Luis F. Prada ⁸,
Alfonso Ramos-Cañón ⁹, Yezid A. Alvarado ¹⁰, Fernando J. Reyes ¹¹, Miguel A. Cabrera ¹²,
Bernardo Caicedo-Hormaza ¹³, José S. Naranjo ¹⁴, Iván F. Otálvaro ¹⁵, Alejandra Gómez-Jiménez ¹⁶,
Mayra A. Galvis ¹⁷, July E. Carmona ¹⁸, Cesar A. García ¹⁹, Alex E. Álvarez ²⁰, Edgardo J. Díaz ²¹,
Julio E. Colmenares ²², Carlos R. Reina ²³, Cristhian C. Mendoza ²⁴, Diego F. Gil ²⁵,
Laura M. Espinosa ²⁶, Eliana Martínez-Rojas ²⁷, Juan G. Bastidas ²⁸, and Jhan P. Rojas ²⁹

¹ PhD in Geotechnics, University of Brasília. Affiliation: Assistant professor, Universidad Militar Nueva Granada. E-mail: juan.ruge@unimilitar.edu.co

² PhD in Civil Engineering, University of Porto. Affiliation: Lecturer researcher, Universidad Militar Nueva Granada. E-mail: fausto.molina@unimilitar.edu.co

³ MSc in Geotechnics, University of Brasília. Affiliation: Lecturer researcher, Universidad Militar Nueva Granada. E-mail: u1102789@unimilitar.edu.co

⁴ PhD in Civil Engineering, Universidade Técnica de Lisboa. Affiliation: Full Professor, Universidad Militar Nueva Granada. E-mail: javier.camacho@unimilitar.edu.co

⁵ PhD in Civil Engineering, Universidad Politécnica de Cataluña. Affiliation: Full professor, Universidad Militar Nueva Granada. E-mail: oscar.reyes@unimilitar.edu.co

⁶ PhD in Civil Engineering, Georgia Institute of Technology. Affiliation: Associate professor, Pontificia Universidad Javeriana. E-mail: jlarrahondo@javeriana.edu.co

⁷ PhD in Civil Engineering, Pontificia Universidad Javeriana. Affiliation: Assistant professor, Pontificia Universidad Javeriana. E-mail: vacca@javeriana.edu.co

⁸ PhD in Geotechnical Engineering, Karlsruhe Institute of Technology. Affiliation: Associate professor, Pontificia Universidad Javeriana. E-mail: lf.pradas@javeriana.edu.co

⁹ PhD in Engineering, Universidad de Los Andes. Affiliation: Full professor, Pontificia Universidad Javeriana. E-mail: a-ramos@javeriana.edu.co

¹⁰ PhD in Construction, Universitat Politècnica de València. Affiliation: Laboratory director at the Engineering Faculty of Pontificia Universidad Javeriana. E-mail: alvarado.y@javeriana.edu.co

¹¹ MEng, Pontificia Universidad Javeriana. Affiliation: Laboratory engineer, Pontificia Universidad Javeriana. E-mail: f_reyes@javeriana.edu.co

¹² PhD in Soil Science, Universität für Bodenkultur, Wien. Affiliation: Associate professor, Universidad de Los Andes. E-mail: m.a.cabrera@tudelft.nl

¹³ PhD in Soil Mechanics and Structures, Ecole Centrale Paris. Affiliation: Full professor, Universidad de Los Andes. E-mail: bcaicedo@uniandes.edu.co

¹⁴ Civil Engineer, Escuela de Ingenieros Militares. Affiliation: Laboratory engineer, Universidad de Los Andes. E-mail: jnaranjo@uniandes.edu.co

¹⁵ PhD in Geotechnics, University of Brasília. Affiliation: Associate professor, Pontificia Universidad Javeriana – Cali. E-mail: ifotalvaro@javerianacali.edu.co

¹⁶ PhD in Geotechnics, University of Brasília. Affiliation: Lecturer professor, Pontificia Universidad Javeriana – Cali. E-mail: alejgomez@javerianacali.edu.co

¹⁷ Civil Engineer, Escuela de Ingenieros Militares. Affiliation: Laboratory Engineer, Pontificia Universidad Javeriana – Cali. E-mail: mayra.galvis@javerianacali.edu.co

¹⁸ MSc in Civil Engineering, Universidad Distrital Francisco José de Caldas. Affiliation: PhD student, Universidad Distrital Francisco José de Caldas. E-mail: jecarmonaa@udistrital.edu.co

¹⁹ PhD in Engineering, Universidad de Los Andes. Affiliation: Full professor, Universidad Distrital Francisco José de Caldas. E-mail: cagarciau@udistrital.edu.co

²⁰ PhD in Materials, Texas A&M University. Affiliation: Full professor, Universidad Industrial de Santander. E-mail: allex.alvarez@uis.edu.co

²¹ Civil Engineer, Universidad del Magdalena. Affiliation: Assistant lecturer, Universidad del Magdalena. E-mail: ediaz@unimagdalena.edu.co

²² PhD, Imperial College of London. Affiliation: Full professor, Universidad Nacional de Colombia. E-mail: jecolmenaresm@unal.edu.co

²³ MSc in Geotechnical Engineering, Universidad Nacional de Colombia. Affiliation: Laboratory engineer, Universidad Nacional de Colombia. E-mail: crreinal@unal.edu.co

²⁴ PhD in Geotechnical Engineering, University of Brasília. Affiliation: Associate professor, Universidad Nacional de Colombia Sede Manizales. E-mail: cmendozab@unal.edu.co

²⁵ MSc in Structural Engineering, Universidad Nacional de Colombia Sede Manizales. Affiliation: Lecturer professor, Universidad Nacional de Colombia Sede Manizales. E-mail: dfgilo@unal.edu.co

²⁶ MSc in Urban Planning, Université Grenoble Alpes. Affiliation: Assistant professor, Universidad Católica de Colombia. E-mail: lespinosa@ucatolica.edu.co

²⁷ PhD in Geotechnics, Universidad Politécnica de Madrid. Affiliation: Universidad Piloto de Colombia. E-mail: elianamartinezrojas@gmail.com

²⁸ PhD in Geotechnics, University of Brasília. Affiliation: Assistant professor, Universidad Piloto de Colombia. E-mail: juan-bastidas@unipiloto.edu.co

²⁹ MSc in Civil Engineering, Universidad de Los Andes. Affiliation: Assistant professor, Universidad Francisco de Paula Santander. E-mail: jhanpierorojas@ufps.edu.co



ABSTRACT

In soil testing, assessing physical properties is essential for accurately characterizing sands. However, testing results can vary depending on the experimental procedures used and their implementation. A round-robin exercise facilitates the simultaneous analysis of the reproducibility and replicability of the standard methods used to characterize the properties of a specific material. This paper presents the outcomes of the first inter-laboratory testing initiative (i.e., a round-robin exercise) aimed at assessing the results variability of the physical characterization of a sandy soil. Guamo sand, widely utilized in local research and engineering projects in Colombia, was the focus of this study. 11 national academic laboratories participated in the program, conducting seven replicates of grain size distribution, solids specific gravity, and maximum and minimum void ratio tests. The data provided by all participants were analyzed and interpreted using statistical techniques. The results revealed significant differences between the data collected for each physical property, which can be attributed to the intrinsic variability of this sand's natural origin and to the use of diverse testing procedures. These comparisons offer valuable practical insights into the discrepancies between the testing methodologies employed by the participants for soil characterization, and they constitute a comprehensive database for future research or practical applications.

Keywords: round-robin testing, laboratory tests, standards and codes of practice, statistical analysis, sands

RESUMEN

En los ensayos de suelos, la evaluación de las propiedades físicas es esencial para caracterizar arenas con precisión. Sin embargo, los resultados de los ensayos pueden variar según los procedimientos experimentales utilizados y su implementación. Un ejercicio tipo *round-robin* facilita el análisis simultáneo de la reproducibilidad y la replicabilidad de los procedimientos estándar utilizados para caracterizar las propiedades de un material específico. Este artículo presenta los resultados de la primera iniciativa de ensayos inter-laboratorios (i.e., un ejercicio *round-robin*) cuyo objetivo fue evaluar la variabilidad de los resultados en la caracterización física de un suelo arenoso. La arena de Guamo, ampliamente utilizada en proyectos de investigación e ingeniería en Colombia, fue el foco de este estudio. 11 laboratorios académicos nacionales participaron en el programa, realizando siete réplicas de ensayos de distribución de tamaño de grano, gravedad específica de sólidos y de relación de vacíos máxima y mínima. Los datos proporcionados por todos los participantes fueron analizados e interpretados utilizando técnicas estadísticas. Los resultados revelaron diferencias significativas entre los datos recopilados para cada parámetro físico, atribuibles a la variabilidad intrínseca del origen natural de esta arena y a la utilización de diversos procedimientos de ensayo. Estas comparaciones ofrecen valiosas perspectivas prácticas sobre las discrepancias entre las metodologías de prueba empleadas por los participantes para la caracterización del suelo y constituyen una base de datos integral para futuras investigaciones o aplicaciones prácticas.

Palabras clave: ensayos *round-robin*, ensayos de laboratorio, estándares y normativa de ensayos, análisis estadístico, arenas

Received: March 14th, 2024

Accepted: September 14th, 2024

Introduction

Round-robin testing (RRT) programs involve collaborative testing conducted by multiple specialized laboratories. The objectives of RRT include assessing the reproducibility and replicability of a particular test or procedure, reviewing new experimental processes, and validating standards for certification or updating purposes in engineering practices. In addition, RRT programs allow assessing the variability of a testing method and improving the characterization of materials and products. Typically, these programs involve various entities, notably universities and research institutes that are specialized or recognized in specific disciplines. RRT execution is the primary strategy to evaluate and identify differences between procedures and results. In addition, the advantage of RRT lies in conducting measurements under the same experimental conditions to determine the properties controlling the behavior of the same type of material [1]. Thus, many collaborations focus on specific tests or parameters – physical, chemical, or mechanical – as shown by [2]. Furthermore, cooperative research and comparison may lead to developing or modifying classical experimental methods or proposing new test procedures [3].

RRTs are organized by one of the participating laboratories or groups responsible for coordinating activities and compiling results. Slight variations in any empirical processes often result in poor reproducibility estimates of collaborative studies [4]. Hence, if critical parameters are identified during RRT, pilot tests should be conducted with specialized setups within the same laboratory. These preliminary tests enable the examination of methods' susceptibility to minor changes in system conditions before formal collaboration [5]. Consequently, collaborative programs between research centers aimed at identifying the accuracy and precision levels of soil analytical methods are becoming increasingly common [6].

In some cases, a set of randomly selected samples is analyzed using an unsystematic sampling method to obtain statistically homogeneous data during RRT development. However, this technique is only applied to data meeting the assumption of variance homogeneity [7]. For soils, which are heterogeneous materials due to their geological formation process, efforts should be made to establish uniform test protocols to ensure consistent characterization procedures. Representative data are employed, and statistical methods

are used to determine significant differences between the test procedures or protocols evaluated in the RRT.

Accurate characterization of soil properties is crucial for both practical engineering applications and academic research. However, the inherent variability of natural soils, coupled with differences in experimental procedures and their implementation, often leads to discrepancies in test results. These inconsistencies can undermine the reliability of soil characterizations, particularly when the results are used as benchmarks for construction projects or research studies. In geotechnical engineering, several RRTs have evaluated the behavior of different soils using various procedures. These RRTs include assessing the liquefaction resistance of Japanese sands through cyclic triaxial tests [8]; the stress-strain behavior of Toyoura sand, Fujinomori remolded clay, and soft sedimentary rock [9]; the shear wave velocity of Toyoura sand, measured while employing bender elements [10]; the measurement and suction control of sand, kaolin, and bentonite mixtures [11]; the critical state line of Coimbra sand [12]; and the critical state line of a gold mine tailings by triaxial compression testing [13]. These RRTs have allowed drawing conclusions regarding the differences between the procedures used by participating laboratories. Additionally, they have contributed to identifying factors inducing bias in the tests and increasing variability between the measured properties in the studied soils.

Guamo sand is a reference sand in Colombia, locally akin to Toyoura and Ottawa sands, that is frequently used to calibrate constitutive models for geotechnical research [14], [15], [16]. This sand is sourced from alluvial deposits in the department of Tolima. Given its alluvial origin, the geotechnical properties of this sand are expected to be dependent on sampling location. However, the inherent variability and statistical metrics of Guamo sand's geotechnical properties have not yet been extensively studied. The objective of this paper is to quantify statistical measures related to the inherent variability of the geotechnical properties of Guamo sand.

As previously noted, there has been a notable absence of RRT programs specifically focused on assessing the physical properties of sands. Addressing this gap, the Geotechnical Research Group of Universidad Militar Nueva Granada (Colombia) initiated the first inter-laboratory testing program to evaluate the reproducibility and accuracy of measurements of key physical properties of a representative national sandy soil, *i.e.*, Guamo sand. This material, extensively used in Colombia for research and practical applications, was selected due to its importance. The physical properties of Guamo sand examined in this study include grain size distribution, specific gravity of solids, and soil packing, as indicated by the maximum and minimum void ratios.

This article presents the outcomes of this pioneering inter-laboratory testing program, conducted as a round-robin exercise, with the following objectives: (i) to outline the organization and execution of the testing program for

evaluating the reproducibility and accuracy of measurements of physical properties of sands; (ii) to compile and present the experimental results from the 11 participating laboratories; and (iii) to quantitatively analyze the differences between the datasets using statistical methods. The findings herein offer valuable insights into the discrepancies between different testing procedures. In addition, the comprehensive database established through this collaborative testing program can serve as a useful resource for validating the physical properties of Guamo sand in future research or practical applications. This initiative underscores the importance of collaborative efforts in advancing the understanding of soil behavior and refining testing methodologies to accurately characterize soil properties.

Materials and methods

Description of Guamo sand

Guamo sand originates in the Luisa River in the department of Tolima, Colombia, and it is characterized as a sandy soil of alluvial origin. This particular sand is extensively used in Colombia for practical and research applications due to its widespread commercial availability and consistent gradation [17], [18]. The practical applications of Guamo sand include its use in assessing soil density and unit weight through the sand-cone method, as outlined in ASTM D1556M-15e1 [19]. Moreover, researchers have used this soil to investigate mechanisms related to the instability of granular soils, as evidenced by [20].

Mineralogically, Guamo sand consists predominantly of quartz, accounting for approximately 99% of its composition [21], [22], with its particles' morphology predominantly characterized as sub-angular to sub-rounded [15], [23], [24]. The friction angle at the critical state (ϕ'_{cs}) of this soil has been reported by [16] and [25] to range between 31 and 34°. Table 1 summarizes the physical parameters documented in previous research studies on Guamo sand. These index properties include the coefficient of curvature (C_c), the coefficient of uniformity (C_u), the average particle size (d_{50}), the effective size (d_{10}), the specific gravity of solids (G_s), the maximum void ratio (e_{max}), and the minimum void ratio (e_{min}). The values presented in this table indicate variability among the reported physical parameters, emphasizing the importance of accurate characterization or establishing a reliable database that contains the properties of this soil.

Organization and methodology of the inter-laboratory testing program

The Geotechnical Research Group of Universidad Militar Nueva Granada (UMNG) launched a collaborative testing program that brought together specialized geotechnical characterization laboratories to estimate and compare the physical parameters of Guamo sand as reported by various institutions. This effort marks the first instance of this work by geotechnical specialists in Colombia. Invitations

Table I. Physical parameters of Guamo sand as reported in the literature

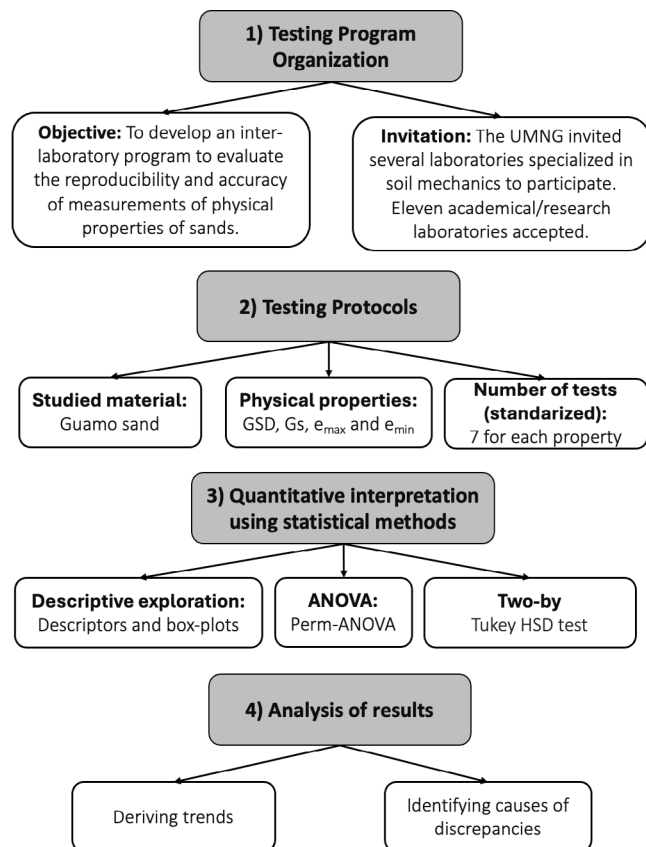
Reference	Cc	Cu	d ₅₀ (mm)	d ₁₀ (mm)	Gs	e _{max}	e _{min}
Arias (2006) [26]	-	2.04	0.51	0.11	-	1.00	0.52
Patiño (2006) [27]	-	0.65	0.56	0.16	2.63	0.83	0.50
Gómez (2010) [28]	-	-	0.43	0.07	2.66	1.00	0.52
Jiménez (2011) [25]	1.21	2.73	0.54	0.14	2.62	0.92	0.55
Camacho-Tauta <i>et al.</i> (2014) [29]	1.35	1.67	0.55	0.75	2.70	0.95	0.77
Bermúdez and Ruiz (2015) [23]	1.05	2.44	0.55	0.15	2.70	1.00	0.59
Tique (2016) [15]	0.63	3.05	0.54	0.24	2.64	0.97	0.60
Dulcey-Leal <i>et al.</i> (2018) [17]	0.69	1.76	0.49	0.19	2.70	0.90	0.63
Molina-Gómez <i>et al.</i> (2019) [21]	1.29	1.23	0.50	0.20	2.70	0.93	0.57
Ruge <i>et al.</i> (2019) [22]	0.93	1.34	0.43	0.23	2.70	0.90	0.58
Ramos-Cañon <i>et al.</i> (2022) [14]	2.44	1.05	0.53	0.21	2.71	1.02	0.59

Source: Authors

were extended to academic and commercial laboratories renowned for their expertise in soil mechanics and geotechnical engineering. While 11 academic laboratories from Colombian universities were accepted based on established criteria, unfortunately, no commercial laboratory responded positively to the invitation. All the data were compiled and statistically assessed by the RRT coordinators.

The coordinating group requested reporting seven replicates for each estimated parameter to assess the repeatability and variability of each test and reduce the probability of type I and II errors associated with random sampling of the replicates analyzed [4], [7]. However, no specific test method was set; on the contrary, it was indicated that each participant was free to choose the test method (standardized or non-standardized) that they deemed relevant to measure each of the aforementioned physical parameters. In addition, it was indicated that each laboratory was responsible for acquiring the material for the collaborative testing program. All laboratories, except Uniandes, UCC, and Unal-B (Table II), used material from the same supplier. For confidentiality and commercial reasons, the names of the material suppliers are not disclosed in this document and are represented by the numbers 1 and 2. Fig. 1 summarizes the methodology of the inter-laboratory testing program.

Table II presents the laboratories that participated in the Guamo sand characterization RRT program and the methods implemented by them to measure physical parameters. The physical characterization of Guamo sand involved an experimental program focused on measuring grain size distribution (GSD) curves (Fig. 1), Gs, e_{max}, and e_{min}. These parameters were selected because they influence the macro-mechanical behavior of sandy soils, as observed by [30], [31], [32], [33], [34], [35], [36], and [37].

**Figure 1.** Methodology of the interlaboratory testing program.

Source: Authors

Table II. Data on participating laboratories and summary of test methods

Laboratory/University (Acronym)	Location	Material supplier	GSD	Gs	$\frac{e_{\max}}{e_{\min}}$
Universidad Militar Nueva Granada (UMNG)	Bogotá DC	1	123-13 ^a	128-13 ^e	136-13 ^h
Pontificia Universidad Javeriana (PUJ-B)	Bogotá DC	1	123-13 ^a	128-13 ^e	136-13 ^h
Universidad de los Andes (Uniandes)	Bogotá DC	2	78-95 ^b	128-13 ^e	92-95 ⁱ
Pontificia Universidad Javeriana (PUJ-C)	Cali	1	D6913 ^c	D854-14 ^f	136-13 ^h
Universidad Distrital Francisco José de Caldas (UDFJC)	Bogotá DC	1	123-07 ^d	128-07 ^g	136-07 ^j
Universidad Industrial de Santander and Universidad del Magdalena	Bucaramanga/Santa Marta	1	123-13 ^a	128-13 ^e	136-13 ^h
Universidad Nacional de Colombia (Unal-B)	Bogotá DC	2	123-13 ^a	128-13 ^e	136-13 ^h
Universidad Católica de Colombia (UCC)	Bogotá DC	2	123-13 ^a	128-13 ^e	136-13 ^h
Universidad Piloto de Colombia (UPC)	Bogotá DC	1	123-13 ^a	128-13 ^e	136-13 ^h
Universidad Nacional de Colombia (Unal-Mz)	Manizales	1	123-13 ^a	128-13 ^e	136-13 ^h
Universidad Francisco de Paula Santander (UFPS)	Cúcuta	1	123-13 ^a	128-13 ^e	136-13 ^h

^a INVIAS (2013). Determinación de los tamaños de las partículas de los suelos INV E-123-13.

^b ICONTEC (1995). Método para determinar por lavado el material que pasa el tamiz 0.075 mm en agregados minerales NTC 78.

^c ASTM International (2021). Standard test methods for soil particle-size distribution (gradation) using sieve analysis ASTM D 6913.

^d INVIAS (2007). Determinación de los tamaños de las partículas de los suelos INV E-123-07.

^e INVIAS (2013). Determinación de la gravedad específica de las partículas sólidas de los suelos y del llenante mineral, empleando un picnómetro con agua INV E-128-13.

^f ASTM International (2014). Standard test methods for specific gravity of soil solids by water pycnometer ASTM D854.

^g INVIAS (2007). Determinación de la gravedad específica de las partículas sólidas de los suelos y del llenante mineral, empleando un picnómetro con agua INV E-123-07.

^h INVIAS (2013). Determinación de las masas unitarias máxima y mínima para el cálculo de la densidad relativa INV E-136-13.

ⁱ ICONTEC (1995). Determinación de la masa unitaria y los vacíos entre partículas de agregados NTC 92.

^j INVIAS (2007). Determinación de las masas unitarias máxima y mínima para el cálculo de la densidad relativa INV E-136-07.

Source: Authors

Results and discussion

Grain size distribution

Fig. 2 presents the GSD curves obtained during the collaborative testing program. A visual inspection of the figure indicates no significant differences between GSD curves, suggesting that the material supplier does not influence these outcomes. However, a visual analysis shows that the results provided by Unal-Mz and UIS-UM have the highest dispersion among all the datasets (Figs. 2j and 2f, respectively). In contrast, the GSD curves from Uniandes, Unal-B, and UMNG show the slightest variations, reflecting higher consistency in their testing procedures (Figs. 2c, 2g, and 2a). Further analysis of the differences in GSD regarding the coefficients of curvature (Cc) and uniformity (Cu) is presented below. Cc and Cu were calculated using Eqs. (1) and (2), respectively:

$$C_u = \frac{d_{60}}{d_{10}} \quad (1)$$

$$C_c = \frac{d_{30}^2}{d_{60} d_{10}} \quad (2)$$

where d_{10} , d_{30} , and d_{60} correspond to the particle sizes associated with 10, 30, and 60% of soil sample passing, respectively.

The results for Cc and Cu indicate that all Guamo sand samples can be classified as poor-graded sand (SP) according to the Unified Soil Classification System. This classification is based on the quantitative criterion that Cu should be less than 6, and Cc should not fall within the range of 1 to 3 [38].

Descriptive exploration

The statistical analysis of the RRT results included a descriptive exploration phase for each physical parameter, which was estimated using several observations ($n = 77$). The descriptors used were the mean value (\bar{X}), the standard deviation (s), the coefficient of variation (COV), the median (\hat{X}), the maximum value ($X_{(1)}$), the minimum value (X_n), the range of values ($X_n - X_{(1)}$), symmetry (\hat{k}_1), kurtosis (\hat{k}_2), and the standard error (ϵ). Table III presents the descriptive exploration results for Cc, Cu, Gs, e_{\max} , and e_{\min} , as reported in the collaborative testing program.

The COV was calculated separately to measure the dispersion variation of each property. The COV values in Table III could be utilized to prioritize the testing of each property and help to meticulously select geotechnical investigation scopes, allowing to perform further tests [39], [40]. These values reveal that the e_{\min} and Gs results exhibit the highest and lowest variability, respectively. The 22% variability in the e_{\min} results stems from the differences between the methods and procedures chosen for characterizing this physical parameter. The data highlight the need for improvements to

testing procedures to assess e_{\min} , where variability is more pronounced. In contrast, the low variability observed in Gs can be attributed to the stable mineralogical composition of Guamo sand, predominantly comprising silica minerals. This minimal variation suggests that, despite the diverse testing environments and techniques used, the inherent material properties of the sand remain consistently measurable, reinforcing the robustness of this parameter for geotechnical evaluations.

Furthermore, the low variation in the Gs results is due to the fact that 90% of the laboratories used the same standard procedure. The coefficients \hat{k}_1 and \hat{k}_2 indicate that the groups of the evaluated physical parameters follow asymmetric trends. For example, Cc, Cu, and Gs are rightward asymmetric due to positive \hat{k}_1 results, while e_{\max} and e_{\min} are leftward asymmetric due to negative \hat{k}_1 results. In addition, the results for \hat{k}_2 showed that the distributions of Cc, e_{\max} , and e_{\min} are leptokurtic, and that the distributions of Cu and Gs are platykurtic. The values of \hat{k}_1 and \hat{k}_2 suggest that, while some parameters exhibit deviations from normality, others indicate a broader spread in the data, which could impact the overall interpretation of the

results. These tendencies provide preliminary insights into the normality of the data (in general, normally distributed data are symmetric), which will be further discussed below.

The descriptive exploration of the results also involved a graphical comparison using boxplots. Fig. 3 presents the boxplots of the physical Guamo sand parameters assessed. In the data presented therein, significant differences between the results reported by the participants are observed. Moreover, the boxplots identify outlier differences (or *outliers*) in all datasets. The outliers vary between participants and intra-parameters, which precludes confirmation of whether the test procedures were correctly applied in the physical characterization of the Guamo sand, as observed in the statistical analysis of granular materials by [39]. These outliers may indicate potential inconsistencies in the experimental procedures – or variations in standard implementation. Hence, more appropriate statistical methods will be implemented in the following sections, in order to quantify the differences between the results and subsequently group them. This approach will help to ensure that the physical properties of Guamo sand are characterized more accurately and consistently.

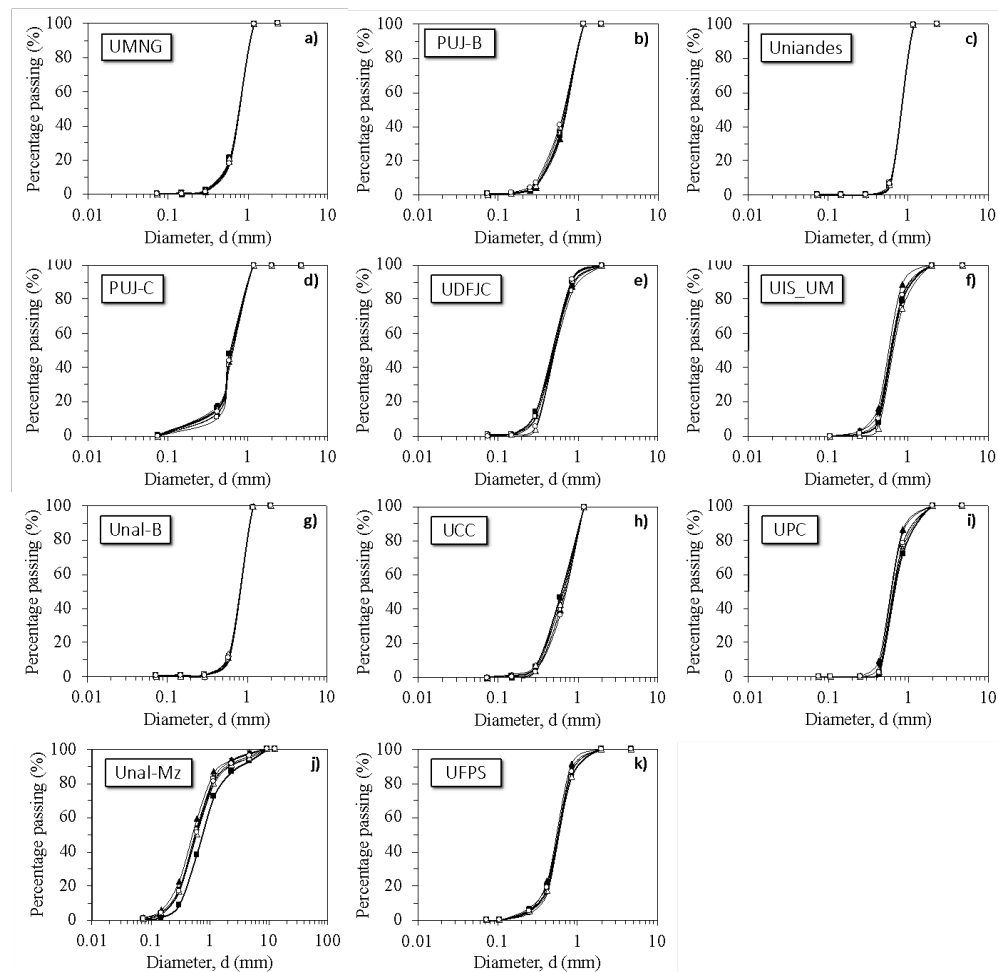


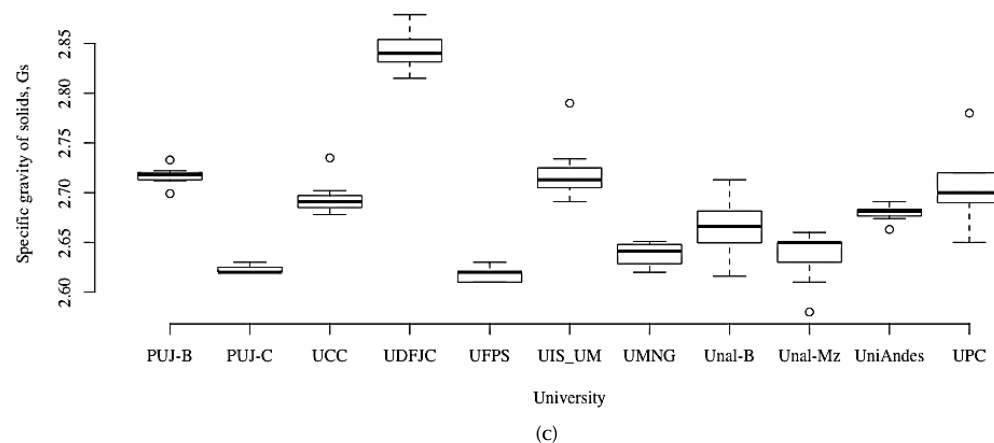
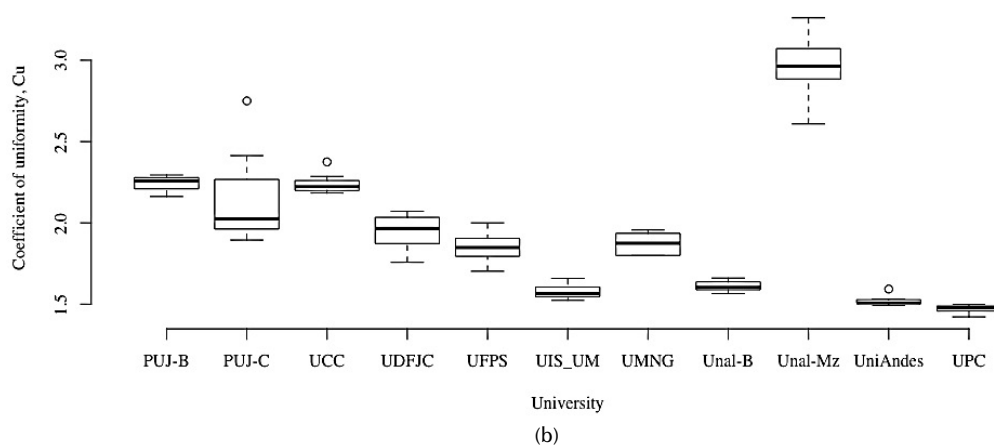
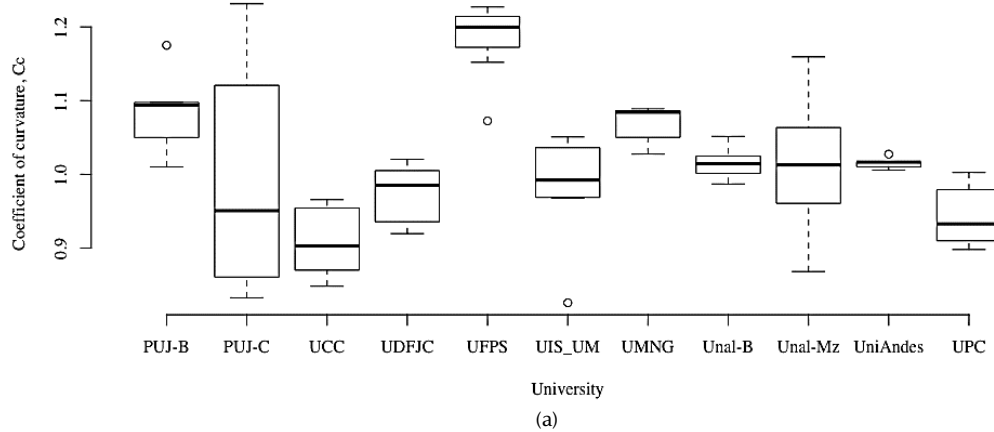
Figure 2. GSDs of Guamo sand as reported by the participants: a) UMNG, b) PUJ-B, c) Uniandes, d) PUJ-C, e) UDFJC, f) UIS_UM, g) Unal-B, h) UCC, i) UPC, j) Unal-Mz, and k) UFPS

Source: Authors

Table III. Descriptors of the physical parameters of Guamo sand

Variable	n	\bar{X}	s	COV	\hat{X}	$X_{(1)}$	X_n	$X_n - X_{(1)}$	\hat{k}_1	\hat{k}_2	ϵ
Cc	77	1.02	0.10	0.10	1.01	0.83	1.23	0.41	0.26	-0.26	0.01
Cu	77	1.95	0.34	0.17	1.88	1.42	3.26	1.84	1.09	0.67	0.05
Gs	77	2.69	0.07	0.03	2.68	2.58	2.88	0.3	1.12	0.89	0.01
e_{\max}	77	0.91	0.06	0.07	0.90	0.77	1.04	0.27	-0.28	-0.01	0.01
e_{\min}	77	0.63	0.14	0.22	0.66	0.35	0.81	0.46	-0.51	-1.09	0.02

Source: Authors



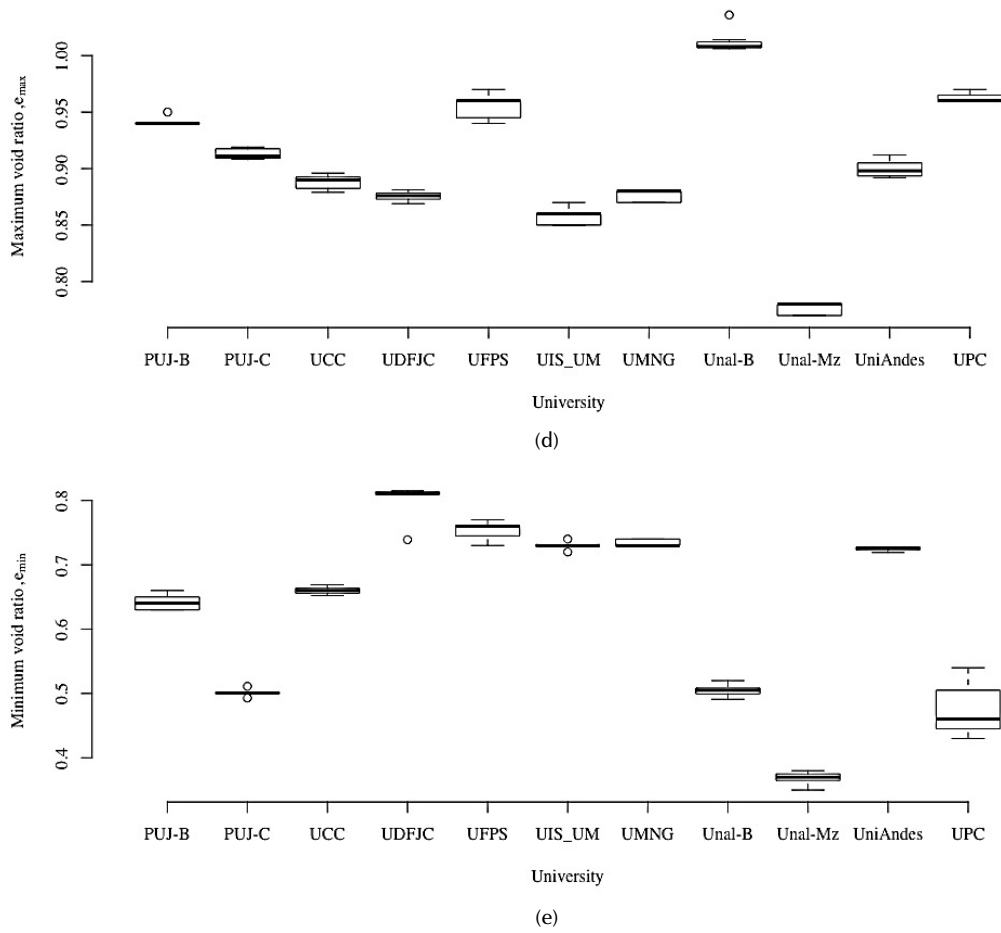


Figure 3. Boxplots for the values of a) the coefficient of curvature, b) the coefficient of uniformity, c) the solids specific gravity, d) the maximum void ratio, and e) the minimum void ratio.

Source: Authors

ANOVA analysis

All participants interpreted their data using RStudio [41], a freely available integrated tool for statistical analysis. A univariate analysis of variance (ANOVA) was employed to compare the results reported by the participants, aiming to assess the statistical equality between the population means of each set of variables (μ) while utilizing the hypothesis (H_0) described in Eq. (3).

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \dots = \mu_n \quad (3)$$

The alternative hypothesis (H_a) in Eq. (4) allowed determining whether there were different means in at least one pair of datasets (or results provided by the participants):

$$H_a : \mu_1 \neq \mu_2 \neq \mu_3 \neq \dots \neq \mu_n \quad (4)$$

In the collaborative testing program, a $g = 11$ group of data was established, with the same number of replicates or observations ($n = 7$), resulting in a total of $N = 77$ observations for each physical parameter. Based on this

configuration, it was assumed that the data for each physical parameter of Guamo sand could be statistically analyzed under the criterion of a balanced experiment, as established by Woodward's design experiments [42]. Eq. (5) describes the response of the ANOVA model when comparing the observed values ($i = 1, 2, 3, \dots, n$) for the datasets reported by each participant ($j = 1, 2, 3, \dots, g$) with regard to each physical parameter studied.

$$y_{ij} = \mu + \tau_i + \varepsilon_{ij} \quad (5)$$

where μ is the mean of all random observations, τ_i represents the group effect, and ε_{ij} corresponds to the model error or residuals. The validation of H_0 involved calculating the mean of each set of values reported by each participant ($\bar{y}_{i.}$) as well as the mean of the total of all values ($\bar{y}_{..}$). This was done by using Eqs. (6) and (7) [43]:

$$\bar{y}_{i.} = \frac{1}{n} \sum_{j=1}^n (y_{ij}) \quad (6)$$

$$\bar{y}_{..} = \frac{1}{N} \sum_{i=1}^g \sum_{j=1}^n (y_{ij}) \quad (7)$$

where $Y_{i.}$ represents the sum of the data for each dataset, and Y_{ij} is the value of each individual observation. In addition, using Eqs. (8) and (9), the ANOVA procedure performed herein considered the degrees of freedom for the results reported by the participants (df_p) and those of the model residuals (df_e).

$$df_p = (g - 1) = (11 - 1) \quad (8)$$

$$df_e = g(n - 1) = 11(7 - 1) \quad (9)$$

The sums of squares and mean squares of each dataset (SS_p and MS_p) were obtained using Eqs. (10) and (11), while those of the residuals (SS_e and MS_e) were calculated via Eqs. (12) and (13).

$$SS_p = n \sum_{i=1}^g (\bar{y}_{i.} - \bar{y}_{..})^2 \quad (10)$$

$$MS_p = \frac{SS_p}{N - 1} \quad (11)$$

$$SS_e = \sum_{i=1}^g \sum_{j=1}^n (y_{ij} - \bar{y}_{..})^2 - n \sum_{i=1}^g (\bar{y}_{i.} - \bar{y}_{..})^2 \quad (12)$$

$$MS_e = \frac{SS_e}{n - N} \quad (13)$$

Then, F_0 (representing the F-value of the F-distribution) was calculated to evaluate H_0 through Eq. (14).

$$F_0 = \frac{MS_p}{MS_e} \quad (14)$$

The last step of the ANOVA involved calculating the upper percentile of the distribution (F_{df_p, df_e}), which is equivalent to F_α . In this study, a significance level of 95% was assumed (i.e., $1 - \alpha = 0.95$) for all statistical analyses. Therefore, if $F_0 > F_\alpha$ or $p\text{-value} < 0.05$, H_0 was rejected, as reported by Hassan *et al.* [44]. Table IV presents the ANOVA results for the physical parameters of Guamo sand.

Table IV. ANOVA results

Data group	Degrees of freedom	Sum of squares	Mean square	F value	p-value
Cc	10	0.38	0.36	7.65	5.6×10^{-8}
Residuals of Cc	66	0.33	0.0049	–	–
Cu	10	13.46	171.53	80.61	2.0×10^{-16}
Residuals of Cu	66	1.1	0.0002	–	–

Data group	Degrees of freedom	Sum of squares	Mean square	F value	p-value
Gs	10	0.29	0.03	55.72	2.0×10^{-16}
Residuals of Gs	66	0.03	0.0002	–	–
e_{\max}	10	0.28	0.03	555	2.0×10^{-16}
Residuals of e_{\max}	66	0.01	0.0001	–	–
e_{\min}	10	1.40	0.14	474	2.0×10^{-16}
Residuals of e_{\min}	66	0.02	0.0002	–	–

Source: Authors

The findings in Table IV suggest notable differences between all data groups in relation to the physical parameters of Guamo sand. However, assessing and validating the assumptions of normality and variance homogeneity of the ANOVA residuals before affirming this conclusion is essential. This study evaluated said assumptions through the Shapiro-Wilk and Levene tests. The results are presented in Tables V and VI.

Table V. Normality test results

Data group	Shapiro-Wilk	p-value
Cc	0.94	0.0014
Cu	0.87	8.46×10^{-7}
Gs	0.92	0.0002
e_{\max}	0.96	0.0111
e_{\min}	0.68	4.62×10^{-7}

Source: Authors

Table VI. Results obtained in the variance homogeneity test

Data group	Degrees of freedom	F value	p-value
Cc	10	4.27	0.0001
Cu	10	2.65	0.0088
Gs	10	1.48	0.1674
e_{\max}	10	0.76	0.6654
e_{\min}	10	3.05	0.0031

Source: Authors

The Shapiro-Wilk and Levene tests revealed that not all datasets met the normality and variance homogeneity criteria. Therefore, in this study, a permutational analysis of variance (perm-ANOVA) was employed to validate the differences between the results obtained for the physical parameters. The perm-ANOVA is a non-parametric technique that does not require validating the normality and homogeneity of variance [45]. It addresses data randomization through permutation in order to rearrange the data N_p times under the H_0 hypothesis of an ANOVA [46]. In other words, the perm-ANOVA allows calculating the number of times that $F_0 > F_\alpha$ by randomly rearranging the data N_p times. This technique can be applied in experiments with relatively few observations [47]. [48] used a perm-ANOVA to compare parameters of gradation curves

in granular materials, demonstrating its applicability to soils. In this study, the perm-ANOVA was applied with $N_p = 5000$ – as recommended by [49] – using the *RVAideMemoire* package [50]. Table VII presents the results.

Table VII. Perm-ANOVA results

Data group	Degrees of freedom	Sum of squares	Mean square	F value	p-value
Cc	10	0.38	0.36	7.65	0.0019
Residuals of Cc	66	0.33	0.0001	–	–
Cu	10	13.46	171.53	80.61	0.0005
Residuals of Cu	66	1.11	0.0002	–	–
Gs	10	0.29	0.03	55.72	0.0009
Residuals of Gs	66	0.03	0.0002	–	–
e_{\max}	10	0.28	0.03	555	0.0042
Residuals of e_{\max}	66	0.01	0.0001	–	–
e_{\min}	10	1.40	0.14	474	0.0031
Residuals of e_{\min}	66	0.02	0.0002	–	–

Source: Authors

The results in Table VII confirm the significant differences between the results obtained by the participants for the physical properties of Guamo sand ($p\text{-value} < 0.05$). This provides clear statistical evidence that the mean values of the physical parameters differ significantly between participants. Such discrepancies highlight the need for improved standardization and uniformity in testing methodologies, in order to ensure more consistent and comparable results in future studies.

One limitation of ANOVA methods is their inability to pinpoint which treatments have distinct means (in this context, the treatments correspond to the participants) [43]. Thereupon, a multiple comparisons analysis was conducted to determine which means deviated from the others concerning each physical parameter. This analysis employed the HSD (honestly significant difference) test introduced by [51]. This test computes the overall mean of all the data and juxtaposes the confidence intervals of the data reported in pairs. Essentially, it compares the mean for each pair of participants. These intervals follow a skewed range distribution (q), as estimated through Eq. (15).

$$q = \frac{\bar{y}_{\max} - \bar{y}_{\min}}{S\sqrt{2/n}} \quad (15)$$

In this equation, $\bar{y}_{\max} - \bar{y}_{\min}$ is the difference between means, and S corresponds to the variance. Eq. (16) outlines the calculation of the confidence limits used in the Tukey HSD test for a significance level of $1 - \alpha$ (currently 95%).

$$\bar{y}_i - \bar{y}_j \pm \frac{q_{\alpha; k; N-k}}{\sqrt{2}} S \sqrt{\frac{2}{n}} \quad (16)$$

for $i, j = 1, \dots, k \quad i \neq j.$

Thus, comparing the confidence intervals allows identifying the participants whose reported values exhibit statistically significant differences. Fig. 4 presents the confidence intervals obtained from the Tukey HSD test for all physical parameters of Guamo sand. Interpreting these results involves approximating the confidence intervals to a common point (zero), making it easier to visually assess which participants' results deviate significantly from the others. When a confidence interval does not overlap with these zero points, it indicates a statistically significant difference from the mean values of the other participants. This visual approach is crucial for highlighting variations in testing methods or material handling across laboratories and underscores the need for standardization to minimize discrepancies, as detailed below.

The multiple comparisons performed using the confidence intervals from the Tukey HSD test indicate that the Cu and Gs datasets show the lowest differences among the RRT participants. It was observed that, for Cu, the results of UFPS differed significantly from those of the other participants. In contrast, for Gs, differences were noted in the results of Unal-Mz and UDFJC when compared to the other datasets. The low discrepancies between Cc and Cu are attributed to the intrinsic variability of the Guamo sand samples tested. The differences in the Gs results are acceptable and fall into the typical ranges of other literature-reported siliceous sands of alluvial origin, which generally vary from 2.6 to 2.7 [31], [33], [34]. However, the results reported for UDFJC fall outside said ranges, revealing that this participant may have applied the wrong testing procedure. This confirms that the Gs results reported by the other RRT participants match the typical values of different sands, indicating consistent characterization.

On the other hand, the e_{\max} and e_{\min} datasets reported the most significant divergence, particularly for e_{\min} . This effect is primarily attributed to the correct application of test procedures and the differences between the standardized methods for estimating these physical parameters. The likely sources of errors include variations in potential energy (controlled by the height of grain fall) and particle crushing when assessing e_{\max} and e_{\min} , respectively. Similar results were reported by [52] when comparing e_{\max} and e_{\min} values for different sands obtained through various test methods. These authors also highlighted the effect of loose sand compaction on e_{\max} results, as well as that of particle crushing on e_{\min} values. The high variability of the datasets may be due to the above-mentioned factors. However, the intrinsic variability of Guamo sand can be identified in the e_{\max} and e_{\min} results [30], [31], [37], [53], so the high variability can be attributed to a combination of procedural differences and

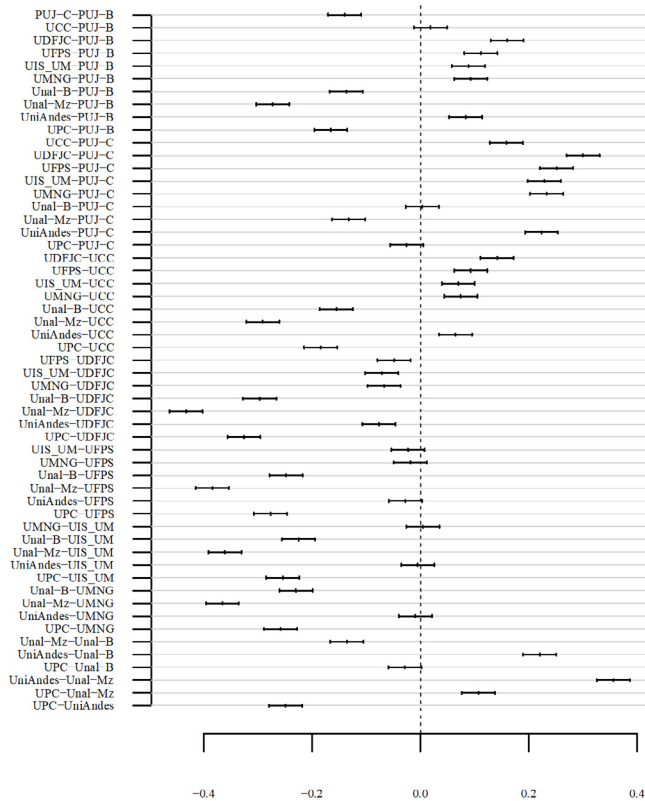


Figure 4. Confidence intervals from the Tukey HDS test for a) the coefficients of curvature, b) the coefficients of uniformity, c) the solids specific gravity, d) the maximum void ratio, and e) the minimum void ratio

Source: Authors

The results indicate that, when the same standardized procedure is consistently applied across different laboratories, the mean values of the measurements tend to align closely, as shown in Fig. 4d. This can also be demonstrated by the G_s assessment and results, which exhibited the lowest variability. This consistency suggests that the discrepancies observed in other parameters, such as e_{\max} and e_{\min} , are primarily due to variations in the testing methods rather than the intrinsic properties of the material. In light of this, the authors strongly recommend adopting a unified testing procedure across all laboratories when conducting physical characterizations of materials like Guamo sand. The likelihood of significant discrepancies can be minimized by standardizing methods, particularly in relation to parameters with higher sensitivity to procedural differences, such as e_{\max} and e_{\min} , leading to more reliable and comparable results across different studies and applications.

The discussion of the results is anchored in the limited references that are most relevant to this study. Although these references do not pertain directly to a joint round-robin exercise, they offer valuable analytical tools because the material under study remains the same. According to a simple descriptive analysis, all the research on Guamo sand shows an acceptable C_c range. Only one outlier is revealed in the study by [14], with a value of 2.44. As for C_u , all the

values obtained are similar, around a median of 1.76. In this study, C_u reported a value of 1.95 on average.

G_s is a very convenient parameter because it has a restricted range of values, i.e., generally between 2.55 and 2.85 in soils. Therefore, the possibility of finding outliers in the exploration is much higher. Almost all previous studies recorded a value of 2.70. In this RRT program, a value of 2.70 was reported. This parameter is the most reliable in performing a complete analysis of our results. According to the literature, the maximum and minimum void ratios also show reasonable and physically acceptable values. Only one of the values for e_{\min} (0.77) reported by [29] revealed itself as an outlier, as the median took a value of 0.59. In our RRT, the estimated value was 0.63.

It is important to note that this study quantifies the uncertainty in the properties of Guamo sand without separating its inherent and epistemic components. Epistemic uncertainty is implicitly represented in sampling frequency, while inherent uncertainty is related to the alluvial deposit's spatial variability [54]. The contribution of each component to the properties' overall uncertainty is outside the scope of this paper.

Regular inter-laboratory comparisons should be implemented as part of ongoing quality controls. These comparisons would enable the early detection of deviations from standardized procedures and provide opportunities for corrective actions before significant discrepancies occur. Moreover, laboratories should maintain detailed records of the procedures followed, including any deviations or adjustments made during testing. A centralized review process can then be implemented to analyze these records, ensuring that any variations are understood and addressed in subsequent modifications of standard testing procedures for the physical characterization of sands.

Summary and conclusions

The collaborative testing program organized by the Geotechnical Research Group of Universidad Militar Nueva Granada was considered successful. The participants were able to estimate the values of the variables involved using the abovementioned procedures, analyzing them through different statistical techniques.

From the experimental data obtained in this collaborative testing program, the following conclusions were drawn:

- The analysis of the physical properties of Guamo sand provided significant insights into its variability and characterization among the participants. The examination of COV values exhibited notable differences, with e_{\min} displaying the highest variability due to methodological disparities. Meanwhile, G_s exhibited minimal variability owing to a stable mineralogical composition, which was dominated by silica minerals.

- The perm-ANOVA results showed significant differences in the physical properties reported by the participants, prompting a multiple comparisons analysis through the Tukey HSD test. The study revealed that the Cu and Gs datasets exhibited the lowest participant discrepancies.
- As highlighted by previous studies, the e_{\max} and e_{\min} datasets exhibited the most significant divergence, which can be attributed to methodological differences and potential errors in test procedures.

Considering that reliability-based geotechnical analysis and design necessitate knowledge of the uncertainty regarding soil properties, the results presented in this paper provide important statistical information that will enable the use of uncertainty propagation methods during the constitutive modeling of Guamo sand.

The findings and conclusions of this collaborative testing program offer valuable insights for future research using Guamo sand and underscore the importance of standardized testing procedures and methodological consistency for geotechnical characterizations.

Acknowledgements

The Geotechnical Research Group of Universidad Militar Nueva Granada would like to thank all the researchers and research groups of the different universities involved in this study, who made this research work possible.

CRedit author statement

Authors 1, 2, and 3: writing (original draft, review, editing). *Authors 1 and 2:* conceptualization, data curation, methodology, supervision, visualization. *Authors 4, 6, 9, 10, and 20:* supervision, resources, investigation, writing (review and editing). *The rest of the authors:* resources, investigation.

Conflicts of interest

The authors declare no conflict of interest related to this research.

References



- [1] H. L. Anderson, A. Kemmler, G. W. H. Höhne, K. Heldt, and R. Strey, "Round robin test on the kinetic evaluation of a complex solid state reaction from 13 European laboratories. Part 1. Kinetic TG-analysis," *Thermochimica Acta*, vol. 332, no. 1, pp. 33-53, 1999. [https://doi.org/10.1016/S0040-6031\(99\)00045-3](https://doi.org/10.1016/S0040-6031(99)00045-3)
- [2] M. Thompson and R. Wood, "International harmonized protocol for proficiency testing of (chemical) analytical laboratories," *J. AOAC Int.*, vol. 76, no. 4, pp. 926-940, 2006. <http://dx.doi.org/10.1351/pac200678010145>
- [3] E. Hund, D. Massart, and J. Smeyers-Verbeke, "Inter-laboratory studies in analytical chemistry," *Analytica Chimica Acta*, vol. 423, no. 2, pp. 145-165, 2000. [https://doi.org/10.1016/S0003-2670\(00\)01115-6](https://doi.org/10.1016/S0003-2670(00)01115-6)
- [4] R. K. Misra, J. F. Uthe, and C. J. Musial, "Multivariate analysis of a round-robin study on the measurement of chlorobiphenyls in fish oil," *Analyst*, vol. 117, no. 7, pp. 1085-1091, 1992. <https://doi.org/10.1039/AN9921701085>
- [5] W. J. Youden and E. H. Steiner, *Statistical Manual of the Association of Official Analytical Chemists: Statistical Techniques for Collaborative Tests*. Rockville, MD, USA: The Association, 1975.
- [6] D. Hanson, J. Kotuby-Amacher, and R.O. Miller, "Soil analysis: Western states proficiency testing program for 1996," *Fresenius J. Anal. Chem.*, vol. 360, no. 3-4, pp. 348-350, 1998. <https://doi.org/10.1007/s002160050707>
- [7] M. Thompson and P. J. Lowthian, "Statistical aspects of proficiency testing in analytical laboratories Part 2. Testing for sufficient homogeneity," *Analyst*, vol. 121, no. 11, pp. 1593-1596, 1996. <https://doi.org/10.1039/AN9962101597>
- [8] S. Tori, F. Tatsuoka, S. Miura, Y. Yoshimi, S. Yasuda, and Y. Makihara, "Cyclic undrained triaxial strength of sand by a cooperative test program," *Soils Found.*, vol. 26, no. 3, pp. 117-128, 1986. https://doi.org/10.3208/sandf1972.26.3_117
- [9] S. Yamashita, Y. Kohata, T. Kawaguchi, and S. Shibuya, "International round-robin test organized by TC-29," in *Advanced Laboratory Stress-Strain Testing of Geomaterials*, 1st ed., R. Kuwano, Ed. London: Routledge, 2001, pp. 6-46
- [10] S. Yamashita, T. Kawaguchi, Y. Nakata, T. Mikami, T. Fujiwara, and S. Shibuya, "Interpretation of international parallel test on the measurement of Gmax using bender elements," *Soils Found.*, vol. 49, no. 4, pp. 631-650, 2009. <https://doi.org/10.3208/sandf.49.631>
- [11] A. Tarantino et al., "Benchmark of experimental techniques for measuring and controlling suction," *Géotechnique*, vol. 61, no. 4, pp. 303-312, 2011. <https://doi.org/10.1680/geot.2011.61.4.303>
- [12] J. A. Santos et al., "Coimbra sand-round robin tests to evaluate liquefaction resistance," in *15th World Conf. Earthquake Eng.*, 2012, pp. 24-28.
- [13] D. Reid et al. "Results of a critical state line testing round robin programme," *Géotechnique*, vol. 1, no. 15, pp. 373, 2020. <http://dx.doi.org/10.1680/jgeot.19.P.373>
- [14] A. Ramos-Cañón, L. F. Prada-Sarmiento, and J. Camacho-Tauta, "Effects of initial stress anisotropy on the onset of undrained instability for Guamo sand," *Rev. Fac. Ing. Univ. Antioquia*, vol. 109, pp. 69-78, 2022. <https://doi.org/10.17533/udea.redin.20221104>
- [15] D. O. Tique, "Estudio experimental de la inestabilidad difusa para la arena del Guamo Tolima," M.S. thesis, Pontif. Univ. Javeriana, Bogotá, 2014.
- [16] D. P. Solaque, "Comparación del ángulo de fricción crítico con el ángulo de reposo-análisis de la influencia de algunos factores en la determinación del ángulo de reposo," MS thesis, Univ. de Los Andes, Bogotá, 2008.
- [17] E. Dulcey-Leal, F. A. Molina-Gómez, and L. A. Bulla-Cruz, "Hydraulic conductivity in layered saturated soils asses-

- sed through a novel physical model", *DYNA*, vol. 85, no. 205, pp. 119-124, 2018. <http://doi.org/10.15446/dyna.v85n205.64473>
- [18] D. F. Gil, C. C. Mendoza, L. R. Vásquez-Varela, and S. Cano, "Physical model of shallow foundation under dynamic loads on sands," *Infrastructures*, vol. 7, no. 11, 2022, art. 147. <https://doi.org/10.3390/infrastructures7110147>
- [19] *Standard Test Method for Density and Unit Weight of Soil in Place by Sand-Cone Method*, ASTM D1556/D1556M-15, American Society for Testing and Material International, 2015.
- [20] O. Jiménez and A. Lizcano, "Liquefaction flow behavior of guamo sand," in *From Fundamentals to Applications in Geotechnics: Proc. of the XV Pan-American Conf. on Soil Mechanics and Geot. Eng.*, Buenos Aires, D. Manzanal and A. Sfriso, Eds., 2015, pp. 470-477.
- [21] F. Molina-Gómez, B. Caicedo, and A. Viana da Fonseca, "Physical modelling of soil liquefaction in a novel micro shaking table," *Geomech. Eng.*, vol. 19, no. 3, pp. 229-240, 2019. <https://doi.org/10.12989/gae.2019.19.3.229>
- [22] J. C. Ruge, F. Molina-Gómez, J. Bastidas, R. P. da Cunha, and I. Otálvaro, "Partially saturation effect on fine sands and earth pressure in a sheetpile wall," *Rev. Int. Métodos Numér. Cál. Diseño Ing.*, vol. 35, no. 4, art. 007, 2019. <http://10.23967/j.rimni.2019.09.007>
- [23] J. F. Bermúdez Cuervo and J. C. Ruiz Acero, "Estudio experimental de la línea de inestabilidad bajo condiciones anisotrópicas de carga no drenada monotónica," M.S. thesis, Pontif. Univ. Javeriana, Bogotá, 2015.
- [24] J. E. Carmona, "Análisis de la compacidad relativa en la construcción de muestras de arena usando técnicas experimentales de pluviación y modelamientos con el método de elementos discretos (DEM)," M.S. thesis, Univ. Dist. Francisco José de Caldas, Bogotá, 2019.
- [25] O. Jiménez, "Análisis del comportamiento de la licuación por flujo en la arena del guamo," M.S. Thesis, Univ. de los Andes, Bogotá, 2011.
- [26] P. A. Arias, "Modelo de comportamiento de suelos granulares-estudio y determinación de sus parámetros," M.S. thesis, Univ. de los Andes, Bogotá, 2006.
- [27] J. C. Patiño, "Parámetros hipoplásticos de la arena del Guamo-Colombia," M.S. thesis, Univ. de los Andes, Bogotá, 2006.
- [28] V. E. J. Gómez, "Cambios de transformación de fase y atractores en materiales granulares," Universidad de los Andes, Tech. Rep. MIC 2010 II 9, 2010
- [29] J. F. Camacho-Tauta, F. A. Molina Gómez, and O. J. Reyes Ortiz, "Preparación de especímenes de arena para ensayos triaxiales mediante un método controlado de compactación," *Rev. Científica Gral. José María Córdova*, vol. 12, no. 14, pp. 185-196, 2014.
- [30] M. Cubrinovski and K. Ishihara, "Maximum and minimum void ratio characteristics of sands," *Soils Found.*, vol. 42, no. 6, pp. 65-78, 2002. https://doi.org/10.3208/sandf.42.6_65
- [31] G.C. Cho, J. Dodds, and J.C. Santamarina, "Particle shape effects on packing density, stiffness, and strength: Natural and crushed sands," *J. Geotech. Geoenvironmental Eng.*, vol. 132, no. 5, pp. 591-602, 2006. [https://doi.org/10.1061/\(ASCE\)1090-0241\(2006\)132:5\(591\)](https://doi.org/10.1061/(ASCE)1090-0241(2006)132:5(591))
- [32] L. M. Wei and J. Yang, "On the role of grain shape in static liquefaction of sand-fines mixtures," *Géotechnique*, vol. 4, no. 9, pp. 740-745, 2014. <http://dx.doi.org/10.1680/geot.14.T.013>
- [33] F. N. Altuhafi, M. R. Coop, and V. N. Georgiannou, "Effect of particle shape on the mechanical behavior of natural sands," *J. Geot. Geoenvironmental Eng.*, vol. 142, no. 12, art. 3, 2016. [https://doi.org/10.1061/\(ASCE\)GT.1943-5606.0001569](https://doi.org/10.1061/(ASCE)GT.1943-5606.0001569)
- [34] C. Ramos, C. Ferreira, F. Molina-Gómez, and A. Viana da Fonseca, "Critical state lines of portuguese liquefiable sands," *E3S Web Conf.*, vol. 92, art. 06003, 2019. https://www.e3s-conferences.org/articles/e3sconf/pdf/2019/18/e3sconf_iscg2019_06003.pdf
- [35] D. Sarkar, M. Goudarzy, and D. König, "An interpretation of the influence of particle shape on the mechanical behavior of granular material," *Gran. Matter*, vol. 21, no. 3, art. 53, 2019. <https://link.springer.com/article/10.1007/s10035-019-0909-3>
- [36] Y. Xiao, L. Long, T. M. Evans, H. Zhou, H. Liu, and A. W. Stuedlein, "Effect of particle shape on stress-dilatancy responses of medium-dense sands," *J. Geot. Geoenvironmental Eng.*, vol. 145, no. 2, art. 3, 2019. [https://doi.org/10.1061/\(ASCE\)GT.1943-5606.0001994](https://doi.org/10.1061/(ASCE)GT.1943-5606.0001994)
- [37] F. Molina-Gómez and A. Viana da Fonseca, "Key geomechanical properties of the historically liquefiable TP-Lisbon sand," *Soils Found.* vol. 61, no. 3, pp. 836-56, 2021. <https://doi.org/10.1016/j.sandf.2021.03.004>
- [38] *Standard Practice for Classification of Soils for Engineering Purposes (Unified Soil Classification System)*, ASTM D2487-06, American Society for Testing and Materials International, 2006.
- [39] F. Molina-Gómez, L. A. Bulla-Cruz, and A. E. Darghan Contreras, "Profiles analysis as a modality of repeated measures for comparing grain size distributions in granular bases," *Measurement J. Int. Measur. Confed.*, vol. 146, pp. 930-937, 2019. https://ui.adsabs.harvard.edu/link_gateway/2019Meas..146..930M/doi:10.1016/j.measurement.2019.07.004
- [40] W. Hassan *et al.*, "Statistical interpolation and spatial mapping of geotechnical soil parameters of District Sargodha, Pakistan," *Bull. Eng. Geol. Environ.*, vol. 82, art. 37, 2023. <https://doi.org/10.1007/s10064-022-03059-2>
- [41] RStudio Team, "RStudio: Integrated Development for R." [Online]. Available: <http://www.rstudio.com/>
- [42] J. A. Woodward, D. G. Bonett, and M. L. Brecht, *Introduction to linear models and experimental design*. San Diego, CA, USA: Academic Press, 1990.
- [43] D. C. Montgomery, *Design and analysis of experiments*, 8th ed. Hoboken, NJ, USA: Wiley, 2013.
- [44] W. Hassan *et al.*, "Geospatial and statistical interpolation of geotechnical data for modeling zonation maps of Islamabad, Pakistan," *Environ. Earth Sci.*, vol. 81, art. 547, 2022. <https://doi.org/10.1007/s12665-022-10669-2>
- [45] P. Good, *Permutation, parametric, and bootstrap tests of hypotheses*, 3rd ed. New York, NY, USA: Springer, 2005.

- [46] A. M. Winkler, G. R. Ridgway, M. A. Webster, S. M. Smith, and T. E. Nichols, "Permutation inference for the general linear model," *NeuroImage*, vol. 92, pp. 381-397, 2014. <https://doi.org/10.1016/j.neuroimage.2014.01.060>
- [47] E. S. Edgington and P. Onghena, *Randomization tests*, 4th ed. Boca Raton, FL, USA: Chapman & Hall, 2007.
- [48] F. Molina-Gómez, L. A. Bulla-Cruza, and A. E. Darghan Contreras, "A novel approach for the control of grain size distributions based on variance analysis," *Constr. Building Mat.*, vol. 285, 2021, art. 122748. <http://dx.doi.org/10.1111/ejss.13354>
- [49] M. J. Anderson, "Permutation tests for univariate or multivariate analysis of variance and regression," *Can. J. Fisheries Aquatic Sci.*, vol. 58, pp. 626-639, 2001. <https://doi.org/10.1139/f01-004>
- [50] M. Hervé, "RVAideMemoire." [Online]. Available: <https://cran.r-project.org/web/packages/RVAideMemoire/RVAideMemoire.pdf>, 2020
- [51] J. Tukey, "Comparing individual means in the analysis of variance," *Biometrics*, vol. 5, no. 2, pp. 99-114, 1949. <https://doi.org/10.2307/3001913>
- [52] T. Lunne et al. "Methods used to determine maximum and minimum dry unit weights of sand: Is there a need for a new standard?" *Can. Geotech J.*, vol. 56, no. 4, pp. 536-553, 2019. <https://doi.org/10.1139/cgj-2017-0738>
- [53] C. S. Chang, J. Y. Wang, and L. Ge, "Maximum and minimum void ratios for sand-silt mixtures", *Eng. Geol.*, vol. 211, pp. 7-18, 2016. <https://doi.org/10.1016/j.enggeo.2016.06.022>
- [54] G. Baecher and J. Christian, *Reliability and statistics in geotechnical engineering*. Hoboken, NJ: Wiley, 2003.

Statistical Study on Domestic Gas Boiler Failures Using Various Software Platforms

Estudio estadístico de fallos en calderas de gas domésticas utilizando diversas plataformas de *software*

Pavel Shcherban ¹ and Reda Abu-Khamdi ²

ABSTRACT

Modern public utilities require a high level of reliability, especially with regard to systems used for heating and hot water supply, such as gas boilers. This paper collected and studied statistical information on the failures of gas boilers of four brands. The data were provided by three service companies that repair gas equipment in Kaliningrad, in the Russian Federation. The specifics of these failures were studied, determining their possible causes, and they were stratified by severity. Service companies typically have to use existing platforms (supplementing them with add-ons) or develop their own software to analyze failure statistics. In this regard, they are interested in the emergence of simple and effective tools for monitoring the quality of gas boiler maintenance and repair work. In this study, we used both the publicly available Scikit-learn library of the Jupyter Notebook environment and a custom program to perform data clustering. The main goal was to conduct a comparative assessment of the reliability of gas boilers of various brands based on the analysis of their failure statistics, as well as to develop a software product that enables such an assessment.

Keywords: gas equipment failures, quality management, equipment reliability, cluster analysis, statistical data processing software

RESUMEN

Las empresas de servicios públicos modernas requieren un alto nivel de fiabilidad, especialmente en lo que respecta a los sistemas utilizados para calefacción y suministro de agua caliente, como las calderas de gas. Este artículo recopiló y estudió información estadística sobre las fallas de calderas de gas de cuatro marcas. Los datos fueron proporcionados por tres empresas de servicio que reparan equipos de gas en Kaliningrado, en la Federación de Rusia. Se analizaron las características específicas de estas fallas, determinando sus posibles causas, y se clasificaron según su gravedad. Las empresas de servicio generalmente deben utilizar plataformas existentes (complementándolas con extensiones) o desarrollar su propio *software* para analizar estadísticas de fallos. En este sentido, están interesadas en el desarrollo de herramientas simples y efectivas para monitorear la calidad del mantenimiento y las reparaciones de calderas de gas. En este estudio, se utilizó tanto la biblioteca de código abierto Scikit-learn del entorno Jupyter Notebook como un programa personalizado para realizar la agrupación de datos. El objetivo principal fue realizar una evaluación comparativa de la fiabilidad de las calderas de gas de diferentes marcas en función del análisis de sus estadísticas de fallos, así como desarrollar un producto de *software* que permita dicha evaluación.

Palabras clave: fallos en equipos de gas, gestión de calidad, fiabilidad de equipos, análisis de conglomerados, software de procesamiento de datos estadísticos

Received: November 5th, 2023

Accepted: October 25th, 2024

Introduction

The reliability of gas equipment is achieved through service and repair work throughout its lifecycle. This is particularly vital for household appliances such as gas boilers. During their operation, various failures may occur due to the violation of installation and usage rules, the quality of gas and water, the stability of the connection to the electrical network, the schedule of diagnostic and service tasks, and the use of non-original spare parts.

In large enterprises of the energy complex, specialized structural divisions address the diagnostics and maintenance of gas boilers. As a rule, they have laboratories and diagnostic resources at their disposal, as well as a significant database of potential failures associated with the equipment

in operation. This allows them to maintain the reliability of gas energy equipment at a high level.

Small companies operating in the market for servicing and repairing domestic gas equipment usually have scattered data on the statistics of gas boiler failures. This is due to the large variety of boiler brands and models, the lack of a structured data storage system, staff turnover, and users

¹ Immanuel Kant Baltic Federal University, the Branch scientific cluster, Institute of High Technologies, Kaliningrad, 236009, Russia. Email: ursa-maior@yandex.ru

² Moscow Institute of Physics and Technology. Faculty of Applied Mathematics and Informatics, Moscow, 117303, st. Kerchenskaya, 1 A, bldg. 1. Email: rabouhamdi@gmail.com



Attribution 4.0 International (CC BY 4.0) Share - Adapt

turning to various service companies when new failures occur (which makes tracking repair history impossible). One significant problem in this regard is the lack of specialized systems for analyzing the statistics of gas power equipment failures.

In small service companies, this often leads to the irrational planning of spare parts purchases, the impossibility of planning the work volume of repair teams, and difficulties in forecasting maintenance activities for gas boilers of various brands (which should be corrected according to local usage conditions, rather than simply copying the provisions specified by the manufacturer). In general, these problems reduce companies' competitiveness in the technical services market.

A number of studies, such as [1] and [2], have already raised these issues and proposed various ways to solve them. In particular, an idea was put forward to restructure the gas boiler maintenance system, introduce passports for assessing technical conditions and repair work, create a common database of failures, develop mobile applications for contacting service companies, and install analytical units on gas boilers in order to track the characteristics of its operation (with the ability to subsequently extract the data). The proposed measures have their strengths and weaknesses. In some cases, they are organizationally complex, costly, and the promotion of such services in the market can be problematic. In solving these issues, the standardization of software and hardware and the establishment of uniform regulatory requirements are important. These proposals assume greater control over the technical conditions of the gas boilers used in everyday life, and, as a result, they increase the quality of technical maintenance and reliability.

It should be noted that the possible ways to improve companies' quality of service are not limited by the aforementioned measures; an alternative is the creation of a software product intended for the analysis of gas boiler failures, hence the necessity of developing a software package that enables the preparation of statistical data and the comprehensive analysis of the causes of failures.

Thereupon, two goals were set in this study: firstly, to develop a prototype software product for analyzing statistical data on the operation of gas boilers, using the cluster analysis method with stratification by failure severity; and secondly, using this program to form the main statistical groups of reliability for each type of boiler (less reliable, more reliable, and standard). These three equipment groups are generally consistent with the theory of reliability. A number of parameters for said groups (distribution density, and frequency) should indicate greater or lesser reliability in boilers of a particular brand.

To test the software product, statistical data on gas boiler failures, as recorded by three service companies in Kaliningrad in 2022, were used. The study also included a comparison of the results of data processing using our own

developed program and the existing scikit-learn library within the Jupyter Notebook interactive computing environment.

This study comprises several sections. The first section is analytical and outlines the collection of data on gas boiler failures as well as a description of their causes. Then, the methodological section presents the principles through which the data are ranked by equipment failure severity using a risk matrix. This section also describes primary data normalization according to the established ranking rule. The software section is based on the outlined methodological apparatus. The resulting array was processed using publicly available software and our own program. Both of these tools were developed to provide a clustering of failures by severity for each gas boiler brand. Finally, the results obtained with various software programs are compared against the data presented in scientific periodicals, and, based on them, the reliability issues of gas boilers are examined, possible reasons for the obtained data are discussed, and directions for further research are defined.

Failures of double-circuit boilers – grouping and pre-processing of statistical data

When thoroughly assessing the technical conditions of equipment, as a rule, one must first determine the prediction strategy and prognostic background and then develop a system of parameters reflecting the nature and structure of the technical object. Then, one must develop normative and search models for the equipment, perform simulations, and evaluate the accuracy of the models (*i.e.*, the rate of equipment failure, the correspondence between the initial and final parameters). The models must be verified using statistics, experimental data, or expert methods. Afterwards, recommendations should be made to optimize decision making with regard to the planning and management of equipment operation based on the obtained models.

However, this method is acceptable either for equipment developers (who provide a warranty for trouble-free operation) or for large service centers responsible for high-tech industrial devices. Small and medium-sized services, which account for the main share of the semi-industrial and domestic equipment repair market, should rather use the most informative parameters that allow assessing technical conditions in a comprehensive and rapid manner, aiming for the operational processing of failure statistics. For the case of double-circuit gas boilers, we selected data on the time until failure of boiler equipment, the efficiency factor of double-circuit boilers, and the severity of the failure that preceded the repair [3]. In general, any set of technical condition parameters can be used for analysis, but the highest informativeness will be obtained if the processed data are complete (*i.e.*, they fully reflect the results of equipment failures), reliable (*i.e.*, they are experimentally verifiable), accurate, holistic, interrelated, measurable, and thus able to reflect the different technical and technological characteristics of the object over time [4].

The data used in this study (mean time until failure, efficiency, and severity of the failure that preceded the repair) were processed and prepared step by step. The sample was formed from four groups of double-circuit gas boilers (Ariston, Buderus, Bosch, and Wisseman), with 25 boilers in each group, which were examined after failure and repair. The first parameter considered was the MTBF, a technical parameter that characterizes reliability of a repaired device, equipment, or technical system [5]. The reliability function can be defined using non-failure operation probability, whose formula is as follows:

$$P(t) = \exp\left(-\frac{t}{T_o}\right) \quad (1)$$

where:

- t – operating time of the machine
- T_o – mean time until failure

The MTBF of a piece of equipment can also be estimated through the failure rate formula [6]. The failure rate (2) is a parameter that determines the reliability of a piece of equipment:

$$\lambda = \frac{1}{MTBF} \quad (2)$$

where:

- λ – failure rate
- MTBF – mean time between failures

However, in general, the MTBF is determined statistically over time (3):

$$T = \frac{\sum_{i=1}^m t_i}{m} \quad (3)$$

where:

- t_i – mean time until failure i
- m – number of failures

It should be noted that this study deals with a MTBF measured from the moment of a machine's first start-up to the first breakdown before repair [7]. Thus, this parameter indicates the non-failure operation time of the new equipment purchased and used by the consumer. This information is extremely important for the consumer to plan their expenses, as well as for the service organization to schedule incoming orders depending on the previous sales volume of certain equipment [8].

We used the efficiency of boilers before failure as the second parameter. In this case, this is an effective analytical parameter because, on the one hand, from the manufacturer and the service company's perspective, it indicates the technical system's dynamics of degradation over time. On the other hand, from the consumer's perspective, it shows the efficiency of the equipment's operation before failure. Boiler efficiency mainly corresponds to the ratio of fuel consumed to heat emitted [9]. This parameter can be calculated through several methods, the first of which is direct:

$$\eta_k = \frac{Q_k}{Q_H^p} \quad (4)$$

where:

η_k – boiler efficiency

Q_k – useful energy transferred to the coolant

Q_H^p – thermal energy released as a result of the chemical reaction of combustion

There is also the reverse method [10]:

$$Q_H^p = Q_1 + Q_2 + Q_3 + Q_4 + Q_5 \quad (5)$$

where:

- Q_H^p – thermal energy released as a result of the combustion reaction
- $Q_1 = D(h_s - h_{fw}) / B$ – heat used for steam generation
 - D – boiler steam capacity (kg/s)
 - B – fuel flow rate per second (kg/s or m³/s)
 - h_s and h_{fw} – enthalpy of the steam and feed water (kJ/kg)
- Q_2 heat loss caused by the exhaust gases in the boiler unit
- Q_3 – chemical heat loss (underburning) caused by the incomplete combustion of the fuel
- Q_4 – mechanical heat loss (underburning) caused by incomplete combustion
- Q_5 – heat loss to the environment through the boiler's external enclosures

The third parameter analyzed was failure severity, based on the complexity of repair work and the frequency of occurrence. In general, failure severity can also be regarded

as the total criticality of the j -th failure of the i -th piece of equipment [11]. It can be calculated as:

$$\sum C m_{ij}^k = \beta_{ij}^k \cdot \alpha_{ij} \cdot \lambda_i \cdot (T_{work})_i \quad (6)$$

where:

- $C m_{ij}^k$ – criticality of the j -th type of failure of the i -th element
- β_{ij}^k – the probability of consequences of a certain category of severity for the j -th type of failure of the i -th element
- α_{ij} – share of the j -th type of failure of the i -th element
- λ_i – failure rate of the i -th element
- $(T_{work})_i$ – mean time until failure of the i -th element

Visually, this parameter can also be represented by a risk matrix (Fig. 1).

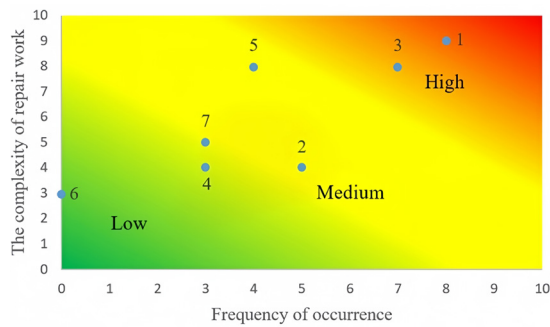


Figure 1. Severity of boiler equipment failure

Source: Authors

This figure includes the following items:

1. Corrosion of gas burners
2. Removal of fouling products
3. Switch failure
4. Breakage of the temperature sensor
5. Gas combustion noise
6. Inoperative burners
7. Inoperative traction sensor

We prepared the initial data on all four types of double-circuit gas boilers according to all the given parameters. Moreover, we identified the relationship between these parameters (which was obvious from a technical and technological perspective), but we did not know its extent or how evident it would be in the groups [12]. In general analytics, it is better to employ cluster analysis, noting that clustering can be performed in different ways. However, before clustering, it was necessary to normalize the initial data, given that, in their original form, which exhibited a high degree of numerical variation, the results were more difficult to formalize and visualize.

Data normalization for subsequent cluster analysis

Normalization is the process of converting data into a single, dimensionless range of values. Depending on the purpose of this process, there are various functions that match the original data of the row with their normalized value. One of these is the scaling function, which maps a value from the range $[\min(X), \max(X)]$ (where $X = x_1, x_2, \dots, x_n$, representing a row of experimental data) to a value from a given range $[a, b]$. If $[a, b] = [0, 1]$, there are several options for the scaling function, e.g., min-max normalization, z-estimation, and scaling relative to a unit vector. Min-max normalization, the simplest of the scaling functions, has the following form:

$f(x_i) = \frac{x_i - \min(X)}{\max(X) - \min(X)}$. Z-estimation allows transforming a data row such that its mean value becomes 0 and its variance becomes 1. Scaling relative to a unit vector implies representing a row of n data as an n -dimensional vector, after which the formula for normalizing the vector ($x' = \frac{x}{|x|}$) is applied based on the selected norm function.

In this study, the purpose of normalization was to bring three-dimensional data, which had various units of measurement and differed in absolute value by orders of magnitude, to relative dimensionless values. To this effect, the simplest and most effective method was a scaling function, i.e., min-max scaling [13]. Accordingly, each component of the three-dimensional data $A = \{(x_i, y_i, z_i), i = 1..n\}$ was represented as a separate data row ($X = \{x_i\}$, $Y = \{y_i\}$, $Z = \{z_i\}$), applying min-max normalization to each of them. As a result, the normalized data rows $X' = \{f(x_i)\}$, $Y' = \{f(y_i)\}$, and $Z' = \{f(z_i)\}$ were obtained. Afterwards, these rows were gathered into a single set, restoring the original three-dimensional row while including the normalized values $A' = \{(f(x_i), f(y_i), f(z_i)), i = 1..n\}$. The dimensionlessness achieved was a result of data normalization using the min-max scaling formula and the division of two values with the same dimensions. This facilitated the analysis, as well as the graphical representation of the data's dependence on the values of other rows, especially when the values differed by orders of magnitude in the original dimensions. This was necessary because the dimensionality of the analyzed data complicated their graphical representation. In general, this procedure can be applied to any kind of data used in cluster analysis.

Comparison of cluster analysis based on the Scikit-learn library with our own custom implementation

There are different types of cluster analysis: the probabilistic approach, which deals with the relationship between each studied object and one of the cluster classes; the logical approach, which builds dendrograms using decision trees; and the hierarchical approach, which requires different types of clusters combined into a single one. In this case, we applied the K-means method [14], as it allows distributing

events into groups both qualitatively and quantitatively. The idea of the method is to reduce the cluster points' sum of quadratic deviations from the clustering centers [15].

As a rule, K-means method involves setting an initial number of clusters [16]. In the case of this study, it can be logically assumed that three main groups of clusters will be formed:

- *Unreliable devices.* These had a high failure rate or exhibited low efficiency and MTBF values, i.e., the equipment failed before its warranty period ended or underperformed before failing.
- *Devices with normative reliability.* For these devices, all the parameters related to failure or performance loss behaved as expected within the warranty period, which, as a rule, entailed no severe consequences.
- *Devices exhibiting a high degree of reliability.* These devices showed low failure severity while remaining in operation longer than specified in the warranty and without significant loss of performance.

For a more accurate analysis, we used two approaches for clustering: the publicly available Scikit-learn package of the Jupyter Notebook environment and our own custom program. We then compared the accuracy of clustering on the same raw data, which were previously normalized. Jupyter Notebook is a development environment that immediately displays the results when executing code [17]. The advantage of this environment is that the code can be segmented and executed in any order. The Scikit-learn library is a Python library that includes clustering and regression analysis methods.

In our custom program, also written in Python, the initial parameters of the clustering process can be set more accurately. The steps of this process are presented in Fig. 2 in the form of flowcharts for both methods.

The first flowchart describes the clustering algorithm of the Scikit-learn library. After initialization, the data are entered into the first block, after which they are delivered to the *k_means* function. This function performs clustering according to the K-Means algorithm and writes the result

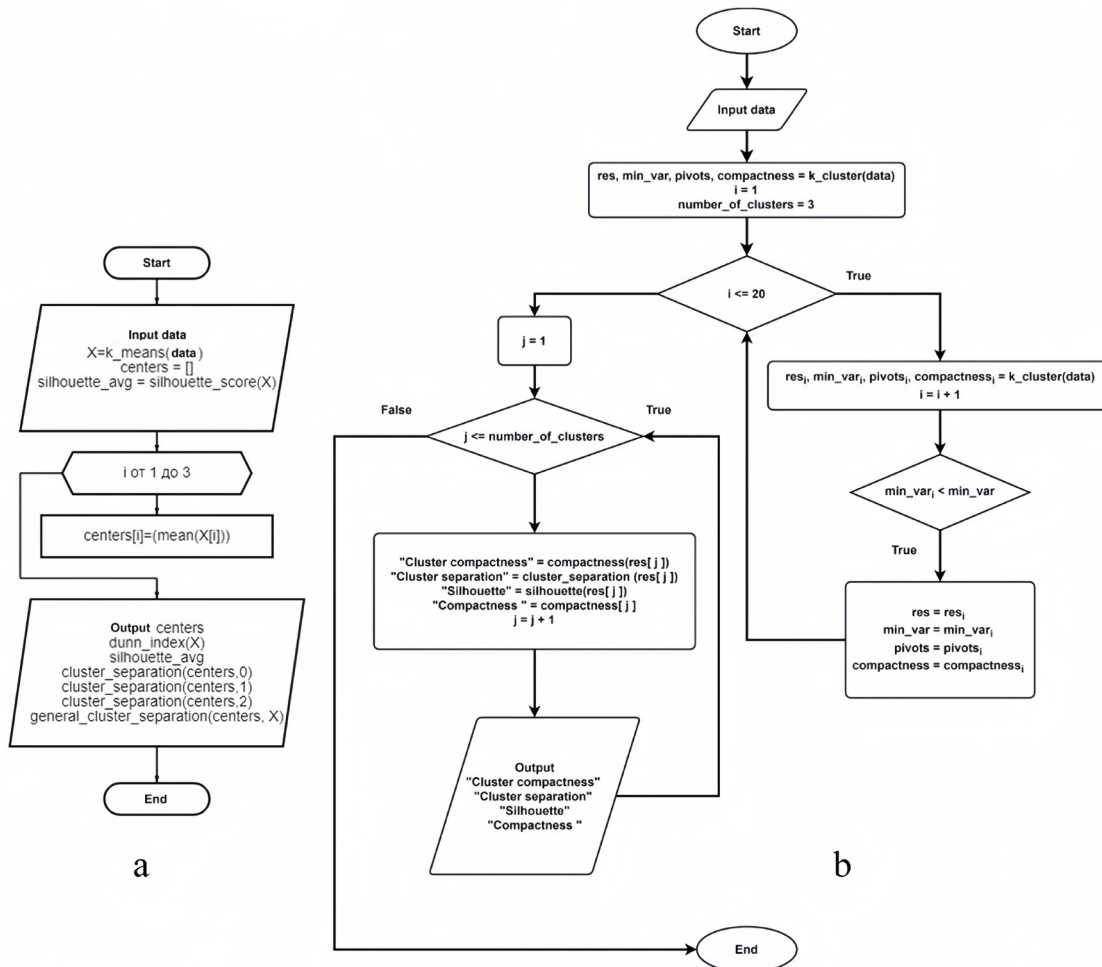


Figure 2. Clustering using the k-means method via a) the Scikit-learn library of the Jupyter Notebook environment and b) our custom program for cluster analysis

Source: Authors

into the X variable, after which an empty array of centers is created, with the purpose of storing the coordinates of the cluster centers. The *silhouette_score* function calculates the average clustering silhouette and writes the result into the *silhouette_avg* variable. Next, within the loop, the centers are calculated as the average point in the cluster using the mean function, after which they are recorded in the array. In the last block, before the end, the results are obtained, namely the cluster centers, the Dunn index, the silhouette, the separability of each individual cluster, and the total separability.

The second flowchart describes the algorithm of our program. The first block after initialization is data entry. The clustering algorithm is executed using the *k_means* function, which accepts the data as input and returns the following values:

- res_i : an array containing the three clusters
- min_var_i : the sum of all clusters' compactness scores
- $pivots_i$: an array containing the central point of each cluster
- $compactness_i$: an array containing the compactness score for each cluster

The process was repeated 20 times, selecting the clustering with the lowest sum of quadratic deviations. Afterwards, the algorithm estimated the following parameters for each cluster: compactness, separability, silhouette, and the Dunn index. Based on the two presented algorithms for the four groups of normalized data, we conducted a

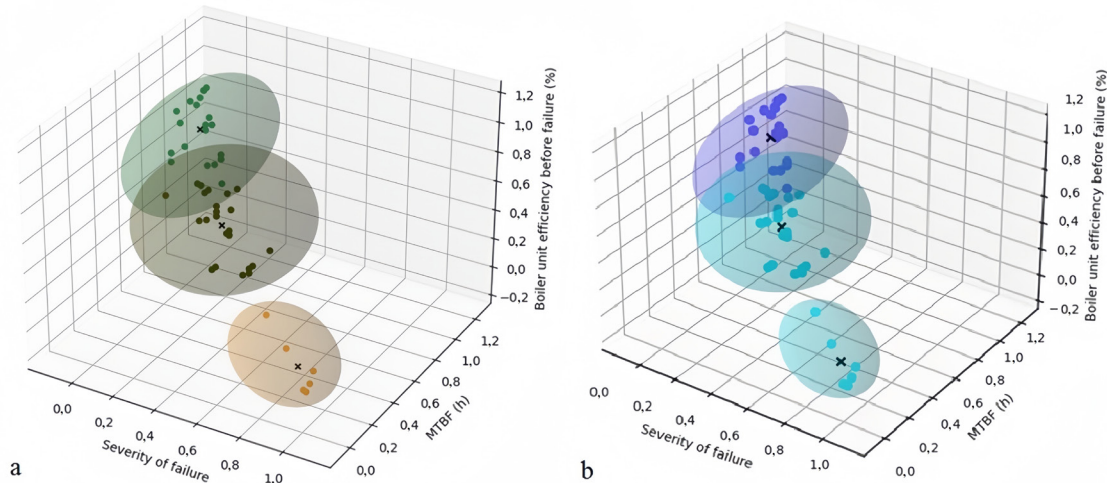


Figure 3. Clustering of Ariston boiler failure data, as obtained a) using the Scikit-learn library in Jupyter Notebook and b) using our own program
Source: Authors

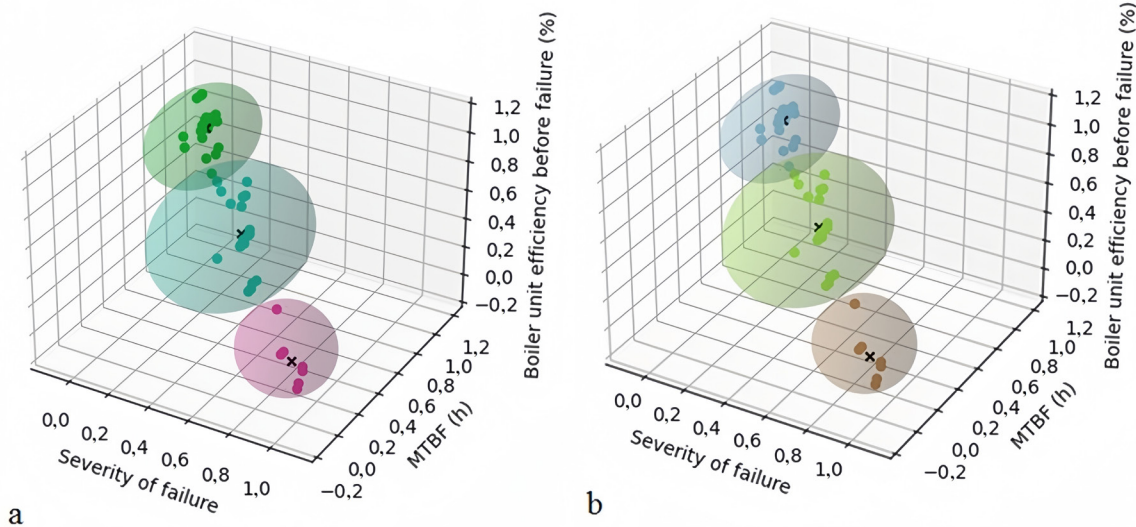


Figure 4. Clustering of Bosch boiler failure data, as obtained a) using the Scikit-learn library in Jupyter Notebook and b) using our own program
Source: Authors

cluster analysis. The results of this analysis are presented in the form of graphs (Figs. 3 to 6). In this case, the visual differences are minimal. Each cluster center is fixed, and the data group is clearly delineated. By comparing the results obtained, the visual differences between algorithms are not as clear as those between the sizes of the clusters for the boiler brands. We initially assumed three main groups of clusters: unreliable devices, devices with normative reliability, and devices exhibiting a high degree of reliability. In the figures, the first category is closest to the reader, the second is in the middle of the clustering cube, and the third is in the upper left corner. In the case of Ariston boilers, the largest clusters correspond to devices exhibiting high and normative reliability (Fig. 3). As for Bosch boilers (Fig. 4), the largest cluster is in the normative reliability zone, and, for the Buderus and Viessmann boilers (Figs. 5 and 6), the largest clusters are in the low reliability zone.

Even a visual assessment of the results of the cluster analysis provides a lot for technicians and managers. In our case, we identified the need to study the causes of the increased failure frequency and severity of Buderus and Viessmann boilers. Measures should also be taken to shift the failure distribution statistics from the zone of low reliability towards those of high and normative reliability. However, a visual assessment was obviously not enough to determine the effectiveness of data clustering by means of the two studied approaches. In order to determine exactly what method is more accurate and efficient in processing the data, several key clustering parameters had to be calculated. These structurally indicated the density of the clusters obtained, their separability, and the closeness of the links between the grouped events [18], making it possible to compare the results obtained for each of the four failure datasets and for both approaches.

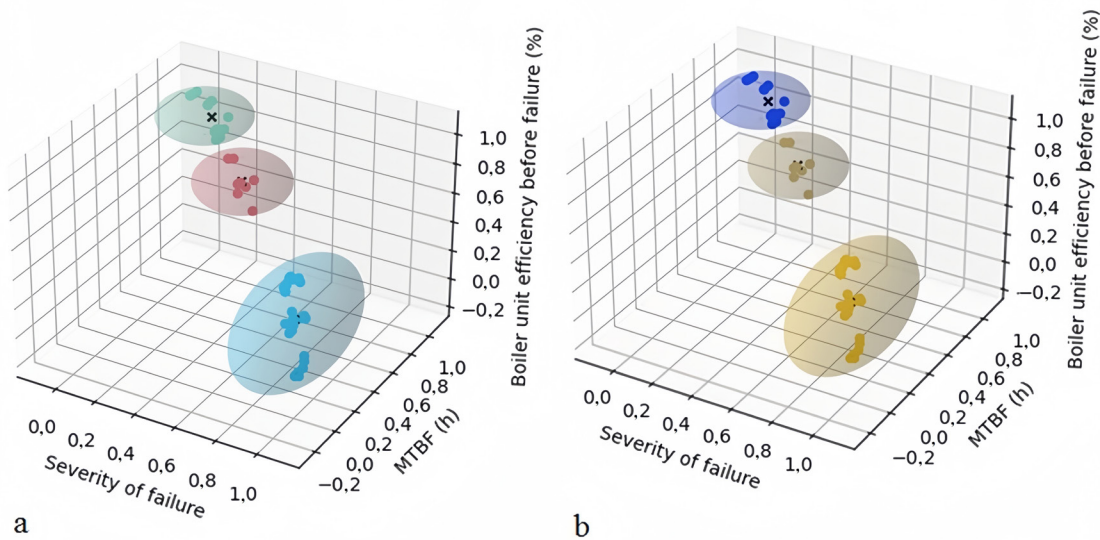


Figure 5. Clustering of Buderus boiler failure data, as obtained a) using the Scikit-learn library in Jupyter Notebook and b) using our own program
Source: Authors

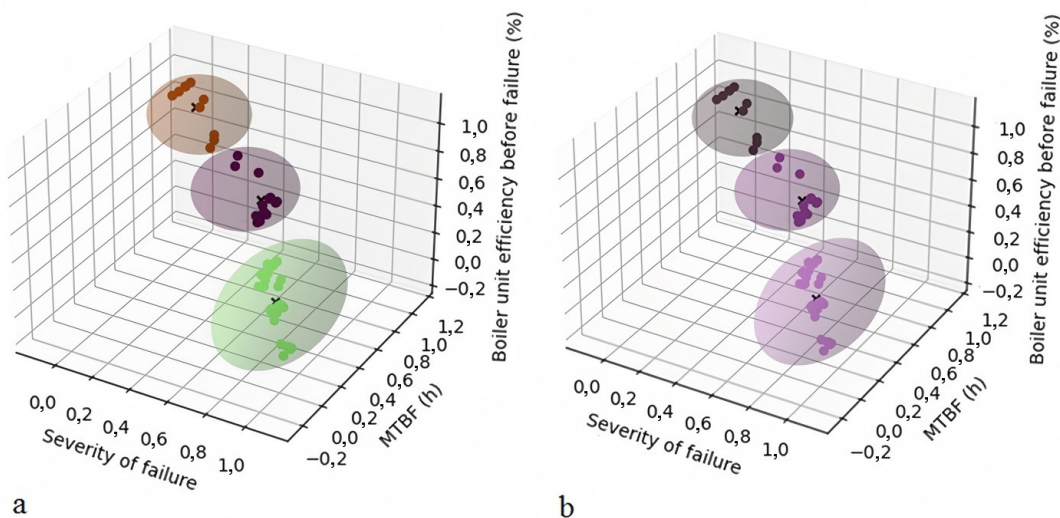


Figure 6. Clustering of Viessmann boiler failure data, as obtained a) using the Scikit-learn library in Jupyter Notebook and b) using our own program
Source: Authors

To make comparisons, we used the Dunn index, a metric for evaluating clustering algorithms. In these contexts, the goal is to identify sets of clusters that are compact, exhibit little variance between elements, and are well separated. The mean values of the clusters must be distant from each other when compared to the variance within them. For a given cluster distribution, a higher Dunn index indicates better clustering [19]. This can be calculated in various ways, such as

$$D(C) = \frac{\min_{C_k \in C} \left\{ \min_{C_l \in C \setminus C_k} \{ \delta(C_k, C_l) \} \right\}}{\max_{C_k \in C} \{ \Delta(C_k) \}} \quad (7)$$

where:

- C_k, C_l – clusters from set C
- C – set of clusters
- δ – inter-cluster distance
- Δ – cluster diameter

The Euclidean norm was used as a distance function in this and all subsequent estimations. We considered the distance between clusters to be the minimum distance between their points.

We also used the silhouette index, which indicates how similar an object is to its cluster when compared to other clusters. The silhouette index can be calculated as follows:

$$il(C) = \frac{1}{N} \sum_{C_k \in C} \sum_{x_i \in C_k} \frac{b(x_i, c_k) - a(x_i, c_k)}{\max \{ a(x_i, c_k), b(x_i, c_k) \}} \quad (8)$$

where:

- x_i – element of the cluster
- a, b – average distance of point x_i
- N – number of elements in the set C

Cluster separability is evidently an important parameter, as it indicates the degree of overlap between clusters and the distance from each other in space. In general, it confirms the correctness of the initial hypothesis regarding the calculated number of clusters according to their remoteness from each other. Separability is determined using the following formula:

$$BSS = n \cdot \sum_{j=1}^M \left(\bar{x}_j - \bar{x} \right)^2 \quad (9)$$

where:

- n – a cluster number
- M – a number of clusters

- j – an index

Another important parameter is cluster compactness, which evaluates the distance between cluster elements. This property can be expressed through the distance between the elements in the cluster, the cluster's density, or the volume occupied by the cluster in the multidimensional space. Compactness (10) is calculated as follows:

$$WSS = \sum_{j=1}^M \sum_{i=1}^{|C_j|} \left(x_{ij} - \bar{x}_j \right)^2 \quad (10)$$

The obtained data on cluster centers, silhouettes, cluster separability, compactness, and the Dunn index are summarized in Table 1. The results presented therein show that the use of our own program allowed for a more accurate organization of the data in clusters. This is especially important for large sample sizes. Thus, in most cases, the results of statistically processing all considered parameters are higher by the hundredths. This indicates a clearer clustering, which has to do with the specific characteristics of Jupyter Notebook [20]. In particular, the Scikit-learn library was originally written as a universal tool for mathematical data analysis, which is why, in applying individual information processing methods, it can yield slightly larger deviations [21].

It can be stated that, for the analysis of small samples with an accuracy of one hundredth of a unit, it is possible to employ clustering via Scikit-learn. In other cases, to study large amounts of statistical data with the purpose of assessing the quality of maintenance and repair, it is better to create custom clustering software.

Discussion

Regardless of the clustering program used, an analysis of the results (Figs. 3 to 6 and Table 1) allows drawing several important technical conclusions. This section discusses them and compares them against facts previously presented in scientific periodicals. The shift in the density function of the failures is most visible in gas boilers of the Buderus and Viessmann brands (within the data sample, these boilers failed more often, exhibiting sharp drops in efficiency and total equipment failure).

It is clear that the reliability indicator is higher for Bosch and especially Ariston boilers. This is evidenced in the smaller size of the lower right cluster of events in each of the graphs. This cluster shows the number of boilers of a certain brand from the sample which exhibited failures requiring complex repairs.

Similar conclusions regarding Bosch and Ariston boilers' increased time until failure and greater reliability have been presented in other publications [22, 23]. It is also worth noting that, in comparison, the technical data sheets of the

Table I. Summary of the clustering results

	Ariston		Bosch		Buderus		Viessmann	
	Jupiter	Our program	Jupiter	Our program	Jupiter	Our program	Jupiter	Our program
Dunn index	0.195	0.200	0.259	0.260	0.223	0.220	0.235	0.240
Silhouette index	0.509	0.841	0.628	0.965	0.627	0.923	0.594	0.911
Separability	0.497	0.498	0.528	0.528	0.502	0.503	0.478	0.478
Compactness	2.043	2.043	1.463	1.463	1.62	1.62	1.639	1.639
Cluster centers	0.938	0.938	0.528	0.528	0.156	0.156	0.846	0.846
	0.069	0.069	0.383	0.383	0.855	0.855	0.250	0.250
	0.113	0.113	0.609	0.609	0.951	0.951	0.172	0.172
	0.432	0.432	0.116	0.116	0.872	0.872	0.580	0.580
	0.354	0.354	0.817	0.817	0.256	0.256	0.619	0.619
	0.644	0.644	0.913	0.913	0.197	0.197	0.576	0.576
	0.106	0.106	0.937	0.937	0.431	0.431	0.100	0.100
	0.751	0.751	0.073	0.073	0.608	0.608	0.866	0.866
	0.908	0.908	0.114	0.114	0.750	0.750	0.930	0.930

Source: Authors

Buderus and Viessmann boilers requested longer periods between diagnostics and maintenance work (*i.e.*, six months longer), which could lead to the untimely detection and elimination of early-stage defects, which could subsequently become more critical [24].

The separation of clusters is clearly observed in the case of the Buderus and Viessmann boilers. This may be due to differences in assembly quality (*i.e.*, separation of the upper left cluster of a more reliable group) or to an abrupt transition of boilers from a standard condition to serious failure (*i.e.*, gap between the central and lower right cluster). Such a sharp transition (cluster separability) is more widely associated in scientific research [25, 26] with the peculiarities of gas distribution networks' wear than with the insufficient reliability of the equipment itself. This is also more likely to occur in countries where the gas transmission system mainly consists of steel pipes (which is more typical for the countries of the former COMECON) [27].

To verify this fact, additional information is required, particularly concerning the frequency of boiler failures depending on the distribution network type (plastic, steel). The materials of these brands' gas burners may require additional reliability testing.

It can only be stated with certainty that there is a group of factors (external or internal) that leads to the emergence of such a sharp separation in the case of the central and right lower clusters for the Buderus and Viessmann brands boilers. Perhaps more clarity on this issue will be obtained through further research.

The slight separability of the central and upper right clusters for the Bosch and Ariston boilers indicates a higher quality of the assembled equipment, the majority of which has either the declared characteristics or an additional strength margin. This is also confirmed by the increased time until failure of the aforementioned brands [23], as well as by the fact that smart work management systems were used in these products first [28]. In general, these factors contribute to the accumulation of the largest clusters in the upper right corner and in the center (high and normative reliability).

In addition to the technical component (*i.e.*, the reliability and failure rate of double-circuit gas boilers of various brands), there are slight differences in the results of data clustering using Jupyter Notebook and our program. As stated earlier, this is due to the heterogeneity and greater breadth of the Jupyter Notebook libraries used. Nevertheless, the use of two different programs with a common methodological basis allowed verifying the calculations.

However, despite the correctness of the applied methods and the presented results, the most accurate conclusions regarding the studied boilers could only be obtained by conducting a broader statistical analysis.

It is necessary to mention that the data samples presented in this study are small (25 boilers of each brand) and were taken from one region, *i.e.*, the Kaliningrad region of the Russian Federation (even though the data were provided by different companies carrying out repairs). It is possible that, with a larger dataset, the reliability indicators might paint a different picture. A larger dataset could eliminate bias in the

recorded values, which might occur due to both regional influence (usage and temperature conditions as well as the quality of gas purification) and the quality of the supplied batches of this equipment. The influence of these factors and the analysis of larger databases are issues for further research.

The desirability of such an approach to assessing the quality of using equipment, especially under different external conditions, is indicated by [29] and [30].

Conclusions

To process the results of the observations of service companies on the failure of gas boilers of various brands, this study proposes an approach that allows assessing the severity of equipment failure based on frequency and repair complexity. A cluster analysis was conducted while considering the equipment's time until failure as well as its efficiency. Four brands were analyzed: Bosch, Ariston, Buderus, and Viessmann. Clustering was carried out using our own software and the publicly available Scikit-learn library of Jupyter Notebook. As a result, a convenient method for processing statistical information on gas boiler failures was obtained, which allows verifying the results through two software products prior to visualization.

Studying the data on gas boiler failure statistics in Kaliningrad, as recorded by three service companies in 2022, allows drawing a number of analytical conclusions. It was found that, for the studied samples, three groups of clusters are clearly formed: devices with increased reliability, devices with normative reliability, and unreliable devices. The most reliable gas boilers were of the Ariston brand, which exhibited lowest failure frequency and severity. This was also the case with Bosch boilers, which showed a smaller share of severe failures. Buderus and Viessmann boilers are less reliable. This may be due to issues in software or hardware, the specifics of the production technology, or the specifics of service work (i.e., lower frequency). The efficiency parameter of the Ariston and Bosch boilers before failure was also higher.

The results show that Buderus and Viessmann boilers require more frequent monitoring by service companies. Furthermore, considering the frequency and severity of the reported failures, service companies are advised to pay attention to the quality of their repair work, use original spare parts, and ensure their timely availability. A more detailed understanding of the results, as well as their confirmation based on the statistics of other regions' gas boiler failures could be provided in further research. We propose conducting a comparative analysis of the results with data from other cities, as this will allow determining the influence of natural factors. These parameters could vary and have a significant impact on equipment reliability. Secondly, it is necessary to compare data not only based on the brand, but also on individual models, or to link them

to batches from manufacturing plants (equipment reliability can also vary from model to model and from batch to batch, even within the same manufacturer).

The obtained results can be used, on the one hand, as practical information in planning the activities of service companies regarding diagnostics and maintenance work (e.g., in determining the volume and type of the work planned). On the other hand, the presented research materials and the proposed methodological and software apparatus should serve as a good basis for conducting further studies on the reliability of gas boilers of different brands during operation.

CRedit author statement

Shcherban, P. Conceptualization, investigation, methodology, project administration, writing (review and editing).

Abu-Khamdi, R. Data curation, software, validation, visualization, writing (original draft).

Conflicts of interest

The authors declare no competing interests.

References

- [1] S. Liu, "Corrosion failure analysis of the heat exchanger in a hot water heating boiler," *Eng. Fail. Anal.*, vol. 142, no. 1, pp. 106847, Jan. 2022. <https://doi.org/10.1016/j.engfailanal.2022.106847>
- [2] F. Samanlıoğlu, "Evaluation of gas-fired combi boilers with HF-AHP-MULTIMOORA," *Appl. Comput. Intell. Soft Comput.*, vol. 2022, no. 1, pp. 1–10, Jan. 2022. <https://doi.org/10.1155/2022/9225491>
- [3] K. Simic, "Modelling of a gas-fired heating boiler unit for residential buildings based on publicly available test data," *Energy Build.*, vol. 253, no. 1, art. 111451, Jan. 2021. <https://doi.org/10.1016/j.enbuild.2021.111451>
- [4] K. Ritosa, "A probabilistic approach to include the overall efficiency of gas-fired heating systems in urban building energy modelling," *J. Phys. Conf. Ser.*, vol. 2069, no. 1, art. 012105, Jan. 2021. <https://doi.org/10.1088/1742-6596/2069/1/012105>
- [5] Y. Chang, "Statistics and analysis of test data of industrial boiler approved products in 2020," in *Proc. Int. Conf. Adv. Manuf. Technol. Manuf. Syst. (ICAMTMS)*, vol. 12309, pp. 604–610, Jan. 2022. <https://doi.org/10.1117/12.2645720>
- [6] V. Prabhu and D. Chaudhary, "Machine learning-enabled condition monitoring models for predictive maintenance of boilers," in *Proc. 4th Int. Conf. Recent Dev. Control Autom. Power Eng. (RDCAPE)*, Nov. 2021, pp. 426–430. <https://doi.org/10.1109/RDCAPE52977.2021.9633534>
- [7] A. R. Aikin, "The process of effective predictive maintenance," *Tribol. Lubr. Technol.*, vol. 77, no. 2, pp. 34–40, Feb. 2021

- [8] S. Mohanty, K. C. Rath, and O. P. Jena, "Implementation of total productive maintenance (TPM) in the manufacturing industry for improving production effectiveness," in *Industrial Transformation*, 1st ed. Boca Raton, FL, USA: CRC Press, 2022, pp. 45–60. <https://doi.org/10.1201/9781003229018-3>
- [9] N. Xodjiev *et al.*, "Analysis of the resource-saving method for calculating the heat balance of the installation of hot-water heating boilers," in *AIP Conf. Proc.*, vol. 2432, no. 1, art. 020019, Jan. 2022. <https://doi.org/10.1063/5.0090455>
- [10] W. Ma *et al.*, "Energy efficiency indicators for combined cooling, heating, and power systems," *Energy Convers. Manag.*, vol. 239, art. 114187, Jan. 2021. <https://doi.org/10.1016/j.enconman.2021.114187>
- [11] M. Larbi Rebaiaia and D. Aït-Kadi, "Maintenance policies with minimal repair and replacement on failures: Analysis and comparison," *Int. J. Prod. Res.*, vol. 59, no. 23, pp. 6995–7017, Dec. 2021. <https://doi.org/10.1080/00207543.2020.1832275>
- [12] J. Leukel, J. González, and M. Riekert, "Adoption of machine learning technology for failure prediction in industrial maintenance: A systematic review," *J. Manuf. Syst.*, vol. 61, pp. 87–96, Jan. 2021. <https://doi.org/10.1016/j.jmsy.2021.08.012>
- [13] C. Acuña-Soto, V. Liern, and B. Pérez-Gladish, "Normalization in TOPSIS-based approaches with data of different nature: Application to the ranking of mathematical videos," *Ann. Oper. Res.*, vol. 296, no. 1, pp. 541–569, Jan. 2021. <https://doi.org/10.1007/s10479-018-2945-5>
- [14] A. Chabane, S. Adjerid, and I. Meddour, "Dependability analysis in systems engineering approach using the FMECA extracted from the SysML and failure modes classification by K-means," *Int. J. Dyn. Cont.*, vol. 10, no. 3, pp. 981–998, Jul. 2022. <https://doi.org/10.1007/s40435-021-00855-8>
- [15] F. Aksan *et al.*, "Clustering methods for power quality measurements in virtual power plants," *Energies*, vol. 14, no. 18, p. 5902, Sep. 2021. <https://doi.org/10.3390/en14185902>
- [16] W. Peiyi *et al.*, "Analysis and research on enterprise resumption of work and production based on K-means clustering," in *Proc. 6th Int. Conf. Big Data Anal. (IC-BDA)*, pp. 169–174, Dec. 2021. <https://doi.org/10.1109/ICBDA51983.2021.9403217>
- [17] L. Quaranta, F. Calefato, and F. Lanubile, "KGTorrent: A dataset of Python Jupyter notebooks from Kaggle," in *Proc. IEEE/ACM 18th Int. Conf. Mining Softw. Repos. (MSR)*, May 2021, pp. 550–554. <https://doi.org/10.1109/MSR52588.2021.00072>
- [18] H. Fangohr, T. Kluyver, and M. Dipierro, "Jupyter in computational science," *Comput. Sci. Eng.*, vol. 23, no. 2, pp. 5–6, Mar. 2021. <https://doi.org/10.1109/MCSE.2021.3059494>
- [19] C. J. Weiss and A. Klose, "Introducing students to scientific computing in the laboratory through Python and Jupyter notebooks," in *ACS Symp. Ser. Teach. Program. Across Chem. Curric.*, Jan. 2021, pp. 57–67. <https://doi.org/10.1021/bk-2021-1387.ch005>
- [20] E. Mazur, P. Shcherban, and V. Mazur, "Research and evaluation of the operating characteristics of used ship engine oil using the process parameter matrix method," *FME Trans.*, vol. 51, no. 4, pp. 497–503, Dec. 2023. <https://doi.org/10.5937/fme2304497M>
- [21] J. Piazentin Ono, J. Freire, and C. T. Silva, "Interactive data visualization in Jupyter notebooks," *Comput. Sci. Eng.*, vol. 23, no. 2, pp. 99–106, Mar. 2021. <https://doi.org/10.1109/MCSE.2021.3052619>
- [22] P. Shcherban, A. Sokolov, and R. A. Hamdi, "Study of failure statistics of cavitators in the fuel oil facilities through the application of regression and cluster analysis," *Proc. Eng. Sci.*, vol. 5, no. 1, pp. 39–48, Jan. 2023. <https://doi.org/10.24874/PES05.01.004>
- [23] J. Zhao *et al.*, "Root cause analysis of a cracked primary heat exchanger in a gas wall-mounted boiler," *Eng. Fail. Anal.*, vol. 153, art. 107583, Jan. 2023. <https://doi.org/10.1016/j.engfailanal.2023.107583>
- [24] R. Oliver, A. Duffy, and I. Kilgallon, "Statistical models to infer gas end-use efficiency in individual dwellings using smart metered data," *Sustain. Cities Soc.*, vol. 23, pp. 1–10, Nov. 2016. <https://doi.org/10.1016/j.scs.2016.01.009>
- [25] Z. Liao, M. Swainson, and A. L. Dexter, "On the control of heating systems in the UK," *Build. Environ.*, vol. 40, no. 3, pp. 343–351, Mar. 2005. <https://doi.org/10.1016/j.buildenv.2004.05.014>
- [26] A. V. Gubarev, "Determination of the thermotechnical measures of the condensing water heating boiler's high-temperature section," in *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 791, no. 1, art. 012011, Jan. 2020. <https://doi.org/10.1088/1757-899X/791/1/012011>
- [27] M. Bukurov, S. Bikić, and B. Marković, "Efficiency and management of gas boilers in public buildings in Vojvodina," *J. Process. Energy Agric.*, vol. 20, no. 2, pp. 87–92, Jun. 2016.
- [28] J. P. Xu, X. Q. Ma, and H. Shen, "Remote control system design for domestic wall-mounted gas boilers based on GSM modem," *Adv. Mater. Res.*, vol. 650, pp. 559–564, Mar. 2013. <https://doi.org/10.4028/www.scientific.net/AMR.650.559>
- [29] M. S. Mahmud *et al.*, "A survey of data partitioning and sampling methods to support big data analysis," *Big Data Min. Anal.*, vol. 3, no. 2, pp. 85–101, Jun. 2020. <https://doi.org/10.26599/BDMA.2019.9020015>
- [30] A. M. Ikotun *et al.*, "K-means clustering algorithms: A comprehensive review, variants analysis, and advances in the era of big data," *Inf. Sci.*, vol. 622, pp. 178–210, Jan. 2023. <https://doi.org/10.1016/j.ins.2022.11.139>

Heuristics-Based Energy Demand Forecasting with Scarce Data in the Department of Huila, Colombia

Pronóstico de la Demanda de Energía Basado en Heurística con Datos Escasos en el Departamento del Huila, Colombia

Juan J. Cuenca¹, Diego Palacios-Castro², and Rodolfo García³

ABSTRACT

Within the framework of the energy transition, electrical distribution grid operators require effective tools to predict the demand of individual users. These tools are necessary for an adequate planning of future generation resources and infrastructure modernization. However, understanding future electricity needs poses a significant challenge, especially in emerging economies, where historical data are manually collected on a monthly or bi-monthly basis and exhibit a significant amount of missing information. In response to the above, this work proposes a novel heuristics-based method for medium-term energy demand forecasting with scarce data. Qualitative and quantitative information was abstracted into a mathematical model representing the trend and noise components of historical energy consumption observations. In addition, external factors were considered as an additional layer for the mathematical model, in order to account for events that could not be foreseen by merely using the dataset. A train-test data split was proposed to iteratively search for the best parameters to predict electricity demand with respect to different categorical divisions of users (social stratum, rural or urban location, and municipality). For testing and validation, real historical data were used, as provided by the regional distribution system operator (DSO) of the department of Huila, Colombia. The results suggest a trade-off between accuracy and computational intensity, as well as the fact that a non-categorical approach leads to the algorithm with the best performance (average absolute error of 1.61%) at a low computational cost.

Keywords: demand forecasting, energy, heuristics, scarce data

RESUMEN

En el marco de la transición energética, los operadores de redes de distribución eléctrica requieren herramientas efectivas para predecir la demanda de usuarios individuales. Estas herramientas son necesarias para un planeamiento adecuado de los recursos futuros de generación y la modernización de la infraestructura. Sin embargo, entender las necesidades futuras de electricidad constituye un desafío significativo, especialmente en economías emergentes donde los datos históricos son recolectados manualmente en períodos mensuales o bimensuales y presentan una cantidad significativa de información faltante. En respuesta a esto, se propone un novedoso método basado en heurística para el pronóstico de la demanda de energía en el mediano plazo con datos escasos. Se abstraen información cualitativa y cuantitativa en un modelo matemático que representa las componentes de tendencia y ruido en observaciones históricas de consumo de energía. Adicionalmente, se consideraron factores externos como capa adicional para el modelo matemático, en aras de dar cuenta de eventos que no podrían ser previstos solamente con el conjunto de datos. Se propuso una división de datos de entrenamiento y prueba con el fin de buscar iterativamente los mejores parámetros para predecir la demanda de electricidad respecto a diferentes divisiones categóricas de usuarios (estrato social, ubicación rural o urbana y municipio). Para realizar pruebas y validaciones, se utilizaron datos históricos reales proporcionados por el operador del sistema de distribución (OSD) regional del departamento del Huila, Colombia. Los resultados sugieren que hay una compensación entre precisión e intensidad computacional, y que un enfoque no categórico resulta en el algoritmo con un mejor desempeño (error absoluto promedio de 1.61 %) y un bajo costo computacional.

Palabras clave: pronóstico de la demanda, energía, heurística, datos escasos

Received: June 16th 2023

Accepted: September 6th 2024

Introduction

The electricity sector is undergoing significant changes due to the energy transition. The inclusion of distributed energy resources with uncertain behavior, e.g., solar photovoltaic (PV), wind turbines, etc. (Cuenca and Hayes, 2022), the electrification of heat and transport (Mehigan *et al.*, 2022), and the development of new local energy markets and price schemes (Cuenca, Jamil, and Hayes, 2023) are changing the way we produce, transport, use, and trade electrical energy.

Within this changing paradigm, there are numerous new challenges for the operation and planning of electrical distribution networks. One of these challenges is the

forecasting of energy demand. Forecasting requires the use of historical data to determine a hypothetical future

¹Electrical engineer, Escuela Colombiana de Ingeniería, Colombia. MEng Electrical Engineering, Auckland University of Technology, New Zealand, PhD Electrical Engineering, University College Cork, Ireland. Affiliation: Researcher in Electrical and Energy Engineering at EHS Ltda., Colombia, contractor at Centro Internacional de Física (CIF), Colombia, and assistant professor of Smart Control for Energy Systems at CentraleSupélec - IETR Lab (UMR CNRS 6164), France. E-mail: juan.cuenca@ehs.com.co

²Affiliation: Head of the Systems and Organization Office, Electrificadora del Huila SA ESP, Colombia. E-mail: diego.palaciosc@electrohuila.co

³Electrical engineer, Universidad Nacional de Colombia. MSc Economic Sciences, Universidad Nacional de Colombia. PhD Engineering, Universidad Nacional de Colombia. Affiliation: Mega-projects and innovation researcher, Centro Internacional de Física (CIF), Colombia. E-mail: rgarciasi@unal.edu.co



Attribution 4.0 International (CC BY 4.0) Share - Adapt

state. This is useful for researchers, regulators, system operators, and utilities to understand how to schedule future generation resources and infrastructure upgrades (Hemmati, Hooshmand, and Taheri, 2015). In the energy sector, numerous algorithms have been developed to this effect (Klyuev *et al.*, 2022). Previous research argues that, with sufficient historical data, it is possible to approximately predict energy customer behavior.

Traditionally, historical data on energy consumption have been collected on a monthly or multi-monthly basis. The utility dispatches personnel to physically visit the customers' energy meters and record their consumption for the last billing cycle (Bimenyimana and Asemota, 2018). This is changing as we are moving towards the digital era: modern smart energy meters include measuring capabilities on smaller time steps (in the order of seconds or minutes) and utilize wired or wireless communication. This information on consumption is transmitted to the utility in close-to-real time (Bimenyimana and Asemota, 2018). Numerous nations are pioneering the rollout of smart metering, providing the necessary inputs for the functioning of forecasting algorithms in the literature.

Nonetheless, in emerging economies like Colombia, smart meter rollout is still at an early stage, and historical data on energy consumption are still collected according to tradition (*i.e.*, monthly or multi-monthly). This is especially the case with rural areas, where access difficulties may further delay the installation of smart meters. Within this frame of data scarcity, it is important to develop alternative methods for forecasting energy customer demand. In late 2021, the regional distribution system operator (DSO) of the department of Huila (Colombia), *i.e.*, Electrohuila SA ESP, opened a call for data scientists, researchers, and data enthusiasts to provide solutions to the issue of energy forecasting with scarce data. A total of 25 proposals were submitted to the Hackathon Opita Challenge call (ElectroHuila S.A. E.S.P., 2021). This manuscript reports on the most effective one of these methods.

The purpose of this study is threefold: (i) to provide a framework for data processing to abstract qualitative consumption patterns into numerical inputs; (ii) to describe the algorithm search methodology and the train-test data split in order to develop an effective heuristics-based method for forecasting electricity demand that leverages scarce real data provided by the DSO; and (iii) to report on the results of the implementation and describe potential use cases in Colombia and abroad.

Literature review

Electricity consumption forecasting has been a relevant research topic for many years. As early as 1910, experimental studies on electrical installations and the application of mathematical methods related to probability theory to calculate the future energy requirements of customers were carried out (Bunn and Farmer, 1985). Since then, and with the development of computer technology, there have been studies on the application of technocenosis, fuzzy set, game, pattern recognition, cluster analysis, and decision theories. There are numerous reviews describing the history of electricity consumption forecasting methods, which the reader is encouraged to consult (vom Scheidt *et*

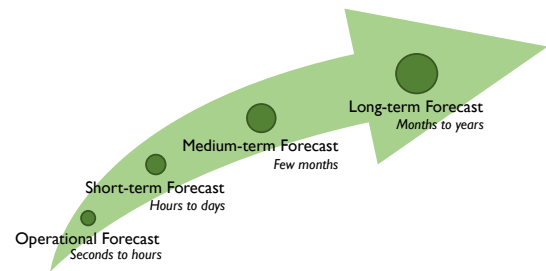


Figure 1. Classification of forecasts by lead time

Source: Authors, based on Klyuev *et al.* (2022)

al., 2020; Wei, Li, Peng, Zeng, and Lu, 2019; Ghoddusi, Creamer, and Rafizadeh, 2019; Biel and Glock, 2016).

Resulting from their review of the literature, (Klyuev *et al.*, 2022) recommended that, when designing forecasting methods, it is important to report not only quantitative estimates, but also qualitative features, as well as the specific conditions that make methods applicable.

Out of the numerous ways to categorize forecasting models, the most useful for the purpose of this work is by lead time (*i.e.*, the period of anticipation) (Klyuev *et al.*, 2022). As shown in Figure 1, depending on the prediction horizon, forecasts can be classified as operational, short-, medium-, and long-term (Hong and Fan, 2016). As will be discussed ahead, due to data availability, the remainder of this literature review will focus on medium-term forecasting methodologies.

Medium-term energy consumption forecasting is useful for the utility and system operators to schedule infrastructure upgrades and maintenance, plan electricity pricing, and measure the technical and economic performance of their grids (Klyuev *et al.*, 2022). Furthermore, a good knowledge of future demand is useful when defining flexibility strategies like price signaling for demand response (Honarmand, Hosseinneshad, Hayes, Shafie-Khah, and Siano, 2021). Different approaches to the issue of energy demand forecasting are available in the literature, a selection of which is presented below.

Using deep machine learning algorithms, social and climatic factors were considered to predict the energy demand of six buildings in a city district (Yuce, Mourshed, and Rezgüi, 2017). The factors were weighed to account for differential influences, and multiple regression analyses were performed. The results of this study suggest that it is possible to achieve increased forecasting accuracy in some seasons given the selected climatic factors.

The G-, Z-, and GZ-methods from statistics and time series theory were used to predict consumption by technocenosis objects while considering their individual and/or system properties (Gnatyuk, Polevoy, Kivchun, and Lutsenko, 2020). In their work, the authors introduce the autoregressive moving average (ARMA) model, the time series decomposition (TVRD) model, and the singular spectrum analysis (SSA) model.

A feature extraction algorithm was used in Meng, Niu, and Sun (2011), in which forecasting was carried out by applying

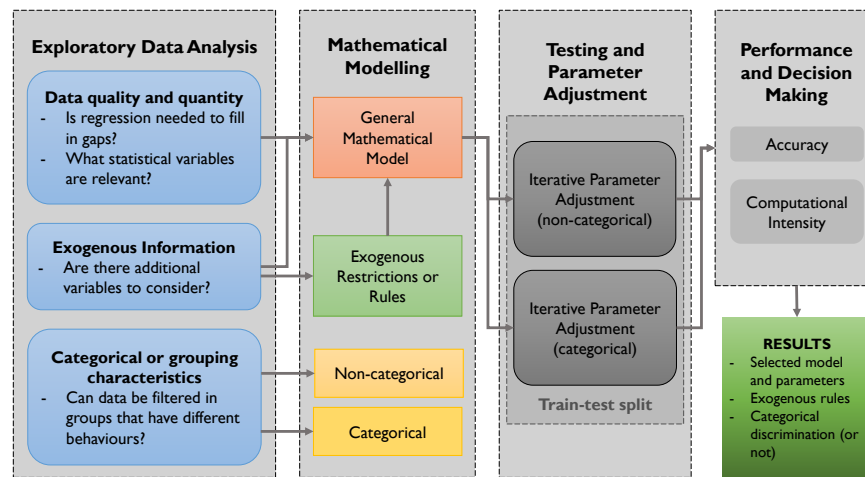


Figure 2. Structure of the proposed algorithm to select the forecasting model with scarce demand data

Source: Authors

a discrete waveform and decomposing power consumption data into a trend, a periodic component, and noise.

In [Amber et al. \(2017\)](#), a simple mathematical equation was conceived to determine future electricity usage via a multiple regression technique, considering variables such as building time, the temporal index, and surrounding temperature. These variables were found to have an important influence on energy consumption.

Shallow neural networks were used in [Shumilova, Gottman, and Starceva \(2008\)](#) to address the studied problem. A three-layered perceptron with a sigmoid activation function was proposed. The input layer had 24 inputs (*i.e.*, neurons) including the power consumption of the previous month, the maximum and minimum monthly demand, the average number of hours of daylight, the monthly temperature, the number of holidays in the previous and target months, and the geographical location. The hidden layer included five neurons, and the third layer contained one neuron that gave one predicted value of consumption as the output. The authors also integrated adaptive feedback into their model to improve performance and ultimately concluded that the most accurate result is provided by a fuzzy neural network.

Classical probabilistic approaches, intelligent algorithms, and hybrid methods are used in the literature to understand trends and noise components of present demand and to ultimately forecast energy needs. As suggested in [Klyuev et al. \(2022\)](#), it is important to consider the frequency and extent of the original demand dataset to decide on the forecasting tool to be used. Monthly or multi-monthly consumption data spanning numerous years are enough to apply classical approaches, but the accuracy of intelligent algorithms depends largely on the availability of large and detailed data sources.

Considering social development, exogenous factors, the stochastic nature of human behavior, and complex macroeconomic conditions, the medium-term forecasting of energy consumption is a difficult task. The literature recommends identifying monthly electricity consumption trends using data from numerous consecutive years while considering periodic components (*e.g.*, seasonal temperature variations) and quantifying sources of noise.

The above highlights the gaps in the literature on medium-term forecasting that this manuscript addresses: data scarcity, the inclusion of complex periodic components, and noise identification.

Methodology

The forecasting tool resulting from this work was obtained using a modular algorithm to select a performing model, including its structure, parameters, and rules. As seen in [Figure 2](#), an initial data assessment is followed by the mathematical definition of the model and its variables and additional rules. Two different approaches can be applied to the mathematical model, *i.e.*, categorical or non-categorical, where the model finds or does not find different parameters for different groups of data. A train-test data split is used to analyze the performance of different parameters and structures, ultimately aiming to contrast models in terms of accuracy and computational intensity for a final decision. This section presents the details of the methodology employed in our work.

Exploratory data analysis

The first module considers a statistical analysis of the historical data available. The purpose of this initial assessment is to determine which parameters, data curation approaches, exogenous factors, and general characteristics of the dataset *should* be included in the mathematical model. First, it is important to evaluate whether the available historical data have significant missing portions that must be filled. Depending on the quantity or quality of the dataset, data curation approaches might be relevant, as discussed in [Dong and Peng \(2013\)](#).

Second, it is important to consider whether there are correlations between the available data and certain parameters. The following questions should be asked:

- Are there significant variations in consumption depending on the time of the year (*e.g.*, the residential electricity consumption is lower in June because it coincides with the holiday season)?

- Is temperature variation relevant (e.g., the warmest month may be associated with extra electricity consumption for air conditioning)?
- Is population density relevant (e.g., densely populated areas have different consumption patterns compared to rural settings)?

These questions must be framed within the socio-geo-demographic context of the case study: for different locations, social or economic contexts, different questions can be asked.

Third, we propose considering exogenous factors as part of the initial data analysis through the following questions:

- Are there any macroeconomic correlations (e.g., energy consumption may be correlated to changes in energy prices)?
- Is it possible for consumption patterns to drastically change due to an exogenous event (e.g., when a household is vacated, the energy consumption suddenly drops until there is a new resident)?

Lastly, provided that the dataset includes additional information (i.e., not only on historical consumption), performing a comparative analysis of different categorical or grouping characteristics is very useful. If the dataset is grouped by the type of user (e.g., residential, industrial, or commercial), it is possible to assess the differences in the consumption of each category. Ultimately, this illustrates the need for a differential approach that addresses the categorical nature of the dataset.

Mathematical modeling

The above-presented exploratory data analysis above sheds light on the statistical variables that may be subjected to study and are useful to build the mathematical model. Among others, values like the mean, median, variance, and standard deviation of the population or a sample can be useful to define a mathematical model.

Once identified, different mathematical relations can be drawn between variables (linear, polynomial, exponential, logarithmic, etc.). The selection of these relations is reserved to the model designer and involves subjective criteria, given the stochastic nature of the problem. As suggested in [Lindsey \(2004\)](#), mathematical model selection for stochastic problems like demand forecast is important, but secondary to the correct selection of parameters. In this vein, and for the purpose of this heuristics-based study, a linear combination of variables was selected as the mathematical model, which is shown in Equation (1). Here, a , b , and x were the selected variables, and α , β , and χ the corresponding parameters.

$$f(a, b, \dots, x) = (\alpha \times a) + (\beta \times b) + \dots + (\chi \times x) \quad (1)$$

An intermediary block to test different mathematical relations can be added to the proposed algorithm (Figure 2). Instead of arriving at the general mathematical model directly from the exploratory data analysis, it is possible to create a loop to select an adequate formula from several

candidates (e.g., linear, polynomial, exponential, etc.) by means of a data sample. Nonetheless, given the added complexity of that approach, and since this additional block would still be influenced by the subjective criteria of the model designer, its inclusion will be addressed in future work.

After model selection, the additional rules, exceptions, or constraints resulting from the exploratory data analysis above can be superposed. It is important to correctly define and document them in terms of origin, relevance, and scope.

The resulting model can be applied to different samples of the population (i.e., groups or categories). Running the selected model and adjusting parameters while considering the entire population corresponds to the non-categorical model. In contrast, the categorical model involves separately adjusting parameters for each category or group.

Testing and parameter adjustment

The above-presented mathematical model provides generality, but it is necessary to identify the parameters that will better predict the energy demand. This subsection proposes an iterative search algorithm to identify the set of parameters for a better forecast.

Train-test split approach: To evaluate different parameters for the mathematical formulation, we propose extracting a sample of the entire dataset that corresponds to the most up-to-date observations. The larger portion of the dataset (i.e., the oldest observations) is used as input to train the mathematical model, and the small sample is used to test the accuracy of the predictions. This is known in the literature as a *train-test data split*, a common technique for evaluating the performance of machine learning algorithms ([Medar, Rajpurohit, and Rashmi, 2017](#)).

Error calculations: The train-test split makes it possible to evaluate a prediction using the existing dataset. To quantify the accuracy of the forecast, the percent error of the prediction $\epsilon_{\omega}^{\%}$ for the set of parameters ω is calculated. This is done through the average error between the corresponding forecast number $x_{c,\omega}^{for}$ from the training data and the mathematical model in Equation (1), and the actual observations x_c^{obs} from the testing sample, using Equation (2) for all customers c .

$$\epsilon_{\omega}^{\%} = \sum_{\forall c} \left(\frac{x_{c,\omega}^{for} - x_c^{obs}}{x_c^{obs}} \right) \times 100 \quad (2)$$

Sensitivity analysis: We selected the secant numerical method to find zeros in a discontinuous function. In this search algorithm, given two initial guesses ω_0 and ω_1 , it is possible to iteratively obtain the best-performing value of each parameter in the set ω with a tolerance τ . The equations for the secant search algorithm are as follows:

$$\omega_j = \omega_{j-1} - \epsilon_{\omega_{j-1}}^{\%} \times \frac{\omega_{j-1} - \omega_{j-2}}{\epsilon_{\omega_{j-1}}^{\%} - \epsilon_{\omega_{j-2}}^{\%}} \quad (3)$$

$$\tau \geq \omega_j - \omega_{j-1} \quad (4)$$

Alternatively, the parameter search can be performed using an incremental approach, a bisection method, or inverse quadratic interpolation ([Allen and Isaacson, 2019](#)).

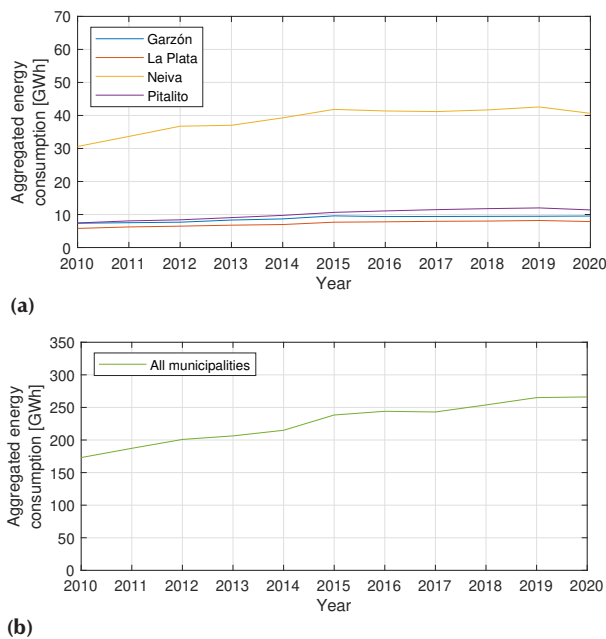


Figure 3. Evolution of the aggregated energy demand in a) the four largest cities of Huila and b) the entire department

Source: Authors

Performance and decision-making

The categorical model is expected to perform better in terms of accuracy and exhibit increased computation times. This is because classification and grouping require additional computation steps, and individually analyzing every group or category also requires additional power and memory compared to the entire population. Given this trade-off between accuracy and processing times, the model designer must evaluate their context, in order to decide between a slower but more accurate categorical model or a non-categorical one with fast solutions but increased error.

Results and simulation

To test the proposed methods, the regional DSO of Huila provided a dataset containing monthly energy consumption readings from its 159 039 electricity customers in the region and other parts of the country over a 10-year period. This section presents the data analysis and simulation results. It was reported that 44 data scientists presented 25 solutions to the prediction problem during the Hackathon Opita Challenge (ElectroHuila S.A. E.S.P., 2021), and that the solution reported in this manuscript resulted in the best forecast after being tested on new data.

First, it is necessary to consider the big picture and analyze the aggregated energy demand for the department of Huila. Considering the entire population, the energy demand grows on a yearly basis (Figure 3). This growth is explained by the population increase and industrialization, and it has been extensively discussed in the literature. Based on this, a forecasting tool can simply identify the slope of the corresponding curves (i.e., the growth rate) and apply it to existing observations in order to predict the demand. Nonetheless, the smooth behavior observed above is only evident when a significant amount of customers is aggregated. To extend on this, Figure 4 shows the average

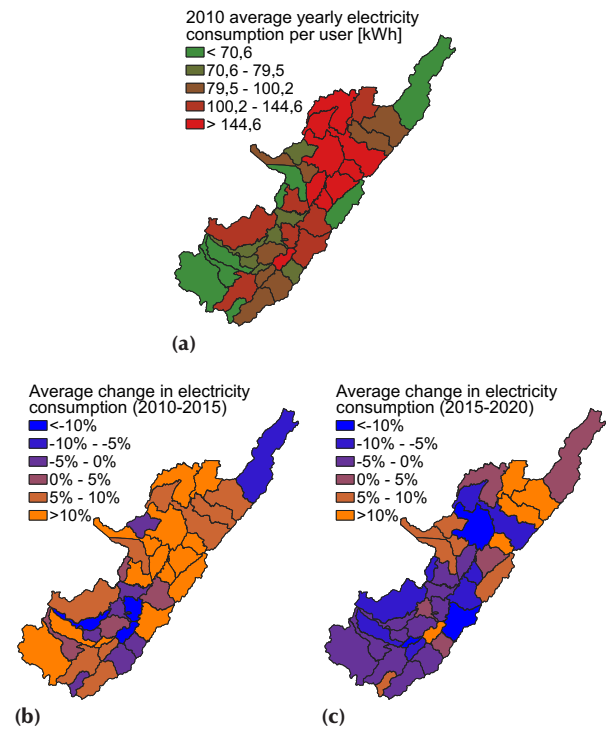


Figure 4. Yearly energy demand in the department of Huila: a) 2010 average per-customer electricity demand, b) change in electricity consumption for the 2010-2015 period, and c) change in electricity consumption for the 2015-2020 period

Source: Authors

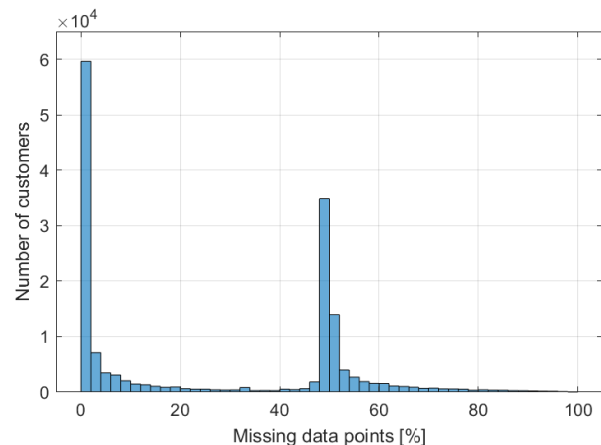


Figure 5. Histogram of missing data. Number of customers with a percentage of missing data points.

Source: Authors

per-customer energy demand for 2010 as well as the 5-year growth seen in 2015 and 2020.

In Figure 4, note that the average per-customer variation in energy consumption cannot be represented by a global increase rate from Figure 3. In the same time horizon, municipalities exhibit increases or decreases amounting to the global aggregated increase. Moreover, when comparing the two time horizons, a municipality can have demand increases over a period and decreases over the next, a behavior that is not reflected in the aggregated curves in Figure 3. The global rate might be useful for a system-wide estimation of future demand, but the stochastic nature of per-customer demand requires the use of a more detailed approach.

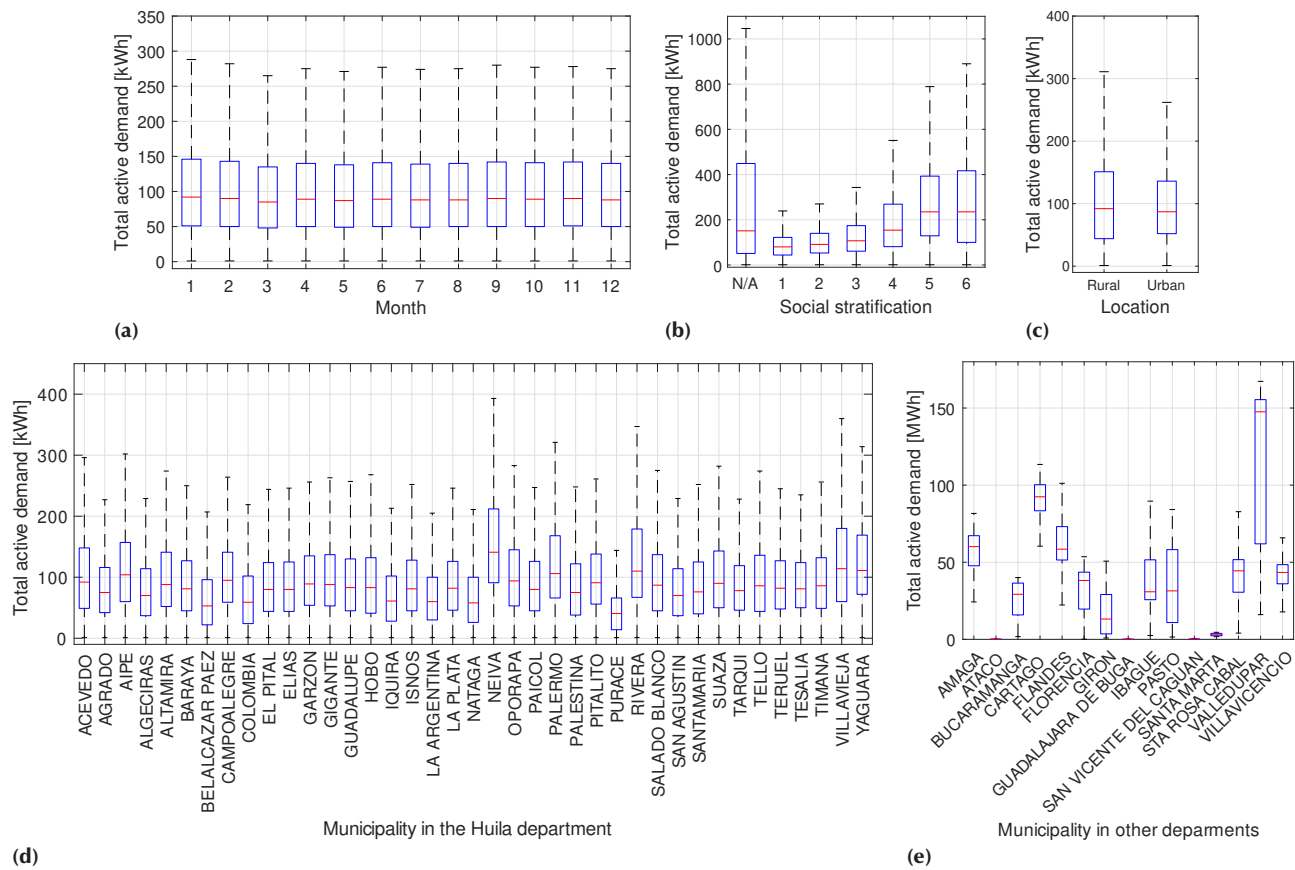


Figure 6. Categorical analysis of the population. Boxplots showing energy demand a) by month, b) by socioeconomic stratum, c) by classification (urban or rural), d) by municipality within the department, and e) by municipality outside the department. The red line represents the median, the blue box represents the 25th and 75th percentiles, and the whiskers represent the lower and upper adjacent points. Outliers are not included.

Source: Authors

Data description and curation

After organizing and filtering the information, it was necessary to curate the dataset. The main modification involved identifying and eliminating repeated data points: out of the initial 17.8 million monthly observations, approximately 29.7% were repeated values. Considering that the amount of customers over a 10-year period should amount to over 19 million observations, 36.4% of the data points were estimated to be missing. This may be due to some new customers appearing later, or some old ones disappearing before the end of the 10-year horizon. Figure 5 presents a histogram with the percentage of missing data points as a proportion between the dates of the first and last observations.

There are two peaks in Figure 5 at 0-5 and 50-55%. This suggests that (i) most customers have at least some data points missing, and (ii) more than 35 000 customers have bimonthly observations, in contrast with the initially reported monthly dataset.

Furthermore, a significant number of customers with monthly observations have between 5 and 20% of missing data points, and those with bi-monthly observations have 55-80%. This highlights the importance of an approach that can deal with data scarcity.

Once the dataset had been homogenized, a categorical analysis was performed. The DSO differentiates customers by social stratum, location (*i.e.*, rural or urban), and municipality. The following questions must be considered:

- Is income disparity relevant?
- Is population density relevant?
- Can data be filtered in groups that exhibit different behaviors?

To evaluate this, Figure 6 shows the boxplots of all the data per category, as well as a monthly consumption analysis.

A relatively small difference between the energy consumption for different times of the year is observed. The months with the largest and smallest average consumptions are January and March, reporting medians of 92 and 85 kWh, respectively. Regarding the proportions of the quartiles in Figure 6a, it is important to note that there are significant variation values, as represented by the whiskers of the boxplot. That is to say that, when considering the entire population, the data variation is significant.

As for the different categories in Figure 6, it is clear that a customer belonging to one or the other exhibits a significantly different behavior. This suggests the value of employing a segregated approach that considers the categorization of the population. The most striking categorical difference is observed in social stratification (Figure 6b), where the median, the 25th and 75th percentiles, and the upper adjacent points of customers categorized as low-income (*i.e.*, social strata 1 and 2) are three times lower than those of high-income customers (social strata 5 and 6). Users that are not categorized into a social stratum are understood to be commercial, industrial,

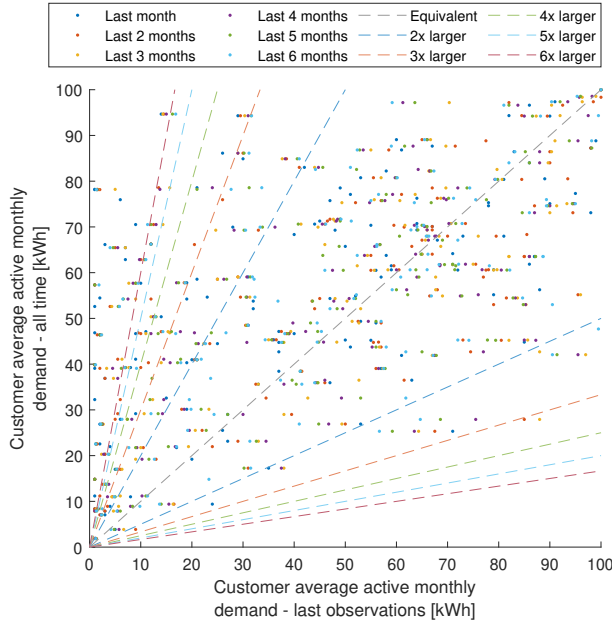


Figure 7. Scatter plot to determine recent changes in consumption patterns. The y-axis presents the average consumption of a customer based on all observations, and the x-axis corresponds to the last few months.

Source: Authors

or agricultural in nature, which explains the high variability in their energy consumption.

Different municipalities exhibit distinct energy consumption patterns. Neiva, the capital of Huila, reports the highest median energy consumption and has the largest observation variability. In addition, note the difference in scales between Figures 6d and 6e. Customers located outside Huila exhibit consumption patterns many times larger than those within the department. While this is not clarified in the documentation of the original dataset, it may be due to the existence of large-scale customers outside the department which are not physically connected to the grids of the DSO but employ it as a service provider.

The last consideration of the data analysis was whether there were exogenous factors to consider in the mathematical models. The assumption made in this work was that consumption patterns may change in the last observations. Figure 7 shows a scatter plot of the average values for the last one to six observations and the average value of all observations, given a random sample of 100 customers.

In this Figure, the diagonal grey-dashed line represents an equivalent value between the last observations and the all-time averages, which means that values close to the diagonal correspond to customers whose consumption patterns have not changed in the last few months. In contrast, the other color-dashed lines (i.e., navy, orange, green, blue, and burgundy) represent how many times larger or smaller the short-term average is relative to the all-time value. Data points located between the blue-dashed line and the orange-dashed one counting counter-clockwise from the diagonal show short-term observations that are two to three times smaller than the all-time average, suggesting a change in consumption patterns.

Note that a significant number of points is located counter-clockwise from the burgundy-dashed line, representing

Algorithm 1 Parameter selection pseudo-code

```

1: Get  $C$  ▷ Number of categories
2: Get  $\omega_0$  and  $\omega_1$  ▷ Two initial parameter guesses
3: Get  $\tau$  ▷ Error tolerance
4: Get  $i_{max}$  ▷ Maximum iterations allowed
5:  $P_{train} = P \cap P_{test}$  ▷ Split population for training-testing
   Calculate forecast for initial guesses (Equation (1)):
6:  $x_{\omega_0}^{for} \leftarrow f(P_{train}, \omega_0)$ 
7:  $x_{\omega_1}^{for} \leftarrow f(P_{train}, \omega_1)$ 
8:  $x_{test}^{obs} \leftarrow P_{test}$  ▷ Get observations from testing population
   Calculate error for initial guesses (Equation (2)):
9:  $\epsilon_0 \leftarrow g(x_{\omega_0}^{for}, x_{test}^{obs})$ 
10:  $\epsilon_1 \leftarrow g(x_{\omega_1}^{for}, x_{test}^{obs})$ 
11: for  $c \leftarrow 1$  to  $C$  do ▷ Do this for each category
12:   for  $i \leftarrow 1$  to  $i_{max}$  do ▷ Do this for each iteration
     Calculate new guess for next iteration (Equation (3)):
13:      $\omega_i \leftarrow h(\omega_{i-1}, \omega_{i-2}, \epsilon_{i-1}, \epsilon_{i-2})$ 
14:      $T \leftarrow \omega_i - \omega_{i-1}$  ▷ Calculate tolerance (Equation (4))
15:     if  $T \leq \tau$  then
16:        $\omega_c \leftarrow \omega_i$  ▷ Store parameters for category  $c$ 
17:       break ▷ Stop the for loop, parameters found
18:     else
19:        $x_{\omega_i}^{for} \leftarrow f(P_{train}, \omega_i)$  ▷ Equation (1) for  $\omega_i$ 
20:        $\epsilon_i \leftarrow g(x_{\omega_i}^{for}, x_{test}^{obs})$  ▷ Equation (2) for  $\omega_i$ 
21:     end if
22:   end for
23: end for

```

short-term observations more than six times smaller than the all-time average. This procedure was followed for five different 100-customer random samples, with equivalent results. While there is no additional information, this evinces that changes in consumption patterns are an important factor for energy forecasting.

The lower portion of Figure 7 (i.e., counting clockwise from the grey-dashed line) includes significantly fewer relative observations. This means that, while it is common for the average consumption to be significantly reduced in the last few months compared to all observations, the opposite is not often the case. Consumption patterns change towards a net decrease, which may be due to short periods of residential vacancy or industry stall periods. Considering that these low-consumption periods are expected to be short in the scale of the dataset (a few months of the 10-year period), any increase in consumption caused by re-occupancy is diluted by a longer average occupancy over previous months.

Based on the exploratory analysis of the information provided by the DSO, we decided to include the monthly median and standard deviation of the user in the mathematical model. These variables represent the trends and noise, respectively. The categorical and non-categorical models were tested to assess the trade-off between accuracy and computational intensity. Each parameter associated with these variables is presented in the next subsection. In addition, an exogenous rule to account for changes in consumption patterns was included: if the average energy consumption of a customer for the last six observations was more than six times larger/smaller than the average of all their observations, a change in consumption pattern

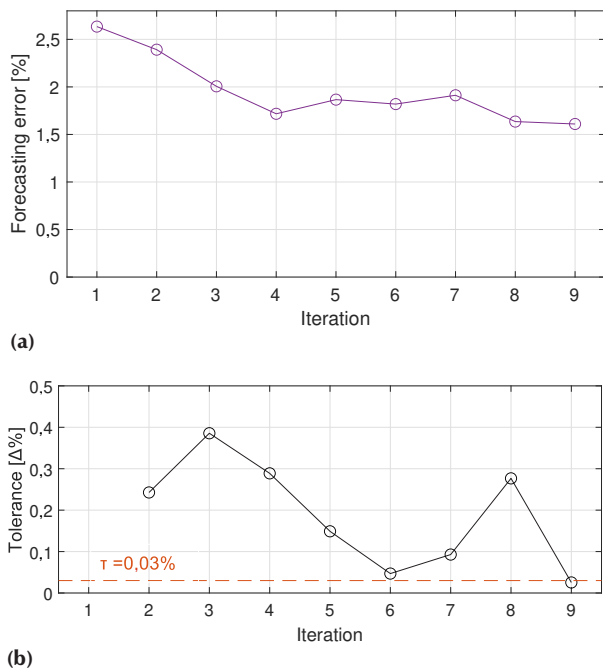


Figure 8. Results for the non-categorical parameter selection algorithm: a) forecast error for the train-test split and b) its corresponding tolerance

Source: Authors

was assumed { these values represent the most critical change observed in Figure 7. In this case, the forecast from the model was replaced by the average of these last six observations.

Parameter adjustment

The parameters for each categorical model (and its relevant categories) were calculated using an iterative search algorithm based on the secant method, whose pseudo-code is presented in Algorithm 1. This algorithm was executed for every categorical division outlined in the previous sections.

Depending on how the population is split, the runtime of Algorithm 1 can be significant. We recommend carefully selecting the size of the testing population. This is especially important for categorical models, as the algorithm includes a for loop to calculate the parameters of each category. Computational intensity is therefore linked to the testing population size and the number of categories.

The least computationally intensive scenario corresponded to the non-categorical model, whose parameters were calculated for the entire population. To illustrate this, Figure 8 presents the error and tolerance of each iteration in obtaining the non-categorical parameters. The testing portion of the population was January, February, and March 2020, and the training portion included every previous observation (i.e., from January 2010 until December 2019).

The behavior in Figure 8 suggests that the search algorithm correctly identifies a solution with a low forecasting error given the selected tolerance. However, note that iteration 6 almost fulfilled the condition $\tau = 0.03\%$. The discontinuous nature of the error function from Equations (1), (2), and (3) leads to locally minimum tolerances and, hence, to premature solutions. Future developments of this work

should include additional iterations to confirm whether there is a nearby local minimum with a better performance.

Algorithm selection

The algorithm was coded in the SciLab 6.1.1 open-access software and run using a desktop with 16 GB RAM and an Intel Core i5-8400 CPU. The accuracy and computational intensity resulting from the train-test split analysis are shown in Table 1.

Table 1. Comparison of the results obtained from the train-test split for different categories

Categorical Division	Num. Categ.	Total Iter.	Run time [h] Train	Run time [h] Test	Train-test Error [%]
None	1	9	14.5	1.73	1.61
Social stratum	7	39	61.4	5.72	1.59
Location	2	15	23.13	2.56	1.64
In/out of Huila	2	23	36.88	3.14	1.58
Municipality	54	342	457.2*	37.6	1.43

* This long parameter search was run for two categories at a time.

Source: Authors

The previously hypothesized trade-off between computational intensity and accuracy was confirmed by the results in Table 1. The best-performing algorithm involves segregating parameters for each municipality, but this comes at a very high computational cost: 19 days of training time. Assigning equal value to accuracy and computational intensity, it is possible to compare the relative computational cost per error unit. In the design constraints, the DSO specified that computational intensity was an important decision variable, so the non-categorical model was selected, with an average error of 1.61% in the train-test simulation and a total training time of 14.5 hours.

The methods presented in this work provide the DSO with the flexibility to decide which categorical division to select according to its particular needs regarding accuracy and computational intensity. This is reserved as a decision variable for future applications of this work.

Discussion

The review by Klyuev *et al.* (2022) shows that the absolute percentage error of medium-term electricity demand predictions oscillates between 7.95 and 15.92%. For the sake of comparison, the absolute percent errors reported by other authors are compared against our results in Table 2. The authors referenced therein did not report execution times, which are difficult to benchmark in any case because each study uses different population sizes, has varying degrees of missing data, and considers more or less historical data.

This tool can be extended for application by other regional and national DSOs. Considering the computational constraints, it is possible to obtain a more accurate solution via the train-test split of historical data. This is also determined by the purpose of the forecast (e.g., if it aims to schedule and prioritize infrastructure upgrades, accuracy is more important than computational intensity) (Migliavacca *et al.*, 2021).

The work of Schafer (1999) suggests that missing data amounting to 5% or less can be tolerated in statistical

Table 2. Benchmarking against other methods in the literature

Reference	Method	Dataset (granularity)	Reported Error [%]
Proposed	Heuristics	10 years (1-2 month)	1.61
(Yuce <i>et al.</i> , 2017)	Deep ML	1.5 years (not reported)	4.5-8.8
(Meng <i>et al.</i> , 2011)	Feature extraction	2 years (1 month)	2.7-2.8
(Amber <i>et al.</i> , 2017)	Multiple regression	5 years (not reported)	8.6-9.8

* ML = machine learning.

Source: Authors

analysis. Furthermore, Bennett (2001) argues that more than 10% of missing data points causes any statistical analysis to be biased. Thereupon, as future work, we propose the use of regression to fill large gaps in energy consumption datasets (e.g., there are more than 36% missing data points in the set used in this work). Numerous approaches for data regression are suggested in Dong and Peng (2013) in relation to the missing data mechanism or its origin.

The historical consumption dataset used in this article includes one temporal (month), two spatial (municipality and urban/rural location), and one socioeconomic feature (social stratum) for individual consumers. Note that the social stratum is an official classification implemented by the Colombian government to group dwellings with similar social and economic characteristics within a delimited area, and it does not include useful information about individual household characteristics (Chica-Olmo, Sánchez, and Sepúlveda-Murillo, 2020). In this regard, we propose treating social and economic factors independently and on a larger spectrum, i.e., economic factors such as family employment, household income, debt capacity, and savings; and social/demographic factors like family size, behavior towards the environment, and population density, among others. These can be used to build and select models that better represent individual behavior (Yuce *et al.*, 2017). Due to lack of data, this will be studied in future work.

Future applications of this work could include the definition of dynamic tariffs that account not only for generation resource availability but also for forecast demand scenarios (Ochoa, Dent, and Harrison, 2010). Flexibility resources, demand response, and infrastructure upgrades could be optimally planned if the future demand is known (Capitanescu, Ochoa, Margossian, and Hatziaargyriou, 2015). The hosting capacity of distribution networks largely depends on operational states that rely on an accurate representation of future demand (Ochoa *et al.*, 2010). Ultimately, a good demand prediction is key to assess the reliability and resilience of modern distribution grids (Escalera, Hayes, and Prodanović, 2018).

Conclusions

This manuscript presents a historical demand data processing framework for medium-term electricity consumption forecasting. Qualitative and quantitative patterns were abstracted to build a mathematical model, which was

later tested given different categorical divisions. The best-performing method was selected while considering the trade-off between accuracy and computational intensity. This method was tested using real data from the regional DSO of the department of Huila, Colombia. This DSO reported that applying the selected algorithm resulted in a forecast that was at least 3% more accurate than other solutions regarding the real customer demand (i.e., data not used to train or test the algorithm).

Through heuristics, simple statistical quantities of the population and samples of it can be used to build a robust model with accurate outputs. It was found that it is necessary to account for exogenous factors. In this case, changes in consumption patterns played an important role in forecasting energy demand.

Opportunities for future work include filling the gaps in the dataset, especially considering that a significant amount of customers exhibit a bimonthly cycle of observations. Moreover, additional categories could be created for data filtering. By mixing two categories, subcategories that better inform customer behavior could be obtained (e.g., by mixing social stratification and the type of location, consumption patterns can be better represented). However, this would be limited by computation time constraints, as it represents the inclusion of subcategories, which would considerably increase the size of the problem.

Acknowledgments

The authors would like to give special thanks to Electrohuila SA ESP for providing the data used in this work for testing and validation, and for funding and organizing the Hackathon Opita Challenge of 2021, from which the work presented herein is a major output. Special thanks to M. Rouillé-Tamayo for her support, as well as to Centro Internacional de Física and EHS Ltda. in Colombia for their financial and infrastructural support in writing this manuscript.

Author contributions

Cuenca J. conceived the idea and did the background research. Palacios-Castro D. collected the data. García R. provided critical feedback. Cuenca J. led the writing process and wrote the main part of the manuscript, to which all authors contributed.

Conflicts of interest

There are no conflicts of interest to declare.


References

- Allen, M., and Isaacson, E. (2019). *Numerical analysis for applied science*. Wiley. <https://books.google.be/books?id=PpB9cj0xQAQC>.
- Amber, K. P., Aslam, M. W., Mahmood, A., Kousar, A., Younis, M. Y., Akbar, B., ... Hussain, S. H. (2017). Energy consumption forecasting for university sector buildings. *Energies*, 10(10). <https://doi.org/10.3390/en10101579>
- Bennett, D. A. (2001). How can i deal with missing data in my study? *Australian and New Zealand Journal of Public Health*, 25(5), 464-469. <https://doi.org/10.1111/j.1467-842X.2001.tb00294.x>

- Biel, K., and Glock, C. H. (2016). Systematic literature review of decision support models for energy-efficient production planning. *Computers & Industrial Engineering*, 101, 243-259. <https://doi.org/10.1016/j.cie.2016.08.021>
- Bimenyimana, S., and Asemota, G. N. O. (2018). Traditional vs smart electricity metering systems: A brief overview. *Journal of Marketing and Consumer Research*, 46, 1-7. <https://www.iiste.org/Journals/index.php/JMCR/article/view/42505/43773>. (Accessed: 2023-03-02)
- Bunn, D., and Farmer, E. (1985). *Comparative models for electrical load forecasting*. Wiley. <https://www.osti.gov/biblio/6256333>. (Accessed: 2023-03-02)
- Capitanescu, F., Ochoa, L. F., Margossian, H., and Hatziaargyriou, N. D. (2015). Assessing the potential of network reconfiguration to improve distributed generation hosting capacity in active distribution systems. *IEEE Transactions on Power Systems*, 30(1), 346-356. <https://doi.org/10.1109/TPWRS.2014.2320895>
- Chica-Olmo, J., Sánchez, A., and Sepúlveda-Murillo, F. H. (2020). Assessing colombia's policy of socio-economic stratification: An intra-city study of self-reported quality of life. *Cities*, 97, 102560. <https://www.sciencedirect.com/science/article/pii/S0264275119312995>. <https://doi.org/10.1016/j.cities.2019.102560>
- Cuenca, J. J., and Hayes, B. P. (2022). Non-bias allocation of export capacity for distribution network planning with high distributed energy resource integration. *IEEE Transactions on Power Systems*, 37(4), 3026-3035. <https://doi.org/10.1109/TPWRS.2021.3124999>
- Cuenca, J. J., Jamil, E., and Hayes, B. P. (2023). Revenue-based allocation of electricity network charges for future distribution networks. *IEEE Transactions on Power Systems*, 38(2), 1728-1738. <https://doi.org/10.1109/TPWRS.2022.3176186>
- Dong, Y., and Peng, C.-Y. J. (2013, May 14). Principled missing data methods for researchers. *SpringerPlus*, 2(1), 222. <https://doi.org/10.1186/2193-1801-2-222>
- ElectroHuila S.A. E.S.P. (2021). *Hackatón OpitaChallenge*. <https://reto.electrohuila.com.co/>. (Accessed: 2023-03-02)
- Escalera, A., Hayes, B., and Prodanović, M. (2018). A survey of reliability assessment techniques for modern distribution networks. *Renewable and Sustainable Energy Reviews*, 91, 344-357. <https://doi.org/10.1016/j.rser.2018.02.031>
- Ghoddusi, H., Creamer, G. G., and Rafizadeh, N. (2019). Machine learning in energy economics and finance: A review. *Energy Economics*, 81, 709-727. <https://doi.org/10.1016/j.eneco.2019.05.006>
- Gnatyuk, V. I., Polevoy, S. A., Kivchun, O. R., and Lutsenko, D. V. (2020, apr). Applying the potentiating procedure for optimal management of power consumption of technocenose. *IOP Conference Series: Materials Science and Engineering*, 837(1), 012001. <https://doi.org/10.1088/1757-899X/837/1/012001>
- Hemmati, R., Hooshmand, R.-A., and Taheri, N. (2015). Distribution network expansion planning and dg placement in the presence of uncertainties. *International Journal of Electrical Power & Energy Systems*, 73, 665-673. <https://doi.org/10.1016/j.ijepes.2015.05.024>
- Honarmand, M. E., Hosseinneshad, V., Hayes, B., Shafie-Khah, M., and Siano, P. (2021). An overview of demand response: From its origins to the smart energy community. *IEEE Access*, 9, 96851-96876. <https://doi.org/10.1109/ACCESS.2021.3094090>
- Hong, T., and Fan, S. (2016). Probabilistic electric load forecasting: A tutorial review. *International Journal of Forecasting*, 32(3), 914-938. <https://doi.org/10.1016/j.ijforecast.2015.11.011>
- Klyuev, R. V., Morgoev, I. D., Morgoeva, A. D., Gavrina, O. A., Martyushev, N. V., Efremenko, E. A., and Mengxu, Q. (2022). Methods of forecasting electric energy consumption: A literature review. *Energies*, 15(23). <https://doi.org/10.3390/en15238919>
- Lindsey, J. K. (2004). *Statistical analysis of stochastic processes in time*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511617164>
- Medar, R., Rajpurohit, V. S., and Rashmi, B. (2017). Impact of training and testing data splits on accuracy of time series forecasting in machine learning [conference paper]. In *2017 international conference on computing, communication, control and automation (icccubea)*. <https://doi.org/10.1109/ICCUBEA.2017.8463779>
- Mehigan, L., Zehir, M. A., Cuenca, J. J., Sengor, I., Geaney, C., and Hayes, B. P. (2022). Synergies between low carbon technologies in a large-scale mv/lv distribution system. *IEEE Access*, 10, 88655-88666. <https://doi.org/10.1109/ACCESS.2022.3199872>
- Meng, M., Niu, D., and Sun, W. (2011). Forecasting monthly electric energy consumption using feature extraction. *Energies*, 4(10), 1495-1507. <https://doi.org/10.3390/en4101495>
- Migliavacca, G., Rossi, M., Siface, D., Marzoli, M., Ergun, H., Rodríguez-Sánchez, R., ... Morch, A. (2021). The innovative flexplan grid-planning methodology: How storage and flexible resources could help in debottlenecking the european system. *Energies*, 14(4). <https://doi.org/10.3390/en14041194>
- Ochoa, L. F., Dent, C. J., and Harrison, G. P. (2010). Distribution network capacity assessment: Variable dg and active networks. *IEEE Transactions on Power Systems*, 25(1), 87-95. <https://doi.org/10.1109/TPWRS.2009.2031223>
- Schafer, J. L. (1999). Multiple imputation: A primer. *Statistical Methods in Medical Research*, 8(1), 3-15. <https://doi.org/10.1191/096228099671525676>
- Shumilova, G., Gottman, N., and Starceva, T. (2008). Forecasting of electrical loads in the operational management of electric power systems based on neural network structures. *KNC UrO RAS: Syktyvkar, Russia*, 85.
- vom Scheidt, F., Medinová, H., Ludwig, N., Richter, B., Staudt, P., and Weinhardt, C. (2020). Data analytics in the electricity sector. a quantitative and qualitative literature review. *Energy and AI*, 1, 100009. <https://doi.org/10.1016/j.egyai.2020.100009>
- Wei, N., Li, C., Peng, X., Zeng, F., and Lu, X. (2019). Conventional models and artificial intelligence-based models for energy consumption forecasting: A review. *Journal of Petroleum Science and Engineering*, 181, 106187. <https://doi.org/10.1016/j.petrol.2019.106187>
- Yuce, B., Mourshed, M., and Rezgui, Y. (2017). A smart forecasting approach to district energy management. *Energies*, 10(8). <https://doi.org/10.3390/en10081073>

Genetic Algorithm-Based Optimization of Solar Photovoltaic Integration and Demand Response for CO₂ Reduction in Indian Coal Power

Optimización basada en algoritmos genéticos de integración de energía solar fotovoltaica y respuesta a la demanda para la reducción de CO₂ en la energía de carbón de la India

Vivek Saxena ¹ and Saurabh Kumar Rajputo ²

ABSTRACT

In 2022, global coal combustion contributed significantly to global pollution, producing 15.22 billion metric tons of carbon dioxide (CO₂). This research addresses the urgent challenge of mitigating CO₂ emissions in Indian coal power plants by strategically deploying solar photovoltaic (PV) systems and integrating demand response mechanisms. The imperative to reduce greenhouse gas emissions from coal-based electricity generation underscores the critical context of climate change. Emphasizing the vital role of integrating renewable energy-based distributed generators into the existing coal infrastructure, this study positions solar PV technology as a promising solution. Optimal solar PV system allocation is achieved through the implementation of the genetic algorithm technique. Factors such as solar resource availability, electricity demand patterns, and the CO₂ intensity associated with coal power generation are considered in this process. The primary research objective is twofold: to minimize CO₂ emissions and maximize the integration of solar PV systems while mitigating power losses. The proposed approach considers the intermittent nature of solar power and the dynamic characteristics of demand. Rigorous testing on an IEEE 33-bus system powered by the studied coal power plant reveals a substantial 29.31% reduction in CO₂ generation following the implementation of the proposed strategy. This research represents a decisive step towards fostering a more sustainable and environmentally friendly energy landscape. Our study's outcomes offer valuable insights for policymakers and stakeholders in the energy sector, providing a robust foundation for the advancement of environmentally conscious practices within the coal power industry.

Keywords: bi-level optimization, distribution network, power quality, renewable energy

RESUMEN

En 2022, la combustión global de carbón contribuyó significativamente a la contaminación mundial, produciendo 15.22 mil millones de toneladas métricas de dióxido de carbono (CO₂). Esta investigación aborda el desafío urgente de mitigar las emisiones de CO₂ en las plantas de energía de carbón en India mediante el despliegue estratégico de sistemas solares fotovoltaicos (FV) y la integración de mecanismos de respuesta a la demanda. La necesidad imperiosa de reducir las emisiones de gases de efecto invernadero derivadas de la generación eléctrica a base de carbón subraya el contexto crítico del cambio climático. Destacando el papel esencial de integrar generadores distribuidos basados en energías renovables en la infraestructura de carbón existente, este estudio posiciona la tecnología solar FV como una solución prometedora. La asignación óptima de sistemas solares FV se logra mediante la implementación de la técnica de algoritmo genético. En este proceso se consideran factores como la disponibilidad de recursos solares, los patrones de demanda eléctrica y la intensidad de CO₂ asociada a la generación de energía por carbón. El objetivo principal de la investigación es doble: minimizar las emisiones de CO₂ y maximizar la integración de sistemas solares FV mientras se mitigan las pérdidas de energía. El enfoque propuesto tiene en cuenta la naturaleza intermitente de la energía solar y las características dinámicas de la demanda. Pruebas rigurosas en un sistema IEEE de 33 nodos alimentado por la planta de energía de carbón estudiada revelan una reducción sustancial del 29.31 % en la generación de CO₂ tras la implementación de la estrategia propuesta. Esta investigación representa un paso decisivo hacia la promoción de un panorama energético más sostenible y respetuoso con el medio ambiente. Los resultados de nuestro estudio ofrecen valiosos conocimientos para los formuladores de políticas y las partes interesadas del sector energético, proporcionando una base sólida para el avance de prácticas ambientalmente responsables dentro de la industria de energía a base de carbón.

Palabras clave: optimización bifásica, red de distribución, calidad de energía, energía renovable

Received: September 22nd, 2023

Accepted: September 13th, 2024

¹ Associate professor, Electrical and Electronics Engineering Department, ABES Engineering College, Ghaziabad, U.P., India. Email: vvsaxena1234@gmail.com

² Assistant professor, Centre for Internet of Things, Madhav Institute of Technology and Science, Gwalior, India. Email: saurabh9march@gmail.com



Attribution 4.0 International (CC BY 4.0) Share - Adapt

Acronyms

$P_{L(t)}$	Power transmission losses
$P_{i(t)}$	Real power at the i^{th} node at any time t
$P_{j(t)}$	Real power at the j^{th} node at any time t
$Q_{i(t)}$	Reactive power at the i^{th} node at any time t
$Q_{j(t)}$	Reactive power at the j^{th} node at any time t
$V_{i(t)}$	Voltage at the i^{th} node at any time t
$V_{j(t)}$	Voltage at the j^{th} node at any time t
r_{ij}	Resistance of the branch between the i^{th} and j^{th} nodes
$\delta_{i(t)}$	Angle of the voltage at the i^{th} node
$\delta_{j(t)}$	Angle of the voltage at the j^{th} node
$P_R(t)$	Reverse power at time t
$I_{G(t)}$	Current from the grid at time t
$V_{G(t)}$	Voltage of the grid at time t
I_S	Designated reverse current limit
$V_D(t)$	Penalty for voltage deviations
$V_{\text{Max.}}$	Maximum permissible voltage at the nodes
V_{Min}	Minimum permissible voltage at the nodes
$P_{Gi(t)}$	Real power generation at the i^{th} node for the time period t
$P_{Di(t)}$	Real power demand for the time period t
$Q_{Gi(t)}$	Reactive power generation at the i^{th} node for the time period t
$Q_{Di(t)}$	Reactive power demand for the time period t
$P_{in,i(t)}$	Nonreceptive load at time t
$P_{el,i(t)}$	Receptive load at time t
E_i^{Total}	Energy demand per day
$L_{d,i(t)}$	Load per hour for the time period t
$P_{DG,i}$	Real power injection by DG
P_{DG}^{max}	Maximum value of real power generation by DG
$I_{ij(t)}$	Current flowing between the i^{th} and j^{th} nodes at t
I_{ij}^{max}	Maximum permissible current value
Y_{ij}	Admittance matrix between the i^{th} and j^{th} nodes
θ_{ij}	Angle of impedance between the i^{th} and j^{th} nodes
I_{sm}	Solar PV system current
$S_r(t)$	Solar irradiance at time t
S_r^r	Rated value of solar irradiance for PV systems

Introduction

Electricity plays a pivotal role across diverse sectors, underpinning industrial processes, construction activities, and daily life. However, its production predominantly relies on energy sources like coal, natural gas, uranium, and renewable resources such as solar, wind, and hydropower. The environmental footprint of electricity generation is substantial, notably contributing to global CO₂ emissions. The appropriate selection of energy generation technology is paramount in mitigating these environmental consequences, especially considering that coal, a high carbon emitter, occupies one end of the spectrum.

China, a global industrial powerhouse, relies heavily on coal-based energy, with the power industry accounting for a substantial 40% of the annual CO₂ emissions. Critiquing current carbon emission calculations, a study proposed an innovative radial basis function neural network model for enhanced accuracy in predicting emissions from coal-fired power plants, thus addressing a crucial aspect of environmental impact assessment (Cheng *et al.*, 2023).

Coal-fired captive power plants, which are crucial for industrial cost savings, are confronted with significant carbon emissions. To quantify and improve low-carbon development, The study by Ma *et al.* (2022) introduced source-network-load interactive evaluation indicators, providing a scientific approach to this issue. This encourages grid participation, supports sustainable energy utilization, and aligns with global efforts towards environmental stewardship.

Samanta *et al.* (2015) proposed a partial repowering approach for a 250 MW CPP, which involved the removal of two coal mills, the utilization of a pressurized combustion chamber, and the optimization of waste heat. This innovative strategy resulted in a substantial 30.7% increase in capacity and efficiency, along with a significant reduction in CO₂ emissions (26.5%).

In China, a pivotal shift is occurring in the CPP, driven by internal improvements and external factors such as increased renewable energy integration. Zhang *et al.* (2020) conducted a comprehensive study, covering 99.7% of the operational plants, and emphasized the potential for a 265 Mt CO₂eq reduction by 2020. This study highlights regional variations and recommends tailored post-2020 decarbonization strategies, showcasing the complexity of policy effectiveness in mitigating emissions.

Addressing the challenge of harmful gas emissions from coal power plants, Smaisim *et al.* (2023) explored the integration of renewable sources such as molten carbonate fuel cells and solar farms. Their simulation results demonstrated promising energy outputs and reduced environmental impact, highlighting the potential of integrating renewables to enhance overall efficiency.

In the field of CO₂ capture, some authors have investigated energy-saving mechanisms through chemical absorption, revealing two waste heat recovery techniques while showcasing significant energy savings of 9.32 and 8.71% through optimization and heat recovery (Akbari *et al.*, 2022; Eslami *et al.*, 2011).

The urgency for developing clean coal technologies was underscored by Hanak *et al.* (2015), who modeled the substitution of ammonia in CO₂ capture within a supercritical coal-fired plant. The study revealed efficiency penalties ranging from 8.7 to 10.9%, emphasizing the importance of ongoing efforts to meet the EU 2050 greenhouse gas reduction target.

Distributed generators (DGs) and demand response (DR) are pivotal segments in implementing smart grid systems. DGs can be categorized into renewable and non-renewable sources, with renewable DG including solar PV systems, wind turbines, hydroelectric plants, and biomass generators, relying on naturally replenished resources. The non-renewable category includes sources like diesel generators, natural gas turbines, and fuel cells, which depend on finite fossil fuel reserves. This distinction is crucial when evaluating sustainable options for distributed energy generation.

The optimal allocation of renewable energy based DG within the distribution network (DN) hinges on factors such as load location, solar resource availability, and DN capacity. DR, characterized by consumers adjusting electricity consumption in response to cost variations, offers benefits such as peak demand reduction, enhanced grid reliability, and decreased reliance on costly infrastructure investments (Saxena *et al.*, 2021a; Yaghoubi *et al.*, 2022).

There are several methodologies to analyze how DR influences the optimal placement of solar PV systems in the DN. Incorporating DR and solar PV systems into smart grids makes it possible to optimize PV placement, enhance grid performance, and reduce infrastructure strain. The interplay between DR and solar PV generation constitutes a crucial area of research aiming for more sustainable and efficient energy systems (Saxena *et al.*, 2022a; Eslami *et al.*, 2012).

Zhong *et al.* (2021) argued that achieving sustainable development necessitates a shift towards a low-carbon economy and a reduction in energy consumption, addressing the mounting energy crises and environmental imperatives. The adoption of DGs is gaining momentum, offering not only economic benefits but also heightened system adaptability. Optimal grid-connection strategies for DGs can lead to diminished carbon emissions and reduced operating costs. In this vein, this paper introduces an algorithm designed to determine the location and scale of DGs within distribution networks to facilitate low-carbon practices.

Power prediction in PV applications plays a crucial role in optimizing energy management and integration with the grid (Al-Dahidi *et al.*, 2018). Accurate PV power forecasting

helps to balance supply and demand, improving the stability of power systems and reducing reliance on fossil fuels. Various machine learning techniques have been employed to enhance prediction accuracy (Al-Dahidi *et al.*, 2019; Alrabai *et al.*, 2022). These models account for factors like irradiance, temperature, and weather conditions, significantly impacting PV performance. The integration of such predictive models is vital for the efficient planning and operation of renewable energy systems.

Yoon *et al.* (2022) demonstrated a sophisticated system that collaborates seamlessly with distributed energy generation and storage technologies to efficiently deliver energy to users, addressing energy consumption demands. The study of model design and outcomes for regulating energy supply with a focus on carbon reduction encompasses three distinct perspectives. Firstly, an operational model is presented for an extensive examination of small-scale energy generation with minimal carbon footprints, particularly within the context of energy storage system-integrated, distributed power. Secondly, a novel supply system, attuned to fluctuations in energy demand, is outlined within a scheduling framework. Furthermore, the aim is to attain both energy self-sufficiency and carbon neutrality. This objective is pursued by overseeing and governing carbon emissions at the urban scale, facilitated by the optimized operation of DR.

In PV applications, power prediction is crucial for optimizing energy storage systems, scheduling maintenance, and enhancing overall energy production efficiency (Al-Ghussain *et al.*, 2023). Hybrid models combining machine learning with physical models are also gaining attention, as they incorporate real-time data and predictive analytics (Ayadai *et al.*, 2022). Moreover, advancements in data-driven techniques enable more precise short-term and long-term power forecasting, which is critical for both grid operators and energy providers (Al-Dahidi *et al.*, 2024). PV power prediction plays a key role in maximizing renewable energy utilization and mitigating the challenges associated with variable solar output, contributing to more sustainable and reliable energy systems.

Viana *et al.* (2018) discussed the potential of DR and DG in facilitating sustainable DN through the active involvement of end-users. The study suggested the need for regulatory changes, e.g., implementing optional time-of-use tariffs, to simultaneously enhance both DR and PVDG.

Shirazi *et al.* (2021) pointed out that the forthcoming era of intelligent microgrids holds the promise of heightened stability and resilience, in conjunction with the astute planning, precise control, and adept administration of DGs comprising wind, solar, and diesel generators. The synergistic utilization of wind and solar energy sources emerges as a potent avenue for mitigating pollution. However, orchestrating the management, scaling, and strategic siting of DGs within the electricity grid represents a formidable endeavor that is replete with challenges. This paper introduces a multi-objective model grounded

in gray wolf optimization, purpose-built to enable the adequate placement of DG within intelligent microgrids (MG). The primary objectives encompass the reduction of financial costs alongside environmental impacts, spanning greenhouse gas emissions and overall pollution levels. Validation using the IEEE 30 system robustly attested to the efficacy of the proposed model, simulation outcomes unequivocally demonstrated the cost-effectiveness of the approach, effectively pinpointing the optimal DG location while entailing the smallest environmental footprint conceivable.

Wang *et al.* (2021) suggested that, in the face of escalating climate concerns, attaining carbon peak and carbon neutrality has emerged as a paramount endeavor. The pervasive integration of distributed energy resources is exerting its influence across all sectors of society. Effectively harnessing and adeptly managing this environmentally beneficial supply plays a pivotal role in curtailing carbon emissions. To this effect, this study underscores the vital importance of dispersed energy resources and their proficient management in CO₂ reduction. By incorporating the computation of carbon emission flow into the optimization of distributed energy management, the objective is to enhance emission flow dynamics, ultimately leading to more significant reductions in CO₂ emissions.

Statistics show that the energy sector, primarily due to coal burning, is responsible for 70% of greenhouse gas (GHG) emissions. To address this issue, distribution companies should utilize low-emission-coefficient generators as DG units. This reduces reliance on thermal plants, leading to decreased coal burning and GHG emissions. Therefore, a robust computational algorithm is needed for optimal DG utilization in active distribution systems which focuses on reducing emissions (Lakshmi 2023).

The Central Electricity Authority (CEA) reported a significant CO₂ emission rate of 0.975 tCO₂/MWh for coal-based power plants (CEA, 2023). The optimal placement of renewable DGs in the DN offers a promising avenue to reduce these emissions. In this context, solar PV systems have gained prominence as a pollution-free form of electricity generation. The lifetime CO₂ emission for PV modules is estimated at 0.053 kg per 1 kWh of electrical energy (Rajput, 2022).

The CEA reported changes from the financial year (FY) 2000-01 to 2021-22. During this period, there was a substantial increase in coal-based capacity additions up to FY 2015-16, followed by a significant decline. In a similar vein, within the Indian grid, hydro-based capacity additions decreased starting from 2017-18.

Despite global efforts to transition towards cleaner energy sources, there is a notable research gap with regard to the reduction of CO₂ emissions from coal-based electricity generation in India. The unique challenges and opportunities posed by the integration of solar PV systems and demand

response mechanisms into existing coal infrastructure constituted the focal point of this research.

The hypothesis underlying this study posits that the strategic deployment of solar PV systems, coupled with the integration of demand response mechanisms, can significantly reduce CO₂ emissions in Indian coal power plants. The hypothesis is grounded in the belief that renewable energy-based DGs, specifically solar PV technology, can play a pivotal role in transforming the environmental impact of coal power.

The primary research objectives are twofold:

To minimize CO₂ emissions. To develop and implement strategies to minimize CO₂ emissions from Indian coal power plants through the strategic deployment of solar PV systems.

To maximize solar PV integration. To optimize the integration of solar PV technology into the existing coal infrastructure while mitigating power losses and considering the intermittent nature of solar power and the dynamic characteristics of demand.

This paper is structured to provide a systematic exploration of the research topic. It unfolds through distinct sections, starting with a systematic literature review that lays the foundations of the research by surveying the existing knowledge in a methodical manner. The subsequent section states the problem, intricately detailing the formulation of the fitness function and constraints that govern the research inquiry. This establishes a clear framework for the study.

The paper then transitions into a discussion on the test system, where the methodologies selected for experimentation and analysis are outlined. The methodology section describes the systematic approach undertaken to address the research objectives, ensuring transparency and reproducibility. Implementation details follow, providing insights into the practical application of the proposed methodologies.

The core of the paper lies in the results section, where the findings are presented and analyzed. This section encapsulates the outcomes of the research, validating the proposed methodologies and shedding light on their efficacy. Following this, the paper culminates in a robust conclusion and discussion. The conclusion succinctly summarizes the key insights and contributions, while the discussion section delves into the broader implications of the findings, opening avenues for future research.

Problem statement

A substantial reduction in CO₂ emissions can be achieved by reducing reliance on electricity generated from conventional coal-fired power plants (CPPs). In light of this, this paper sets forth the following objectives to actualize the proposed framework:

Power loss minimization

The efficient operation of a power system hinges on the minimization of transmission losses within the distribution network. These losses stem from the inherent resistance in transmission wires, leading to voltage drops and energy dissipation during electricity transmission from the source to the consumers. Hence, one of our primary objectives is to minimize power losses, which is defined as follows (Meena *et al.* 2018):

$$\mathcal{L}_1 = \sum_{t=1}^{24} P_{L(t)} \quad (1)$$

$$P_{L(t)} = \sum_{i=1}^N \sum_{j=1}^N \alpha_{ij(t)} (P_{i(t)} P_{j(t)} + Q_{i(t)} Q_{j(t)}) + \beta_{ij(t)} (Q_{i(t)} P_{j(t)} - P_{i(t)} Q_{j(t)}) \quad \forall t \quad (2)$$

where

$$\alpha_{ij(t)} = r_{ij} \cos(\delta_{i(t)} - \delta_{j(t)}) / V_{i(t)} V_{j(t)}$$

and

$$\beta_{ij(t)} = r_{ij} \sin(\delta_{i(t)} - \delta_{j(t)}) / V_{i(t)} V_{j(t)}$$

Managing reverse power flow

Reverse power flow occurs when DG units generate surplus power beyond the local demands, feeding excess power back into the grid. This phenomenon can introduce stability and safety concerns within the DN, including voltage fluctuations and potential equipment damage. To mitigate these issues, our objective is to manage reverse power flow efficiently:

$$\mathcal{L}_2 = \sum_{t=1}^{24} P_{R(t)} \quad (3)$$

$$P_{R(t)} = \begin{cases} 0, & \text{if } I_{G(t)} \geq I_s \\ \text{Re}(V_{G(t)} I_{G(t)}^*) & \text{if } I_{G(t)} < I_s. \end{cases} \quad (4)$$

Nodal voltage deviation control

Nodal voltage deviation signifies the variance between actual and standard voltage levels at specific nodes within the power system. Voltage deviations can result from various factors, including load fluctuations, reactive power flow, and line losses. These deviations can disrupt the system's efficiency, increase losses, and potentially damage equipment. To address this issue, our objective is to control voltage deviations:

$$\mathcal{L}_3 = \left(1 + \sum_{t=1}^{24} V_{D(t)} \right) \quad (5)$$

$$V_{D(t)} = \begin{cases} |V_{\text{Min}} - V_{i(t)}| & \text{if } V_{i(t)} < V_{\text{Min}}. \\ 0 & \text{if } V_{\text{Min}} \leq V_{i(t)} \leq V_{\text{Max}}. \\ \ell & \text{if } V_{i(t)} > V_{\text{Max}}. \end{cases} \quad (6)$$

where ℓ is the unacceptable value.

Fitness function

To optimize these objectives, a fitness function encompassing distinct objective functions with weighting factors is required. Below is the fitness function (Υ_1) for level-1 optimization:

$$\min(\Upsilon_1) = \varphi \times M \times \mathcal{L}_3 \quad (7)$$

where $M = \mathcal{L}_1 + \mathcal{L}_2$, and φ denotes the daily to yearly conversion product.

DR planning and DG scheduling are considered at level 2 of the optimization objectives. The objective function for level 2 of the optimization problem is as follows:

$$\min(\Upsilon_2) = M \times \mathcal{L}_3 \quad (8)$$

where Υ_2 is the fitness function for level 2.

Demand response aggregator

A demand response aggregator (DRA) plays a pivotal role in managing energy consumption during peak demand periods. It collaborates with energy consumers to reduce electricity usage during peak hours and sells the saved energy back to grid operators or utilities. The DRA encourages consumers through incentives like discounted electricity rates. It employs various technologies and strategies, including automated demand response systems and smart thermostats, in order to optimize the DR process (Saxena *et al.* 2021b).

DRAs play a vital role in assisting grid operators to effectively manage peak demand, mitigate energy costs, and enhance system reliability. By offering incentives to energy consumers for curtailing their energy usage during peak periods, DRAs contribute to balancing the electricity supply and demand, thereby minimizing the need for additional generation capacity. The following factors outline the constraints associated with DR that are given careful consideration:

$$P_{i(t)} = (P_{Gi(t)} - P_{Di(t)}) \quad \forall t, i \quad (9)$$

$$Q_{i(t)} = (Q_{Gi(t)} - Q_{Di(t)}) \quad \forall t, i \quad (10)$$

$$P_{Di(t)} = (P_{in,i(t)} + P_{el,i(t)}) \quad \forall t, i \quad (11)$$

$$\sum_{i=1}^N \sum_{t=1}^{24} (P_{in,i(t)} + P_{el,i(t)}) \times \Delta t = E_i^{Total} \quad (12)$$

$$P_{el,i}^{min} \leq P_{el,i(t)} \leq \min\left((C - P_{in,i(t)}), P_{el,i}^{max}\right) \forall t \quad (13)$$

$$P_{el,i}^{max} = \mu \sum_{t=1}^{24} L_{d,i(t)} \quad (14)$$

where C and μ are the contract load and the DR penetration rate, respectively.

Objective constraints

The aforementioned objective functions are subject to constraints that account for technical and operational considerations:

Solar PV output constraint:

$$0 \leq P_{DG,i} \leq P_{DG}^{max} \quad \forall i \quad (15)$$

Feeder thermal limit constraint:

$$I_{ij(t)} \leq I_{ij}^{max} \quad \forall t, i, j \quad (16)$$

Real and reactive power constraints:

$$P_{i(t)} = V_{i(t)} \sum_{j=1}^N V_{j(t)} Y_{ij} \cos(\theta_{ij} + \delta_{j(t)} - \delta_{i(t)}) \quad \forall t, i \quad (17)$$

$$Q_{i(t)} = -V_{i(t)} \sum_{j=1}^N V_{j(t)} Y_{ij} \sin(\theta_{ij} + \delta_{j(t)} - \delta_{i(t)}) \quad \forall t, i \quad (18)$$

Demand modeling

Demand modeling is expressed as

$$P_{D,i(t)} = \Omega_{i(t)} P_{D,i}^0 \quad \forall t, i \quad (19)$$

$$Q_{D,i(t)} = \Omega_{i(t)} Q_{D,i}^0 \quad \forall t, i \quad (20)$$

where $\Omega_{i(t)}$ is the assigned load factor for the time period t .

Modeling the solar PV output

Solar power generation is influenced by factors such as solar panel characteristics, tilt angle, and solar irradiance. Assuming that other variables remain constant during the specified period, the current relative to the rated voltage is calculated as follows:

$$I_{sm(t)} = \begin{cases} I_{sm} & \text{if } S_{r(t)} \geq S_r^r \\ I_{sm} \times S_{r(t)} / S_r^r & \text{if } S_{r(t)} < S_r^r \end{cases} \quad (21)$$

In summary, this research framework encompasses multiple objectives, from minimizing power losses and managing reverse power flow to controlling nodal voltage deviation. These objectives are optimized through a fitness function with weighted factors. Additionally, the role of the DRA in managing peak energy consumption is emphasized. The framework adheres to various constraints, ensuring that both technical and operational considerations are met while modeling the demand and solar PV output to support the optimization process.

Optimization approach

The genetic algorithm (GA) is a heuristic optimization method that draws inspiration from the natural selection process. This approach is especially efficient in addressing complex, nonlinear, and non-differentiable optimization problems. The GA maintains a population of potential solutions and uses genetic operators such as crossover and mutation to progressively develop improved solutions.

The parameters and settings for the GA optimization approach in this study are outlined in [Table 1](#), with the latter explained below:

Population size: The number of individuals in the population. This determines the diversity of solutions explored in each generation.

Crossover probability: The likelihood of crossover (recombination) occurring. This parameter controls the rate at which genetic material is exchanged between individuals.

Mutation probability: It introduces small random changes to individual solutions, helping to explore new regions of the solution space.

Number of generations: The total number of iterations or generations that the algorithm will run. It influences how thoroughly the solution space is explored.

Table 1. Simulation parameters for the genetic algorithm approach

Parameters	Level 1	Level 2
Population size	50	100
Crossover probability	0.8	0.8
Mutation probability	0.01	0.01
Number of generations	100	150

Source: Authors

These parameters were selected based on a comprehensive review of the literature, specifically drawing on the work of [Saxena et al. \(2022b, 2023\)](#). They are tailored to suit the specific characteristics and requirements of the optimization problem at levels 1 and 2. These parameters are meant to efficiently converge towards high-quality solutions within the defined computational resources.

Test system

The multilevel optimization technique proposed in this study was applied to the IEEE 33-bus system depicted in Figure 1 (Baran *et al.* 1989). The power supply for the grid was sourced from the Indian CPP. This research involved a thorough investigation and analysis of the effects of DR technologies, aiming to discern the most effective power transmission strategies while accommodating diverse conditions and constraints. The central goal was to elevate the efficiency of power distribution.

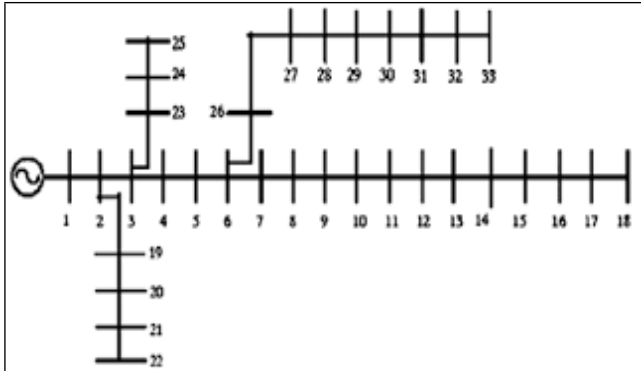


Figure 1. IEEE 33-bus system
Source: Baran *et al.* (1989)

The proposed optimization methodology was implemented using MATLAB software.

Furthermore, critical system parameters such as the base voltage, the nominal active demand, the nominal reactive demand, the power losses, V_{min} , V_{max} , and P_{DG}^{max} were set at 12.66 kV, 3715 kW, 2300 kVAR, 202.7 kW, 0.95 p.u., 1.05 p.u., and 2 MW, respectively. These values served as the foundational metrics for the optimization process.

Results

Case 1

This analysis explores the base scenario to assess the effectiveness of integrating solar PV technology into a 33-bus radial distribution system. The study customized the objective functions to align with daily consumption patterns, as outlined by Tazi *et al.* (2019). To quantify the system's performance, the annual energy losses were derived by averaging the losses recorded each day. The demand patterns reveal that the lowest demand occurs around 5:00 a.m., while peak demand is observed at 8:00 p.m. This variation in demand is crucial for understanding the system's performance across different times of the day. The results, detailed in Tables 2, 3, and 4, indicate the following key metrics for the base scenario: highest demand: 5397.73 kW; lowest demand: 5397.73 kW; minimum mean voltage: 0.978178 p.u.; annual energy losses: 1426 MWh; daily CO₂ emissions: 75446.33 kg.

These values highlight the system's performance and the potential impact of integrating solar PV technology, providing a comprehensive overview of the benefits and challenges associated with this approach.

Case 2

This subsection delves into the optimization of PV installations, aiming to assess the impact of integrating DGs into a DN. The optimization analysis revealed significant advancements in several key power quality parameters.

One of the most notable findings was the substantial 21.8% reduction in annual energy losses. This result emphasizes the critical role of DGs in enhancing energy efficiency by minimizing the amount of wasted energy within the network. The integration of DGs directly contributes to a more efficient and sustainable energy system by reducing the dependence on traditional power sources and cutting energy losses. In addition to energy losses reduction, our study also observed a marked improvement in voltage stability throughout the network. The minimum mean voltage experienced a significant increase from 0.978178 to 0.99634 p.u., indicating a stronger and more stable voltage profile across the system. This enhancement in voltage stability is crucial for maintaining the reliability of power distribution and ensuring consistent energy delivery to consumers.

Table 3 offers a thorough analysis of the optimal sizing and strategic placement of solar PV systems to maximize energy generation and distribution efficiency. These findings are critical in understanding how renewable energy sources can be integrated into power networks. In Figures 2, 3, and 4, the impact of DGs is further explored in relation to the network's demand patterns, voltage profiles, and active power losses, providing visual insights. These Figures collectively illustrate how the system's performance improves with the integration of solar PV technology, revealing enhanced voltage stability and reduced power losses.

A particularly notable outcome of the study is the environmental impact achieved through the integration of solar PV systems. The 26.51% reduction in CO₂ emissions attests to the effectiveness of renewable energy solutions in curbing GHG emissions. This reduction not only highlights the potential of solar PV in mitigating the environmental footprint of energy production but also underscores its role in improving the overall efficiency of the DN. In addition to its environmental benefits, this integration enhances the sustainability of the energy system, showcasing how renewable energy can contribute to a more resilient and eco-friendly power infrastructure.

Case 3

This subsection investigates the effectiveness of the DR approach by analyzing two different levels of demand elasticity in the absence of DG coordination. Demand

elasticity is measured through DR rates, which reflect how responsive consumers are to shifts in electricity prices or grid demand. In this case, two DR rates (10 and 20%) are considered, assuming that DG resources are not available or deployed. The results reveal that the DR approach has a significant impact on reducing peak electricity demand. Specifically, a 10% DR rate results in a 14.72% reduction in peak demand, while a 20% DR rate achieves an 18.32% reduction. This illustrates how increasing the DR rate leads to more substantial peak demand reductions, helping to alleviate stress on the electrical grid during periods of high demand.

In addition to reducing peak demand, DR also effectively cuts down annual energy losses, with reductions of 5.96% at a 10% DR rate and 8.2% at a 20% DR rate. These results show the potential of DR to improve overall grid efficiency by minimizing energy waste. Moreover, DR reduces active power losses, contributing to the stability and reliability of the grid. However, while it enhances demand flexibility and reduces energy losses, it also causes an increase in the peak-to-valley difference, which denotes the variation between peak and off-peak electricity demand. This indicates that DR shifts demand but may amplify load fluctuations, which could affect grid stability in some scenarios.

Interestingly, this study demonstrates that DR can be highly effective even in the absence of DG resources, implying that DR alone can play a vital role in grid management. Figures 5 to 8 provide visual insights into how DR rates of 10 and 20% impact key factors such as demand, voltage profiles, and active power losses. Nevertheless, the analysis also reveals that the influence of DR rates on the voltage profile and CO₂ emissions of the system is relatively insignificant. This suggests that, while DR helps with demand management and energy losses reduction, its direct impact on voltage stability and emission reductions is minimal, especially in the absence of DG integration. Nonetheless, DR remains a valuable tool for managing energy consumption and improving grid performance.

Table 2. Effect of the coordination of DR with optimally integrated solar PV on demand

Case no.	Type of case	Maximum demand value (kW)	Reduction in the maximum demand value (%)	Demand span (kW)	% maximum loss mitigation at 20:00 h
1	Base case	6519	-	5397.73	-
2	DG	6519	-	6016.39	-
3	DR@10%	5559.3	14.7216	4166.14	25.496
	DR@20%	5324.7	18.3204	3730.6	30.9725
4	DG+DR@10%	5375.66	17.5386	4322.87	29.807
	DG+DR@20%	4794.57	26.45	3540.31	42.3061

Source: Authors

Case 4

A comprehensive analysis was conducted to assess the integration of DGs into disaster recovery coordination and planning while considering system limitations. In this scenario, DGs were incorporated into DR scheduling, factoring in these constraints. The combination of higher DR rates and smaller DGs significantly enhanced system performance, leading to a notable reduction in annual energy losses. Compared to earlier examples, the lowest mean voltage increased by 29.03-33.31% depending on the DR rate applied. As DGs help to close the gap between maximum and minimum demand, the load profile becomes more balanced and flattened. Figures 9 to 11 demonstrate the impact of DGs at a 10% DR rate on demand, voltage, and active power losses, while Figures 12 to 14 highlight the effects of a 20% DR rate on the same parameters. Notably, DR rates of 10 and 20% reduce CO₂ emissions by 29.32 and 24.87%, respectively, when diesel generators are included in the system.

Table 3. Outcomes of the coordination of DR with optimally integrated solar PV systems

Case no.	Type of case	DG location (bus no., kW)	Demand/day (kWh)	Yearly losses (MWh)	Daily losses (kWh)	Losses reduction/year (%)	PV utilization (%)	Average voltage level (p.u.)
1	Base case		73 474	1426	3906.84			0.978
2	DG	14(1343)-30(1706)-25(1078)	52 681	1115	3054.79	21.80	69.44	0.996
3	DR@10%		73 472	1341	3673.97	5.96		0.978
	DR@20%		73 411	1309	3586.30	8.20		0.978
4	DG+DR@10%	15(1163)-7(1876)-33(904)	50 684	1012	2772.60	29.03	66.33	0.996
	DG+DR@20%	18(418)-29(1820)-11(1636)	54 498	951	2605.47	33.31	64.56	0.996

Source: Authors

Table 4. Outcomes of the coordination of DR with optimally integrated solar PV systems

Case no.	Type of case	Required energy from CPP/Day (kWh)	Energy by DG/Day (kWh)	CO ₂ emissions from solar PV systems (kg)	Daily energy losses (kWh)	Energy supplied from CPP/day (kWh)	Daily CO ₂ emissions (kg) by the CPP	Total CO ₂ emissions/day (Kg)	% reduction in CO ₂ emissions/day
1	Base case	73 474			3906.85	77 380.85	75 446.33	75 446.33	
2	DG	52 681	20 793	1102.029	3054.79	55 735.79	54 342.4	55 444.43	26.51
3	DR@10%	73 472			3673.97	77 145.97	75 217.32	75 217.32	0.30
	DR@20%	73 411			3586.30	76 997.3	75 072.37	75 072.37	0.50
4	DG+DR@10%	50 684	22 790	1207.87	2772.60	53 456.6	52 120.19	53 328.06	29.32
	DG+DR@20%	54 498	18 976	1005.728	2605.48	57 103.48	55 675.89	56 681.62	24.87

Source: Authors

Discussion

The proposed strategy for mitigating CO₂ emissions in Indian CPPs through optimal solar PV systems allocation and DR coordination was examined alongside the findings from key literature reports. [Samanta et al. \(2015\)](#) demonstrated a 30.7% increase in plant capacity and efficiency through a partial repowering approach, while [Smaisim et al. \(2023\)](#) highlighted the benefits of renewable integration in coal power plants. [Zhang et al. \(2015\)](#) focused on CO₂ capture with chemical absorption, achieving a 9.32-8.71% reduction in equivalent work consumption. Additionally, [Hanak et al. \(2015\)](#) modeled ammonia substitution in CO₂ capture, revealing efficiency penalties of 8.7-10.9%. Contrasted with these studies, our paper introduces a novel approach, directly implementing solar PV and DR mechanisms in the context of Indian CPPs. The outcomes demonstrate a noteworthy 29.32% reduction in CO₂ emissions and a substantial 69.44% penetration of renewable energy, offering a distinctive and effective strategy for emissions reduction and renewable integration within the coal-based electricity generation sector. This study not only contributes to existing literature on emissions reduction strategies but also presents a pioneering application of solar PV technology and DR coordination within Indian CPPs.

Conclusions

- This paper addresses CO₂ emissions reduction within Indian CPPs.
- The strategic allocation of solar PV systems and the synchronization of demand response are the key methods presented.
- This study uses GA to optimize solar PV placement, considering factors like solar resource availability, electricity demand trends, and CO₂ intensity from coal power.

- The goal is to minimize CO₂ emissions, maximize solar PV integration, and reduce power losses while addressing the intermittent nature of solar energy and dynamic demand.
- The approach was validated on the IEEE 33-bus system, yielding a 29.31% reduction in CO₂ emissions.
- The results demonstrate the potential of solar PV integration and DR in reducing emissions and enhancing climate change mitigation.
- DGs are effective in reducing energy losses but pose challenges like voltage elevation and reverse power flow as their penetration increases.
- DR helps to stabilize load profiles by reducing peak and off-peak demand gaps and alleviating system stress.
- DR also enhances demand normalization, especially with lower solar PV penetration levels.
- Case studies show significant reductions in maximum demand (26.78%), annual energy losses (34.5%), and DG penetration (67.76%).

CRedit author statement

Vivek Saxena was in charge of data collection, conceptualization, implementation, analysis, drafting, and review. Saurabh Kumar Rajput revised and reviewed the article.

Conflicts of interest

The authors declare no conflict of interest.

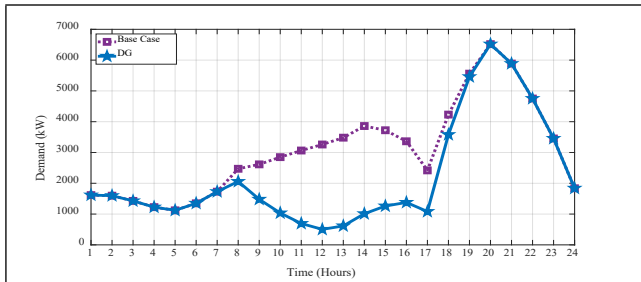


Figure 2. Impact of DG on the demand curve
Source: Authors

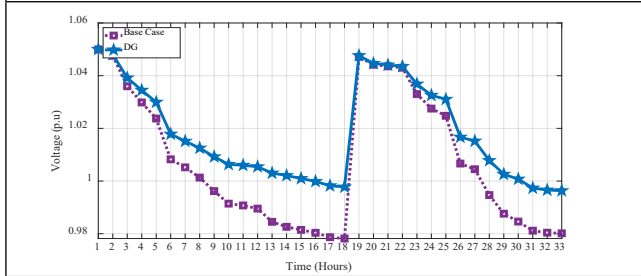


Figure 3. Impact of DG on the voltage curve
Source: Authors

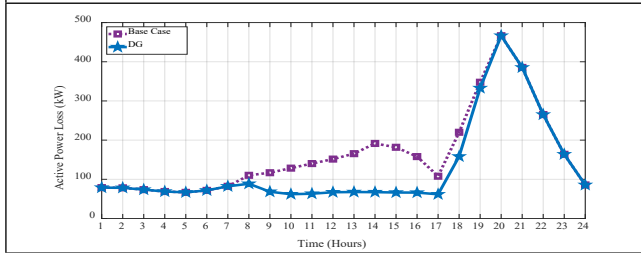


Figure 4. Impact of DG on power losses
Source: Authors

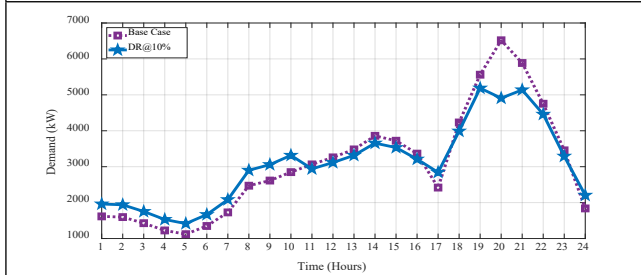


Figure 5. Impact of DR@10% on the demand curve
Source: Authors

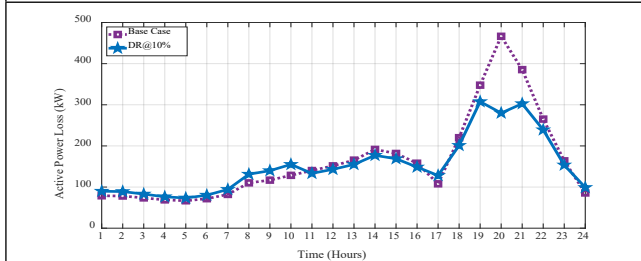


Figure 6. Impact of DR@10% on power losses
Source: Authors

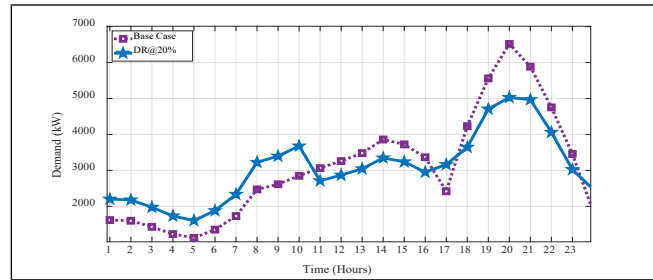


Figure 7. Impact of DR@20% on the demand curve
Source: Authors

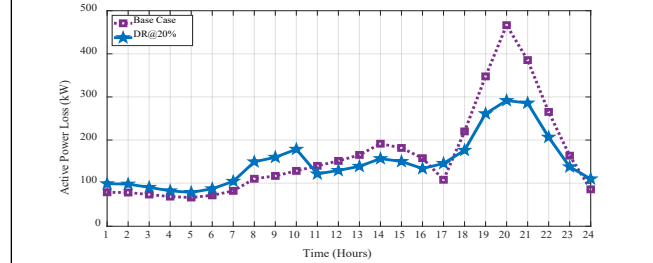


Figure 8. Impact of DR@20% on power losses
Source: Authors

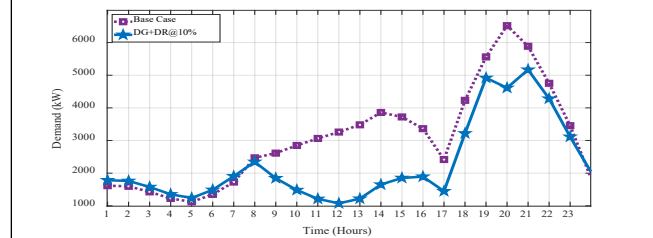


Figure 9. Impact of DG+DR@10% on the demand curve
Source: Authors

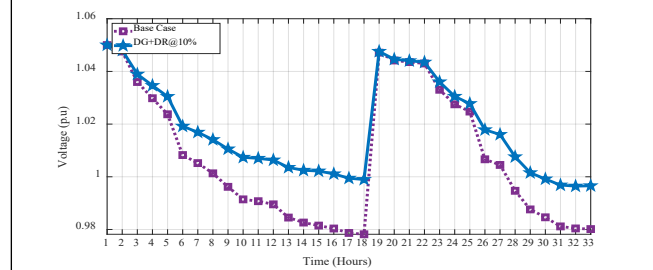


Figure 10. Impact of DG+DR@10% on the voltage curve
Source: Authors.

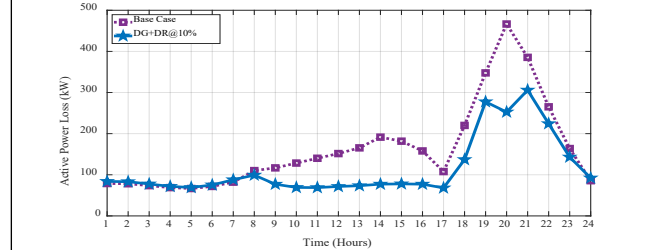


Figure 11. Impact of DG+DR@10% on power losses
Source: Authors

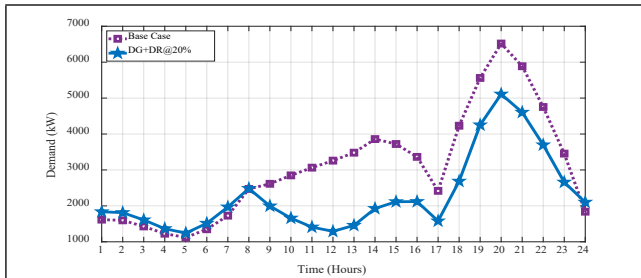


Figure 12. Impact of DG+DR@20% on the demand curve
Source: Authors

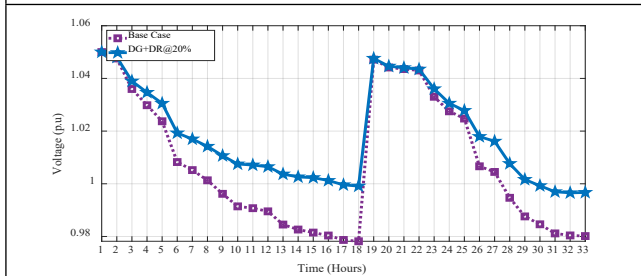


Figure 13. Impact of DG+DR@20% on the voltage curve
Source: Authors

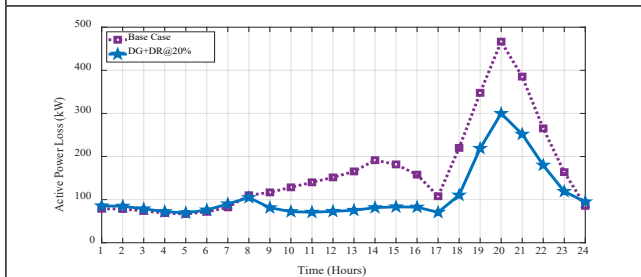


Figure 14. Impact of DG+DR@20% on power losses
Source: Authors

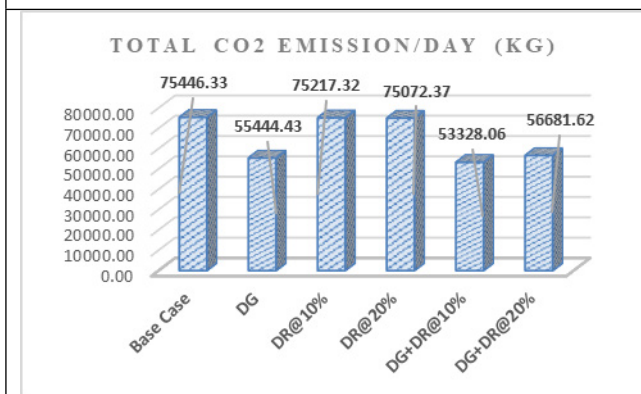


Figure 15. Total CO₂ emissions per day in different cases
Source: Authors

Data availability

Not applicable.






References

- Akbari, E., Molljafari, M., Ridha Al-Kafaji, H. M., Alkattan, H., Abotalef, M., Eslami, M., and Palani, S. (2022). Improved salp swarm optimization algorithm for damping controller design for multimachine power system. *IEEE Access*, 10, 82910-82922. <https://doi.org/10.1109/ACCESS.2022.3196851>
- Al-Dahidi, S., Adeeb, J., Alrbai, M., Ayadi, O., and Qawasmeh, B. R. (2018). Extreme learning machines for solar photovoltaic power predictions. *Energies*, 11, 2725. <https://doi.org/10.3390/en1102725>
- Al-Dahidi, S., Ayadi, O., Alrbai, M., and Adeeb, J. (2019). Ensemble approach of optimized artificial neural networks for solar photovoltaic power prediction. *IEEE Access*, 7, 81741-81758. <https://doi.org/10.1109/ACCESS.2019.2923905>
- Al-Dahidi, S., Hammad, B., Alrbai, M., and Al-Abed, M. (2024). A novel dynamic/adaptive K-nearest neighbor model for the prediction of solar photovoltaic systems' performance. *Results in Engineering*, 22, 102141. <https://doi.org/10.1016/j.rineng.2024.102141>
- Al-Ghussain, L., Darwish Ahmad, A., Abubaker, A. M., Alrbai, M., Ayadi, O., Al-Dahidi, S., and Akafuah, N. K. (2023). Techno-economic assessment of photovoltaic-based charging stations for electric vehicles in developing countries. *Energy Sources, Part A: Recovery, Utilization, and Environmental Effects*, 45(1), 523-541. <https://doi.org/10.1080/15567036.2023.2171517>
- Alrbai, M., Abubaker, A. M., Darwish Ahmad, A., Al-Dahidi, S., Ayadi, O., Hjouj, D., and Al-Ghussain, L. (2022). Optimization of energy production from biogas fuel in a closed landfill using artificial neural networks: A case study of Al Ghabawi Landfill, Jordan. *Waste Management*, 150, 218-226. <https://doi.org/10.1016/j.wasman.2022.07.011>
- Baran, M. E., and Wu, F. F. (1989). Network reconfiguration in distribution systems for loss reduction and load balancing. *IEEE Transactions on Power Delivery*, 4(2), 1401-1407. <https://doi.org/10.1109/61.25627>
- CEA (2023). *Emission: 2021-22*. https://cea.nic.in/wp-content/uploads/baseline/2023/01/Approved_report_emission_2021_22.pdf
- Cheng, D., Xiao, J., Sun, L., Liu, Y., Feng, Y., and Kang, Y. (2023). Research on CO₂ emission prediction of coal-fired power plant based on RBF neural network [Conference paper]. 2023 International Conference on Smart Electrical Grid and Renewable Energy (SEGRE), Changsha, China. <https://doi.org/10.1109/SEGRE58867.2023.00075>
- Eslami, M., Shareef, H., and Mohamed, A. (2011). Optimization and coordination of damping controls for optimal oscillations damping in multi-machine power system. *International Review of Electrical Engineering*, 6(4), 1984-1993. <https://research.uaeu.ac.ae/en/publications/optimization-and-coordination-of-damping-controls-for-optimal-osc>
- Eslami, M., Shareef, H., Mohamed, A., and Khajehzadeh, M. (2012). A survey on flexible AC transmission systems (FACTS). *Przeglad Elektrotechniczny*, 88(1A), 1-11. <https://api.semanticscholar.org/CorpusID:56347323>
- Hanak, D. P., Biliyok, C., and Manovic, V. (2015). Efficiency improvements for the coal-fired power plant retrofit with CO₂ capture plant using chilled ammonia process. *Applied Energy*, 151, 258-272. <https://doi.org/10.1016/j.apenergy.2015.04.059>
- Lakshmi, G. V. N., Jayalaxmi, A., and Veeramsetty, V. (2023). Optimal placement of distributed generation based on DIS-

- CO's financial benefit with loss and emission reduction using hybrid Jaya-Red Deer optimizer. *Electrical Engineering*, 105, 965-977. DOI: <https://doi.org/10.1007/s00202-022-01709-y>
- Ma, L., Wang, R., Fan, J., Wang, J., Peng, D., and Zhao, H. (2022). Index and evaluation method of source-network-load interaction for low-carbon development of coal-fired power plants [Conference paper]. 2022 4th International Conference on Electrical Engineering and Control Technologies (CEECT), Shanghai, China. <https://doi.org/10.1109/CEECT55960.2022.10030609>
- Meena, N. K., Parashar, S., Swarnkar, A., Gupta, N., and Niazi, K. R. (2018). Improved elephant herding optimization for multiobjective DER accommodation in distribution systems. *IEEE Transactions on Industrial Informatics*, 14(3), 1029-1039. <https://doi.org/10.1109/TII.2017.2748220>
- Rajput, S. K., and Dheer, D. K. (2022). Integration of 100-kWp PV with low-voltage distribution power system in composite climate: Performance and energy metrics analysis. *International Journal of Ambient Energy*, 1, 8176-92. <https://doi.org/10.1080/01430750.2022.2092775>
- Samanta, S., and Ghosh, S. (2015). A techno-economic analysis of partial repowering of a 210 MW coal fired power plant. *Advances in Energy Research*, 3(3), 167-179. <https://doi.org/10.12989/eri.2015.3.3.167>
- Saxena, V., Kumar, N., and Nangia, U. (2021a). Smart grid: A sustainable smart approach. *Journal of Physics: Conference Series*, 1, 012042. <https://doi.org/10.1088/1742-6596/2007/1/012042>
- Saxena, V., Kumar, N., and Nangia, U. (2021b). Analysis of smart electricity grid framework unified with renewably distributed generation. Advances in smart communication and imaging systems. *Lecture Notes in Electrical Engineering*, 721, 735-751. https://doi.org/10.1007/978-981-15-9938-5_68
- Saxena, V., Kumar, N., Nangia, U. (2022a). An impact assessment of distributed generation in distribution network. In M. Pandit, M. K. Gaur, P. S. Rana, and A. Tiwari (Eds.), *Artificial Intelligence and Sustainable Computing. Algorithms for Intelligent Systems* (pp. 333-346). Springer. https://doi.org/10.1007/978-981-19-1653-3_26
- Saxena, V., Kumar, N., and Nangia, U. (2022b). Recent trends in the optimization of renewable distributed generation: A review. *Ingeniería e Investigación*, 42(3), e97702. DOI: <https://doi.org/10.15446/ing.investig.97702>
- Saxena, V., Kumar, N., and Nangia, U. (2023). An extensive data-based assessment of optimization techniques for distributed generation allocation: Conventional to modern. *Archives of Computational Methods in Engineering*, 30, 675-701. <https://doi.org/10.1007/s11831-022-09812-w>
- Shirazi, H., Ghiasi, M., Dehghani, M., Niknam, T., Garpachi, M. G., and Ramezani, A. (2021). Cost-emission control based physical-resilience oriented strategy for optimal allocation of distributed generation in smart microgrid [Conference paper]. 2021 7th International Conference on Control, Instrumentation and Automation (ICCIA), Tabriz, Iran. <https://doi.org/10.1109/ICCIA52082.2021.9403561>
- Smaisim, G. F., Abed, A. M., and Alavi, H. (2023). Analysis of pollutant emission reduction in a coal power plant using renewable energy. *International Journal of Low-Carbon Technologies*, 18, 38-48. <https://doi.org/10.1093/ijlct/ctac130>
- Tazi, K., Abbou, F. M., and Abdi, F. (2019). Load analysis and consumption profiling: An overview. In B. Hajji, et al. (Eds.), *Proceedings of the 1st International Conference on Electronic Engineering and Renewable Energy. ICEERE 2018. Lecture Notes in Electrical Engineering* (vol. 519, pp. 80-88). Springer. https://doi.org/10.1007/978-981-13-1405-6_80
- Viana, M. S., Manassero, G., and Udaeta, M. E. M. (2018). Analysis of demand response and photovoltaic distributed generation as resources for power utility planning. *Applied Energy*, 217, 456-466. <https://doi.org/10.1016/j.apenergy.2018.02.153>
- Wang, X., Huang, G., Qiao, Z., and Li, X. (2021). Contribution of distributed energy resources management to peak carbon dioxide emissions and carbon neutralization [Conference paper]. 2021 IEEE Sustainable Power and Energy Conference (iSPEC) (pp. 2101-2107), Nanjing, China. <https://doi.org/10.1109/iSPEC53008.2021.9735717>
- Yaghoubi, M., Eslami, M., Noroozi, M., Mohammadi, H., Kamari, O., and Palani, S. (2022). Modified salp swarm optimization for parameter estimation of solar PV models. *IEEE Access*, 10, 110181-110194. <https://doi.org/10.1109/ACCESS.2022.3213746>
- Yoon, G., Shin, H., Jung, H., Huang, Y., Lee, T., and Park, S. (2022). Intelligent energy supply management model for distributed energy-based reduce carbon emission [Conference paper]. 2022 IEEE 5th Student Conference on Electric Machines and Systems (SCEMS). <https://doi.org/10.1109/SCEMS56272.2022.9990770>
- Zhang, Li, J., Tian, Y., and Cheng, Y. (2020). Reduction of carbon emissions from China's coal-fired power industry: Insights from the province-level data. *Journal of Cleaner Production*, 242, 118518. <https://doi.org/10.1016/j.jclepro.2019.118518>
- Zhong, X., Heping, P., Wenxiong, M., Yong, W., and Le, L. (2021). Locating and sizing optimization algorithm of distributed generation in distribution networks for low-carbon operation [Conference paper]. 2021 International Conference on Power System Technology (POWERCON), Haikou, China. <https://doi.org/10.1109/POWERCON53785.2021.9697698>

Effect of Glycerin and Urea on the Synthesis of Potato and Cassava Starch-Based Biopolymers: Hardness, Micrography, and Thermogravimetric Analyses

Efecto de la glicerina y la urea en la síntesis de biopolímeros a base de almidón de papa y yuca: análisis de dureza, micrografía y termogravimetría

Daniel N. Mariño Quintana ¹, Diana P. Sanabria Chaparro ², Hugo F. Salazar ³,
Hugo F. Castro Silva ⁴, and Ricardo A. Paredes ⁵

ABSTRACT

Petroleum-derived polymers used in everyday products generate large amounts of waste and negative environmental impacts due to their slow decomposition. To address this issue, other options have been studied, such as biopolymers based on starch, a polysaccharide formed by chains of amylose and amylopectin that can be linked together by heat and water to form a polymeric matrix similar to petroleum-derived plastics. In this study, glycerin and urea were used as plasticizing additives to improve the flexibility of biopolymers. The objective was to expand knowledge on biopolymers and their potential applications as an alternative to petroleum-derived plastics. A quantitative and qualitative factorial experimental design was implemented which consisted of three factors: the percentage of starch type, the percentage of plasticizer type, and the total percentage of plasticizers. Once the biopolymers were synthesized, they were characterized through thermal stability tests using thermogravimetric, hardness, and micrography analysis. The results indicated that glycerin increases the flexibility of the biopolymer, while urea increases its hardness. The working temperature of the biopolymer is below 130 °C, causing no disintegration of the material. In addition, it was identified that the presence of gas inclusions, both internal and superficial, can significantly affect the mechanical properties of the biopolymers. In conclusion, it was demonstrated that starch-based biopolymers with plasticizing additives have the potential to become a viable and sustainable alternative to petroleum-derived plastics in everyday products.

Keywords: biopolymers, glycerin, urea, starch, TGA, hardness, micrograph

RESUMEN

Los polímeros derivados del petróleo que se utilizan en productos de uso cotidiano generan grandes cantidades de residuos y un impacto ambiental negativo debido a su lenta descomposición. Para combatir este problema, se han estudiado otras opciones como los biopolímeros a base de almidón, un polisacárido formado por cadenas de amilosa y amilopectina que pueden enlazarse mediante calor y agua para formar una matriz polimérica similar a los plásticos derivados del petróleo. En este estudio se utilizó glicerina y urea como aditivos plastificantes para mejorar la flexibilidad de los biopolímeros. El objetivo fue ampliar el conocimiento sobre los biopolímeros y sus posibles aplicaciones como alternativa a los plásticos derivados del petróleo. Se implementó un diseño experimental factorial cuantitativo y cualitativo que constaba de tres factores: el porcentaje del tipo de almidón, el porcentaje del tipo de plastificantes y el porcentaje total de plastificantes. Una vez sintetizados los biopolímeros, se caracterizaron mediante ensayos de estabilidad térmica con TGA, dureza y micrografía. Los resultados indicaron que la glicerina aumenta la flexibilidad del biopolímero, mientras que la urea aumenta su dureza. La temperatura de trabajo del biopolímero se encuentra por debajo de los 130 °C, sin generar desintegración en el material. Además, se identificó que la presencia de inclusiones gaseosas, tanto internas como superficiales, puede incidir significativamente en las propiedades mecánicas de los biopolímeros. En conclusión, se demostró que los biopolímeros a base de almidón con aditivos plastificantes tienen el potencial de convertirse en una alternativa viable y sostenible a los plásticos derivados del petróleo en productos de uso cotidiano.

Palabras clave: biopolímeros, glicerina, urea, almidón, TGA, dureza, micrografía

Received: May 19th, 2023

Accepted: December 5th, 2024

¹ Industrial engineer, Universidad Pedagógica y Tecnológica de Colombia. PMS Project Management Specialist, Universidad del Bosque, Colombia. Affiliation: Researcher, Universidad Pedagógica y Tecnológica de Colombia. Email: daniel.quintana01@uptc.edu.co

² Industrial engineer, Universidad Pedagógica y Tecnológica de Colombia. Master's student in Metallurgical Engineering, Universidad Pedagógica y Tecnológica de Colombia. Affiliation: Researcher, Universidad Pedagógica y Tecnológica de Colombia. Email: diana.sanabria01@uptc.edu.co

³ Industrial engineer, Universidad Industrial de Santander, Colombia. MS in Industrial Engineering, Universidad Distrital, Colombia. Affiliation: Full professor, Universidad Pedagógica y Tecnológica de Colombia. Email: hugo.salazar@uptc.edu.co

⁴ Industrial engineer, Universidad Pedagógica y Tecnológica de Colombia. MS in Industrial Engineering, Universidad de los Andes, Colombia. PhD in Project Management, Universidad EAN, Colombia. Affiliation: Researcher and PhD professor, Universidad Pedagógica y Tecnológica de Colombia. Email: hugofernando.castro@uptc.edu.co

⁵ Industrial engineer, Universidad Pedagógica y Tecnológica de Colombia. Master's student in Metallurgical Engineering, Universidad Pedagógica y Tecnológica de Colombia. Affiliation: Researcher and professor, Universidad Pedagógica y Tecnológica de Colombia. Email: ricardo.paredes@uptc.edu.co



Attribution 4.0 International (CC BY 4.0) Share - Adapt

Introduction

Prior to the 19th century, plastic was not widely used [1], [2]. It was during the 20th century when interest in its use increased and its demand grew, given the need to develop advanced materials [3]–[5]. The production and manufacture of plastic products entails lower costs compared to other materials, which has significantly influenced its production and applicability in various industrial sectors [6], [7].

Petrochemical polymers are materials characterized by their easily moldable molecular structure, flexibility, and corrosion resistance. Single-use petrochemicals, also known as *disposable products*, constitute a significant portion of everyday products and are manufactured in large quantities due to their high demand [3], [8]. However, their final disposal in landfills and ecosystems generates a substantial environmental impact, as they can take decades or even hundreds of years to degrade [9]–[11].

Various research approaches are currently being explored to address the issue of plastics and their replacement. One widely studied option is biopolymers, which include starch, a polysaccharide or *glucose polymer* that occurs in granular form and serves as an energy source for plants and animals. The polymeric chains of amylose and amylopectin that make up starch vary according to their source [12]–[14]. Under controlled humidity and temperature conditions, these chains can cross-link to form a polymer matrix suitable for the production of bioplastic films [15]–[19].

In this context, the use of industrial waste from processing companies handling potato (*Solanum tuberosum*) and cassava (*Manihot esculenta*) has been proposed. These carbohydrate-rich foods are widely cultivated and processed worldwide [20]. According to the Food and Agriculture Organization (FAO), the annual potato production exceeded 300 million tons in 2016 [21]. For cassava and other tubers, global processing is estimated to exceed 320 million tons, generating 5–30% waste, such as peels and pulp, at each production stage. This approach reduces the impact of using food crops in biopolymer production, thereby promoting a more sustainable and ethical resource utilization [20].

Peels, common byproducts in the processing of these tubers, represent an affordable raw material that is available in large volumes and has potential applications in biopolymer production. In the potato industry, processing generates between 70 000 and 140 000 tons of peels per year, which are generally discarded or used in low-value products such as animal feed [21]. Similarly, cassava processing yields up to 1.6 tons of peels and 280 tons of bagasse with high moisture content, which are often wasted without adequate treatment, ending up in landfills and watercourses [22].

The starch composition of these crops varies: potatoes contain 17–24% amylose and 76–83% amylopectin, whereas cassava contains 16–22% amylose and 81–83% amylopectin [23]. These differences are crucial in the gelatinization process,

which determines the formation and properties of the polymer matrix in bioplastics [24]. Leveraging these industrial sources of starch not only contributes to sustainability but also offers specific properties in the resulting bioplastic films.

Thus, this study aims to create biopolymers from potato and cassava starch, with the addition of plasticizers to enhance their flexibility and physical properties. According to the free volume theory [18], the addition of plasticizers with flexible chains increases the free volume of molecules, resulting in more flexible biopolymers [25], [26]. Additionally, it has been found that biopolymers synthesized solely with water and starch tend to be brittle, making the addition of plasticizers crucial to enhance their flexibility [27], [26].

Plasticizers are widely used to enhance the flexibility and processability of plastic materials, with their selection being a crucial aspect that depends on chemical and physical compatibility with the polymer matrix [28]. Among the most commonly used plasticizers are urea, sorbitol, and glycerol, given their effectiveness in modifying the mechanical properties of biopolymers [29], [30]. Other compounds, such as ethylene glycol, diethylene glycol, and triethylene glycol, are recognized for their ability to reduce the rigidity of the polymer matrix [31].

Additionally, more specialized plasticizers like triethyl citrate have demonstrated significant potential in applications requiring greater thermal and chemical stability [32], and biodegradable materials such as chitosan and polylactic acid have been investigated as alternative plasticizers for sustainable biopolymer production, standing out for their renewable origin and reduced environmental impact [33].

For this research, we used glycerol and urea. To date, the extent to which plasticizers affect material properties when added to a combined potato and cassava thermoplastic starch polymeric matrix for the production of biodegradable and compostable films has not been investigated [34].

Methodology

The methodology used in this study was divided into four phases, namely extraction, polymerization, fabrication, and characterization.

Extraction

Starch from potato and cassava was extracted according to the method described by [35], which involved crushing the raw materials in an aqueous medium. The resulting wet mass with lumps was diluted in water and agitated to separate the starch granules from unwanted materials [35]. The mixture was then filtered, and the resulting liquid was decanted to obtain the starch together with other impurities in the form of sediment. Three washes were performed with distilled water to remove these impurities, and the resulting wet starch was dried in an electric oven at 60 °C for 24 h [36].

Polymerization

The entire polymerization process was conducted while following the proportions outlined in the experimental design. This design, summarized in Table I, was developed to assess the impact of three primary factors on biopolymer synthesis: % starch type, % plasticizer type, and % total plasticizer, all on a basis of 10 g total starch per sample. The factor levels considered potato/cassava starch combinations (100/0, 60/40, 40/60, and 0/100), urea/glycerin ratios (100/0, 60/40, 40/60, and 0/100), and total plasticizer percentages (5, 10, and 20%). For instance, for the 40/60 potato/cassava combination, 4 g of potato starch and 6 g of cassava starch were used. With a 5% total plasticizer level, corresponding to 0.5 g, and a 40/60 urea/glycerin ratio, 0.2 g of urea and 0.3 g of glycerin were incorporated. This approach was replicated for all combinations, enabling a comprehensive evaluation of the resulting properties.

Table I. Experimental design factors

Factors				% Total Plasticizer		
				5%	10%	20%
% Potato/ Cassava	100/0	% Glycerin/ Urea	100/0	P1	P2	P3
			60/40	P4	P5	P6
			40/60	P7	P8	P9
			0/100	P10	P11	P12
	60/40	% Glycerin/ Urea	100/0	P13	P14	P15
			60/40	P16	P17	P18
			40/60	P19	P20	P21
			0/100	P22	P23	P24
	40/60	% Glycerin/ Urea	100/0	P25	P26	P27
			60/40	P28	P29	P30
			40/60	P31	P32	P33
			0/100	P34	P35	P36
	0/100	% Glycerin/ Urea	100/0	P37	P38	P39
			60/40	P40	P41	P42
			40/60	P43	P44	P45
			0/100	P46	P47	P48

Source: Authors

For each factor combination, four specimens were synthesized per sample in order to ensure the repeatability and reliability of the data obtained. The physical-mechanical properties of the biopolymer were measured from these specimens. The results of these measurements were averaged and used for statistical and graphical analysis, allowing for the identification of trends and relationships between the evaluated factors.

A full factorial design was employed in the statistical analysis, enabling the evaluation of interactions between factors and their combined effect on biopolymer properties. This robust approach facilitated the interpretation of the results and allowed for precise comparisons between different experimental combinations.

This methodology enabled a comprehensive analysis of the interactions between factors and their impact on biopolymer

properties, providing reliable data for interpreting the results and drawing conclusions.

The polymerization method used in this study involved the addition of plasticizing agents in an aqueous medium, following the methodology of [37], with modifications to suit the additives and equipment used in the process. On a hot plate at 100 °C, a homogeneous mixture of water, starch, and urea was prepared in predetermined proportions. Continuous stirring was maintained to prevent lump formation and ensure uniformity. Heating caused the starch granules to swell, allowing water to penetrate the amorphous regions of the starch, thus fragmenting the granular structure [38].

In the formulations containing glycerin, this component was added when the mixture reached 40 °C, allowing polymer chains to form before the gelatinization process began. This approach was based on the findings of [39], who reported that starch gelatinization starts between 40 and 45 °C. Throughout the procedure, constant stirring was also maintained to prevent lump formation and ensure a homogeneous mixture.

Acetic acid was subsequently added when the mixture reached 60 °C, in order to completely disrupt the starch granules and thus optimize the efficiency of polymer chain formation. The temperature was continuously monitored to ensure it did not exceed 80 °C, with heat adjustments made on the hot plate as needed. The mixture was stirred until a fully homogeneous sol-gel was achieved and removed from the hot plate after 40 min [40].

The pH of the mixture was carefully controlled throughout the process to prevent variations that could compromise the integrity of the starch granules. According to [41], pH changes can affect the starch gelatinization process, potentially influencing granule stability. No significant pH alterations were observed during the tests. However, in the event of an unexpected decrease, the procedure would need to be restarted and the amount of acetic acid adjusted in order to restore optimal conditions.

Molding

The bioplastic specimens were produced using the casting method, following the methodology proposed by [42]. In this technique, the biopolymer is poured in a sol-gel state onto a mold coated with a release agent. The resulting sheet is then dried at 50 °C for 24 h in an electric oven before being demolded and cut into the desired specimen shape.

Material characterization

Thermogravimetric analysis (TGA) is a quantitative analytical technique that monitors the mass of a sample as the temperature increases in an oven under a stable or changing gas flow [43]. This technique allows detecting weight

changes that occur in a sample subjected to heat treatment, providing valuable information about its thermal stability, maximum working temperature, moisture loss temperature, disintegration temperature, and ash generation temperature [44].

Optical microscopy was used to study the surface morphology of starch-based biopolymer sheets. The samples were analyzed using a Nikon SMZ800 stereo microscope equipped with a SONY video camera, with magnifications of 10x, 20x, and 60x. The inspection of the micrographs focused on the conformation and appearance of the films, providing information on their structural characteristics and potential performance.

Hardness is defined as a material's resistance to plastic deformation from penetration or scratching by another solid body. In this project, the Shore scale was used, as it is widely employed to evaluate the hardness of soft and semi-rigid materials like plastics and elastomers [45]. Measurements were conducted using a PC-511/A digital durometer (type A) equipped with an indenter, following the parameters specified by the Shore A scale.

Results and discussion

TGA

Fig. 1 illustrates the weight loss as a function of time and temperature for biopolymer samples containing different plasticizer types and concentrations (glycerin and urea). Overall, the thermogravimetric behavior was similar among samples with equivalent plasticizer content, although specific differences were observed at each stage of the thermal analysis.

During the initial phase, spanning temperatures between 0 and 100 °C, weight loss corresponds to the evaporation of water absorbed or physically bound to the biopolymer. Samples with higher plasticizer content exhibited faster weight loss, indicating that these additives increase moisture retention by interacting with the starch chains. This trend reflects the hygroscopic nature of the plasticizers used in the polymer matrix.

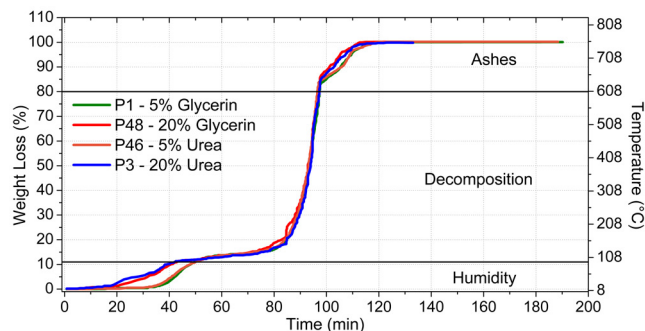


Figure 1. TGA of the biopolymers

Source: Authors

In the second phase, at temperatures ranging from 150 to 500 °C, the thermal degradation of polymer chains and plasticizing additives takes place. The onset temperature of decomposition varies depending on the plasticizer type and concentration. Despite these variations, the samples followed a uniform degradation pattern within the specified range. This behavior suggests that, while plasticizers influence the thermal stability of the material, they do not significantly alter the decomposition mechanism.

Finally, at temperatures exceeding 600 °C, the samples reach a residual state corresponding to ash formation. At this stage, the decomposition of organic components is complete, and differences in residual weight are associated with the initial composition of the biopolymer and the type of plasticizer used.

On a broader scale, it is noted that the onset temperatures of degradation in this study (approximately 130 °C) are lower than those reported in previous research, such as [46] and [47], which indicate onset temperatures between 150 and 250 °C. However, the weight loss patterns are consistent with the characteristic values for starch-based biopolymers. This suggests that the studied samples can be safely processed at temperatures below 130 °C, while ash formation consistently occurs above 600 °C, in line with prior studies. The graph clearly illustrates these phases and highlights the influence of plasticizer concentration and type at each stage of the TGA.

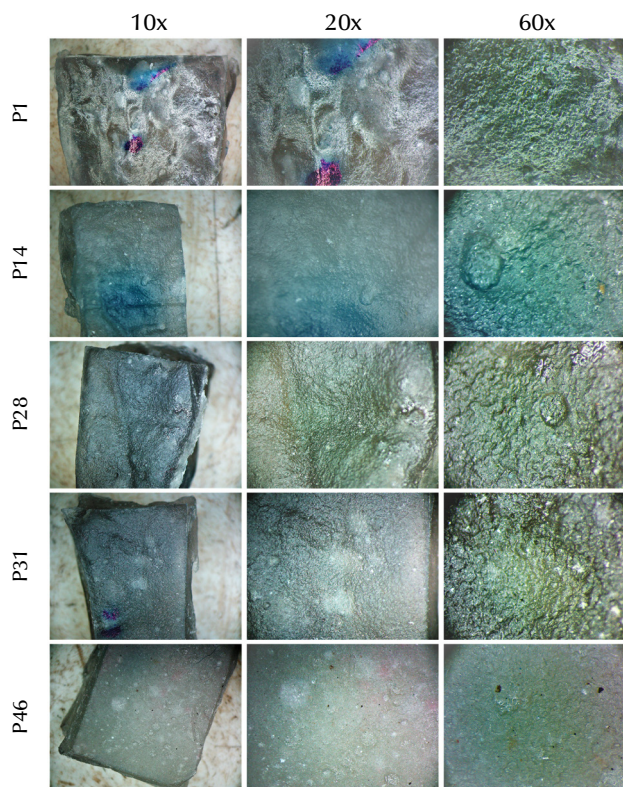


Figure 2. Biopolymer micrographs 10x, 20x, and 60x

Source: Authors

Optical microscopy

Micrographs were obtained from various tests performed on biopolymer samples, which, as previously noted, were formulated with different proportions of the evaluated variables: starch type (% potato/cassava), plasticizer type (% glycerin/urea), and total plasticizer concentration. The micrographs, captured at magnifications of 10x, 20x, and 60x, are presented in Fig. 2. The analyzed samples included P1 (% potato/cassava: 100/0, % glycerin/urea: 100/0, total plasticizer: 5%), P14 (% potato/cassava: 60/40, % glycerin/urea: 100/0, total plasticizer: 10%), P28 (% potato/cassava: 40/60, % glycerin/urea: 60/40, total plasticizer 5%), P31 (% potato/cassava: 40/60, % glycerin/urea: 40/60, total plasticizer: 5%), and P46 (% potato/cassava: 0/100, % glycerin/urea: 0/100, total plasticizer: 5%).

At 10x and 20x, a non-uniform surface conformation was observed in the biopolymer samples, with noticeable bubbles on the surface and within the formed materials. This phenomenon suggests that, during the synthesis process, the agitation used to homogenize the material introduced air bubbles which could not fully escape the matrix before solidification due to the high viscosity of the system.

At 60x, an irregular surface texture was observed, associated not only with the presence of bubbles, but also with other structural imperfections related to the manufacturing technique employed, *i.e.*, casting. While effective for producing biopolymers, this method is influenced by multiple variables that are challenging to control, such as solidification speed, plasticizer distribution within the matrix, and environmental conditions, which hinder the attainment of a fully homogeneous surface.

These imperfections could potentially be minimized by manufacturing the biopolymers via alternative processing techniques such as injection molding or thermoforming. These methodologies allow for greater precision in controlling material distribution and processing conditions, thereby reducing bubble formation and improving surface uniformity. Injection molding, for instance, ensures consistent pressure and homogeneous components distribution, while thermoforming facilitates more uniform solidification through controlled molds and temperatures. This suggests that optimizing the manufacturing technique is essential to enhancing the final quality of biopolymers and their suitability for end-use applications.

Hardness

The data presented in Figs. 3 and 4 demonstrate an inverse relationship between the glycerin content and the hardness of the biopolymer. An increase in the percentage of glycerin incorporated into the polymer matrix leads to a reduction in material hardness, accompanied by greater variability in the data for different biopolymer samples. Additionally, Fig. 4 highlights the significant impact of increasing the overall percentage of plasticizers on data dispersion across

various biopolymer samples. This behavior can be attributed to the effect of plasticizers on material hardness, wherein glycerin imparts greater flexibility. Moreover, biopolymer compositions with higher urea concentrations exhibit reduced data variability and increased hardness.

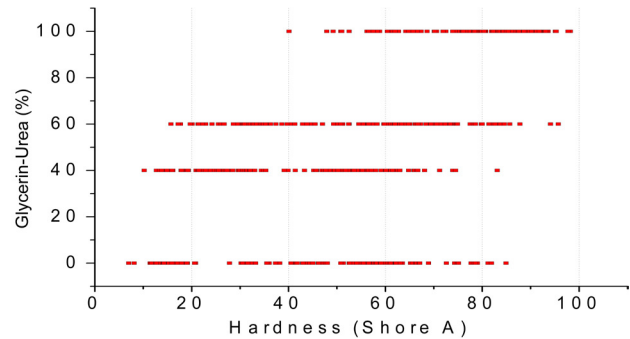


Figure 3. Scatter plot: hardness vs. glycerin-urea (%)
Source: Authors

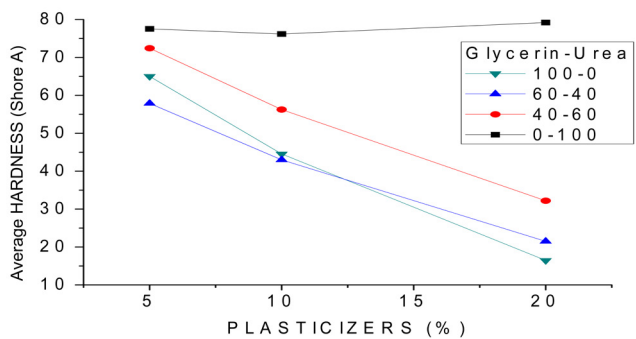


Figure 4. Linear graph: plasticizer hardness vs. glycerin-urea (%)
Source: Authors

Fig. 5 illustrates a clear decrease in material hardness as the ratio of total plasticizer to total starch increases, indicating that incorporating plasticizers enhances material flexibility and reduces rigidity, significantly impacting biopolymer hardness. In other words, the material's hardness is affected by the proportion of plasticizers used during production.

In the study titled *Physical characterization of biopolymers with starch from potato and cassava waste polymerized in water* [48], biopolymers synthesized from starch without the addition of plasticizers reached a maximum hardness of 78.9 on the Shore A scale. Comparing these results to our findings reveals that an increase in the glycerin content within the polymer matrix reduces hardness, with values as low as 33.2 on the Shore A scale.

Fig. 6 explores the interaction between plasticizers and the potato and cassava starch ratios, showing that biopolymer hardness depends on the combination of these factors. Biopolymers with higher glycerin proportions (100/0) exhibit significantly lower hardness values across all starch combinations, confirming glycerin's predominantly plasticizing effect. Conversely, systems with higher urea content (0/100) tend to exhibit greater hardness, particularly

in compositions with cassava starch (0/100), suggesting that urea promotes the formation of a stiffer matrix when interacting with cassava starch.

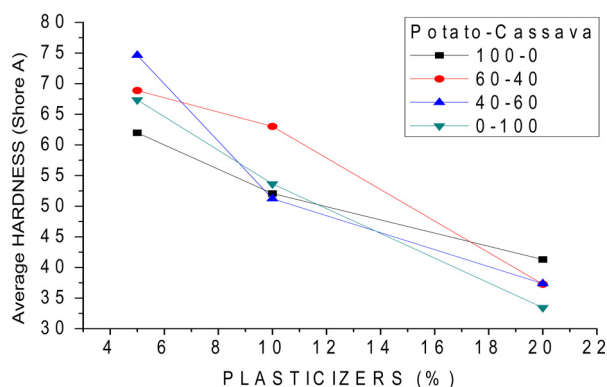


Figure 5. Linear graph: plasticizer hardness vs. potato-cassava (%)
Source: Authors

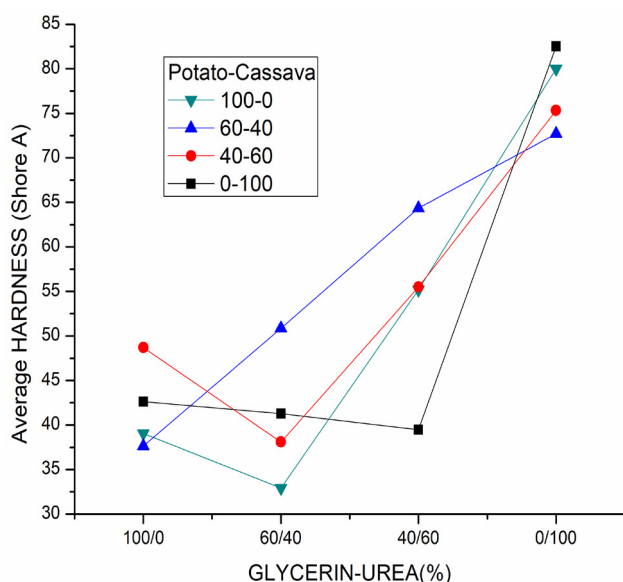


Figure 6. Linear graph: glycerin-urea hardness vs. potato-cassava (%)
Source: Authors

Furthermore, the lines corresponding to different starch combinations (100/0, 60/40, 40/60, and 0/100) indicate that the cassava content increases hardness in the presence of urea, while potato starch displays a more stable behavior under variations in plasticizer content. These differences suggest that the molecular structure and the interaction between polymers and plasticizers are key to tailoring the mechanical properties of the material.

The study titled *Analysis of factors affecting hardness behavior in biopolymers based on potato and banana peels: A factorial experimental evaluation* [49] also elucidates the relationship between increasing glycerin content and decreasing hardness, while noting that urea has a lower influence on material hardness.

Table II. Hardness data for petroleum-derived plastics

Material	Hardness (Shore A)	
HDPE – High-density polyethylene	98,5	[50]
PE – Polyethylene	95,13	[51]
PET – Polyethylene terephthalate	95,43	[51]
PP – Polypropylene	79,1	[52]
PS – Polystyrene	42,5	[53]
PTFE – Polytetrafluoroethylene	94,13	[51]

Source: Authors

The biopolymers synthesized in this study exhibit hardness values ranging from 30 to 85 on the Shore A scale depending on the type and percentage of the plasticizers incorporated into the polymer matrix. When compared to conventional plastics (Table II) such as HDPE (98,5), PE (95,13), PET (95,43), and PTFE (94,13), the biopolymers display lower hardness values. Still, they are competitive with PP (79,1), and, in certain combinations, they even surpass PS (42,5). This demonstrates that, by carefully adjusting the proportions and types of plasticizers, biopolymers with mechanical properties tailored for various applications can be achieved. In particular, this design flexibility suggests that these biopolymers could be synthesized to replace single-use plastics, offering a more sustainable and environmentally friendly alternative.

Conclusions

The addition of plasticizers to the polymeric matrix is known to provide materials with diverse physical-mechanical properties, making them potential substitutes for different petroleum-derived plastics. During the characterization of the biopolymers, the data revealed that the 60/40 potato/cassava starch ratio yielded superior biopolymer hardness. Regarding the plasticizer type, urea imparted significant hardness. However, in high proportions, it rendered the material fragile. In contrast, glycerin considerably reduced the hardness but provided flexibility to the biopolymer.

By combining urea and glycerin plasticizers in different proportions, hardness and plasticity properties can be tailored. Therefore, variations in the plasticizer type allow controlling the flexibility and hardness of the material, enabling the diversification of products. The TGA confirmed that the biopolymers can be processed at less than 130 °C for subsequent heat treatment to improve their properties while avoiding bioplastic disintegration.

The micrograph results showed that minimizing the bubbles in the biopolymers during the molding process can help to control the factors that affect mechanical properties, potentially yielding increased resistance to plastic deformation and a uniform material. Thus, optimizing the molding method is essential for achieving desired properties in biopolymers.

CRediT author statement

Daniel Nicolás Mariño Quintana, Diana Paola Sanabria, and Ricardo Alfonso Paredes conceived the project, defined its general structure, and designed the methodology. Daniel Nicolás Mariño Quintana and Diana Paola Sanabria were responsible for managing and conducting the research, including collecting the data, creating graphical visualizations, and drafting the initial manuscript. Ricardo Alfonso Paredes, Hugo Felipe Salazar, and Hugo Fernando Castro Silva supervised the project, validated the results, and performed a critical review of the draft. Finally, Daniel Nicolás Mariño Quintana and Diana Paola Sanabria edited the final manuscript.

References

- [1] S. C. Rasmussen, "From parkesine to celluloid: The birth of organic plastics," *Angew. Chem.*, vol. 133, no. 15, pp. 8090-8094, 2021. <https://doi.org/10.1002/anie.202015095>
- [2] G. Murray Tortarolo, M. Serrano Medrano, and M. A. Miranda Ackerman, "Una inundación global de plásticos," *Rev. Dig. Univ.*, vol. 22, no. 4, 2021. <https://doi.org/10.22201/cuaieed.16076079e.2021.22.4.10>
- [3] L. Desidery and M. Lanotte, "Polymers and plastics: Types, properties, and manufacturing," in *Plastic Waste for Sustainable Asphalt Roads*. Sawston, Cambridge, UK: Woodhead Publishing, 2022, pp. 3-28. <https://doi.org/10.1016/B978-0-323-85789-5.00001-0>
- [4] R. G. Prieto-Ortiz, "Contaminación ambiental por plásticos durante la pandemia y sus efectos en la salud humana," *Rev. Colombiana Cirugía*, vol. 38, no. 1, pp. 22-29, 2023. <https://doi.org/10.30944/20117582.2203>
- [5] M. E. Callapez, "History of the beginnings of the plastics industry in Portugal," *ChemTexts*, vol. 7, no. 3, art. 17, 2021. <https://doi.org/10.1007/s40828-021-00134-1>
- [6] U. Salahuddin, J. Sun, C. Zhu, M. Wu, B. Zhao, and P. X. Gao, "Plastic recycling: A review on life cycle, methods, misconceptions, and techno-economic analysis," *Adv. Sust. Syst.*, vol. 7, no. 7, art. 2200471, 2023. <https://doi.org/10.1002/adsu.202200471>
- [7] M. Mierzwa-Hersztek, K. Gondek, and M. Kopeć, "Degradation of polyethylene and biocomponent-derived polymer materials: An overview," *J. Polym. Environ.*, vol. 27, pp. 600-611, 2019. <https://doi.org/10.1007/s10924-019-01368-4>
- [8] M. D. M. López-Fernández and A. J. Franco-Mariscal, "Indagación sobre la degradación de plásticos con estudiantes de secundaria," *Educ. Quím.*, vol. 32, no. 2, pp. 21-36, 2021. <https://doi.org/10.22201/fq.18708404e.2021.2.76553>
- [9] N. Mohanan, Z. Montazer, P. K. Sharma, and D. B. Levin, "Microbial and enzymatic degradation of synthetic plastics," *Front. Microbiol.*, vol. 11, art. 580709, 2020. <https://doi.org/10.3389/fmicb.2020.580709>
- [10] A. S. Al Hosni, J. K. Pittman, and G. D. Robson, "Microbial degradation of four biodegradable polymers in soil and compost demonstrating polycaprolactone as an ideal compostable plastic," *Waste Manag.*, vol. 97, pp. 105-114, 2019. <https://doi.org/10.1016/j.wasman.2019.07.042>
- [11] S. N. Idris et al., "The degradation of single-use plastics and commercially viable bioplastics in the environment: A review," *Environ. Res.*, vol. 231, art. 115988, 2023. <https://doi.org/10.1016/j.envres.2023.115988>
- [12] J. Zhu et al., "New insights into amylose and amylopectin biosynthesis in rice endosperm," *Carbohydr. Polym.*, vol. 230, art. 115656, 2020. <https://doi.org/10.1016/j.carbpol.2019.115656>
- [13] C. Li, Y. Hu, T. Huang, B. Gong, and W.-W. Yu, "A combined action of amylose and amylopectin fine molecular structures in determining the starch pasting and retrogradation property," *Int. J. Biol. Macromol.*, vol. 164, pp. 2717-2725, 2020. <https://doi.org/10.1016/j.ijbiomac.2020.08.123>
- [14] C. C. Villa Zabala, "An overview on starch structure and chemical nature," in *Starch-Based Nanomaterials*. Cham, Germany: Springer, 2020, pp. 3-9. https://doi.org/10.1007/978-3-030-42542-5_2
- [15] M. Flórez, P. Cazón, and M. Vázquez, "Selected biopolymers' processing and their applications: A review," *Polymers*, vol. 15, no. 3, art. 641, 2023. <https://doi.org/10.3390/polym15030641>
- [16] A. Avellán, D. Díaz, A. Mendoza, M. Zambrano, Y. Zamora, and M. A. Riera, "Obtención de bioplástico a partir de almidón de maíz (*Zea mays* L.)," *Rev. Colón Cienc. Tecnol. Negoc.*, vol. 7, no. 1, pp. 1-11, 2020. https://revistas.up.ac.pa/index.php/revista_colon_ctn/article/view/1134
- [17] R. De Gracia, E. González, K. Vega, V. Montenegro, and N. Marín, "Elaboración de bioplástico a partir de almidón de granos," *Rev. Inic. Cient.*, vol. 10, no. 1, pp. 16-20, 2024. <https://doi.org/10.33412/rev-ric.v10.1.3998>
- [18] A. A. Ledesma-Ugshiña, V. M. Dalgo-Flores, L. M. Flores-Fiallos, and G. C. Chango-Lescano, "Bioplásticos de almidón de maíz y quinua para uso como envolturas alimenticias biodegradables," *Dom. Cien.*, vol. 7, no. 4, pp. 39-56, 2021. <https://doi.org/10.23857/dc.v7i4.2080>
- [19] J. Sears and J. Darby, *The technology of plasticizers*. Hoboken, NJ, USA: Wiley, 1982.
- [20] L. L. D. R. Osorio, E. Flórez-López, and C. D. Grande-Tovar, "The potential of selected agri-food loss and waste to contribute to a circular economy: Applications in the food, cosmetic and pharmaceutical industries," *Molecules*, vol. 26, no. 2, art. 515, 2021. <https://doi.org/10.3390/molecules26020515>
- [21] A. Javed, A. Ahmad, A. Tahir, U. Shabbir, M. Nouman, and A. Hameed, "Potato peel waste—Its nutraceutical, industrial and biotechnological applications," *AIMS Agric. Food*, vol. 4, no. 3, art. 807, 2019. <https://doi.org/10.3934/agr-food.2019.3.807>
- [22] V. Florencia, O. V. López, and M. A. García, "Exploitation of by-products from cassava and ahipa starch extraction as filler of thermoplastic corn starch," *Compos. Part B: Eng.*, vol. 182, art. 107653, 2020. <https://doi.org/10.1016/j.compositesb.2019.107653>
- [23] A. Gamage et al., "Applications of starch biopolymers for a sustainable modern agriculture," *Sustainability*, vol. 14, no. 10, art. 6085, 2022. <https://doi.org/10.3390/su14106085>

- [24] H. Han *et al.*, "Insight on the changes of cassava and potato starch granules during gelatinization," *Int. J. Biol. Macromol.*, vol. 126, pp. 37-43, 2019. <https://doi.org/10.1016/j.ijbiomac.2018.12.201>
- [25] A. Chaos *et al.*, "Plasticization of poly (lactide) with poly (ethylene glycol): Low weight plasticizer vs. triblock copolymers. Effect on free volume and barrier properties," *J. Appl. Polym. Sci.*, vol. 137, no. 28, art. 48868, 2020. <https://doi.org/10.1002/app.48868>
- [26] H. D. Ozeren, M. Guivier, R. T. Olsson, F. Nilsson, and M. S. Hedenqvist, "Ranking plasticizers for polymers with atomistic simulations: PVT, mechanical properties, and the role of hydrogen bonding in thermoplastic starch," *ACS Appl. Polym. Mater.*, vol. 2, no. 5, pp. 2016-2026, 2020. <https://doi.org/10.1021/acsapm.0c00191>
- [27] S. Agarwal, S. Singhal, C. B. Godiya, and S. Kumar, "Prospects and applications of starch-based biopolymers," *Int. J. Environ. Anal. Chem.*, vol. 103, no. 18, pp. 6907-6926, 2023. <https://doi.org/10.1080/03067319.2021.1963717>
- [28] M. P. Arrieta, "Influence of plasticizers on the compostability of polylactic acid," *J. Appl. Res. Technol. Eng.*, vol. 2, no. 1, pp. 1-9, 2021. <https://doi.org/10.4995/jarte.2021.14772>
- [29] A. A. Mohammed *et al.*, "Effect of various plasticizers in different concentrations on physical, thermal, mechanical, and structural properties of wheat starch-based films," *Polymers*, vol. 15, no. 1, art. 63, 2022. <https://doi.org/10.3390/polym15010063>
- [30] M. Alonso-González, M. Felix, and A. Romero, "Influence of the plasticizer on rice bran-based eco-friendly bioplastics obtained by injection moulding," *Ind. Crops Prod.*, vol. 180, art. 114767, 2022. <https://doi.org/10.1016/j.indcrop.2022.114767>
- [31] J. M. Aguilar, C. Bengoechea, E. Pérez, and A. Guerrero, "Effect of different polyols as plasticizers in soy-based bioplastics," *Ind. Crops Prod.*, vol. 153, art. 112522, 2020. <https://doi.org/10.1016/j.indcrop.2020.112522>
- [32] S. C. Teixeira, R. R. A. Silva, T. V. de Oliveira, P. C. Stringheta, M. R. M. R. Pinto, and N. D. F. F. Soares, "Glycerol and tri-ethyl citrate plasticizer effects on molecular, thermal, mechanical, and barrier properties of cellulose acetate films," *Food Biosci.*, vol. 42, art. 101202, 2021. <https://doi.org/10.1016/j.fbio.2021.101202>
- [33] M. D. R. Salazar-Sánchez, J. A. Cañas-Montoya, H. S. Villada-Castillo, J. F. Solanilla-Duque, R. Rodríguez-Herrera, and F. Ávalos-Belmontes, "Biogenerated polymers: An environmental alternative," *Dyna*, vol. 87, no. 214, pp. 75-84, 2020. <https://doi.org/10.15446/dyna.v87n214.82163>
- [34] R. Thakur *et al.*, "Starch-based films: Major factors affecting their properties," *Int. J. Biol. Macromol.*, vol. 132, pp. 1079-1089, 2019. <https://doi.org/10.1016/j.ijbiomac.2019.03.190>
- [35] L. Monroy, D. Fonseca, and C. E. Rodríguez, "Efecto del método de extracción del almidón de *Canna indica* L. sobre sus propiedades físico-químicas," *Rev. Lasallista Investig.*, vol. 16, no. 2, pp. 44-52, 2019. <https://doi.org/10.22507/rli.v16n2a4>
- [36] D. P. Sanabria *et al.*, "Synthesis of starch powder from different organic wastes: A green approach to a valuable material," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 1154, no. 1, p. 012041, 2021. <https://doi.org/10.1088/1757-899X/1154/1/012041>
- [37] R. Jamali, A. Babaei-Ghazvini, E. Nazari, M. Panahi, I. Shababi-Ghahfarrokhi, and A. R. Moradi, "Surface characterization of biodegradable nanocomposites by dynamic speckle analysis," *Appl. Surf. Sci. Adv.*, vol. 16, art. 100429, 2023. <https://doi.org/10.1016/j.apsadv.2023.100429>
- [38] V. Vamadevan and E. Bertoft, "Observations on the impact of amylopectin and amylose structure on the swelling of starch granules," *Food Hydrocoll.*, vol. 103, art. 105663, 2020. <https://doi.org/10.1016/j.foodhyd.2020.105663>
- [39] G. M. Ulfa, W. D. R. Putri, K. Fibrianto, R. Prihatiningtyas, and S. B. Widjanarko, "The influence of temperature in swelling power, solubility, and water binding capacity of pregelatinised sweet potato starch," *IOP Conf. Ser. Earth Environ. Sci.*, vol. 475, no. 1, art. 012036, 2020. <https://doi.org/10.1088/1755-1315/475/1/012036>
- [40] C. V. V. Martínez, X. S. Z. Murillo, M. H. D. Demera, G. A. B. Briones, and C. A. C. Palacios, "Almidones de cáscara de yuca (*Manihot esculenta*) y papa (*Solanum tuberosum*) para producción de bioplásticos: propiedades mecánicas y efecto gelatinizante," *Rev. Bases Cien.*, vol. 6, no. 2, pp. 137-152, 2021. https://doi.org/10.33936/rev_bas_de_la_ciencia.v6i2.3293
- [41] N. P. Stritzler and C. M. Rabotnikof, *Nutrición y alimentación de rumiantes en la región semiárida central argentina*. Santa Rosa, La Pampa, Argentina: EdUNLPam, 2019.
- [42] M. R. Chuiza-Rojas, A. I. Rodríguez-Basantes, and H. L. Brito-Molina, "Producción de láminas de plástico biodegradables a partir del almidón de arracacia xanthorrhiza," *Dom. Cienc.*, vol. 6, no. 2, pp. 981-994, 2020. <https://doi.org/10.23857/dc.v6i2.1261>
- [43] S. N. Saadatkah *et al.*, "Experimental methods in chemical engineering: thermogravimetric analysis—TGA," *Can. J. Chem. Eng.*, vol. 98, no. 1, pp. 34-43, 2020. <https://doi.org/10.1002/cjce.23673>
- [44] Z. U. Zango, "Measurement of thermal stability of polymeric cable wires using thermogravimetric analytical technique," *Int. J. Energy Eng.*, vol. 10, no. 1, pp. 16-21, 2020. <https://doi.org/10.5923/j.ijee.20201001.03>
- [45] *Standard test method for rubber property—Durometer hardness*, ASTM D2240, American Society for Testing and Materials, 2015.
- [46] F. J. Bernal-Ávila, O. H. P. Cuervo, and N. A. S. Guerrero, "Apósitos a base de almidón: revisión sistemática sobre métodos de caracterización fisicoquímica y su potencial en la cicatrización de heridas," *Cienc. Tecnol. Agropec.*, vol. 25, no. 2, 2024. https://doi.org/10.21930/rcta.vol25_num2_art:3612
- [47] L. A. C. Sares, G. A. Cabrera, D. A. Toro, and V. P. B. Bravo, "Efecto del glicerol como plastificante en películas de almidón de maíz modificado," *J. Sci. Res. Rev. Cien. Inv.*, vol. 8, no. 4, pp. 186-204, 2023. <https://doi.org/10.5281/zenodo.10045595>
- [48] D. N. Quintana, D. P. Sanabria, H. F. Castro, K. L. Roa, R. A. Paredes, and Y. C. Vargas, "Physical characterization of biopolymers with starch from potato and cassava organic wastes polymerized in water," *J. Phys. Conf Ser.*, vol. 2139, no. 1, art. 012015, 2021. <https://doi.org/10.1088/1742-6596/2139/1/012015>

- [49] H. F. Castro, H. I. Farfán, I. V. Abril, R. A. Paredes, and K. L. Roa, "Study of factors affecting hardness behavior of biopolymers based on potato and plantain peels: a factorial experimental evaluation," *J. Phys. Conf. Ser.*, vol. 1938, no. 1, art. 012009, 2021. <https://doi.org/10.1088/1742-6596/1938/1/012009>
- [50] H. Ahmad and D. Rodrigue, "High-performance wood-reinforced crosslinked high-density polyethylene composites," *Polym. Eng. Sci.*, vol. 64, no. 6, pp. 2459-2475, 2024. <https://doi.org/10.1002/pen.26702>
- [51] G. Galo Silva, M. L. da C. Valente, L. Bachmann, and A. C. dos Reis, "Use of polyethylene terephthalate as a prosthetic component in the prosthesis on an overdenture implant," *Mater. Sci. Eng. C*, vol. 99, pp. 1341-1349, 2019. <https://doi.org/10.1016/j.msec.2019.01.136>
- [52] J. Lai, K. Ramash, C. T. Ratnam, R. Bains, N. A. S. Abdul Samat, and S. Sar-ee, "Physico-mechanical properties polypropylene/ethylene-propylene diene monomer (PP/EPDM) binary blends," *J. Appl. Res. Technol.*, vol. 20, no. 5, pp. 546-553, 2022. <https://doi.org/10.22201/icat.24486736e.2022.20.5.1419>
- [53] M. Karaoui, R. Hsissou, M. Alami, and M. Assouag, "Thermal, flow, and mechanical properties of composites based on polystyrene (PS) and snail shell powder (SSP) bio-filler (PS/SSP)," *Iran. Polym. J.*, vol. 32, no. 5, pp. 621-631, 2023. <https://doi.org/10.1007/s13726-023-01151-2>

Investigating the Effects of Cutting Methods for Aluminum Metallic Foams

Investigación de los efectos de los métodos de corte para espumas metálicas de aluminio

Altug Bakirci ¹, Tayfun Sigirtmac ², Mustafa Cemal Cakir ³, and Agah Uguz ⁴

ABSTRACT

The quality of foam metal materials, which are commonly used in industrial applications due to their unique properties, increasingly relies on secondary processing. Metal foams are produced as plates or in the desired shape via direct or indirect foaming. Given their intended use, the primary challenge lies in determining how to cut them with the necessary precision and join them with sufficient strength. However, the most difficult aspect is cutting them in the required shapes and combining them with fixed or removable securing mechanisms. This work involved cutting two sample types: a 10 mm thick AlMgSi foam with a density of 0.5 g/cm³, using a laser cutter, and a 19 mm thick AlMgSi sandwich structure with a 1 mm thick aluminum outer plate via wire electric discharge machining (WEDM). In addition, the results of manual cutting and angle grinding, which are often utilized in production, were analyzed through scanning electron microscopy. Under certain suboptimal conditions, laser cutting caused aluminum to dissolve into the cavities and form burrs beneath the cutting edge. In contrast, when accurate and undistorted cellular architecture is essential, WEDM is very efficient, even though it is 200 times slower than laser cutting. Hand-sawing caused cellular fractures and frequent dispersion, so it is suitable for applications that do not necessitate accuracy.

Keywords: aluminum foam metals, secondary operations, laser cutting, wire EDM cutting

RESUMEN

La calidad de los materiales de espuma metálica, que se utilizan comúnmente en aplicaciones industriales debido a sus propiedades únicas, depende cada vez más del procesamiento secundario. Las espumas metálicas se producen en placas o en la forma deseada mediante espumación directa o indirecta. Dado su uso previsto, el principal desafío radica en determinar cómo cortarlas con la precisión necesaria y unir las con la fuerza suficiente. Sin embargo, el aspecto más difícil es cortarlas en las formas requeridas y combinarlas con mecanismos de sujeción fijos o removibles. Este trabajo involucró el corte de dos tipos de muestras: una espuma de AlMgSi de 10 mm de grosor con una densidad de 0.5 g/cm³, utilizando un cortador láser, y una estructura tipo sándwich de AlMgSi de 19 mm de grosor con placa exterior de aluminio de 1 mm de grosor mediante mecanizado por electroerosión por hilo (WEDM). Además, los resultados del corte manual y del rectificado angular, que a menudo se utilizan en la producción, se analizaron a través de microscopía electrónica de barrido. En ciertas condiciones subóptimas, el corte láser hizo que el aluminio se disolviera en las cavidades y formara rebabas debajo del borde de corte. En contraste, cuando una arquitectura celular precisa y no distorsionada se hace esencial, el WEDM es muy eficiente, a pesar de ser 200 veces más lento que el corte láser. El corte manual causó fracturas celulares y dispersión frecuente, por lo que es adecuado para aplicaciones que no requieran precisión.

Palabras clave: espumas metálicas de aluminio, operaciones secundarias, corte por láser, corte EDM por hilo

Received: October 23rd, 2023

Accepted: November 11th, 2024

Introduction

Foam metals are typically porous metal structures, primarily made of aluminum, that contain gas-filled voids. Metallic foams surpass polymer foams in a variety of engineering characteristics; they are more resilient, stable at higher temperatures, and combustible, producing no hazardous byproducts during combustion. In addition, their recycling poses no contamination or pollution concerns.

To achieve a more stable production of foam metals, it is now feasible to use secondary post-production techniques outside of mass production, rather than relying on the more difficult modulation of pore and cell size during manufacturing.

¹ PhD(c) in Mechanical Engineering, Bursa Uludag University, Turkey. Affiliation: Researcher in Bursa Uludag University Mechanical Engineering, 16059 Turkey. Email: altugbakirci@uludag.edu.tr

² PhD in Mechanical Engineering, Bursa Uludag University, Turkey. Affiliation: Managing Director in Kromas, Istanbul, Turkey. Email: t.sigirtmac@kromas.com

³ PhD in Mechanical Engineering, University of Bath, Engineering Department, Mechanical Engineering, United Kingdom. Affiliation: Prof. Dr. in Bursa Uludag University Mechanical Engineering, 16059 Turkey. Email: cemal@uludag.edu.tr

⁴ PhD in Mechanical Engineering, University Of Oxford-St. Catherine's College, Materials Science, United Kingdom. Affiliation: Prof. Dr. in Bursa Uludag University Mechanical Engineering, 16059 Turkey. Email: uguz@uludag.edu.tr



Attribution 4.0 International (CC BY 4.0) Share - Adapt

This makes it imperative to examine removable and non-removable connection options such as cutting, drilling, machining, welding, and fastener connection. It is crucial for structures to be demountable and durable, particularly for construction purposes. The pervasive application and utility of foam metals are a result of connection methods that enable the creation of more intricate shapes or the attainment of thicknesses that cannot be produced even through casting.

The secondary operations of metal foams are crucial to their applicability, so they are of great importance. There are five main operations in this category: cutting, forming, joining, finishing, and coating [1].

In cutting procedures, it is preferable for the cells of the metal foam to retain their shape. The most prevalent techniques for cutting metal foams are band sawing, circular sawing, wire sawing, diamond sawing, chemical milling, wire electric discharge machining (WEDM), laser cutting, and water-jet cutting. While each strategy offers a number of advantages, it also has certain disadvantages and obstacles. With WEDM techniques, for instance, it is possible to prevent cell degradation [2], but the operation's cost and pace must also be considered.

Wire sawing is more effective than circular-sawing, band-sawing, and EDM at minimizing macroscopic surface roughness (*i.e.*, the quality of the walls/struts in relation to their distance from the cutting plane), according to the study on secondary processes on foams [2].

Wire sawing, circular sawing, wire EDM, and band sawing have been employed to obtain both high and low surface roughness values for 10 PPI foam. The WEDM cutting method produces significantly smaller contact points, while wire sawing-cutting exhibits the largest surface contact area, almost equal to the nominal value [3]. In a study involving the laser cutting of 9 mm-thick closed-cell aluminum foams, thermal issues were observed. In spite of this, it was determined that laser cutting allowed for burr-free and parallel-sided cutting without cell damage [4].

In a previous study on the effect of cutting on near-eutectic silumin (AlSi9) metal foam, the edge geometry was determined [5]. The experiment included circular saws, band saws, WEDM, water jets, and thermal cutting techniques (laser and air plasma). Each of the examined approaches had a unique effect on the aspects of the section plane. The problems arose from the cellular structure of the foams and their minimal cell wall thickness (no more than 0.3 mm), while the presence of voids in the materials allowed cutting residues to accumulate in the voids. WEDM was determined to be the most advantageous method for cutting aluminum foam since it provides the highest edge surface quality for welding. Furthermore, among techniques that do not result in the partial ablation of the material, WEDM and water jets with an abrasive agent produce the best results. Regarding

the thermal cutting procedures, a highly concentrated laser beam yields the most effective results [5].

In their review, [6] pointed out that, despite the volume of existing research, the quality characteristics of laser-cut metal foam have not been ascertained. This was evident in the summary of input process parameters and quality responses resulting from the cutting process. The authors provided a qualitative report on the extremely low dross attachment at the bottom of laser-cut foam, but the abnormalities generated on its kerf wall were significantly different from the striations found on laser-cut metal sheets. In addition, the authors cited research that utilized bending instead of laser cutting, reporting that, as the thickness increases, problems arise [6]. Consequently, laser cutting research must continue to reach success, particularly when dealing with high thickness. Assembling individual metal foams or metal foam sandwich panels may entail several difficulties. In general, these obstacles originate from the deformability of the foam and the limited surface area that is available for joining.

[7] cut closed-cell metal foams with a laser while trying different parameters, noting that the gas trapped in the closed cells emerged during cutting and disrupted the process. Issues such as spatter dross and kerf formation were reported during their experiments. The authors mentioned that the worst results were observed when cutting with oxygen gas. Nitrogen assist gas created the least amount of dross, whereas argon assist gas provided the lowest kerf width. [8] discussed the importance of having smooth surfaces in foam metal butt welding operations. The goal of this work was to combine metal foams using friction stir and induction welding without deteriorating their cellular structure. It was necessary to perform an additional grinding operation to ensure smooth surfaces.

There are three ways to join sandwich panels: joining two sandwich panels, joining a sandwich panel to another material, and joining the face sheets to the foam core (*i.e.*, adhesive or metallurgical joining). The metallurgical joining of face sheets to the foam core (*i.e.*, brazing) results in greater resistance to tearing than adhesive joining. To join metal foam structures, the literature specifies soldering, brazing, diffusion welding, friction stir welding, laser welding, bonding, and fastening (threaded, riveted, or pinned) as the available methods [9]-[21].

These procedures may produce burrs and other surface flaws, and with metal foams, these issues are somewhat more complicated. During refining, negative burr development (smearing) can be observed in metal foam structures. Failure of subsequent processes, such as joining or coating, may result in defective surfaces. It is important to recall that the functionality, safety, cost, and aesthetic appeal of a product are all dependent on the quality of its margins. Poorly finished edges can result in increased or altered friction, higher wear upon movement or stressed elements, interferences, turbulent flow, and decreased formability

[22], [23]. In a study on the development of a heat transfer, the need for shaped cutting of the foam was mentioned [24]. Moreover, [25] had issues with surface contact for bonding and airflow due to the deformation of a foam metal while cutting a heat sink.

It is not always possible to obtain complex component shapes through molding, which is also an expensive procedure. It would be more practicable to create the required shapes by cutting, punching, machining, bonding, welding, and fastening standard foam materials, among others. Foam materials are not suitable for post-production due to their characteristics; because of their compressibility, they are not suitable for forming, which is made challenging by low tensile strength.

Coating is made difficult by irregularities in cell walls and surface layers. The surface is frequently covered with an oxide layer, which makes coating, brazing, and welding difficult. The presence of melting-stabilizing ceramic particulates also hinders machining.

The outer shell significantly improves the characteristics and appearance of foam material. Consequently, if this layer is not required for functionality, its removal, which would increase production costs, is unnecessary. During the design phase, the most difficult aspect is machining. Nonetheless, the components may require additional machining and drilling. In theory, foam materials are amenable to all conventional processing methods, but it is difficult to achieve a high level of surface quality.

[26] mentioned that geometric accuracy is very important for critical design parts such as crash boxes in automobiles, highlighting the importance of obtaining surface parallelism. In this vein, they examined the integrity of the surface as a result of turning the foam upon reaching different parameter values. If ceramic particles were added to the melt to stabilize the liquid foam, especially in large quantities of SiC particles, significant tool wear could not be avoided [27]. In brittle metals, the typical procedure induces cell wall bending and compression, as well as fractures and tearing. This leads to a lack of quality and sensitivity on the surface. The partial melting of the microscopic pore walls and their subsequent diffusion into the cutting tool has a significant effect on the precision and quality of the cutting surface. Low thermal conductivity and high porosity prevent effective heat convection cooling. Conversely, the melting and bending of the pore walls reinforce the cutting surface; as they press into the pores, the curved walls thicken and densify the affected region.

There is no study that combines methods such as laser and WEDM with traditional techniques such as hand-sawing and grinding saws. In this field, it is important for the use of foam metal to become widespread in small workshops, given its functionality, lightness, and low material requirements, in addition to its environmentally friendly nature and its contribution to reducing the carbon footprint [28]-[31].

Foam metals have different applications in many different areas [32], [33]. Aluminum (Al) foam materials, with superior characteristics compared to traditional materials, are currently vying for a position in the market. To ensure their extensive adoption, it is essential to understand their mechanical properties and post-production suitability. In this study, in addition to secondary processes such as WEDM and laser cutting, which are used to cut Al foam materials, experimental research and SEM observations are conducted in relation to the post-production processes of hand saw cutting and angle grinding, which is commonly used in the industry but has received little attention from the literature. A comparison is made as a result of this visual examination. Given the incidence of different factors on laser cutting and WEDM, several parameters were tested prior to the comparison. The *Materials and method* section provides information on the materials and cutting methods used in this work. The focus of the comparison was on the deterioration of cell wall surfaces after employing different cutting methods.

Materials and methods

Material characterization studies were carried out based on the information provided by the manufacturer. The selection of cutting parameters is detailed later in this document. After cutting, SEM images of the samples were taken. The flowchart of this study is presented in Fig. 1.

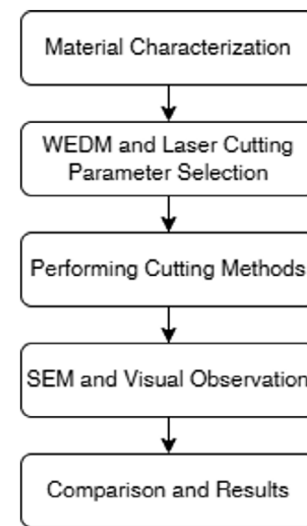


Figure 1. Flowchart of this study
Source: Authors

Materials

Compression tests of Al foam were carried out in accordance with the ISO 13314 standard [34]. To this effect, five 30 x 30 x 30 mm³ specimens were prepared (Fig. 2), and they were compressed at a rate of 20 mm/min. All samples in this research utilized the same foam product.

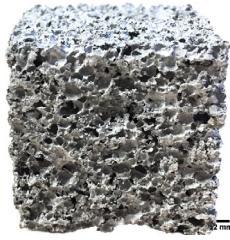


Figure 2. Compression test sample of aluminum foam metal
Source: Authors

Two types of Al foam, sandwich and foam-only, were used as test specimens. The sandwich-structured foam material was produced from a casting foam material (AlMgSi) and 1.2 mm of Al alloy plate. Its total thickness was 19 mm. The other type of foam was AlMgSi foam metal, with a thickness of 10 mm and a closed-cell structure. The length of each sample was 70 mm, the bulk density of the foams was 0.5 gr/cm³, and their modulus of elasticity was 5 GPa (Table I). The chemical properties of the foam metal are outlined in Table II.

Table I. Specifications of the aluminum foam

Specification	Value
Production	Melting with blowing agent
Type	Closed-cell foam
Blowing agent	CaCO ₃
Stabilizing additive	SiC and MgO
Bulk Density	0.500 gr/cm ³
Elastic Modulus	5 GPa
Yield strength	~2 MPa
Poisson ratio	0.3

Source: By courtesy of Alupam

Table II. Chemical composition of the AlMgSi foam

Chemical element	Weight %
Fe	0.50
Si	1.30
Cu	0.10
Mn	0.80
Mg	1.20
Zn	0.20
Ni	0.10
Ti	0.10
Pb	0.05
Sn	0.05
Cr	0.25

Source: By courtesy of Alupam

Laser cutting

The variables that determine laser cutting quality are cutting speed, gas type, gas pressure, cutting power, focal length,

nozzle diameter, and nozzle-to-sheet distance. In our experiments, the effects of nozzle diameter, focal length, gas type, and nozzle-to-sheet distance were kept constant because they are negligible. It is worth adding that the material and the desired cutting technique determine the gas type. In our work, nitrogen gas was utilized, given the difficulties in regulating the oxygen reaction and its influence on the cutting quality of Al foam. The diameter of the nozzle is a material thickness-related characteristic. There are specific nozzle diameter values for removing slag at specific cutting thicknesses and obtaining the appropriate gas flow. The distance between the nozzle and the sheet determines the laser beam's focal point, i.e., the region of the material on which the beam will be focused following lens selection [6], [35]-[37].

Since nitrogen-based laser cutting does not result in an exothermic reaction, the laser beam was focused on the bottom surface of the material. In this context, cutting speed, gas pressure, and power were considered to be the most significant variables, whereas the gas type and the focus point were kept constant due to the method and the material type used. To conduct our cutting experiments, three distinct parameters were varied, with cutting speed and surface quality serving as comparison criteria. Attempts were made to adjust power and pressure parameters, hoping to address issues such as the inability to conclude the operation at high cutting speeds and the inability to remove molten material from the environment at low speeds.

The cutting process of the AlMgSi closed-cell foams was carried out while considering various parameter values (Table III). Using a Durma Hybrid CNC laser cutter with a maximal output of 3.3 kW, CO₂ laser cutting was performed.

Table III. Parameters used in laser cutting

Parameters	1. Level	2. Level	3. Level
Cutting speed (m/min)	1	3	5
Laser power (kW)	3	3	3.3
Assist gas pressure (bar)	17	15	16
Assist gas type	N ₂	N ₂	N ₂
Nozzle diameter (mm)	1	1	1
Distance between the nozzle and the work piece (mm)	1	1	1
Focal length (mm)	5	5	5

Source: Authors

Wire EDM cutting

In WEDM cutting, also referred to as *wire erosion*, there are specific requirements for certain material categories. For materials such as Al and steel, we considered the manufacturer-specified fundamental parameters. As a variable parameter, only the cutting speed could be altered

when necessary. WEDM cutting is a comparatively slow procedure that, by its very nature, allows obtaining fine surface qualities. Sparks and chemical reactions need a certain amount of time to eliminate fragments from a substance. This is the most significant disadvantage of this method.

Using a Makino EV64 CNC wire erosion cutting machine and a Master Brass 0.25 mm diameter CuZn37 wire, sandwich plates with a total thickness of 19 mm were cut into 30 x 67 mm pieces. Tables IV and V list the cutting parameters for the sandwich plates and the closed-cell foam.

Table IV. WEDM cutting parameters for the sandwich foam material

Parameters	1. Level	2. Level
Wire diameter (mm)	0.25	0.25
Wire material	Brass	Brass
Cutting speed (mm/min)	15	22

Source: Authors

Table V. WEDM cutting parameters for the 10 mm closed-cell foam

Parameters	1. Level	2. Level	3. Level	4. Level
Wire diameter (mm)	0.25	0.25	0.25	0.25
Wire material	Brass	Brass	Brass	Brass
Cutting speed (mm/min)	13	16	22	25

Source: Authors

Hand-sawing and grinding saw cutting

In this study, hand-sawing and angle grinding, which are commonly preferred for metal cutting in small workshops, were evaluated for cutting Al foam. Depending on the operator's attention, strength, and dexterity, the results of using these instruments may vary. This cutting equipment is frequently used because it is convenient, saves time, is easily accessible, and requires no specialized personnel. The effect of cutting foam metals with such tools during maintenance, repair, or production is a crucial concern, as these materials can be utilized in a variety of applications.

In this research, an 11 000 rpm Makita angle grinder with 720 W of power was utilized. The cutting disc had a diameter of 115 mm, a thickness of 1 mm, and aluminum oxide as abrasive material. The blade length of the handsaw used was 300 mm, and it was made of HSS and had 24 teeth per inch.

Results and discussion

The compression test revealed a modulus of elasticity of 5 GPa and a yield stress of around 2 MPa, confirming the accuracy of the manufacturer's data. The results of the test conducted on the samples are illustrated in Fig. 3.

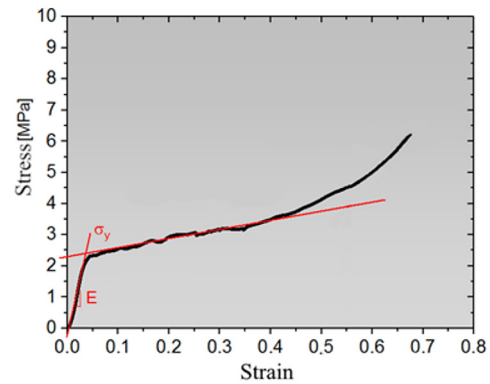


Figure 3. A stress strain graph was observed during the compression testing of foam metal

Source: Authors

During the laser cutting procedure, the molten metal adheres to the cell walls and disrupts the structure of the cells. In this context, changing parameters becomes essential. [4] studied a single-parameter experiment, with very poor outcomes in laser-cut foam. This may be misleading when comparing against other cutting techniques, underscoring the relevance of parametric work in the laser cutting of foam metal. Due to the high speed of 5 m/min in level 3, the bottom surface of the foam metal was not completely decomposed. The anticipated power of 3.3 kW, which is comparatively high, was deemed enough for a speed of 5 m/min. High-speed cutting was expected to sever the walls without slag formation or heat accumulation. Nonetheless, a subpar surface and inadequate cutting led to the rejection of this proposal. Due to the cutting speed, there is insufficient time for the material to dissolve and be removed from the environment, preventing the completion of the process.

The speed value of 3 m/min in level 2 was enough for laser cutting. Here, the nitrogen pressure was decreased by 1 bar, and the power was decreased to 3 kW (less power was required because the speed was reduced and the nitrogen pressure was adjusted to 15 bar, since this gas removes molten metal and serves as a coolant). Even if the cut surface has burrs, it can be used in non-sensitive applications or after a secondary grinding or sanding.

In level 1 laser cutting, the cutting speed was 1 m/min, and, during the experiments, the molten Al could not be removed from the environment. This was due to the slow progression of heat density in the cutting zones. Since the speed was slow, we chose a high gas pressure, knowing that excessive melting would occur due to excessive heat accumulation. We expected the fast flow to cool the surface. However, the high gas pressure was not sufficient to cool the surface and created a slag problem. Despite the high pressure (17 bar), the molten metal filled the foam cavities, and the quality of the surface decreased significantly. Under these conditions, it is not possible to combine the foam material with the cut surfaces, or to align the cut surfaces and use them as reference. The molten metal adhered to the cutting surfaces, and, since it could not be removed from the environment,

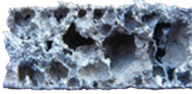
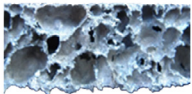

it cooled and solidified, compromising the uniformity of the surface. This suggests that a slower process cannot increase the quality of the surface. However, this does not mean that the speed can be increased as the power increases. Similarly, increasing the gas pressure does not mean that the material will be cut more smoothly and that the slag will be removed.

Table VI shows the resulting cut surfaces for each level. In their parametric study of metal foam laser cutting, [23] stated that the gas was unable to remove the dross well, which was associated with the porous structure under the impact of gas pressure. Furthermore, low power and high cutting speeds were seen to be beneficial in their research. In this study, we discovered that, when using excessively high speeds, the power is insufficient and the operation cannot be completed. When the cutting speed is regulated with respect to the power, more unspoiled cell walls may be obtained. Similarly, it cannot be claimed that removing dross by increasing gas pressure results in a smooth cell wall structure, yet it has been seen that decreasing gas pressure improves this aspect.

Even after the parameters were modified, the laser beam varied slightly as it passed through the air space between the foam metal and the other cell wall, resulting in an inhomogeneous cut between the layers. The greatest advantage of laser cutting is the ability to reach fast cutting speeds (3-5 m/min). Still, poor surface quality in the cut areas of the outer walls, the adhesion of molten Al to the cell walls, and the degradation of the porous structure are among the disadvantages of this method. [5] cited tapering during cutting as an application constraint in their review work, and they discovered laser-cutting research to use a maximum thickness of 10 mm.

During the wire erosion cutting operation, no degradation of the cut surfaces was observed. When compared to other cutting techniques, the quality was deemed to be very good, as the surface retained the shape of the cell walls, could be mounted with direct contact, and did not require a second operation. The surface could also serve as a surface of reference. In a similar study employing EDM during cutting, [2] discovered that the cell walls were not distorted. When filler material was employed, the surface became flat and the cells were closed, indicating good assembly capabilities. As a result, the authors avoided recommending EDM for filler material connection procedures. Moreover, it should be noted that the cutting speed of EDM was approximately 200 times slower than that of laser cutting, which constitutes the greatest disadvantage of this method. Fig. 4 depicts the sandwich panel that was sliced using EDM.

Table VI. Surfaces resulting from different laser cutting parameter levels

Laser cutting macro view		
Level 1	Level 2	Level 3
		

Source: Authors

Samples of sandwich Al foam were cut at different speeds (15 and 22 mm/min) at wire erosion levels 1 and 2. In both instances, the cutting process was completed without incidents. For wire erosion levels 1 (13 mm/min), 2 (16 mm/min), and 3 (22 mm/min), the cutting surface provided an excellent cut of the 10 mm thick closed cell; there was no cell deformation. Throughout the cutting process, there were no difficulties. At level 4 (25 mm/min), the wire broke, bringing the operation to a halt. The speed increase was expected to make a significant difference in the cut surface. However, excessive speed only caused the wire to break faster and the process time to increase. There was no significant difference in the surfaces.



Figure 4. Surface obtained via the wire erosion cutting of sandwich plates

Source: Authors

Among other factors, electrode deterioration depends on the material's electrical and thermal conductivity, melting temperature, and WEDM electrical signals. Wire fracture occurred when the cutting speed exceeded a certain threshold during the process, which varied depending on the alloys, pore structure, and electrical conductivity. This value represents the maximum velocity for the cut's thickness. Fig. 5 illustrates how WEDM and laser cutting alter the cell structure. The best quality obtained for laser cutting was considered.

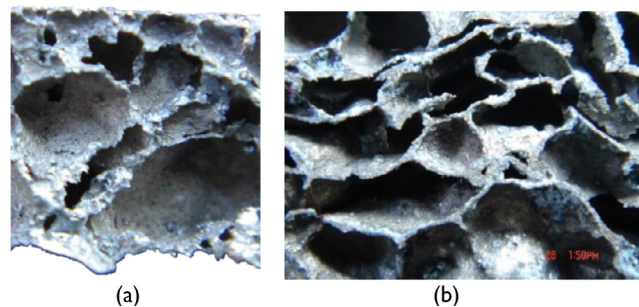


Figure 5. Comparison between a) laser cutting and b) WEDM cutting

Source: Authors

The handsaw cutting experiment demonstrated the role of mechanical force in the fragmentation and crushing of cell walls. Crushed and broken cells were observed since the pressures generated during the process forced the cell walls inward, and hand-sawing is a continuous process involving a frequency-induced mechanical strain. On the other hand, despite the lack of fragmentation, the angle grinding drove the cell walls inward and left visible traces in the cut regions (Figs. 6 and 7). For comparison, after the tests using a circular saw in angle grinding [2] and [4], it was discovered that the pore structure had deteriorated, and the pores had become partially clogged.

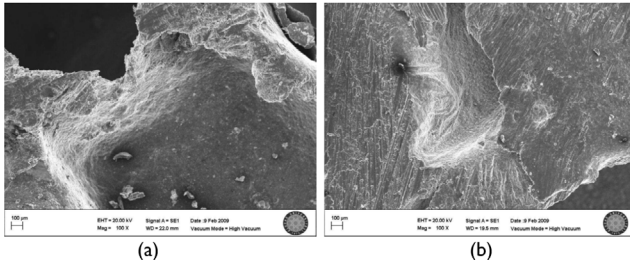


Figure 6. SEM images (x100): a) manual sawing, b) angle grinding
Source: Authors

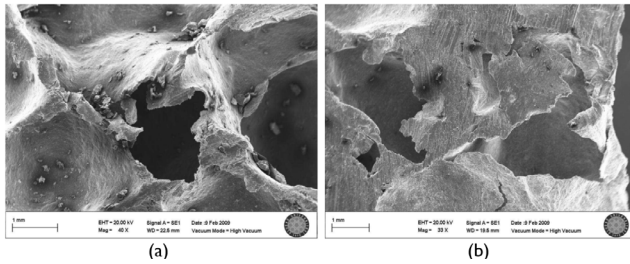


Figure 7. SEM images: a) manual sawing (x40), b) angle grinding cutting (x33)

Despite the changing parameter values, there was no deterioration or change in surface quality. Wire breakage was observed above certain cutting speeds, and the cutting operation could not be completed. Cutting of exceptional quality was accomplished through wire erosion. During this process, a highly clean and smooth pore structure was generated due to the absence of metal melting (which is common after a thermal operation such as laser cutting) and mechanical stress on the cell walls. The cell structure did not change as a result of wire erosion cutting. The anomalies that occur during manufacture include variations in pore size and inhomogeneous distribution. No surface quality degradation was noticed even after increasing the speed increase up to the wire breakage point. Figs. 8 and 9 show the surface patterns of the speed-dependent SEM images.

Due to thickness and focusing challenges during the laser operations, molten Al flowed over itself because of canonicalization and thermal effects, resulting in the closure of pores and the formation of a burr with rounded edges. Certain issues were resolved by modifying the settings, but the burr and conical surface remained. As previously indicated, throughout the WEDM tests, no damage to the cells was seen. Fig. 10 compares the relative cell deformation yielded by the different cutting techniques.

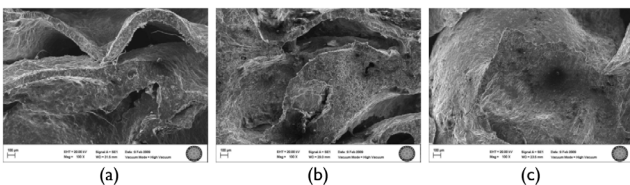


Figure 8. WEDM cutting SEM images (x100): a) V = 13 mm/min; b) V = 16 mm/min; c) V = 22 mm/min
Source: Authors

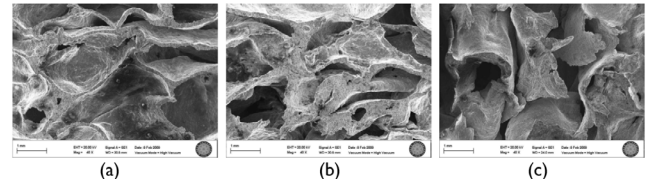


Figure 9. WEDM cutting SEM images (x40): a) V = 13 mm/min; b) V = 16 mm/min; c) V = 22 mm/min

Source: Authors

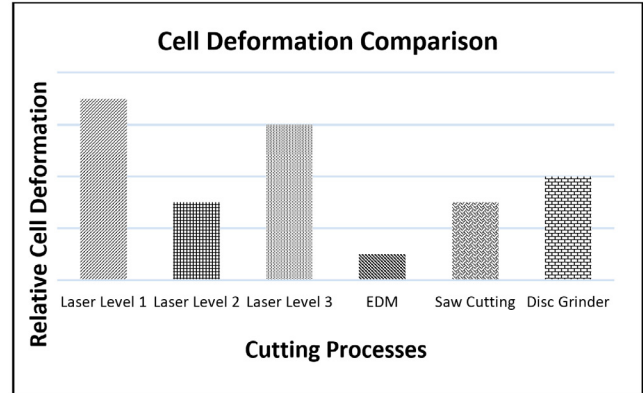


Figure 10. Comparison of relative cell deformation results

Source: Authors

One of the most significant aspects to consider when selecting the cutting technique is where the post-cut parts will be used. For non-precise mass production applications, laser (thermal) or saw (mechanical) cutting is acceptable, but wire erosion is ideal for applications demanding surface quality, shape, and measurement precision.

These values are regarded as average. Deviations of up to 6 s were observed in the hand-sawing of the sandwich plate, whereas a deviation of 2 s was noted during angle grinding. Both methods limited the variations in foam cutting to 1-2 s. The most suitable speeds for achieving the best surface integrity were 3 m/min for laser cutting and 22 mm/min for the WEDM method. The latter was 136 times slower than the former in terms of processing times, and the cutting speed of the grinding discs was competitive. Nonetheless, it relied on the operator's proficiency in executing shaped cuts. Similarly, the duration of cutting may fluctuate based on the operator's weariness and proficiency. The processing times of the studied cutting methods are presented in Table VII.

Table VII. Processing times of the studied cutting methods

Cutting method	Foam-only processing time (s)	Sandwich processing time (s)
WEDM	190.9	190.9
Laser	1.4	-
Grinding saw	3.2	5.8
Handsaw	13.4	33.8

Source: Authors

Conclusion

The final state of foam metals is dictated by the use of secondary techniques beyond mass production, as well as by the challenging control of pore and cell size. As a result, it is critical to investigate removable and non-removable connection methods such as cutting, drilling, machining, welding, and bolt connection.

This study examined the subsequent operations on aluminum foam metals. Closed-cell foam samples were processed via CO₂ laser cutting technology, WEDM, hand-sawing, and angle grinding while also studying their machinability. The results were compared in terms of the methods used and the resulting surfaces. Structural changes were evaluated using a scanning electron microscope (SEM). The following conclusions can be drawn from this study:

- WEDM can cut foam metal cell walls with minimum deformation compared to other methods.
- The parameters need to be adjusted to achieve low cell wall deformation during laser cutting.
- Hand-sawing and angle grinding break the cell wall surface.

Laser cutting may be used to quickly and precisely cut aluminum metal foams (especially for intricate shapes) up to 19 mm thick, as long as the planned applications do not require a high degree of accuracy. A saw cannot be used to cut a star form, but a laser can cut any shape. Although slower than laser cutting, EDM can be employed in thick and delicate applications where the pore structure must be maintained. Hand-sawing and angle grinding can be used in applications where fractures on the outer cell surface are not important. Following these operations, the most important factors to consider are the components' application areas, cell structure, and pore size.

Future research could explore a method that utilizes digital image processing, machine learning, and artificial intelligence to optimize laser cutting settings and rapidly adjust them based on foam thickness [38]. Water-guided laser cutting technology, which has shown remarkable precision, could be also studied [39], [40], comparing it against WEDM – attempting to achieve a comparable accuracy at the speed of laser cutting would be an intriguing endeavor. Furthermore, other studies evaluate the peeling and bonding strength outcomes of foams with surfaces cut by various techniques.

Acknowledgments

The authors thank ALUPAM A.Ş. for supplying the aluminum foam.

CRediT author statement

All authors: conceptualization, methodology, formal analysis, investigation, writing (original, draft preparation, writing, reviewing, and editing), data curation, supervision, project administration, resources, funding acquisition.

Conflicts of interest

The authors declare no conflict of interest.

References

- [1] C. Ensarioglu, A. Bakirci, H. Koluk, and M. C. Cakir, "Metal foams and their applications in aerospace components," in *Materials, Structures and Manufacturing for Aircraft*, Springer Cham, 2022, ch. 2, pp. 27–64. https://doi.org/10.1007/978-3-030-91873-6_2.
- [2] A. Changdar and S. S. Chakraborty, "State-of-the-art manufacturing of metal foams and processing—A review," in *Adv. Form., Mach. Autom.*, 2023, pp. 127–142. https://doi.org/10.1007/978-981-19-3866-5_11
- [3] P. De Jaeger, C. T'Joel, H. Huisseune, B. Ameel, S. De Schampheleire, and M. De Paepe, "Assessing the influence of four cutting methods on the thermal contact resistance of open-cell aluminum foam," *Int. J. Heat. Mass Transf.*, vol. 55, no. 21–22, pp. 6142–6151, 2012. <https://doi.org/10.1016/j.ijheatmasstransfer.2012.06.033>
- [4] B. S. Yilbas, *The laser cutting process*. Amsterdam, Netherlands: Elsevier, 2018. <https://doi.org/10.1016/C2016-0-04438-8>
- [5] S. Krajewski and J. Nowacki, "Preparation of Aluminium Foam Edges for Welding," *Adv. Mat. Sci.*, vol. 13, no. 3, 2013. <https://doi.org/10.2478/adms-2013-0012>
- [6] A. Changdar and S. S. Chakraborty, "Laser processing of metal foam - A review," *J. Manuf. Process.*, vol. 61, pp. 208–225, 2021. <https://doi.org/10.1016/j.jmapro.2020.10.012>
- [7] M. S. Raza, S. Datta, and P. Saha, "Thermal and morphological analysis of the interaction effect of different assist gases with gas-filled closed-cell aluminium foam during laser cutting process," *Proc. Inst. Mech. Eng. B J. Eng. Manuf.*, vol. 237, no. 13, pp. 1958–1969, Nov. 2023. <https://doi.org/10.1177/0954405420917252>
- [8] A. Rabiei, J. Cance, and Z. Chacko, "A study on welding of porous metals and metallic foams," *Adv. Eng. Mater.*, vol. 26, no. 4, Feb. 2024. <https://doi.org/10.1002/adem.202301430>
- [9] J. Zhang, Q. Qin, X. Han, and W. Ai, "The initial plastic failure of fully clamped geometrical asymmetric metal foam core sandwich beams," *Compos. B Eng.*, vol. 87, pp. 233–244, 2016. <https://doi.org/10.1016/j.compositesb.2015.10.027>
- [10] M. Khokhlov, D. Ishchenko, and J. Khokhlova, "Peculiarities of forming diffusion bimetallic joints of aluminum foam with a monolithic magnesium alloy," *J. Mag. Alloys*, vol. 4, no. 4, pp. 326–329, 2016. <https://doi.org/10.1016/j.jma.2016.11.001>

- [11] N. Chen, Y. Feng, J. Chen, B. Li, F. Chen, and J. Zhao, "Vacuum brazing processes of aluminum foam," *Xiyou Jinshu Cailiao Yu Gongcheng/Rare Metal Mat. Eng.*, vol. 42, no. 6, pp. 1118–1122, 2013. [https://doi.org/10.1016/s1875-5372\(13\)60072-7](https://doi.org/10.1016/s1875-5372(13)60072-7).
- [12] C. Yao, Z. Hu, F. Mo, and Y. Wang, "Fabrication and fatigue behavior of aluminum foam sandwich panel via liquid diffusion welding method," *Metals*, vol. 9, no. 5, pp. 1–11, 2019. <https://doi.org/10.3390/met9050582>
- [13] P. Peng et al., "High-performance aluminium foam sandwich prepared through friction stir welding," *Mater. Lett.*, vol. 236, pp. 295–298, 2019. <https://doi.org/10.1016/j.matlet.2018.10.125>
- [14] M. Bušić, Z. Kožuh, D. Klobčar, and I. Samardžić, "Friction stir welding (FSW) of aluminium foam sandwich panels," *Metalurgija*, vol. 55, no. 3, pp. 473–476, 2016. <https://hr-cak.srce.hr/153688>
- [15] J. Banhart, C. Schmoll, and U. Neumann, "Light-weight aluminium foam structures for ships," *Proc. Conf. Mat. Ocean. Environ.*, vol. 1, no. January 1998, pp. 55–63, 1998.
- [16] J. Nowacki, S. Krajewski, and J. Grabian, "Problems of aluminum foam soldering," *Przegląd Spawalnictwa – Weld. Tech. Rev.*, vol. 86, no. 1, 2014. <https://doi.org/10.26628/ps.v86i1.114>
- [17] T. Bernard, H. W. Bergmann, C. Haberling, and H. G. Haltenwanger, "Joining technologies for Al-foam-Al-sheet compound structures," *Adv. Eng. Mater.*, vol. 4, no. 10, pp. 798–802, 2002.
- [18] B. Castanie, C. Bouvet, and M. Ginot, "Review of composite sandwich structure in aeronautic applications," *Comp. P. C: Open Acc.*, vol. 1, 2020. <https://doi.org/10.1016/j.jcomc.2020.100004>
- [19] J. Feldhusen, C. Warkotsch, and A. Kempf, "Development of a mechanical technology for joining sandwich elements," *J. Sand. Struct. Mater.*, vol. 11, no. 6, pp. 471–486, 2009. <https://doi.org/10.1177/1099636209105378>
- [20] A. M. Joesbury, "New approaches to composite metal joining", PhD dissertation, School Aersp., Trans. Manuf. Cranfield Univ., Cranfield, UK, 2015. <http://dspace.lib.cranfield.ac.uk/handle/1826/10009>.
- [21] L. Wan, Y. Huang, T. Huang, Z. Lv, and J. Feng, "Interfacial behavior and mechanical properties of aluminum foam joint fabricated by surface self-abrasion fluxless soldering," *J. Alloys Comp.*, vol. 671, pp. 346–353, 2016. <https://doi.org/10.1016/j.jallcom.2016.01.246>
- [22] J. Liska, K. Kun, and K. Liska, "MMC materials ultrasonic machining and its economic aspects," *Proc. Eng.*, vol. 149, pp. 245–256, 2016. <https://doi.org/10.1016/j.proeng.2016.06.663>
- [23] J. Liska et al., "Evaluation of material structure changing after ultrasonic milling of aluminum foam by Computed Tomography (CT)," *16th IMEKO TC10 Conf. 2019 Test. Diag. Insp. Comp. Value Chain Qual. Safe.*, pp. 45–49, 2019. <https://www.imeko.org/publications/tc10-2019/IMEKO-TC10-2019-005.pdf>
- [24] J. Shi, H. Du, Z. Chen, and S. Lei, "Review of phase change heat transfer enhancement by metal foam," *Appl. Therm. Eng.*, vol. 219, art. 119427, Jan. 2023. <https://doi.org/10.1016/j.applthermaleng.2022.119427>
- [25] P. Samudre and S. V. Kailas, "Thermal performance enhancement in open-pore metal foam and foam-fin heat sinks for electronics cooling," *Appl. Therm. Eng.*, vol. 205, art. 117885, Mar. 2022. <https://doi.org/10.1016/j.applthermaleng.2021.117885>
- [26] Ç. Bolat, B. Ergene, U. Karakılınc, and A. Gökşenli, "Investigating on the machinability assessment of precision machining pumice reinforced AA7075 syntactic foam," *Proc. Inst. Mech. Eng. C. J. Mech. Eng. Sci.*, vol. 236, no. 5, pp. 2380–2394, Mar. 2022. <https://doi.org/10.1177/09544062211027613>
- [27] H. P. Degischer and B. Kriszt, eds., *Handbook of Cellular Metals - Production, Processing, Applications*. Hoboken, NJ, USA: Wiley, 2002.
- [28] V. Thulasikanth and R. Padmanabhan, "Fabrication of sustainable closed-cell aluminium foams using recycled fly ash and eggshell powder," *Mater. Today Commun.*, vol. 37, p. 107302, Dec. 2023. <https://doi.org/10.1016/j.mtcomm.2023.107302>
- [29] S. Sunder Sharma, S. Yadav, A. Joshi, A. Goyal, and R. Khatri, "Application of metallic foam in vehicle structure: A review," *Mater. Today Proc.*, vol. 63, pp. 347–353, 2022. <https://doi.org/10.1016/j.matpr.2022.03.201>
- [30] M. Sharaf, M. S. Yousef, and A. S. Huzayyin, "Year-round energy and exergy performance investigation of a photovoltaic panel coupled with metal foam/phase change material composite," *Renew. Energy*, vol. 189, pp. 777–789, Apr. 2022. <https://doi.org/10.1016/j.renene.2022.03.071>
- [31] M. S. Yousef, M. Sharaf, and A. S. Huzayyin, "Energy, exergy, economic, and enviroeconomic assessment of a photovoltaic module incorporated with a paraffin-metal foam composite: An experimental study," *Energy*, vol. 238, art. 121807, Jan. 2022. <https://doi.org/10.1016/j.energy.2021.121807>
- [32] R. Thiagarajan and M. Senthil kumar, "A review on closed cell metal matrix syntactic foams: A green initiative towards eco-sustainability," *Mat. Manuf. Proc.*, vol. 36, no. 12, pp. 1333–1351, Sep. 2021. <https://doi.org/10.1080/10426914.2021.1928696>
- [33] M. Madgule, C. G. Sreenivasa, and A. V. Borgaonkar, "Aluminium metal foam production methods, properties and applications- a review," *Mater. Today Proc.*, vol. 77, pp. 673–679, 2023. <https://doi.org/10.1016/j.matpr.2022.11.287>
- [34] *Mechanical testing of metals — Ductility testing — Compression test for porous and cellular metals*, ISO 13314:2011 2011.
- [35] O. Cavusoglu, "The 3D surface morphological investigation of laser cutting process of 2024-T3 aluminum alloy sheet," *Optik*, vol. 238, art. 166739, 2021. <https://doi.org/10.1016/j.jileo.2021.166739>
- [36] M. S. Raza, S. Datta, J. Singh, and P. Saha, "Fibre laser cutting of thick closed cell aluminium foam: Morphological analysis and its parametric optimisation," *Int. J. Prec. Tech.*, vol. 8, no. 2/3/4, art. 279, 2019. <https://doi.org/10.1504/ijptech.2019.10022603>

- [37] P. S. Mohammad Shahid Raza, S. Datta, and J. Singh, "Fibre laser cutting of thick closed cell aluminium foam," in *Proc. 10th Int. Conf. Prec. Meso, Micro Nano Eng.*, 2017, pp. 400-403. <https://doi.org/10.1504/ijptech.2019.10022603>.
- [38] G. Doğan, A. Özkiş, and M. H. Arslan, "A new methodology based on artificial intelligence for estimating the compressive strength of concrete from surface images," *Ing. Inv.*, vol. 44, no. 1, art. e99526, Jan. 2024. <https://doi.org/10.15446/ing.investig.99526>
- [39] M. Paksoy, H. Çandar, and N. F. Yılmaz, "The advancement of waterjet-guided laser cutting system for enhanced surface quality in AISI 1020 steel sheets," *Materials*, vol. 17, no. 14, p. 3458, Jul. 2024. <https://doi.org/10.3390/ma17143458>
- [40] C. Zhao *et al.*, "Multi-focus water-jet guided laser: For improving efficiency in cutting superalloys," *J. Manuf. Process*, vol. 119, pp. 729–743, Jun. 2024. <https://doi.org/10.1016/j.jmapro.2024.04.002>

Non-linear Dynamic Behavior Identification of a Quadcopter F450 Using an Artificial Neural Network-Based NARX Model

Identificación del comportamiento dinámico no lineal de un *quadcopter* F450 utilizando un modelo NARX basado en redes neuronales artificiales

Howard E. Sifuentes¹, Carlos A. Rocha², and Edgar A. Manzano³

ABSTRACT

A quadcopter drone is an extremely complex, multi-variable, highly nonlinear, and underactuated system characterized by its six degrees of freedom controlled by only four actuators as inputs. This highlights the importance of employing advanced algorithms for its identification. Therefore, this research aimed to use a nonlinear neural network model to identify the dynamic behavior of a quadcopter based on the commercially available F450 frame. Data acquisition involved four experiments in a controlled environment for both roll and pitch angles, recording the signal duty cycles and the quadcopter's attitude. Then, the selected non-linear autoregressive neural network model with exogenous inputs (N-NARX) model was trained using the acquired data along with the Levenberg-Marquardt algorithm. Afterwards, the response of the quadcopter's actual attitude angles from the validation dataset was analyzed against the predicted values generated by the neural model, obtaining an 89.44% fit with an RMSE of 2.25% for the roll angle and an 89.29% fit with an RMSE of 2.20% for the pitch angle. Both attitude angles were subjected to a statistical cross-correlation validation to assess their relationship at different time lags, observing a solid settling within the confidence bands at a 95% level. It was concluded the proposed neural network model can effectively capture the quadcopter's nonlinear dynamics.

Keywords: drone, system identification, underactuation, neural network, roll, pitch

RESUMEN

El dron *quadcopter* es un sistema extremadamente complejo, multivariable, altamente no lineal y subactuado, caracterizado por sus seis grados de libertad controlados por solo cuatro actuadores como entradas. Esto resalta la importancia de emplear algoritmos avanzados para su identificación. Por lo tanto, esta investigación tuvo como objetivo utilizar un modelo de red neuronal no lineal para identificar el comportamiento dinámico de un *quadcopter* basado en el *frame* comercial F450. La adquisición de datos involucró cuatro experimentos en un entorno controlado para los ángulos de alabeo y cabeceo, registrando los ciclos de trabajo de la señal y la actitud del *quadcopter*. Luego, se entrenó el modelo seleccionado de red neuronal autoregresiva no lineal con entradas exógenas (N-NARX), utilizando los datos adquiridos junto con el algoritmo de Levenberg-Marquardt. Posteriormente, se analizó la respuesta de los ángulos de actitud reales del *quadcopter* en el conjunto de datos de validación frente a los valores predichos generados por el modelo neuronal, obteniendo un ajuste del 89.44 % con un RMSE del 2.25 % para el ángulo de alabeo y un ajuste del 89.29 % con un RMSE del 2.20 % para el ángulo de cabeceo. Ambos ángulos de actitud fueron sometidos a una validación estadística de correlación cruzada para evaluar su relación en diferentes desfases temporales, observándose una sólida estabilización dentro de las bandas de confianza al nivel del 95 %. Se concluyó que el modelo de red neuronal propuesto puede capturar de manera efectiva las dinámicas no lineales del *quadcopter*.

Palabras clave: dron, identificación de sistema, subactuado, red neuronal, alabeo, cabeceo

Received: June 24th 2023

Accepted: September 27th 2024

Introduction

During the last decade, there has been an increase in the implementation of new technologies for unmanned aerial vehicles (UAVs) in multiple fields, e.g., logistics, surveillance and monitoring, public security, among others (Macrina *et al.*, 2020). Many cases can be highlighted, such as the design and construction of the S4 Éhecatl UAV by Hydra Technologies in Mexico, intended for the surveillance and monitoring of hazardous terrain, or the use of a UAV by the Geophysical Institute of Peru in a volcano monitoring project, which flew over the Ubinas volcano in Moquegua over 6000 m above sea level (Saito, 2019).

A quadcopter is a type of UAV that lifts and propels itself with four motors, which control its stability and mobility by varying their rotational speeds (Ahmad *et al.*, 2020).

The most commonly used controller in quadcopters is the classic PID (proportional-integral-derivative) control system, given its ease of implementation and parameter adjustment (Cedro and Wieczorkowski, 2019). Nevertheless, quadcopters pose a significant challenge in terms of control due to their complex nature, as their dynamic model is nonlinear, multivariable, strongly coupled, and underactuated, a result of the six degrees of freedom being

¹ Mechatronics Engineering, Universidad Nacional de Trujillo, Perú. E-mail: howardsg.unt@gmail.com

² Mechatronics Engineering, Universidad Nacional de Trujillo, Perú. E-mail: crocha@unitru.edu.pe

³ Electronics Engineering, Universidad Nacional del Altiplano, Perú. MSc Mechatronics Engineering, Pontificia Universidad Católica del Perú, Perú. Affiliation: Assistant professor, Departamento de Ingeniería Mecatrónica, Universidad Nacional de Trujillo, Perú. E-mail: emanzano@unitru.edu.pe



Attribution 4.0 International (CC BY 4.0) Share - Adapt

controlled by only four actuators (Zhang *et al.*, 2014). Therefore, it would be convenient to implement nonlinear control algorithms to ensure flight stability and robustness against uncertainties such as external airflows or model errors (Santoso *et al.*, 2018).

An appropriate quadcopter controller has to adapt to changes in dynamics and handle uncertainty; it is necessary to identify an accurate dynamic model, as its performance depends on it. To obtain this model, two methods can be used. The first involves modeling the system from a physical analysis using Newton's laws and Euler angles. The second method is called *system identification*, and it involves analyzing input and output data from experimental flights (Pairan and Shamsudin, 2017).

The identification method based on experimental data can be parametric or non-parametric; in the former, a reduced number of parameters are fitted to a predefined mathematical model, whereas, in the latter, more flexible techniques such as neural networks are used to fit the data without imposing a prior mathematical structure (Atteia *et al.*, 2021). Neural network-based identification methods allow obtaining an appropriate dynamic model for quadcopters since they can adapt to the complexity of the system, adjust to any variation, and obtain more precise results in real-time control (Rosales *et al.*, 2019).

System identification can be achieved through linear approaches such as Autoregressive exogenous models (ARX), Autoregressive Moving Average models with exogenous variables (ARMAX), and Output Error models (OE). Belge *et al.* (2020) found the OE method to be the linear model with the highest estimation accuracy for a fixed-wing UAV under different noise variations. In a comparative analysis between a linear and a nonlinear Model Predictive Controller (MPC) for trajectory tracking in a Micro Air Vehicle (MAV), the latter showed better performance regarding disturbance rejection capabilities, step response, tracking performance, and computational effort (Kamel *et al.*, 2017).

Recent studies have devised methods to effectively quantify uncertainties in identifying nonlinear systems, with a focus on cascade or block-oriented approaches such as Volterra series and Wiener and Hammerstein models (Xavier *et al.*, 2021). Similarly, there are black-box models, which have been primarily used in the analysis of electrical and control systems. These models do not incorporate prior knowledge, but instead rely on a structure flexible enough to capture all relevant physics in the measured data. Typical approaches include the Nonlinear Autoregressive exogenous model (NARX), the Nonlinear Autoregressive Moving Average model with exogenous inputs (NARMAX), artificial neural networks, fuzzy networks, statistical learning theory, and kernel methods (Noel and Kerschen, 2017).

In Altan *et al.* (2018), the trajectory tracking of a hexacopter was studied with a NARX neural network controller aimed at the transportation of physical cargo packages. Moreover, this approach was compared against a PID controller using performance indicators such as the Mean Squared Error (MSE) and the Root Mean Squared Error (RMSE), with the NARX controller yielding much more robust and stable results.

In addition, the implementation of a tensor network B-splines algorithm with the NARX model for nonlinear system identification demonstrated high efficiency and accuracy through numerical experiments in Single Input and Single Output nonlinear systems (SISO), surpassing the RMSE of other nonlinear methods during both prediction and simulation processes (Karagoz and Batselier, 2020).

In a detailed performance comparison between a PID and an advanced controller utilizing NARX neural networks, various indicators such as the MSE and RMSE were utilized (Karakaya and Goren, 2022). The findings revealed a similar performance for both controllers in yaw tests. However, the advanced controller yielded superior results in altitude, roll, and pitch tests, exhibiting an enhanced performance.

In this vein, Ozbek *et al.* (2015) identified the issue of modeling uncertainties in UAVs during a survey of quadrotor aerial robot control, and, in their quadcopter modeling review, Pairan *et al.* (2020) concluded that neural network-based identification is a more suitable option due to its high accuracy, cost, and resource availability. They proposed future research using Radial Basis Functions (RBF) for activation.

The most recent proposals in this field include the works by Ullah *et al.* (2020, 2021, 2022), who presented two sliding mode control designs for UAVs. These proposals were evaluated using mathematical models and simulations, with promising results. However, the progress of these works could be accelerated through more accurate models with the help of system identification.

Consequently, the design of a NARX neural network identification algorithm could be considered in order to obtain a robust model of the nonlinear systems featured by quadcopters. This would ensure proper functioning while avoiding total equipment losses associated with flight control failures.

Nonetheless, the way to ensure system identification is to consider data representing the system's non-linear behavior. Thus, input/output quadcopter data that genuinely exhibit a nonlinear behavior must be verified for use in the training and validation of quadcopter system identification models.

In light of the above, the objective of this research was to identify the non-parametric model that captures the dynamic behavior of a F450 quadcopter, with the purpose of representing the inherent nonlinearities of the system using NARX neural networks. The model was trained with experimental data acquired from the quadcopter's sensors, and a nonlinear data validation was conducted to ensure suitable data for system identification.

System identification methods

ARX

The ARX model, a linear input/output model, is commonly used due to its computationally efficient parameter estimation (Galrinho *et al.*, 2017). This is accomplished directly through the use of a least squares algorithm, which handles linear relationships between parameters and prediction errors (Tu *et al.*, 2020). However, if the actual noise in the process differs from the assumed conditions,

the parameters may deviate from their optimal values, and consistency issues may occur (Nelles, 2020).

As described by Equation (1), the ARX model establishes a relationship between the future value of the dependent output signal $y(k)$, the preceding values of the output signal, and an independent or exogenous input signal $u(k)$. In this context, na and nb represent the model orders, whereas $e(k)$ corresponds to the error of the system, indicating the discrepancy between the predicted and actual output values (Billings, 2013).

$$y(k) = [b_1 u(k-1) + \dots + b_{nb} u(k-nb)] - [a_1 y(k-1) + \dots + a_{na} y(k-na)] + e(k) \quad (1)$$

NARX

The NARX model represents a nonlinear generalization of the ARX model that substitutes the linear relationship described in Equation (1) with an unknown nonlinear function $F(\cdot)$, as outlined in Equation (2), where n_u and n_y represent the input and output memory orders, respectively (Nelles, 2020).

$$y(k) = F[y(k-1), y(k-2), \dots, y(k-n_y), u(k-d), u(k-d-1), \dots, u(k-d-n_u)] + e(k) \quad (2)$$

N-NARX

The neural network NARX representation is referred to as N -NARX. It is a dynamic neural network with a closed-loop architecture that consists of an input layer, a hidden layer, and an output layer. In addition, it incorporates lagged inputs and outputs through explicit integration or a recurrent procedure, as depicted in Figure 1.

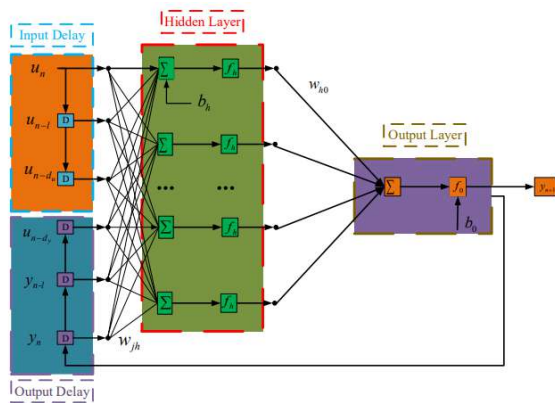


Figure 1. A recurrent single-hidden-layer NARX network
Source: Wei et al. (2020)

Materials and methods

This work was conducted while following the methodology depicted in Figure 2, which involved a systematic review of the available scientific literature to obtain relevant and up-to-date information.

After conducting the literature review, the focus shifted towards implementing the workspace, specifically a test bench, and building a prototype F450 quadcopter. With this

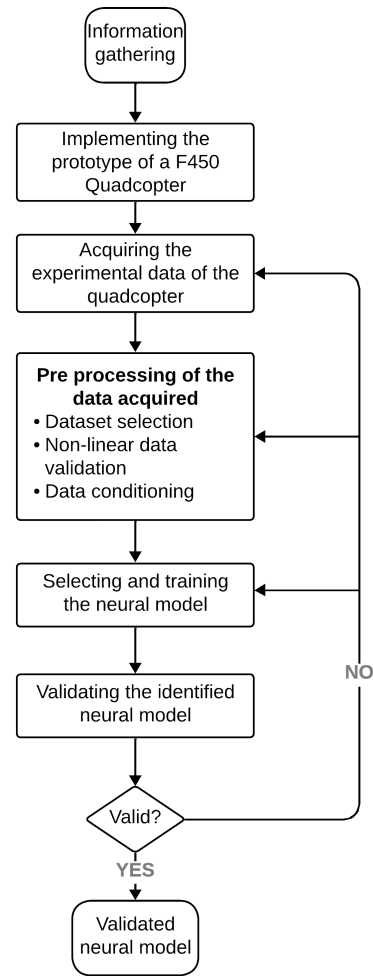


Figure 2. Methodology flowchart
Source: Authors

test bench, experimental data on the roll and pitch angle of the quadcopter were acquired. The input and output data obtained from the quadcopter were pre-processed to reduce noise. Afterwards, a neural network model was selected, its parameters were defined, and it was trained using the conditioned data. Finally, the trained model was validated using performance indicators such as the MSE and the RMSE, as well as through cross-correlation with data that were not used during training.

Implementation

A quadcopter was developed using a commercial F450 frame. The prototype was equipped with four brushless motors, four Electronic Speed Controllers (ESCs), four propellers, and a Lithium Polymer (LiPo) battery. These components were connected to an Arduino UNO hardware development board, along with an MPU9250 IMU sensor, in order to enable precise control of the system's orientation (Figure 3).

The data acquisition process involved the use of specialized equipment, i.e., a test bench specifically designed to collect data on the quadcopter's dynamic behavior. Figure 4 shows the CAD assembly of the test bench with a single degree of freedom (or one axis), which was used to analyze two of the three attitude angles of the F450 quadcopter: roll and pitch.

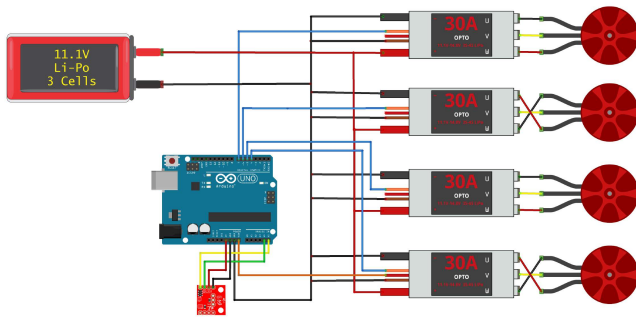


Figure 3. Electronic diagram of the drone
Source: Authors



Figure 4. CAD assembly of a one-degree-of-freedom test bench
Source: Authors

Data acquisition

The data acquisition architecture relied on the test bench, which combined the use of an Arduino board and a computer running the MatLab software. Furthermore, to ensure ease of resetting in case of failure, the Arduino board was located outside the quadcopter.

Figure 5 presents the equipment used for data acquisition alongside a block diagram illustrating the control system of the quadcopter's initial programming, which includes a PID control algorithm with deliberately unsynchronized gains, causing the quadcopter to exhibit an oscillatory motion. The goal of this was to gather crucial insights into the quadcopter's dynamic performance. This motion was critical in capturing the necessary data for subsequent identification processes. Additionally, an algorithm called *Motor Mixer* was implemented to distribute the output of the PID controller into four Pulse-Width Modulation (PWM) signals that served as inputs for the quadcopter's four Electric Propulsion Units (EPUs).

Subsequently, four experiments were conducted for each angle, with each lasting 40 seconds and featuring a sampling time of 10 ms, in order to gather a comprehensive range of data on the quadcopter's dynamic behavior.

The input and output data obtained from Experiment 1 with the roll angle are presented in Figures 6 and 7, respectively.

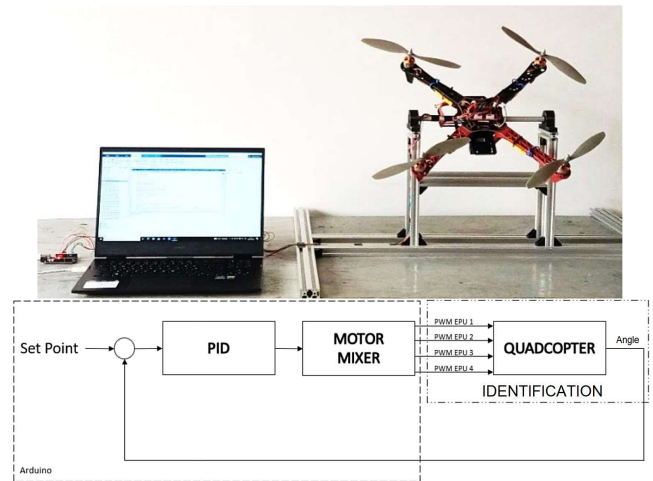


Figure 5. Equipment for data acquisition and block diagram of the control system
Source: Authors

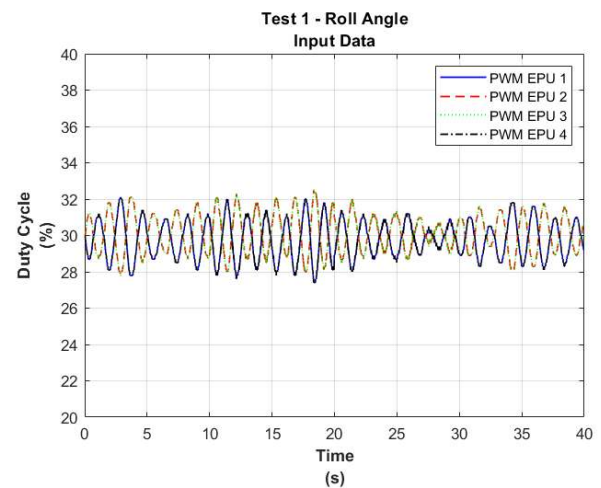


Figure 6. Input data for Experiment 1 – Roll angle
Source: Authors

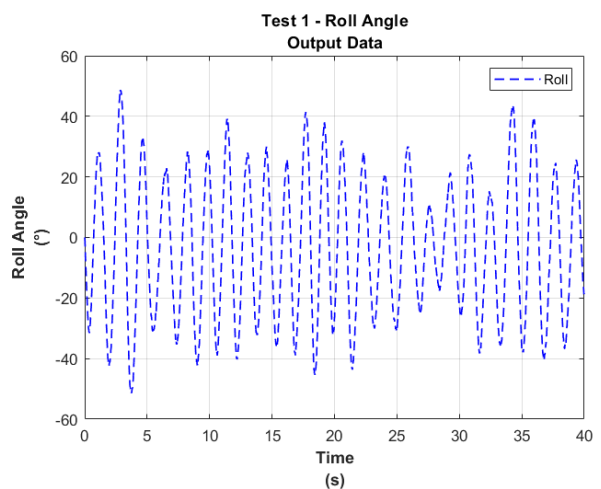


Figure 7. Output data from Experiment 1 – Roll angle
Source: Authors

Data pre-processing

Dataset selection:

The selection of the dataset considered the type of model to be used for the identification process, i.e., a neural network. To train the model, it was important for the data to contain the maximum possible information on the system's dynamics. To ensure this, the dominant frequency from the dynamic system's output data was taken into account alongside its maximum amplitude range.

From the datasets collected during the four experiments conducted for each angle, only two were selected: one for the identification process and the other for cross-validation. Dataset selection for the training process was based on the experiment whose output data had a higher dominant frequency and maximum amplitude range. For the cross-validation process, however, the experiment with lower dominant frequency and a higher maximum amplitude range in its output data would be selected.

The output data for the roll angle data from Experiment 1 are presented in Figure 8, showing a maximum amplitude range of -51.42 to 48.77° .

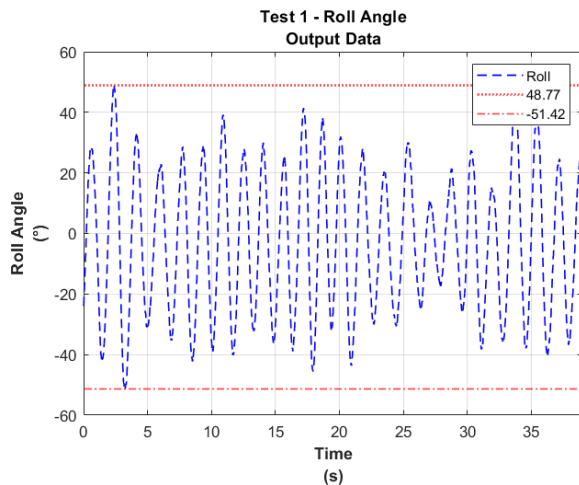


Figure 8. Amplitude of output data from Experiment 1 – Roll angle
Source: Authors

The output pitch angle data from Experiment 1 are shown in Figure 9, indicating a maximum amplitude range of -50.22 to 44.19° .

The Fourier transform was applied to calculate the frequency spectrum of the signal and analyze its dominant frequency. The entire frequency domain of the data was limited to the interval from 0 Hz up to twice the dominant frequency.

The frequency spectrum of the roll angle for Experiment 1 is depicted in Figure 10, indicating a dominant frequency of 0.61523 Hz. The frequency spectrum of the pitch angle from Experiment 1 is presented in Figure 11, revealing a dominant frequency of 0.6665 Hz.

Therefore, based on the criterion of higher dominant frequency, the output data from the first experiments involving the roll and pitch angles were selected as the training dataset for the corresponding neural network model. Similarly, following the criterion of lower dominant frequency and lower amplitude, the data from Experiment 3 were selected for the cross-validation of both angles.

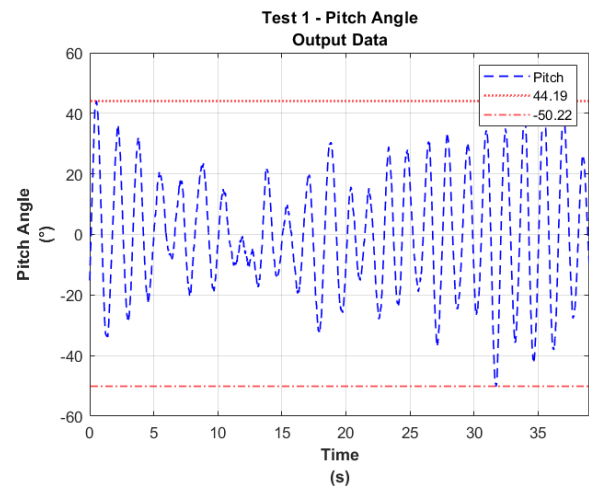


Figure 9. Amplitude of output data from Experiment 1 – Pitch angle
Source: Authors

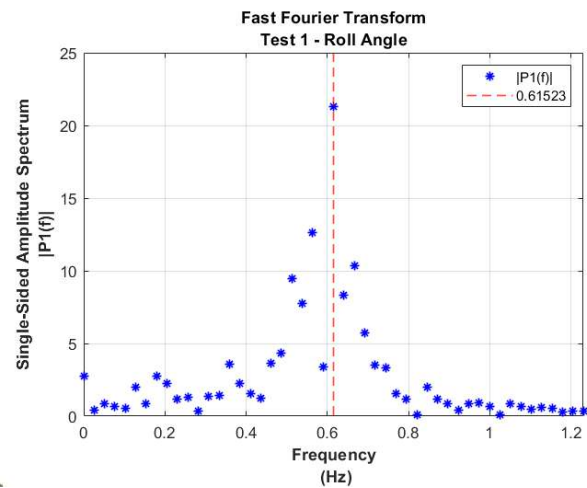


Figure 10. Frequency spectrum of the output data from Experiment 1 – Roll angle
Source: Authors

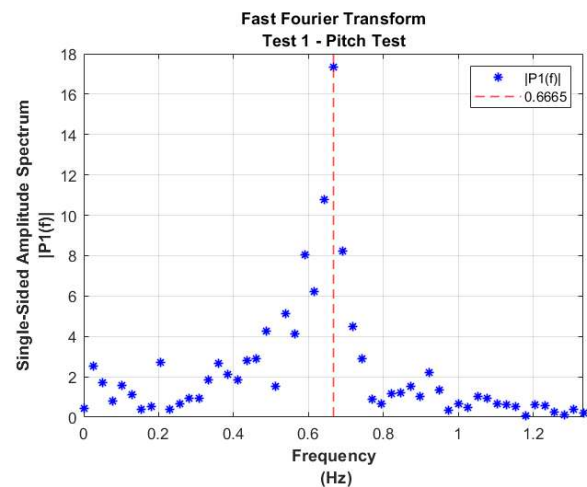


Figure 11. Frequency spectrum of the output data from Experiment 1 – Pitch angle
Source: Authors

Nonlinear data validation:

In order to validate the nonlinear relationship between input and output data (ensuring that the nonlinear behavior of the system would be obtained), two evaluation criteria were considered:

1. The quotient between the RMS value of each wave of the output's oscillatory signal with respect to the RMS value of each wave of the oscillatory signal of the pitch or roll torque.
2. A higher-order correlation function of the system's output signal, as long as the input signal satisfied the nonlinear conditions established by (Billings, 2013). This is described in Equation 3.

$$\begin{cases} \phi_{y'y^2}(\tau) = 0, (\tau = 0, 1, \dots) & \Leftrightarrow \text{linear} \\ \phi_{y'y^2}(\tau) \neq 0, (\tau = 0, 1, \dots) & \Leftrightarrow \text{non-linear} \end{cases} \quad (3)$$

where $\phi_{y'y^2}$ is the cross-correlation between the average level output and its power squared, and τ denotes the torque.

Data conditioning:

Data conditioning was performed for both the input and output of the system. In the case of the input, an internal numerical rounding took place in the Arduino micro-controller during the execution of the Motor Mixer algorithm. On the other hand, the output data exhibited measurement variations as a consequence of the quadcopter's vibrations and the presence of white noise in the signal. These inherent factors contributed to the fluctuations and noise observed in the recorded data.

Figure 12 shows a preliminary demonstration, prior to interpolation, of the one-second sampled stepped input data from the first roll angle experiments.

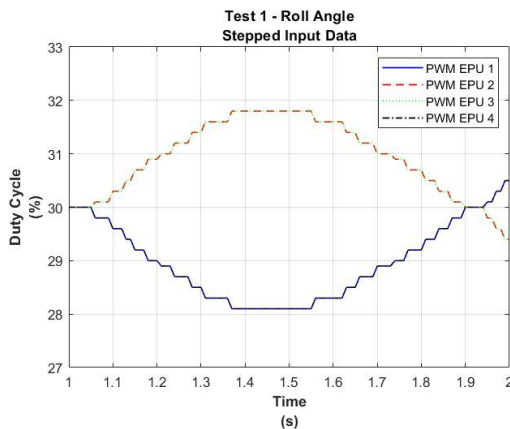


Figure 12. Stepped input data from Experiment 1 – Roll angle
Source: Authors

Afterwards, the acquired stepped input data underwent a smoothing procedure using linear interpolation. This process aimed to obtain a refined dataset that would be more suitable for the subsequent identification.

The interpolated input data from the first set of experiments, with a sample interval of one second, is illustrated in Figure 13 for the roll angle and in Figure 14 for the pitch angle.

It is important to note that this interpolation technique was consistently applied to all the input data gathered throughout the experiments.

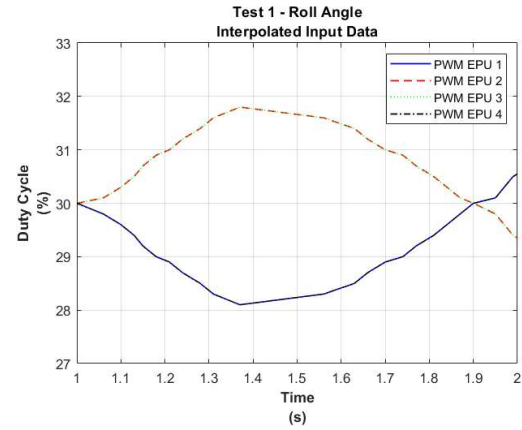


Figure 13. Interpolated input data from Experiment 1 – Roll angle
Source: Authors

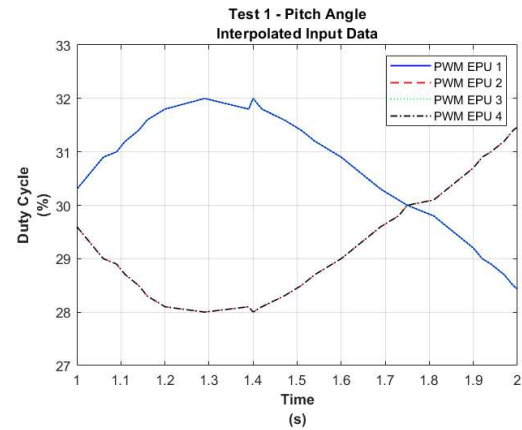


Figure 14. Interpolated input data from Experiment 1 – Pitch Angle.
Source: Authors

The output signals correspond to the roll and pitch tilt angles, which are considered to be the attitude angles of an aircraft with respect to the horizontal plane. Their processing is based on the implementation of the Kalman filter, a robust estimation algorithm that effectively removes inconsistencies and variations in the measurements, thereby ensuring more precise, reliable, and consistent results.

Figure 15 shows a demonstration, prior to filtering, of a one-second sample interval of the output data for the first experiment with the roll angle.

The one-second sample interval of the output data from the first experiments with the roll and pitch angles, as filtered using the Kalman filter, is shown in Figures 16 and 17.

Training the neural network model

Based on Al-Mahasneh et al. (2017), a neural model was chosen for the quadcopter identification process. Comparisons between different neural network approaches were made to determine their usefulness and versatility in this task.

The N-NARX model was selected based on El Dakrory and Tawfik (2016). This model consistently demonstrated its

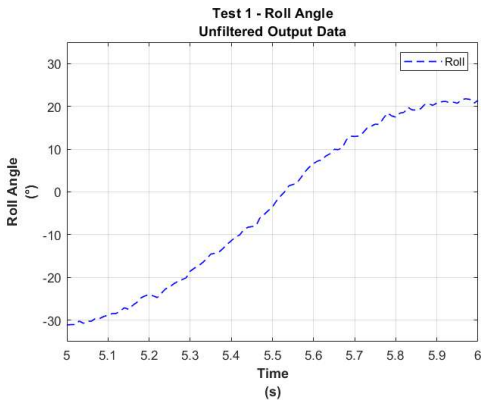


Figure 15. Unfiltered output data from Experiment 1 – Roll Angle
Source: Authors

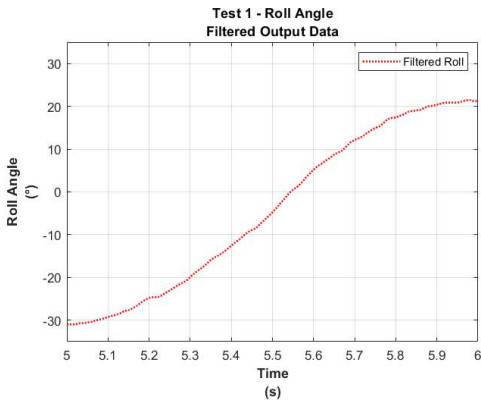


Figure 16. Filtered output data from Experiment 1 – Roll angle
Source: Authors



Figure 17. Filtered output data from Experiment 1 – Pitch angle
Source: Authors

promising performance over other methods, surpassing the capabilities of its linear counterpart, the ARX model.

The system architecture (i.e., the number of layers and neurons) followed the mathematical model obtained by Oktay and Kose (2019), with RBF activation functions stemming from prior work by Pairan and Shamsudin (2017). This was used as a starting point for the training and weight adjustment process. The backpropagation training method and the Levenberg-Marquardt algorithm were also employed to minimize the MSE.

Table 1 presents an overview of the features considered while designing the architecture for the selected neural

network. This includes the network type, the number of layers, and the neuron activation function used.

Table 1. Architecture of the selected neural network

Neural architecture features	Description
Model	Neural network
Neural network type	NARX structure
Network layers	One input layer, one hidden layer with five neurons, one output layer
Neuron activation function	RBF

Source: Authors

In the N-NARX model, one lag was considered for the input and three for the output. This decision was based on the literature review, as illustrated in Figure 18 (Pairan and Shamsudin, 2017). As previously mentioned, an RBF was considered.

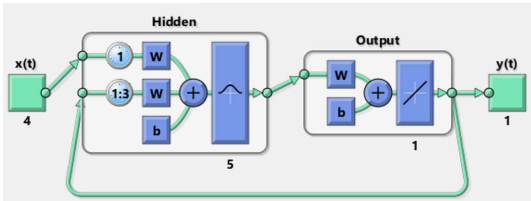


Figure 18. N-NARX model architecture
Source: Pairan and Shamsudin (2017)

After identifying the parameters of the neural network model, the processed roll and pitch angle datasets were trained. Firstly, the roll angle information from Experiment 1 was used, consisting of 3901 input and output data. The N-NARX model's training stopped at 19.48 seconds, achieving a minimized MSE of 4.59.

Meanwhile, for the pitch angle dataset from Experiment 1, also consisting of 3901 input and output data, the training stopped at 2.16 seconds, with a minimized MSE of 1.44.

Results and discussion

Results

Based on the previously mentioned indicators, this study compared the temporal response of the non-parametric quadcopter system's actual attitude angles (from Experiment 1) and those predicted by the N-NARX model.

Two validation steps were implemented:

- Graphical and quantitative validation, showing how closely the neural model's results approximated the real system data, with quantitative metrics such as the MSE, the RMSE, and the fit percentage.
- Cross-validation, using experimental data not employed in training the N-NARX model. This procedure is detailed below.

For the roll angle (Figure 19), the following values were obtained: an MSE of 4.59, an RMSE of 2.14, an RMSE percentage of 2.14%, and a fit percentage of 90.87%.

As for the pitch angle, the comparison is shown in Figure 20, with the following results: a MSE of 1.44, a RMSE of 1.20, a RMSE percentage of 1.28%, and a fit percentage of 93.80%.

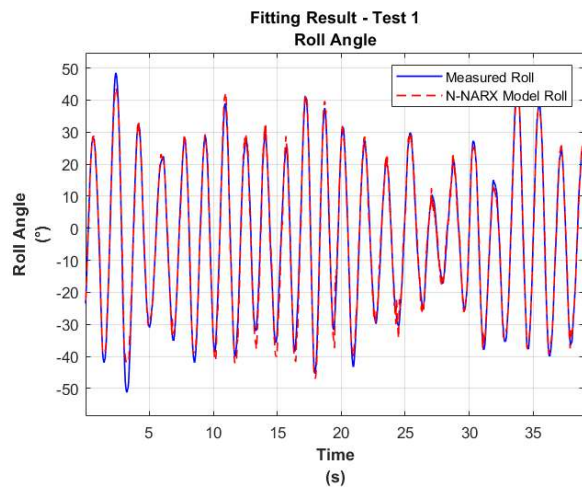


Figure 19. Approximation of the N-NARX model using data from Experiment 1 – Roll angle
Source: Authors

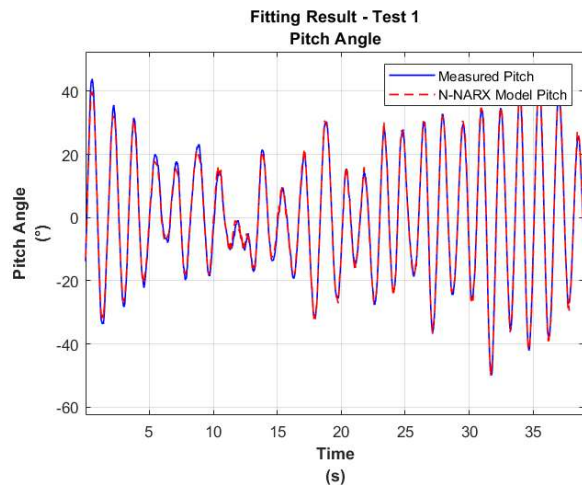


Figure 20. Approximation of the N-NARX model using data from Experiment 1 – Pitch angle
Source: Authors

Afterwards, the selected neural model underwent validation on a separate dataset, specifically that of Experiment 3 (*i.e.*, cross-validation).

For the roll angle (Figure 21), the following values were obtained: a MSE of 5, a RMSE of 2.23, a RMSE percentage of 2.25%, and a fit percentage of 89.44%.

For the pitch angle, as depicted in Figure 22, the performance evaluation yielded the following results: a MSE of 2.70, a RMSE of 1.64, a RMSE percentage of 2.20%, and a fit percentage of 89.29%.

Table 2 presents a quantitative comparison between the performance indicators obtained from Experiment 1, which were used in the training dataset, and those from Experiment 3, *i.e.*, the dataset for cross-validation. This comparison enables an assessment of the model's performance across different datasets.

It is worth mentioning that the results obtained from Experiment 1 evinced a higher level of accuracy and a lower error in predicting both angle values when compared to

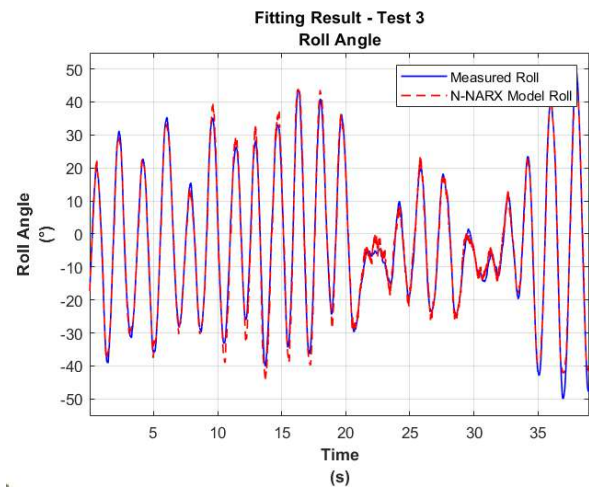


Figure 21. Approximation of the N-NARX model using data from Experiment 3 – Roll angle
Source: Authors

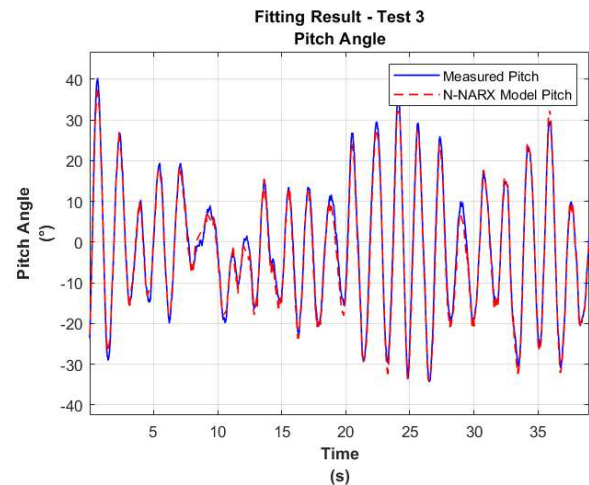


Figure 22. Approximation of the N-NARX model using data from Experiment 3 – Pitch angle
Source: Authors

Table 2. Performance indicators of experiments 1 and 3 from roll and pitch angles

	MSE	RMSE	RMSE(%)	Fit(%)
Roll 1	4.59	2.14	2.14	90.87
Roll 3	5.00	2.23	2.25	89.44
Pitch 1	1.44	1.20	1.28	93.80
Pitch 3	2.70	1.64	2.20	89.29

Source: Authors

the actual data, as the model was trained on the former. However, it is important to note that the N-NARX model showed a remarkable prediction performance even when evaluated on the unseen data from Experiment 3.

A statistical analysis was conducted to assess the behavior of the trained model's predictions and the actual values obtained from Experiment 3 for the roll angle (Figure 23) and the pitch angle (Figure 24).

The results confirm that the simple correlation consistently falls within the established confidence bands, at a 95% level for different time lags. This indicates a strong and

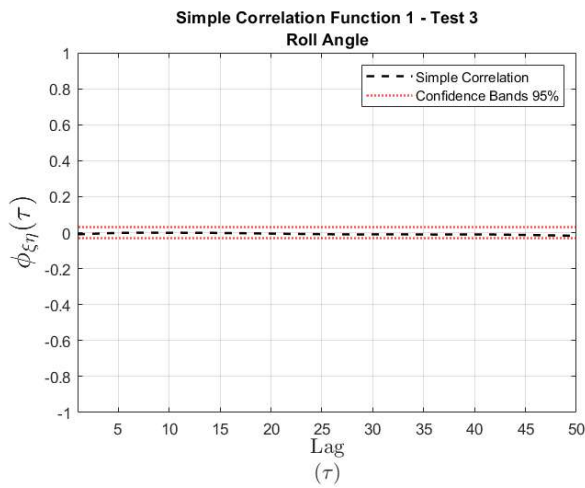


Figure 23. Correlation for Experiment 3 – Roll angle
Source: Authors

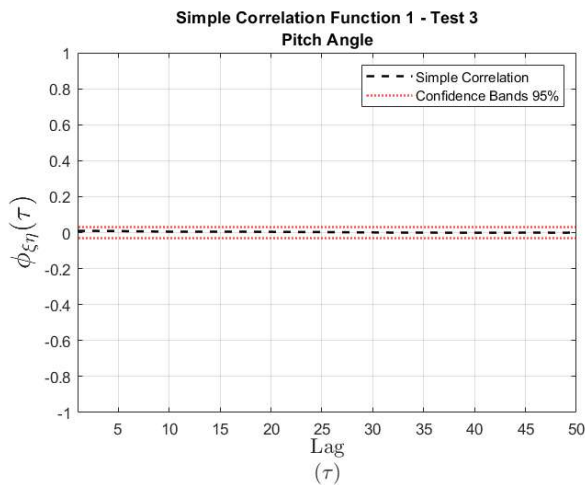


Figure 24. Correlation for Experiment 3 – Pitch Angle
Source: Authors

consistent association between the predicted and actual values, supporting the model's ability to capture the patterns and behaviors of the analyzed nonlinear system.

Discussion

The proposed artificial neural network method showed a high capability to predict attitude angles in a complex nonlinear system model such as a quadcopter. This observation aligns with the results of Muresan *et al.* (2013) and Zhang *et al.* (2014), who reported the model's high performance and accuracy in predicting both roll and pitch angles with a trained nonlinear model, outperforming the capacities of linear counterparts.

In light of the above, a comparison was conducted (Table 3) to assess the RMSE percentages of our method and the 4-neuron Minimal Resource Allocating Network (MRAN) linear model proposed by Pairan and Shamsudin (2017), whose work provided the essential basis for parameter selection during the training of the designed neural network. Note that the trained N-NARX model exhibits a lower RMSE percentage than the MRAN model in both roll and pitch angles.

Table 3. Comparison of the neural models' RMSE percentage

RMSE percentage	N-NARXI (our method)	MRAN (Pairan and Shamsudin 2017)
Roll	2.25 %	10.99 %
Pitch	2.20 %	13.53 %

Source: Authors

Similarly, Table 4 presents a comparison of the fit percentage between our method and the 4-input, 1-output Multiple Input and Multiple Output (MIMO) ARX linear model presented by Salameh *et al.* (2015). Note that the trained N-NARX model significantly outperforms the MIMO ARX model regarding the fit percentage for both roll and pitch angles.

Table 4. Comparison of the neural models' fit percentage

Fit percentage	N-NARX model (our method)	MIMO ARX (Salameh <i>et al.</i> , 2015)
Roll	89.44 %	26.28 %
Pitch	89.29 %	26.28 %

Source: Authors

The results shown in Tables 3 and 4, which compare the designed model with those proposed by other authors, clearly demonstrate the system's nonlinear behavior. Moreover, these findings highlight our model's robustness and its exceptional accuracy in delivering precise predictions, further validating its effectiveness and reliability in real-world applications.

Conclusions

Quadcopter drones are extremely complex and highly nonlinear systems, underscoring the importance of employing advanced algorithms for their identification. Thereupon, in this study, the roll and pitch angles of an F450 prototype were captured using a neural network model. To facilitate experimentation, a test bench was employed, enabling data collection from four tests conducted for each angle. Simultaneously, a real-time signal monitoring system was set up by connecting the Arduino to a computer, enabling the tracking of inputs consisting of four PWM signals sent to the quadcopter's EPU's, while the MPU9250 IMU sensor determined the device's orientation and produced the corresponding output signals.

These experiments revealed inherent variations, and data quality was assessed through amplitude and frequency spectrum analyses, leading to the selection of Experiments 1 and 3 as databases for each angle. For enhanced modeling, the input data were pre-processed using linear interpolation, while the output data underwent Kalman filtering, resulting in smoother signals and improved dynamic capture. Next, the NARX neural network, with a hidden layer comprising five neurons and utilizing RBF activation, was employed. The model was then trained using the dataset derived from Experiment 1, using the Levenberg-Marquardt algorithm. This yielded optimal outcomes.

The neural network model effectively captured the quadcopter's nonlinear dynamics during both the training and validation processes, resulting in an 89.44% fit with an

RMSE of 2.25% for the roll angle, and an 89.29% fit with an RMSE of 2.20% for the pitch angle. Furthermore, the designed neural network model outperformed the proposals of other authors, clearly showcasing its ability to accurately represent the nonlinear behavior of the system.

Future work and scenarios

Considering the work by Ucgun *et al.* (2022), future research could also build on both the NARX model and the Vertical Take-Off and Landing (VTOL) UAV testbed by evaluating free-flight drones. This would allow validating neural network-based control algorithms and their integration with the robust sliding mode controller developed for the testbed. The goal would be to enhance drone stability and performance, comparing the combined approach against traditional control methods and demonstrating improved robustness and tracking in real-world scenarios. This synergy between the NARX model and the testbed could lead to more reliable UAV control systems.

Building on previous research using a NARX neural network model to identify nonlinear quadcopter dynamics, future work could be extended to free-flying drones, capturing a fully dynamic behavior that includes yaw motion. By moving from controlled environments to free flight, a neural network-based system could identify and stabilize the drone in real time, adapting to conditions like wind disturbances or payload changes. Combining the NARX model with traditional methods like PID or model predictive control could help to create a hybrid system, enhancing performance in complex scenarios. Real-time experiments would validate the approach for more robust flight control in unpredictable environments.

In addition, according to Mechali *et al.* (2021), the NARX neural model for UAVs can be extended by integrating it with control strategies such as Continuous Non-singular Terminal Sliding Mode Control (CNTSMC) and Disturbance Observer-Based Control (DOBC). In future research, the use of NARX models to accurately predict and identify the nonlinear dynamics of a free-flying UAV could be explored. Then, the CNTSMC scheme could be applied for robust attitude and position control. By combining the predictive capabilities of the NARX model with CNTSMC's robustness, disturbance rejection (DOBC) and Fixed-time eXtended State Observer (FXESO), a highly accurate and resilient control system for UAVs could be developed which performs well even under uncertainties, nonlinearities, and external disturbances. This approach would allow enhancing trajectory tracking and system stability in real-time scenarios, validating the combined method through both simulations and real-world tests.

Conflicts of interest

All authors declare no conflicts of interest.

CRedit author statement

Howard E. Sifuentes: conceptualization, data curation, formal analysis, funding acquisition, investigation, methodology, software, resources, validation, visualization, writing - original draft.

Carlos A. Rocha: investigation, methodology, writing - original draft, writing (review and editing).

Edgar A. Manzano: project administration, supervision, writing (review and editing).






References

- Ahmad, F., Kumar, P., Bhandari, A., and Patil, P. (2020). Simulation of the quadcopter dynamics with LQR based control. *Materials Today: Proceedings*, 24, 326–332. <https://doi.org/10.1016/j.matpr.2020.04.282>
- Al-Mahasneh, A. J., Anavatu, S. G., and Garratt, M. (2017). *Nonlinear multi-input multi-output system identification using neuro-evolutionary methods for a quadcopter* [Conference paper]. 2017 Ninth International Conference on Advanced Computational Intelligence. <https://doi.org/10.1109/icaci.2017.7974512>
- Altan, A., Aslan, O., and Hacıoglu, R. (2018). *Real-time control based on NARX neural network of hexarotor UAV with load transporting system for path tracking* [Conference paper]. International Conference on Control, Engineering & Information Technology. <https://doi.org/10.1109/ceit.2018.8751829>
- Atteia, G., Mengash, H. A., and Samee, N. A. (2021). Evaluation of using Parametric and Non-parametric Machine Learning Algorithms for Covid-19 Forecasting. *International Journal of Advanced Computer Science and Applications*, 12(10), 0121071. <https://doi.org/10.14569/ijacsa.2021.0121071>
- Belge, E., Kaba, H., Parlak, A., Altan, A., and Hacıoğlu, R. (2020). Estimation of small unmanned aerial vehicle lateral dynamic model with system identification approaches. *Balkan Journal of Electrical & Computer Engineering*, 8(2), 121–126. <https://doi.org/10.17694/bajece.654499>
- Billings, S. (2013). *Nonlinear system identification: NARMAX methods in the time, frequency, and spatio-temporal domains*. John Wiley & Sons.
- Cedro, L., and Wiczorkowski, K. (2019). Optimizing PID controller gains to model the performance of a quadcopter. *Transportation Research Procedia*, 40, 156–169. <https://doi.org/10.1016/j.trpro.2019.07.026>
- El Dakrory, A. M., and Tawfik, M. (2016). *Identifying the attitude of dynamic systems using neural network* [Conference paper]. 2016 International Workshop on Recent Advances in Robotics and Sensor Technology for Humanitarian Demining and Counter-IEDs (RST). <https://doi.org/10.1109/RST.2016.7869856>
- Galrinho, M., Everitt, N., and Hjalmarsson, H. (2017). ARX modeling of unstable linear systems. *Automatica*, 75, 167–171. <https://doi.org/10.1016/j.automatica.2016.09.041>
- Kamel, M., Burri, M., and Siegwart, R. (2017). Linear vs nonlinear MPC for trajectory tracking applied to rotary wing micro aerial vehicles. *IFAC-PapersOnLine*, 50(1), 3463–3469. <https://doi.org/10.1016/j.ifacol.2017.08.849>
- Karagoz, R., and Batselier, K. (2020). Nonlinear system identification with regularized Tensor Network B-splines. *Automatica*, 122, 109300. <https://doi.org/10.1016/j.automatica.2020.109300>

- Karakaya, Ş., and Goren, A. (2022). Performance comparison of PID and NARX neural network for attitude control of a quadcopter UAV. *Journal of Materials and Mechatronics: A*, 3(1), 1–19. <https://doi.org/10.55546/jmm.1010919>
- Macrina, G., Pugliese, L., Guerriero, F., and Laporte, G. (2020). Drone-aided routing: A literature review. *Transportation Research Part C-emerging Technologies*, 120, 102762. <https://doi.org/10.1016/j.trc.2020.102762>
- Mechali, O., Xu, L., Huang, Y., Shi, M., and Xie, X. (2021). Observer-based fixed-time continuous nonsingular terminal sliding mode control of quadrotor aircraft under uncertainties and disturbances for robust trajectory tracking: Theory and experiment. *Control Engineering Practice*, 111, 104806. <https://doi.org/10.1016/j.conengprac.2021.104806>
- Muresan, B., Folea, S., Nascu, I., Ionescu, C. M., and De Keyser, R. (2013). Identification and modeling of the three rotational movements of a miniature coaxial helicopter. *Simulation*, 89(12), 1490–1504. <https://doi.org/10.1177/0037549713504788>
- Nelles, O. (2020). *Nonlinear system identification: From classical approaches to neural networks, fuzzy models, and Gaussian processes* (2nd ed.). Springer Nature.
- Noel, J., and Kerschen, G. (2017). Nonlinear system identification in structural dynamics: 10 more years of progress. *Mechanical Systems and Signal Processing*, 83, 2–35. <https://doi.org/10.1016/j.ymssp.2016.07.020>
- Ozbek, N. S., Onkol, M., and Efe, M. O. (2015). Feedback control strategies for quadrotor-type aerial robots: A survey. *Transactions of the Institute of Measurement and Control*, 38(5), 529–554. <https://doi.org/10.1177/0142331215608427>
- Oktay, T., and Kose, O. (2019). Farklı uçuş durumları için quadcopter dinamik modeli ve simülasyonu. *Eur. J. Sci. Technol.*, (15), 132–142. doi: <https://doi.org/10.31590/ejosat.507222>
- Pairan, M., and Shamsudin, S. (2017). *System identification of an unmanned quadcopter system using MRAN neural* [Preprint]. IOP Conference Series. <https://doi.org/10.1088/1757-899x/270/1/012019>
- Pairan, M. F., Shamsudin, S. S., and Zulkafli, M. F. (2020). Neural network-based system identification for quadcopter dynamic modeling: A review *Journal of Advanced Mechanical Engineering Applications*, 2(1), 20–33. <https://doi.org/10.30880/ijie.2020.02.01.003>
- Rosales, C., Soria, C., and Rossomando, F. G. (2019). Identification and adaptive PID Control of a hexacopter UAV based on neural networks. *International Journal of Adaptive Control and Signal Processing*, 33(1), 74–91. <https://doi.org/10.1002/acs.2955>
- Saito, C. (2019). *Modelamiento dinámico de los parametros de control de vuelo de una aeronave del tipo ala volante utilizando redes neuronales artificiales* [Master's thesis, Pontificia Universidad Católica del Perú]. <https://tesis.pucp.edu.pe/repositorio/handle/20.500.12404/13308>
- Salameh, I., Ammar, E. M., and Tutunji, T. A. (2015). *Identification of quadcopter hovering using experimental data* [Conference paper]. 2015 IEEE Jordan Conference on Applied Electrical Engineering and Computing Technologies. <https://doi.org/10.1109/aeect.2015.7360559>
- Santoso, F., Garratt, M., Anavatti S., and Petersen, I. (2018). Robust hybrid nonlinear control systems for the dynamics of a quadcopter drone. *IEEE Transactions on Systems, Man, and Cybernetics*, 50(8), 3059–3071. <https://doi.org/10.1109/tsmc.2018.2836922>
- Tu, Q., Rong, Y., and Chen, J. (2020). Parameter identification of ARX models based on modified momentum gradient descent algorithm. *Complexity*, 2020, 1–11. <https://doi.org/10.1155/2020/9537075>
- Ucgun, H., Okten, I., Yuzgec, U., and Kesler, M. (2022). Test platform and graphical user interface design for vertical take-off and landing drones. *Romanian Journal of Information Science and Technology*, 25(3), 350–367. <https://www.romjist.ro/abstract-727.html>
- Ullah, S., Mehmood, A., Khan, Q., Rehman, S., and Iqbal, J. (2020). Robust integral sliding mode control design for stability enhancement of under-actuated quadcopter. *International Journal of Control, Automation and Systems*, 18(7), 1671–1678. <https://doi.org/10.1007/s12555-019-0302-3>
- Ullah, S., Mehmood, A., Ali, K., Javaid, U., Hafeez, G., and Ahmad, E. (2021). *Dynamic modeling and stabilization of surveillance quadcopter in space based on integral super twisting sliding mode control strategy* [Conference paper]. 2021 International Conference on Artificial Intelligence (ICAI). <https://doi.org/10.1109/icai52203.2021.9445268>
- Ullah, S., Khan, Q., and Mehmood, A. (2022). Neuro-adaptive fixed-time non-singular fast terminal sliding mode control design for a class of under-actuated nonlinear systems. *International Journal of Control*, 96(6), 1529–1542. <https://doi.org/10.1080/00207179.2022.2056514>
- Wei, M., Ye, M., Li, J. B., Wang, Q., and Xu, X. (2020). State of charge estimation of lithium-ion batteries using LSTM and NARX neural networks. *IEEE Access*, 8, 189236–189245. <https://doi.org/10.1109/ACCESS.2020.3031340>
- Xavier, J., Patnaik, S., and Panda, R.C. (2021). Process modeling, identification methods, and control schemes for nonlinear physical systems { A comprehensive review. *ChemBioEng Reviews*, 8(4), 392–412. <https://doi.org/10.1002/cben.202000017>
- Zhang, X., Li, X., Wang, K., and Lu, Y. (2014). A survey of modelling and identification of quadrotor robot. *Abstract and Applied Analysis*, 2014, 320526. <https://doi.org/10.1155/2014/320526>

Studying the Mechanical Behavior and Strengthening of RCSACC after Exposure to Elevated Temperatures

Estudio del comportamiento mecánico y refuerzo del RCSACC tras su exposición a temperaturas elevadas

Jean Jacques Kouadjo Tchekwagep ¹, Yiping Qiu ², Shifeng Huang ³, Shoude Wang ³, and Xin Cheng ³

ABSTRACT

Rapid calcium sulfoaluminate cement concrete (RCSACC) has received increased attention of late because it can be manufactured with less CO₂ emissions than ordinary Portland cement. In previous studies, RCSACC performed poorly when subjected to elevated temperatures, to which fiber-reinforced concrete (FRC) is a potential alternative. This study investigated the impact of incorporating two types of fibers, i.e., copper-plated steel microfilament (CPM) and shear corrugated steel (SC), on the engineering, mechanical, and microstructural features of RCSACC after exposure to elevated temperatures. Pore size distribution, microstructure, and mechanical properties were tested after exposure to temperatures of 100, 200, and 300 °C. The content of each type of fibers represented 1% of the concrete. The results showed that the mechanical properties were affected by the addition of either type of steel fibers. Adding CPM or SC steel fibers could ensure an adequate resistance of RCSACC when exposed to high temperatures, in addition to improving its residual mechanical behavior, spalling resistance, and ductility after heating. Steel fibers contribute to enhancing both mechanical properties and resistance to heating effects. However, adding steel fibers also appears to increase microstructure damage with heat, reduce workability, entrap air and water, and reduce cracking related to drying shrinkage.

Keywords: mechanical properties, rapid calcium sulfoaluminate cement, copper-plated microfilament fibers, shear corrugated fibers, shrinkage

RESUMEN

Últimamente, el hormigón de cemento sulfoaluminato de calcio rápido (RCSACC) ha recibido una mayor atención porque puede fabricarse con menos emisiones de CO₂ que el cemento Portland ordinario. En estudios anteriores, el RCSACC presentó un mal desempeño cuando se sometió a temperaturas elevadas, para lo cual el hormigón reforzado con fibra (FRC) es una potencial alternativa. Este estudio investigó el impacto de la incorporación de dos tipos de fibras, i.e., microfilamento de acero chapado en cobre (CPM) y acero corrugado (SC), en las características de ingeniería, mecánicas y microestructurales del RCSACC tras su exposición a temperaturas elevadas. Se probaron la distribución del tamaño de los poros, la microestructura y las propiedades mecánicas tras la exposición a temperaturas de 100, 200 y 300 °C. El contenido de cada tipo de fibras representaba el 1 % del hormigón. Los resultados mostraron que las propiedades mecánicas se vieron afectadas por la adición de cualquiera de los dos tipos de fibras de acero. La adición de fibras de acero CPM o SC podría garantizar una resistencia adecuada del RCSACC cuando se expone a altas temperaturas, además de mejorar su comportamiento mecánico residual, su resistencia al desconchado y su ductilidad después del calentamiento. Las fibras de acero contribuyen a mejorar tanto las propiedades mecánicas como la resistencia a los efectos del calentamiento. Sin embargo, la adición de fibras de acero también parece aumentar el daño a la microestructura con el calor, reducir la trabajabilidad, atrapar el aire y el agua, y reducir el agrietamiento relacionado con la contracción por secado.

Palabras clave: propiedades mecánicas, cemento de sulfoaluminato cálcico rápido, fibras de microfilamentos recubiertos de cobre, fibras onduladas de cizallamiento, contracción

Received: March 8th, 2023

Accepted: June 13th, 2024

¹ Ph.D. in Materials Science, University of Jinan, China. Affiliation: Lecturer, Shandong Provincial Key Laboratory of Preparation and Measurement of Building Materials, University of Jinan, China. Email: mse_wanggg@ujn.edu.cn

² Master's Student, University of Jinan, China. School of Materials Science and Engineering, University of Jinan, China. Email: 1019488037@qq.com

³ Postdoctoral Fellow, Hong Kong University of Science and Technology, China. Ph.D. in Materials Science, Wuhan University of Technology, China. Affiliation: Professor, Shandong Provincial Key Laboratory of Preparation and Measurement of Building Materials, University of Jinan, China. Email: mse_huangsf@ujn.edu.cn

⁴ Ph.D. in Materials Science and Engineering, Wuhan University of Technology, China. Affiliation: Professor, School of Materials Science and Engineering, University of Jinan, China. Email: mse_wangsd@ujn.edu.cn

⁵ Ph.D. in Materials Science, Wuhan University of Technology, China. Affiliation: Professor, Shandong Provincial Key Laboratory of Preparation and Measurement of Building Materials, University of Jinan, China. Email: chengxin@ujn.edu.cn



Attribution 4.0 International (CC BY 4.0) Share - Adapt

Introduction

Ordinary Portland cement concrete (OPCC) has been adopted in multiple projects around the world, making it the most widely used material on Earth (Chang *et al.*, 2020). The content of OPCC consists of Portland cement (PC), coarse aggregate, water, and sometimes an admixture (Cheng *et al.*, 2020). Experts do not consider OPCC to be an environmentally friendly material, since its production and manufacture releases a large portion of the world's overall CO₂ emissions (Venkatesh and Wasim, 2011). In this vein, the development and implementation of rapid calcium sulfoaluminate cement (RCSAC) constitute an encouraging option to decrease carbon dioxide emissions, which account for 7% of the global levels in the cement industry alone. Depending on its performance, RCSAC has several applications. Since the 1970s, cement research in China has managed to produce different types of calcium sulfoaluminate (CSA) cement, which are categorized as the *third cement series*. The advantage of their large-scale production lies in an abundance of resources and the reduced research funding required from the Chinese government (Li *et al.*, 2020, 2021; Michels *et al.*, 2016). Thus, China has gradually applied third-series cement in multiple engineering projects. Both China and India have successfully managed to use this binder to build waterproof pipes (Afroughsabe *et al.*, 2017; Augusto *et al.*, 2019).

RCSAC is mainly composed of ye'elimite (C₄A₃S̄) (Afroughsabe *et al.*, 2016; Wang *et al.*, 2020). Increasing the ye'elimite content results in a faster early strength. As the largest and most widely used material in China and India, RCSACC is increasingly being used in current construction projects (Kohoutková *et al.*, 2017). When using this material, many issues can arise, such as those involving safety and sustainable development (Cheng *et al.*, 2020). Materials made with a cement base are generally weak in tension, including RCSACC (Barreto *et al.*, 2019; Chalioris *et al.*, 2018). Cracks appear when RCSACC is heated to high temperatures, as shown by Azzabi *et al.* (1993) and Bjegović *et al.* (2015). Moreover, it may crack when exposed to elevated temperatures, resulting in reduced mechanical properties (Tanyildizi and Yonar, 2016). A solution to this issue could be the addition of shear corrugated (SC) or copper-plated microfilament (CPM) fibers, aiming to increase the properties of RCSACC when exposed to heat (Aguilar *et al.*, 2015; Aravinthan *et al.*, 2018). Fibers have been shown to decrease early-age shrinkage by enhancing the cohesion between the concrete matrices (sand, cement, coarse aggregate, water, and fibers) (Li, 2019). There is ongoing research on the use of fiber in concrete materials, with the aim of reducing the damage inflicted by elevated temperatures on infrastructure. The effectiveness of using fibers depends on the type of fiber used, its elastic modulus, and its aspect ratio (Hong *et al.*, 2017).

The purpose of this article is to investigate the effect of SC and CPM on the behavior of RCSACC after exposure to elevated temperatures. In previous studies, we have

assessed the dehydration effects of increased temperature (Cheng, Huang, *et al.*, 2020a, 2020b). No other works have evaluated the impact of CPM and SC fibers on the features of RCSACC. A fiber content of 1% was used, and 24 RCSACC mixes were made with a 0.4 water-to-cement ratio. Mechanical properties were assessed after exposure to elevated temperatures. SEM and 3D microscopy were employed to explore the microstructural properties of the RCSACC samples used.

Test raw materials

Rapid calcium sulphoaluminate cement

For all preparations, we utilized RCSAC produced by China United Cement Jining Co. Ltd. (Shandong, China). Table 1 presents its chemical composition.

Table 1. Composition of the RCSAC used

Oxide compositions	Weight%	Mineralogical phase compositions	Weight%
CaO	45.2	C ₂ S	18
Al ₂ O ₃	17.5	C ₄ A ₃ S̄	60
SO ₃	15.7	C ₁₁ A ₇ F	4.4
SiO ₂	9.19	CaSO ₄	9.9
Fe ₂ O ₃	2.50	2.5	4.5
MgO	1.90	1.9	Fluorescence
TiO ₂	0.7	0.7	
K ₂ O	0.4	0.4	Others
Na ₂ O	0.4	0.4	3.2
Specific gravity (g/cm ³) 3.1		Specific surface (m ² /kg) 450	

Source: Authors

Table 2 presents the aggregate used in this study (crushed limestone and silica sand).

Table 2. Physical properties of the aggregates

Aggregate type	Maximum size aggregate (mm)	Water absorption	Specific gravity	Fineness modulus
Fine aggregate	4.75	1.1	2.65	2.9
Coarse aggregate	19	0.96	2.74	

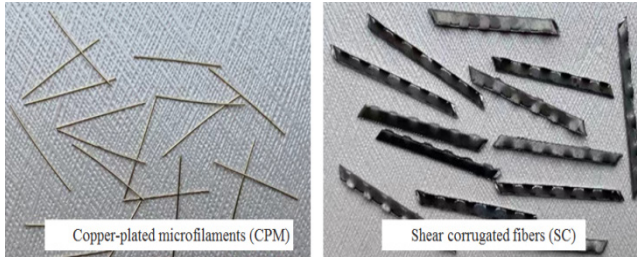
Source: Authors

We employed CPM and SC fibers with aspect ratios of 26.5 and 69.4, respectively. Figure 1 shows these fibers, and Table 3 lists their characteristics.

Table 3. Properties of the steel fibers

Type and shape of fiber	Length L (mm)	Diameter D (mm)	Aspect ratio L/D	Density (g/cm ³)	Tensile strength (N/mm ²)
Copper-plated microfilament (CPM)	13.3	0.5	26.6	7.8	2300
Shear corrugated (SC) fiber	37.5	0.54	69.4	7.8	1050

Source: Authors

**Figure 1.** Two kinds of steel fibers

Source: Authors

RCSACC preparation

Table 4 shows the slump of the RCSACC-CPM and RCSACC-SC preparations. In this work, we established 18 ± 2 cm as the acceptable slump value. After they were produced, the preparations were placed in the molds. Three samples were prepared for each temperature range. After one day, the samples were demolded and cured at 23 ± 2 °C (Gashti et al., 2020) for 672 hours. Then, they were left exposed in the lab for 125 h.

Table 4. Mix proportions of the fiber-reinforced concrete mixes

Mixture ID	W/C	Water	RCSA	FA	CA
		(kg/m ³)			
RCSACC 12 cm (slump)	0.4	216	540	622	727
1% retarder					
RCSACC-CPM					
CPM fiber volume fraction (%)					
0	1	1	0.4	216	540
12.5 cm (slump)	0.4	216	540	622	727
1% retarder					
RCSACC-SC					
SC fiber volume fraction (%)					
0	1	1	0.4	216	540
12 cm (slump)	0.4	216	540	622	727
1% retarder					

CPM: copper-plated microfilament fibers; SC: shear corrugated fibers.

Source: Authors

Then, the samples were placed in an electric furnace for heating at the rate of 4°C/min until they reached the target temperature, which was maintained for 4 h. This time proved to be enough to ensure the effective decomposition of all the important ingredients forming the RCSAC, i.e., Aft and Al(OH)₃ (Aluko et al., 2020; Bai et al., 2019).

Testing properties

Measuring the drying shrinkage

It is important to study the shrinkage behavior of expanded RCSACC for a more pertinent assessment. The Chinese standard GB-T50082 was used for evaluating the drying shrinkage of the concrete. Three specimens of each composition, with a size of 100 × 100 × 515 mm, were used to measure the shrinkage. The test piece was poured and maintained in a molded state, at a temperature of 28 ± 0.5 °C and 75% relative humidity. While this state was maintained, the effective length was measured, and the shrinkage rate was determined. The test result was calculated based on the average of the three specimens.

Surface image analysis

Surface images of the different RCSACC specimens were captured using a high-resolution optical zoom camera. Afterwards, the images underwent a treatment process, with the aid of a software named GelAnalyzer (Ahmed, 2021). Crack images were generated and drawn with the aid of vellum paper, and the Lightroom software helped to clear up the images. Following that, a description of the crack areas was constructed in the ImageJ software.

Ultrasonic testing

Waveform variations can serve as an important indicator of the internal characteristics of each sample, both before and after conditioning (Błaszczyszński et al., 2012; Mehta et al., 2006). In the context of ultrasonic testing, the frequency was set at 50 kHz, and the input voltage at 10 V.

Flexural strength-displacement

In accordance with GB/T50081-2019, the national testing standard for the physical and mechanical properties of concrete, this experiment used specimens with side lengths of 100 × 100 × 400 mm. We calculated the flexural strength (ff) by means of the formula $ff = Fl/bh^2$, where ff is the flexural strength; F is the failure load of the specimen (N); l is the span between supports (mm); b is the cross-section width of the specimen (mm); and h is the cross-section height of the specimen.

Compressive strength

This experiment used standard cube specimens with a side length of 100 × 100 × 100 mm. The percent strength variation (Q) as the temperature increased was defined as follows:

$Q = \left| \frac{\sigma - \sigma_0}{\sigma_0} \right| \times 100\%$. In this formula, σ_0 is the compressive strength of the specimens at ambient temperature (20 °C), and σ is the strength of the sample after thermal exposure.

Mercury porosimetry (MIP) and micromorphology (SEM) analysis

We tested the different samples using MIP and SEM tools, aiming to observe the pore distribution and microcrack patterns. Small squares with a side length of 10 ± 1 mm were taken from the crushed sample for MIP testing. For the SEM observation, we used a diamond saw to create small sections (typically 1-2 cm), focusing on areas of interest (e.g., cracks and fiber-matrix interfaces). Using a vacuum oven, we made sure that the samples were completely dry to prevent water vapor interference during imaging. Finally, we placed the samples in the vacuum chamber of a QUANTAFEG250 field emission scanning electron microscope (SEM) for analysis.

Results

Workability

In accordance with ASTM C143, we determined the workability of the fresh RCSACC-CPM, RCSACC-SC, and RCSACC samples. RCSACC's fresh workability was seen to be negatively affected by the addition of CPM and SC fibers. Figure 2 shows the slump of each blend.

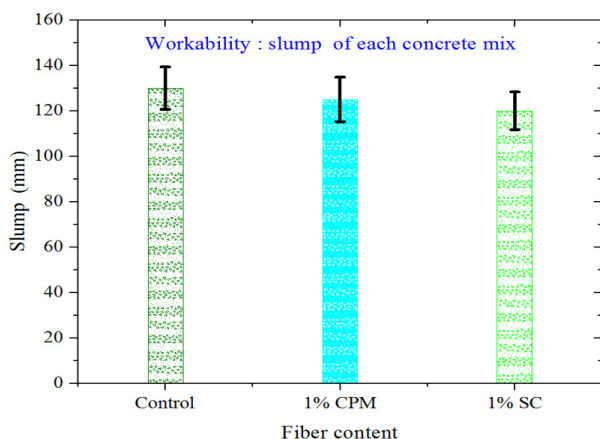


Figure 2. Workability of RCSACC mixes
Source: Authors

Fiber characteristics play an important role in how the fibers blend into the matrix. In this study, the workability observed was found to be relatively consistent across all samples. Moreover, fibers tend to reduce workability because of their large specific area, which adheres some of the available cement paste, thereby reducing the availability of excess paste to reduce inter-particle friction. As for the slump test results, the 1% value obtained more or less meets the workability goal for the fresh RCSACC mixture, as set at the beginning of the study.

Free shrinkage

In the control samples (RCSACC), thin cracks appeared approximately 2 h 30 min after casting, spreading all over the samples. This is a consequence of strong water absorption during setting, which continues throughout curing. In the case of the RCSACC samples with a 1% CPM or SC fiber content, the emergence of the first crack took more than 10 h. Thus, the emergence period for the RCSACC-CMP and RCSACC-SC samples was four times longer than that of the control sample. The free shrinkage strain can be substituted into the following formula to find the stress in the reinforcements:

$$f_{sc} = \frac{\varepsilon_{cs} E_s}{1 + \frac{\alpha A_s}{A_c}} \quad (1)$$

where f_{sc} stands for the stress in the reinforcements, as caused by the shrinkage (N/mm^2); $\alpha = \frac{E_s}{E_{cm}}$, with E_s and E_{cm} being the modulus of elasticity of the fiber and concrete – it is a non-dimensional term that accounts for the non-uniform distribution of stress in the concrete section due to cracking, with values of 0.184 for RCSACC-CMP and 0.123 for RCSACC, and it indicates the shrinkage-related stress in the reinforcements and the relationship between the reinforcement used and the stress in RCSACC; A_s stands for the cross-sectional area of the fiber reinforcement; and A_c stands for the cross-sectional area of the concrete. The tension for both types of reinforced fibers is negligible and could be applicable when the RCSACC is partially restrained by reinforcements. As for the free shrinkage strain, RCSACC must be stretched to cover the strain caused by shrinkage. The restraint values obtained for RCSACC-CMP and RCSACC-SC are greater than the partial restraint regardless of the fiber content. This suggests that the addition of the steel fibers is not helpful to RCSACC exposed to elevated temperatures. Figure 3 illustrates the cracking on the three specimens as the sample shrinks. Here, Figure 3c presents the case of RCSACC-SC, wherein very thin, slight cracks appeared. The assessment focused on the width and area covered by the cracking. This manifestation of cracks is the reason why water was observed on the surface of the different samples.

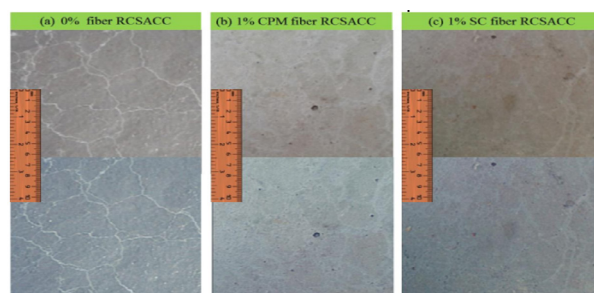


Figure 3. Shrinkage cracking for different RCSACC specimens in a 2.7×2.7 cm area
Source: Authors

The introduction of a 1% fiber content obviously helped to control the crack width in comparison with the control specimen (Figure 4). The crack width was reduced by an average of 30.89% and 51.75% with the introduction of 1% CPM and SC fiber content, respectively. As stated in the ACI 224R.01 standard (ACI Committee, 2011), a width between 0.1 and 0.3 mm is acceptable for the cracks provoked during plastic shrinkage. However, this requirement was not met by the control sample, unlike the 1% fiber-reinforced samples.

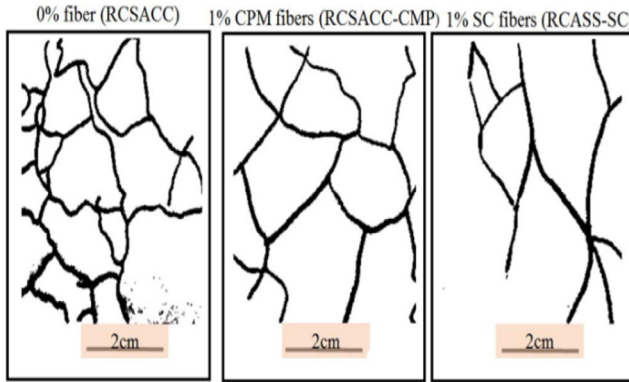


Figure 4. Shrinkage in different types of concrete (drawn in tracing paper from a printed image)
Source: Authors

Table 5 summarizes the characteristics of the cracks found in the samples with CPM and SC fibers.

Table 5. Crack characteristics of the reinforced RSACC

Concrete type	Total cracks (%)	Aver. crack width (mm)	Crack reduction (%)	Fiber content
RCSACC	11.40	1	-	0%
RCSACC-CMP	9.15	0.21	30.89	1%
RCSACC-SC	5.90	0.12	51.75	1%

Source: Authors

Figure 5 shows the shrinkage over 14 days of curing. The shrinkage of the RCSACC-CPM and RCSACC-SC samples differed from that of the RCSACC. The average shrinkage of the fiber-reinforced concrete (FRC) was larger than that of

the control sample. This may be because, when the fiber is added into the mixture, the microstructure is not as dense as that of the RCSACC with natural fine aggregate. Moisture can thus evaporate and migrate out of the sample more easily. The introduction of steel fibers in RCSACC reduces the strain during shrinkage, regardless of the fiber type. SC fibers reduce the shrinkage considerably. The development of RCSACC has two stages: rapid growth during the early hours of curing, when more than 86% of the total shrinkage takes place; and a slow development after 8 h.

Appearance of color and cracks after heating

Figure 6 shows the samples' variations in color, crack spreading, and spalling. In the RCSACC-CPM and RCSACC-SC samples exposed to 100, 200, and 300 °C, no large-scale spalling took place. CPM and SC helped to stop the spalling of the concrete when exposed to heat. In the control sample, the inner pore pressures are not released, leading to large-scale spalling. In contrast, in the samples with fibers, as the temperature rises, canals are formed on the bed of the fibers. These serve as a corridor to let compressed air escape and thereby reduce spalling through decreasing pore pressure.

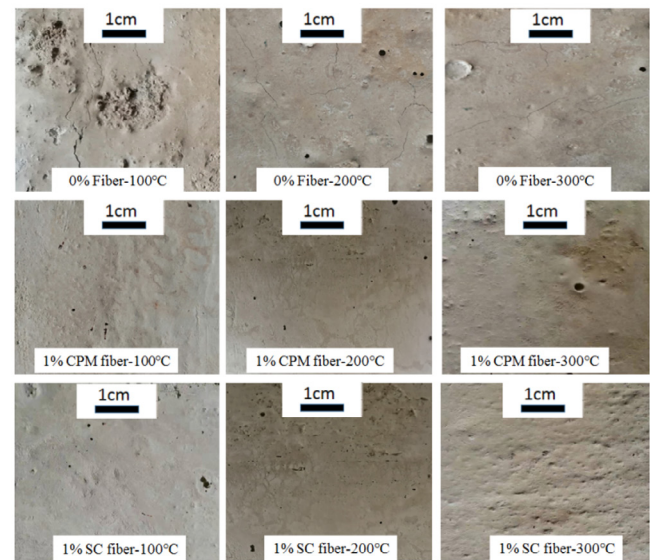


Figure 6. Exterior surface of the different RCSACC samples after exposure to a heat treatment
Source: Authors

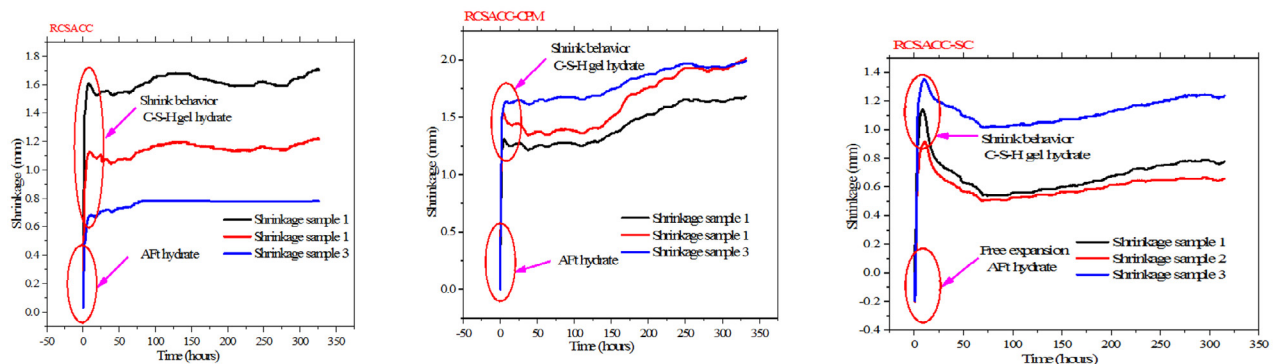


Figure 5. 14 days of free shrinkage in RCSACC with 0%, 1% CPM, and 1% SC fiber content
Source: Authors

Ultrasonic testing

The time required to transmit a pulse and the magnitude values from peak to peak, as obtained from the ultrasonic waveform (UW), are shown in [Figure 7](#).

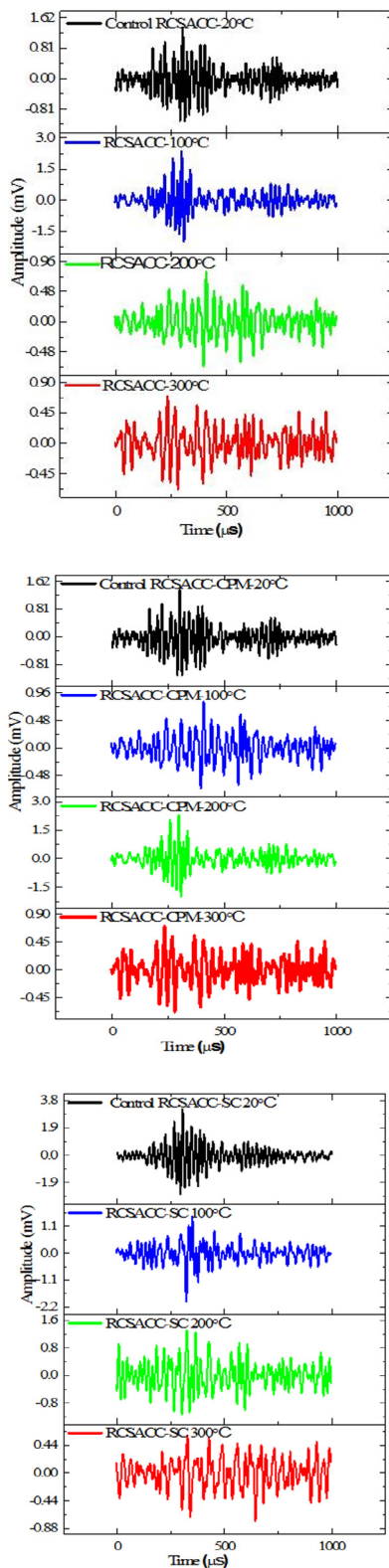


Figure 7. Ultrasonic waveforms

Source: Authors

As a result of exposure to elevated temperatures, the arrival time increased and the magnitude for the RCSACC sample decreased after being exposed to 100 and 200 °C. At 300 °C, the arrival time and magnitude decreased for the RCSACC-CPM and RCSACC-SC samples ([Table 6](#)). This evinces the presence of cracks caused by dehydration as the temperature increases.

Table 6. Ultrasonic wave analysis

	Samples	Aver. arrival time (μs)	Aver. amplitude (mV)	Change in amplitude (mV)
20 °C Before heating	RCSACC	300	3.35	
	CPM	296.7	2.33	
	SC	304	3.19	
100 °C After heating	RCSACC	296.7	2.33	1.02
	CPM	300.4	1.6	0.97
	SC	349.6	1.5	1.69
200 °C After heating	RCSACC	407.9	0.7	2.56
	CPM	408	0.79	1.54
	SC	325.2	1.3	1.89
300 °C After heating	RCSACC	311.9	0.8	2.52
	CPM	234.4	0.69	1.64
	SC	327.9	0.59	2.6

Source: Authors

All the samples, when tested at room temperature, exhibited a short transit time with a large magnitude, suggesting that, before heat exposure, the specimens had greater continuity and density. After the heating treatment, the difference between the exposed and unexposed specimens reflected the internal characteristics of the different RCSACCs. The waveform change could originate from waves spreading and coming into contact with many pores and microcracks. The attenuation was well explained by the presence of microcracks on the surface, which emerged as the temperature increased.

Compressive strength

[Figure 8](#) shows the failure patterns of each specimen under compressive testing. The maximum strength was recorded for the samples at room temperature; they exhibit wide and significant spalling as well as characteristic cracks at higher temperatures. The specimen cracks were significant because they resulted from an applied load. Although lower strength and negligible cracking were observed in the samples heated to 100-300 °C, they performed adequately during the test due to the initial cracking caused by the heat. Moreover, the RCSACC-CPM and RCSACC-SC samples, when heated above 100°C, exhibited an impressively large propagation of cracks and significant spalling, an evidence of the poor compactness between the different component matrices.

Adding CPM and SC fibers to RCSACC enhances some engineering properties after heat exposure. At room

temperature, a 3.6% increase in strength was recorded after adding 1% CPM; at 100 °C, 32.03% of the strength was retained. However, the introduction of 1% SC negatively impacted the sample's compressive strength.

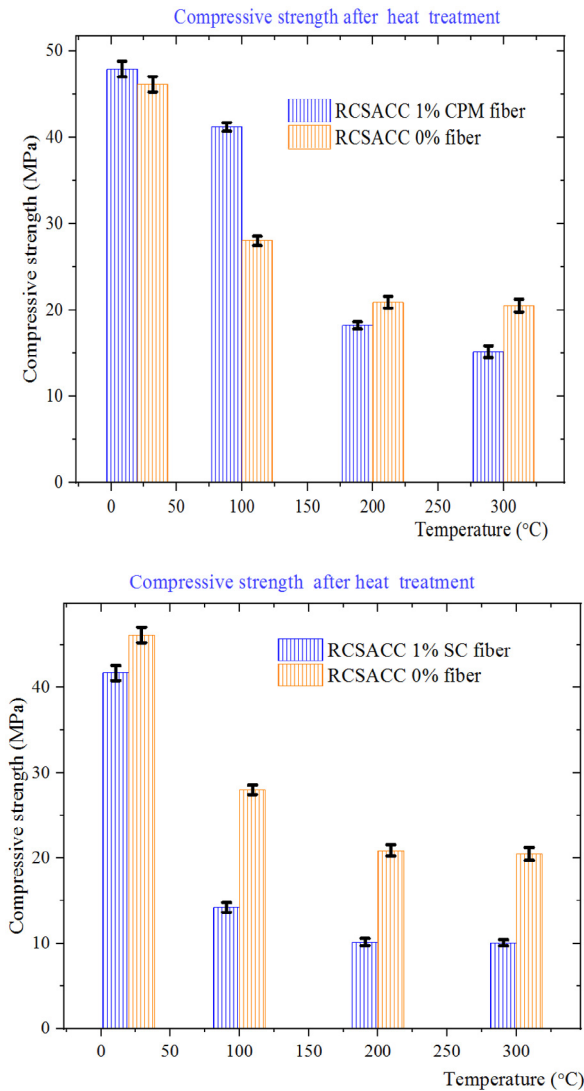


Figure 8. Strength of the fiber-reinforced concrete
Source: Authors

Flexural strength

The cracks observed in the reinforced samples during flexural testing are quite different from those of the control. CPM and SC resist the spread of the cracks uniformly. This suggests that crack propagation will concentrate in the zones with fewer fibers. Figure 9 shows that the FRC containing 1% SC has the highest flexural strength, greater than that of the sample with 1% CPM. The RCSACC-SC specimens showed a better flexural strength capacity because of the enhanced bond formed by the corrugations; as a consequence, the steel fibers could bridge small cracks more quickly. When cracks appeared, the corrugations remained solidly embedded on each side, acting as stress transfer media. When the maximum bond strength was reached, the pull-out

effect fully activated, allowing the adjacent fibers to delay the growth of the cracks. However, there was no significant difference between the strength of the control and that of the FRC with 1% CPM fiber exposed to 300 °C.

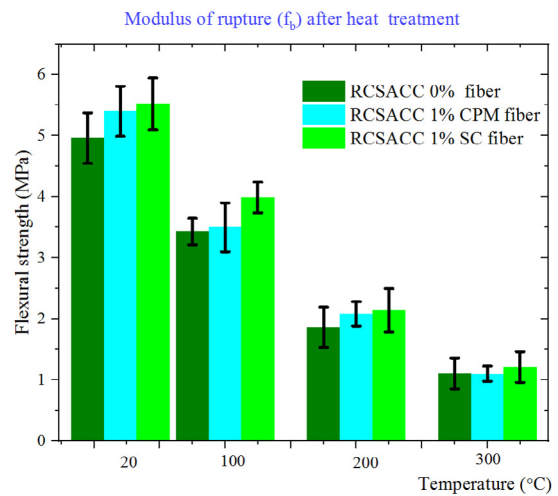


Figure 9. Flexural strength values
Source: Authors

Figure 10 shows the fracture deflection (f_{bb}) of the three types of RCSACC at different temperatures. For RCSACC-CPM, between 20 and 300 °C, there was a gradual difference in f_{bb} . The f_{bb} of the SC fibers varied greatly with temperature. A higher fracture deflection means more and longer cracks, as well as greater damage to the RCSACC sample. CPM fibers have a lower individual volume and can be more evenly distributed in the RCSACC. In contrast, the large volume of SC fibers and their corrugated structure make an even distribution more difficult.

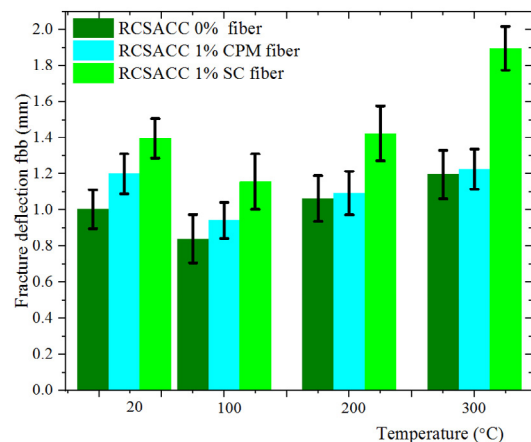


Figure 10. Fracture deflection
Source: Authors

At a normal temperature of 20 °C, the strength of the two FRCs was higher than that of the control sample, and there was little difference in f_{bb} . In a high-temperature environment, the structure of RCSACC-SC was destroyed. It is important to highlight that uneven SC fibers cause a greater f_{bb} . In addition, although their corrugated structure

macroscopically supports the flexural strength of RCSACC, SC fibers' ability to bond with RCSACC is greatly reduced. This indicates, to a certain extent, that the structure of RCSACC-SC is loose, which is another reason for the increase in deflection.

Figure 11 shows the load displacement curves of RCSACC-CPM and RCSACC-SC. The latter (Figure 11b) shows cracks when loaded, and its behavior evidence softening. It has a higher flexural strength than RCSACC-CPM (Figure 11a), exhibiting hardening.

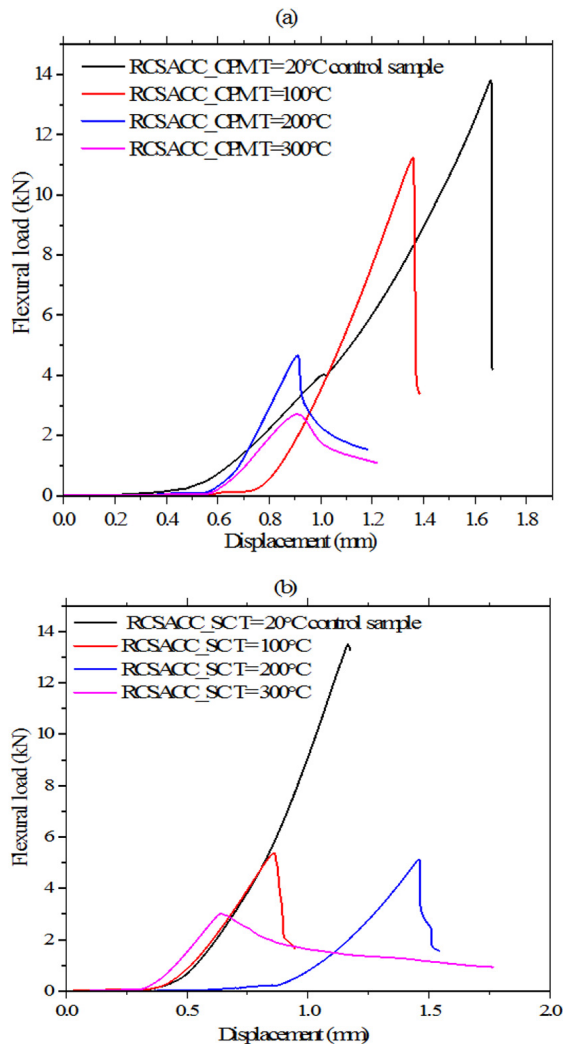


Figure 11. Load displacement curves of the samples with CMP and SC fibers at various temperatures

Source: Authors

MIP and SEM observations

Pore size curves are shown in Figure 12. The pore size of RCSACC-CPM and RCSACC-SC gradually increased with increasing temperature, and high temperatures promoted the deterioration of the microscopic pore structure. The microscopic observation of RCSACC-SC showed a higher mercury intrusion in the 100-1000 nm pore range when compared to RCSACC-CPM, which indicates that the

addition of steel fibers effectively emptied the 200-1000 nm pore size range, resulting in increased porosity.

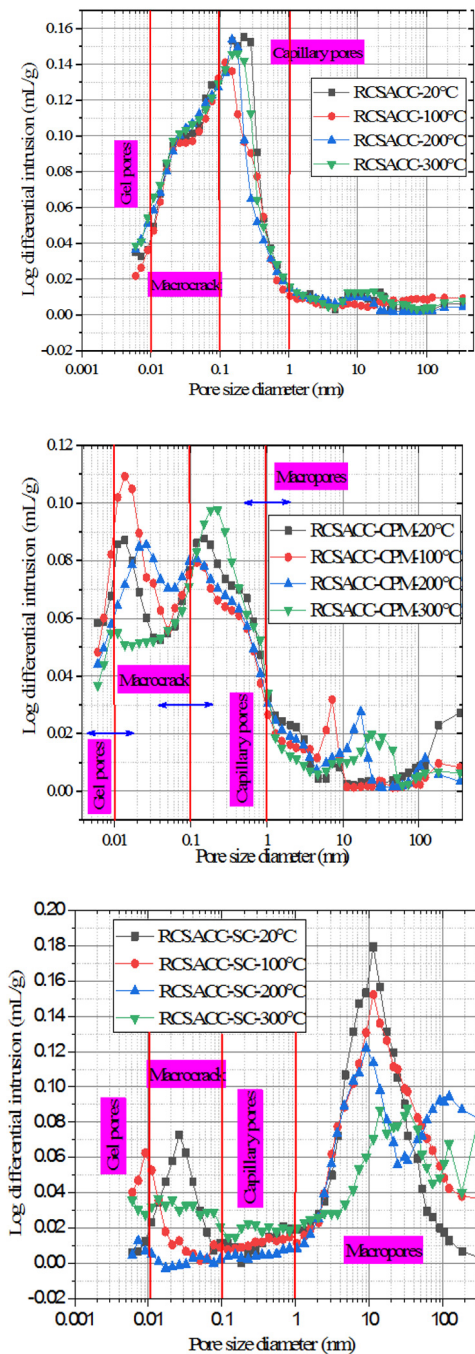


Figure 12. Pore structure results

Source: Authors

The sample reinforced with CPM steel fibers had a more constant increase in pore size distribution (PSD) as the temperature increased, in comparison with the control and the RCSACC-SC sample. Figure 13 shows that RCSACC is more compact at room temperature, which evidences the better homogeneity of the different matrices. Above 100 °C, RCSACC-CPM and RCSACC-SC exhibited a less compact internal structure, with the gradual appearance of macro-

and microcracks and pores as the temperature approached 300 °C. Numerous pores were observed in the different samples at 300 °C.

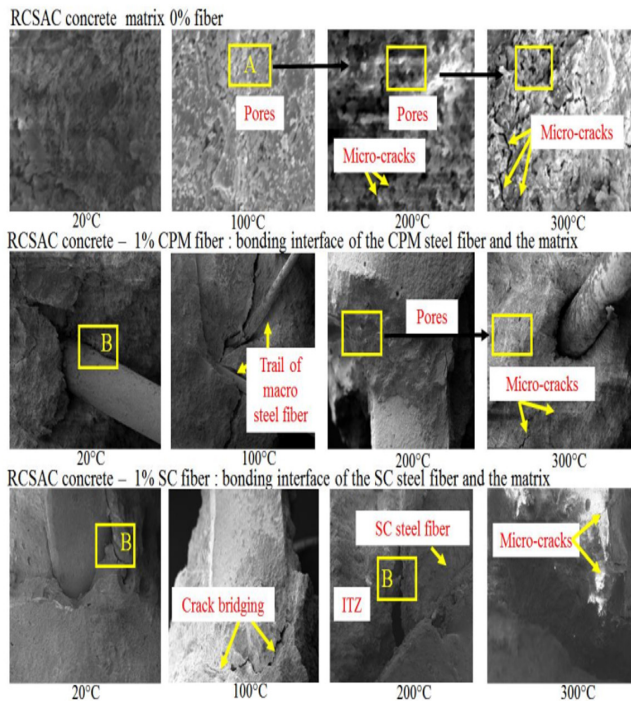


Figure 13. Different RCSACC samples as viewed through SEM
Source: Authors

A large number of pores and microcracks was observed in RCSACC-SC at 100 °C, mainly due to the crystalline transformation of the hydration products. Steel fibers exhibit an interwoven fiber network, which causes a significant increase in the flexural strength of fiber-reinforced RCSACC after high-temperature treatment.

Discussion

Significance of the results

When fiber is added to the mixture, the microstructure is not as dense as that of RCSACC with natural fine aggregate. This differs from literature reports of increased microstructure integrity under dry conditions (Abdulaziz *et al.*, 2020; Kaczmarek *et al.*, 2021), as well as from other works highlighting the increased failure of RCSACC microstructures after high-temperature exposure (Jacek *et al.*, 2021). Increased microstructure failure may affect the strength of the samples and their ability to resist high-temperature exposure (Abbass *et al.*, 2022). Given their ideal aspect ratio, the use of CPM and SC steel fibers in this study facilitated a more effective evaporation and migration of moisture out of the sample in comparison with hooked-end steel fibers, which have a significantly higher failure rate (Bibiana *et al.*, 2017; Ferdous *et al.*, 2018).

In general, fibers can bridge cracks and contribute to mitigating shrinkage cracking (Amin *et al.*, 2022). Our results show the effectiveness of CPM and SC fibers in restraining the shrinkage of the mixture. Adding these materials to RCSACC slightly improved the strength of the sample when compared to other steel fibers with higher aspect ratios (An *et al.*, 2017). Temperature was seen to be associated with changes in the suitability of RCSACC-CPM and RCSACC-SC at multiple scales. However, it was clear that adding a 1% steel fiber content was beneficial for flexural strength and crack expansion (i.e., FRC containing 1% SC exhibited the highest flexural strength). RCSACC-SC showed the highest normalized strengths at all temperatures. This suggests that the brittleness decreased at high temperatures, whereas the ratio of flexural to compressive strength (F/C) increased with increasing temperature.

This study suggests that adaptive SC or CPM fibers may allow RSACC to resist high-temperature exposure. The strength ratio data suggest that the brittleness decreased with increasing temperature regardless of the type of fiber used. The normalized flexural strength generally decreased above 100 °C. For the SC specimens, the normalized flexural strength was much greater than that of the other two types. Our results highlight the potential for increasing the flexural strength of RCSACC by adding 1% SC or CPM fiber content to structures susceptible to heat exposure (Bang *et al.*, 2022). This addition increased the strength at 20 °C. Heating weakened the compressive strength of RCSACC with 1% SC more than it weakened that of RCSACC-CPM. For the latter, heating created pores in the structure (internal voids), which led to the unbonding of the aggregate, CPM fibers, and RCSACC paste that constitute the matrix. This, in turn, weakened, resulting in reduced compressive strength.

The samples exposed to 100 °C exhibited a compressive strength 6.7 MPa lower than that at 20 °C. Previous research (Dong *et al.*, 2022) found that the primary chemical decomposition product is generated after extended exposure to elevated temperatures. This was confirmed by the observed compressive strength reduction and the increase in the number of capillary pores, as well as in the length and width of the microcracks. RCSACC-CPM's flexural strength was lower than that of RCSACC-SC. However, in comparison with the control, there was an increase in flexural strength as the temperature rose. An increase in the load-bearing capacity and the ductility of the sample was observed in the FRC samples after extended heat treatment. RCSACC-SC, with a significant L/D value, increased its ability to resist crack propagation, given the large bonding area of SC fibers with the surrounding matrix (mortar). The specimens made with steel fibers showed a higher displacement resistance capacity and a high tensile strength. This observation agrees with previous research (Abdulkader *et al.*, 2017).

Adding SC fibers or CPM improved the flexural strength at elevated temperatures. The SC fibers increased the

flexural tensile strength of the sample after exposure in comparison with the control. FRC showed a post-exposure flexural traction strength comparable to that of RCSACC before exposure. RCSACC-CPM had a substantially greater load-bearing capacity after the treatment when compared to the control. RCSACC-CPM and RCSACC-SC introduced a pseudo-plastic state, suggesting that they could play a positive role in structures built with RCSACC that are susceptible to heat exposure.

Conclusions

According to the results and the analysis, the following conclusions can be drawn:

1. Ultrasonic wave testing could serve as a quick tool for monitoring and predicting the behavior of RCSACC at high temperatures; after extended conditioning, weakened properties were detected. Ultrasonic wave tests found that the quality of the samples was seriously compromised beyond 100 °C, leaving only RCSACC-SC in suitable condition. With the addition of 1% steel fiber content, the crack width was decreased by 72-93%. In RCSACC-SC, the cracks were almost eliminated, whereas the control showed much wider cracks.
2. The addition of 1% SC reduces the workability and, as a result, the slump of RCSACC. This could be due to the D/L value of SC fibers. The aspect ratio of these fibers caused the RCSACC and sand to wrap around them, creating sturdy fiber-matrix bonds. The ultimate result of this bonding was reduced strength.
3. The aforementioned addition improves the mechanical properties of RCSACC after heat exposure depending on the type of fiber used. At 20 °C, there was a 3.6% increase in compressive strength for RCSACC-CPM, and its flexural strength improved by 8.14%. Overall, a 1% fiber content enhanced the flexural strength of RCSACC.
4. SC and CPM steel fibers aid in transferring stress, which results in increased strength after the treatment when compared to the control. At 100 °C, RCSACC-CPM showed an average compressive strength of 41.2 MPa and a flexural strength of 3.5 MPa, compared to the 28 and 2.43 MPa values obtained by the control. The SC and CPM steel fibers enhanced the samples' ductility.
5. SEM and MIP observations revealed that, at the conditioning temperatures, cracks tended to propagate throughout the RCSACC matrix. The pore size increased as the temperature increased, regardless of the sample, resulting in reduced strength.

Overall, based on these data, a 1% CPM or SC steel fiber content cannot be regarded as an adequate addition, since the paste did not bond sufficiently with these materials, resulting in a more porous concrete with many microcracks, leading to poor performance after conditioning.

Acknowledgements

The authors gratefully acknowledge the financial support received from the Key Laboratory of Green Building Materials (2023GBM03), the Shandong Provincial Higher Education Youth Innovation Team Program of China (2022KJ284), the 111 Project of International Cooperation on Advanced Cement-Based Materials (No. D17001), and the Taishan Scholars Program of China.

CRediT author statement

Jean Jacques Kouadjo Tchekwagep conceived the idea and did the background research. Yiping Qiu and Shifeng Huang collected the data, developed the workflow, and performed the assessment. Shoude Wang and Xin Cheng supervised the research and provided critical feedback. Jean Jacques Kouadjo Tchekwagep wrote the main part of the manuscript, to which all authors contributed.

Conflicts of interest

We declare that there are no conflicts of interest.

References

- ACI Committee (2011). *ACI 214R-11 — guide to evaluation of strength test results of concrete*. American Concrete Institute.
- Abbass, A. A., Abid, S.R., Ali, S. H., Al-Sarray, M. L. J., Murali, G., and Nader, I. A. (2022). Post-high-temperature exposure repeated impact response of steel-fiber-reinforced concrete. *Buildings*, 12, 1364. <https://doi.org/10.3390/buildings12091364>
- Abdulaziz, A., and Yousef R. A. (2022). Strength, durability and shrinkage behaviours of steel fiber reinforced rubberized concrete. *Construction and Building Materials*, 345, 128295. <https://doi.org/10.1016/j.conbuildmat.2022.128295>
- Abdulkader E M., Roland L., and Salem G. N. (2017). Mechanical performance of steel fiber reinforced self-compacting concrete in panels. *Procedia Engineering*, 196, 90-96. <https://doi.org/10.1016/j.proeng.2017.07.177>
- Afroughsabe, V., Biolzi, L., and Ozbakkaloglu, T. (2016). High-performance fiber-reinforced concrete: A review. *Journal of Materials Science*, 51, 6517-6551. <https://doi.org/10.1007/s10853-016-9917-4>
- Aguilar, M. T. P., Bezerra, A. C. S., De Figueiredo, M. A. L., Melo, P. G., Silva, M. J., Oliveira, S. N., Oliveira, L. L. M. S., Resende, D. S., and Silva, N. J. T. (2016). Evaluation of sample preparation parameters in the compressive strength of cementitious composites. *Materials Science Forum*, 869, 93-97. <https://doi.org/10.4028/www.scientific.net/MSF.869.93>
- Ahmed N. E. (2021). EzyGene GelAnalyzer4: A powerful image analysis software for one dimensional gel electrophoresis. *Journal of Genetic Engineering and Biotechnology*, 19, 18. <https://doi.org/10.1186/s43141-020-00114-x>

- Aluko, O. G., Kadir, M. A. A., Yatim, J. M., and Yahya, K. (2020). A review of properties of bio-fibrous concrete exposed to elevated temperatures. *Construction and Building Materials*, 260, 11967. <https://doi.org/10.1016/j.conbuildmat.2020.119671>
- Amin, M. N., Ahmad, W., Khan, K., and Ahmad, A. (2022). Steel fiber reinforced concrete: A systematic review of the research progress and knowledge mapping. *Materials*, 15, 6155. <https://doi.org/10.3390/ma15176155>
- An, L. H., and Ekkehard F. (2017). Influence of steel fiber content and aspect ratio on the uniaxial tensile and compressive behavior of ultra high performance concrete. *Construction and Building Materials*, 153, 790-806. <https://doi.org/10.1016/j.conbuildmat.2017.07.130>
- Aravinthan, T., Ferdous, W., Ghabraie, K., Manalo, A., and Van, E. G. (2018). Evaluation of an innovative composite railway sleeper for a narrow-gauge track under static load. *Journal of Composites for Construction*, 2, 04017050-1-13. [https://doi.com/10.1061/\(ASCE\)CC.1943-5614.0000833](https://doi.com/10.1061/(ASCE)CC.1943-5614.0000833)
- Augusto, C. S. B., Elaine, C. S. C., Maria, T. P. A., Priscilla, S. M., Paulo, R. R. S. J., and Paulo, R. C. (2019). Effect of high temperature on the mechanical properties of steel fiber-reinforced concrete. *Fibers*, 7(12), 100. <https://doi.org/10.3390/fib7120100>
- Azzabi, M., Banthia, N., and Pigeon, M. (1993). Restrained shrinkage cracking in fiber-reinforced cementitious composites. *Materials and Structures*, 26, 405-413. <https://doi.org/10.1007/BF02472941>
- Barreto, R. R., Bezerra, A. C. S., Maciel, P. S., Soares, J. P. R. R., Silva Neto, J. T., and Siqueira Corrêa, E. C. (2019). Thin slabs made of high-performance steel fibre-reinforced cementitious composite: Mechanical behaviour, statistical analysis and microstructural investigation. *Materials*, 20, 3297. <https://doi.org/10.3390/ma12203297>
- Bai, Y., Ferdous, W., Manalo, A., Mendis, P., and Ngo, T. D. (2019). New advancements, challenges and opportunities of multi-storey modular buildings: A state-of-the-art review. *Engineering Structure*, 83, 883-893. <https://doi.org/10.1016/j.engstruct.2019.01.061>
- Bang, Y. L., Jeong-Il C., Se-Eon P., and Yun, Y. K. (2022). Flexural behavior of composite beams of Kagome truss and fiber-reinforced cementitious composites. *Construction and Building Materials*, 361, 129653. <https://doi.org/10.1016/j.conbuildmat.2022.129653>
- Bibiana, L., Facundo, I., Gonzalo, R., Graciela, G., and Raúl Z. (2017). Steel fibers pull-out after exposure to high temperatures and its contribution to the residual mechanical behavior of high strength concrete. *Construction and Building Materials*, 163, 571-585. <https://doi.org/10.1016/j.conbuildmat.2017.12.129>
- Bjegović, D., Baričević, A. R., Pezer, M., Serdar, M., and Štirmir, N. (2015). Shrinkage behaviour of fibre reinforced concrete with recycled tyre polymer fibres. *Civil Engineering Applications of Polymer Composites*, 2015(1), 145918. <https://doi.org/10.1155/2015/145918>
- Błaszczyszński, T., and Przybylska, M. (2015). Steel fibre reinforced concrete as a structural material. *Procedia Engineering*, 11-12, 44-50. <https://doi.org/10.1016/j.pro-eng.2015.10.037>
- Chang, K. H., Wang, W., Wang, H. Y., and Wang, S. Y. (2020). Effect of high temperature on the strength and thermal conductivity of glass fiber concrete. *Construction and Building Materials*, 245, 118387. <https://doi.org/10.1016/j.conbuildmat.2020.118387>
- Chalioris, C. E., Kosmidou, P. M. K., and Karayannis, C. G. (2019). Cyclic response of steel fiber reinforced concrete slender beams; An experimental study. *Materials*, 12(9), 1398. <https://doi.org/10.3390/ma12091398>
- Cheng, X., Che, J., Liu, H., Liu, N., and Zhang, M. (2020). Mechanical performances of concrete produced with desert sand after elevated temperature. *International Journal of Concrete Structures and Materials*, 14, 26. <https://doi.org/10.1186/s40069-020-00402-3>
- Cheng, X., Huang, S., Kouadjo, J. J. T., Mukhopadhyay, A. K., and Wang, S. (2020a). Compressive strength of rapid sulfoaluminate cement concrete exposed to elevated temperatures. *Ceramics-Silikáty*, 64(3), 1-10. https://www.irsm.cas.cz/materialy/cs_content/2020_doi/Tchekwagep_CS_2020_0012.pdf
- Cheng, X., Huang, S., Kouadjo, J. J. T., Mukhopadhyay, A. K., and Wang, S. (2020b). Strengths of sulfoaluminate cement concrete and ordinary portland cement concrete after exposure to high temperatures. *Ceramics-Silikáty*, 64(2), 1-9. https://www2.irsm.cas.cz/materialy/cs_content/2020_doi/Tchekwagep_CS_2020_0019.pdf
- Cheng, X., Huang, S., Kouadjo, J. J. T., Mukhopadhyay, A. K., and Wang, S. (2021). The impact of extended heat exposure on rapid sulfoaluminate cement concrete up to 120°C. *Periodica Polytechnica Civil Engineering*, 65(2), 588-607. <https://doi.org/10.3311/PPci.17122>
- Dong D., Lingchao L., Na C., Piqi Z., Xuecheng W., Yongbo, H., and Zixu Z. (2022). Terneseite-calcium sulfoaluminate cement: Preparation and hydration. *Construction and Building Materials*, 344, 128187. <https://doi.org/10.1016/j.conbuildmat.2022.128187>
- Ferdous, W., Ghazlan, A., Mendis, P., Manalo, A., Ngo, T. D., and Nguyen, K. T. Q. (2018). Effect of fire-retardant ceramic powder on the properties of phenolic-based GFRP composites. *Composites Part B Engineering*, 155, 414-424. <https://doi.org/10.1016/j.compositesb.2018.09.032>
- Gashti, S. H., Sadrmomtazi, A., and Tahmouresi B. (2020). Residual strength and microstructure of fiber reinforced self-compacting concrete exposed to high temperatures. *Construction and Building Materials*, 230, 116969. <https://doi.org/10.1016/j.conbuildmat.2019.116969>
- Hong, S. G., Kang, S. H., Lee, J. H., and Moon, J. (2017). Microstructural investigation of heat-treated ultra-high-performance concrete for optimum production. *Materials*, 10(9), 1106. <https://doi.org/10.3390/ma10091106>
- Jacek, S., Konrad, A. S., Łukasz, K., Mariusz, S., and Sebastian, M. (2021). Impact of elevated temperatures on strength properties and microstructure of calcium sulfoaluminate paste. *Materials*, 14(22), 6751. <https://doi.org/10.3390/ma14226751>
- Kaczmarek, Ł., Miszczak, S., Sodoł, K. A., Stegliński, M., and Szer, J. (2021). The Influence of elevated temperatures on strength properties and microstructure of calcium sulfoaluminate paste. *Materials*, 14, 6751. <https://doi.org/10.3390/ma14226751>

- Kohoutková, A., and Novák, J. (2017). Fiber reinforced concrete exposed to elevated temperature. *Materials Science and Engineering*, 1, 012045. <https://doi.org/10.1088/1757-899X/246/1/012045>
- Li, L. (2019). Stress-rupture of fiber-reinforced ceramic-matrix composites with stochastic loading at intermediate temperatures. Part I: Theoretical analysis. *Materials*, 12(19), 435-458. <https://doi.org/10.1007/s41779-020-00549-y>
- Li, Y., Nguyen, H. T. N., and Tan, K. H. (2021). Shear behavior of fiber-reinforced concrete hollow-core slabs under elevated temperature. *Construction and Building Materials*, 275, 121362. <https://doi.org/10.1016/j.conbuildmat.2020.121362>
- Li, Y., Tan, K. H., and Yang, E. (2020). Flexural behavior of ultra-high performance hybrid fiber reinforced concrete at the ambient and elevated temperature. *Construction and Building Materials*, 250, 118487. <https://doi.org/10.1016/j.conbuildmat.2020.118487>
- Mehta, P. K., and Monteiro, P. J. M. (2006). *Concrete: Micro-structure, properties, and materials* (3rd ed). McGraw-Hill.
- Michels, J., Scherer, J., and Zwicky, D. (2016). Structural strengthening of concrete with fiber reinforced cementitious matrix (FRCM) at ambient and elevated temperature – Recent investigations. *Advances in Structural Engineering*, 17(12), 1785-1799. <https://doi.org/10.1260/1369-4332.17.12.1785>
- Tanyildizi, H., and Yonar, Y. (2016). Mechanical properties of geopolymer concrete containing polyvinyl alcohol fiber exposed to high temperature. *Construction and Building Materials*, 216, 381-387. <https://doi.org/10.1016/j.conbuildmat.2016.09.001>
- Venkatesh, K., and Wasim, K. (2011). Thermal and mechanical properties of fiber reinforced high performance self-consolidating concrete at elevated temperatures. *Cement and Concrete Research*, 41(11), 1112-1122. <https://doi.org/10.1016/j.cemconres.2011.06.012>

Electromechanical Impedance-Based Damage Detection Using Machine Learning Approaches

Detección de daños basada en impedancia electromecánica mediante métodos de aprendizaje automático

Paulo Elias Carneiro Pereira¹, Stanley Washington Ferreira de Rezende², Bruno Pereira Barella³, José dos Reis Vieira de Moura Júnior⁴, and Roberto Mendes Finzi Neto⁵

ABSTRACT

Electromechanical impedance-based structural health monitoring has been the subject of extensive research in recent decades. The method's low cost and ability to detect minor structural damages make it an appealing alternative to other non-destructive techniques. Ongoing research on damage detection approaches continues to be a topic of interest in relation to the electromechanical impedance method. This work proposes the use of the K-Means, Decision Tree, and Random Forest algorithms to distinguish between four structural conditions in an aluminum beam. These techniques were applied to raw impedance data and a dataset reduced via principal components analysis. The findings revealed that the compressed dataset improved the accuracy of all models, except for the Random Forest approach, whose accuracy decreased by 2.9%. The K-Means algorithm was most affected by the reduction in dimensionality, with a 105.9% increase in accuracy. The Decision Tree and Random Forest methods yielded outstanding outcomes, comparable or superior to other state-of-the-art approaches. This makes them a compelling choice for detecting damage using electromechanical impedance data, even when using raw data as the input information.

Keywords: electromechanical impedance method, K-means algorithm, Decision Tree, Random Forest, structural health monitoring

RESUMEN

El monitoreo de la salud estructural basado en la impedancia electromecánica ha sido objeto de investigación exhaustiva en las últimas décadas. El bajo coste del método y su capacidad para detectar daños estructurales menores lo convierten en una alternativa atractiva a otras técnicas no destructivas. La investigación actual sobre enfoques de detección de daños sigue siendo un tema de interés en lo que concierne al método de impedancia electromecánica. En este trabajo se propone utilizar los algoritmos K-Means, Decision Tree y Random Forest para diferenciar entre cuatro condiciones estructurales en una viga de aluminio. Estas técnicas se aplicaron a datos de impedancia en bruto y a un conjunto de datos reducido mediante análisis de componentes principales. Los resultados revelaron que el conjunto de datos comprimido mejoró la precisión de todos los modelos, excepto en el caso del método Random Forest, cuya precisión disminuyó en un 2.9%. El algoritmo K-Means fue el más afectado por la reducción de la dimensionalidad, con un aumento del 105.9% en la precisión. Los métodos Decision Tree y Random Forest produjeron resultados sobresalientes, comparables o superiores a otros enfoques de vanguardia. Esto los convierte en una opción convincente para detectar daños a través de datos de impedancia electromecánica, incluso cuando se utilizan datos en bruto como información de entrada.

Palabras clave: Método de impedancia electromecánica, Algoritmo K-Medias, Árbol de Decisión, Bosque Aleatorio, Control del Estado Estructural.

Received: October 17th 2023

Accepted: May 29th 2024

Introduction

Among the non-destructive methods developed for the structural health monitoring (SHM) of structures and mechanical systems, the method based on electromechanical impedance (EMI) is particularly strategic due to its ability to detect minor structural damage, providing timely data for intervention (Giurgiutiu, 2014).

Damage detection through EMI data is a crucial aspect in the practical application of this technique. To address this issue, various approaches have been devised to enhance damage detection using EMI technology, as demonstrated by Kim and Wang (2019), Fan and Li (2020), Zhou *et al.* (2021), Wang *et al.* (2022), and Nguyen *et al.* (2023).

Machine learning (ML) approaches for damage detection have yielded promising outcomes across various scenarios, with research endeavors delving into both supervised and unsupervised methods to unravel and leverage their potential. Specifically, in the realm of supervised ML

¹Mining engineer, Federal University of Goiás, Brazil. MSc Modeling and Optimization, Federal University of Goiás, Brazil. Affiliation: PhD Candidate, Mechanical Engineering, Federal University of Uberlândia, Brazil. Assistant professor, Federal University of Catalão, Brazil. E-mail: paulo_elias_carneiro@ufu.br, paulo_elias_carneiro@ufcat.edu.br

²Industrial mathematician, Federal University of Goiás, Brazil. MSc Modeling and Optimization, Federal University of Goiás, Brazil. Affiliation: PhD Candidate, Mechanical Engineering, Federal University of Uberlândia, Brazil. E-mail: stanley_washington@ufu.br

³Industrial mathematician, Federal University of Goiás, Brazil. MSc Modeling and Optimization, Federal University of Goiás, Brazil. Affiliation: PhD Candidate, Computer Science, Federal University of Uberlândia, Brazil. E-mail: brunobarella@ufu.br

⁴Mechanical engineer, Federal University of Uberlândia, Brazil. MSc Mechanical Engineering, Federal University of Uberlândia, Brazil. PhD Mechanical Engineering, Federal University of Uberlândia, Brazil. Affiliation: Assistant professor, Federal University of Uberlândia, Brazil. E-mail: zereis@ufu.br

⁵Electrical engineer, Federal University of Uberlândia, Brazil. MSc Electrical Engineering, Federal University of Uberlândia, Brazil. PhD Electrical Engineering, Federal University of Uberlândia, Brazil. Affiliation: Full professor, Federal University of Uberlândia, Brazil. E-mail: finzi@ufu.br



Attribution 4.0 International (CC BY 4.0) Share - Adapt

techniques for damage detection, the predominantly employed algorithms center around neural networks, notably convolutional neural networks (CNNs) (Oliveira et al., 2018a; Rezende et al., 2020; Li et al., 2021; Ai and Cheng, 2023; Du et al., 2023; Nguyen et al., 2023). Other noteworthy neural network architectures include probabilistic neural networks (PNNs) (Na, 2021), deep residual networks (DRN) (Alazzawi and Wang, 2021), fuzzy ARTMAP networks (FAN) (Oliveira et al., 2018b), and extreme learning machines (ELMs) (Djemana et al., 2017). These applications have demonstrated significant efficacy in classifying damage levels, often incorporating considerations for temperature variations.

Non-neural network-based supervised approaches include the random forest (RF) method, which was applied, among other techniques, by Yan et al. (2022) to monitor the early hydration of cement mortar, with promising results. Another application of the RF method developed for EMI-based monitoring was the prediction of impedance signatures in steel structures (Parida et al., 2023).

In parallel, in the realm of unsupervised ML approaches, principal components analysis (PCA) has emerged as a pivotal tool in EMI-based monitoring. The method has found application as a damage index (Malinowski et al., 2021) and has been employed for extracting crucial features from original impedance or admittance signatures (Park et al., 2008; Jiang et al., 2021b; Jiang et al., 2021c; Ai et al., 2022).

In the context of the latter, PCA has been employed in conjunction with various complementary methods, including K-Means (KM) clustering (Park et al., 2008; Jiang et al., 2021b), the support vector machine (SVM) method (Jiang et al., 2021c), and artificial neural networks (ANN) (Ai et al., 2022). These combined approaches have yielded promising results in EMI-based monitoring for damage detection.

Following the application of unsupervised methods in EMI-based monitoring, Perera et al. (2019) used hierarchical clustering and the KM method to detect damages in reinforced concrete (RC) beams.

PCA and KM show promise among the unsupervised methods used in EMI-based monitoring. However, there is a pressing need for more studies to explore the alternative applications of these techniques for effective damage detection. Moreover, the realm of supervised ML methods, including decision trees (DT) and RF, remains relatively uncharted or inadequately examined in EMI-based monitoring.

To address this research gap, our study offers a comprehensive comparative analysis of the KM, DT, and RF methods for damage detection using EMI signals, focusing on their capabilities in classifying four distinct structural conditions and utilizing two types of datasets: unprocessed raw impedance signatures and PCA-compressed datasets. Additionally, the study delves into the impact of dimension reduction on the models' efficacy in predicting damage occurrences.

To the authors' knowledge, the combined study of the DT and RF methods with EMI measurements for damage detection is limited or nonexistent. Hence, this work introduces a novel perspective, wherein the application of these methods can offer fresh insights into approaches for EMI-based damage detection.

Theoretical background

This section presents the theoretical concepts underlying the EMI method, followed by information on the supervised and unsupervised ML methods used to improve damage detection in this work.

Electromechanical impedance method

The EMI method is based on the electromechanical coupling between a structure and a piezoelectric patch, usually made of lead zirconate titanate (PZT), bonded with or embedded into the interrogated structure. The patch is excited by a sinusoidal voltage of varying frequency, resulting in a deformation of the PZT due to the reverse piezoelectric effect as a function of the application of the electric field, whose deformation extends through the local vicinity of the patch (Martowicz and Rosiek, 2013).

The deformation of the patch results in the local deformation of its surroundings, which is returned to the PZT in the form of an electrical signal due to the direct piezoelectric effect. The returned signal is a signature of the structure's health. Any changes in the impedance (or admittance) signals reveal changes in the structural conditions and indicate the occurrence of damage (Sun et al., 1995; Park et al., 2003).

Damage indices are typically used as a quantitative approach to identify damage and evaluate its severity. The developed damage indices include the root mean square deviation (RMSD), the mean absolute percentage deviation (MAPD), the correlation coefficient (CC), the correlation coefficient deviation (CCD), and the changes in peak frequencies (Giurgiutiu, 2014; Fan and Li, 2020; Nomelini et al., 2020; Gonçalves et al., 2021).

The information obtained from EMI-based structural monitoring guides the condition-based maintenance of structures and mechanical systems.

Supervised learning

During supervised learning, the algorithm uses two types of information to predict a new and unobserved state: the input vectors, which are the independent variables, and their corresponding labels; or target variables, which depend on the input vectors (Bishop, 2006; Simeone, 2018; Géron, 2019).

Many algorithms based on supervised learning have been developed. This work used the DT and RF methods to classify the health states in an aluminum beam.

The DT method is a recursive model in which a given number of units are divided into subgroups, and the impurity of a set is reduced by creating subgroups (Rokach and Maimon, 2009; Kim and Upneja, 2014), thereby establishing an effective model for solving classification and regression problems (Liu et al., 2021).

The structure of the method is a tree-like hierarchy consisting of a root, internal or test nodes, and leaves (Liu et al., 2021). The root corresponds to the initial node and has no incoming edges. The internal or test nodes have outgoing edges and form a set of nodes called a *branch*. The leaves correspond to a terminal node, have only one incoming edge, and refer to a subgroup that cannot be further divided,

representing decision outcomes (Koza, 1991; Rokach and Maimon, 2009; Kim and Upneja, 2014; Hong *et al.*, 2018).

During the classification process, the DT algorithm starts at the root node containing all objects in the training dataset, performs a test on an attribute, and generates child nodes according to the value of the objects. The process is repeated recursively in each node and ends when a stopping criterion is met (Jegadeeshwaran and Sugumaran, 2013; Chen *et al.*, 2018; Hong *et al.*, 2018; Fletcher and Islam, 2019; Barros *et al.*, 2023; Loyola-González *et al.*, 2023). The resulting leaf or external nodes represent the outcomes or predicted classes (Chen *et al.*, 2020; Biswal and Parida, 2022).

The DT method generates a single tree structure, which, however, may become trapped in a local optimum (Fletcher and Islam, 2019) or overfit the training set (Ning *et al.*, 2021).

This issue is overcome by using multiple DTs, *i.e.*, the RF method, an ensemble learning algorithm consisting of a set of DTs, each associated with a subsample taken from the training set using bootstrap sampling with replacement, resulting in uncorrelated samples (Chencho *et al.*, 2020; Tang *et al.*, 2021).

Bootstrapping randomly selects samples from a single input training set, generating multiple datasets to train the base classifiers (Loyola-González *et al.*, 2023). Each DT in the forest grows upon the basis of the corresponding bootstrapped dataset and the randomly selected features at each node, resulting in a collection of prediction results (Bergmayr *et al.*, 2023).

The results obtained from each tree are averaged or determined by majority voting and applied to regression and classification problems, respectively (Lim *et al.*, 2021; Dinh *et al.*, 2023). This process adds diversity to the DTs, preventing the model from finding only a local optimum and overfitting the training set (Fletcher and Islam, 2019).

Unsupervised learning

Unlike supervised ones, unsupervised methods can handle unlabeled data sets and learn the specific properties and mechanisms underlying data generation, thereby helping to understand the data structure (Simeone, 2018; Géron, 2019).

From a dataset with a given number of unlabeled observations, the goal is to learn valuable properties of the distribution, which depend on the application, as is the case of density estimation, clustering, dimensionality reduction, feature extraction, and new samples generation (Simeone, 2018).

Clustering is one of the most fundamental unsupervised learning problems. It consists of dividing the original data into clusters, each of which comprises similar items, while the clusters are different from each other (Alelyani *et al.*, 2014), thereby providing important insights into the structure of the dataset.

Among the clustering methods is the KM algorithm, a heuristic technique that groups points close to each other based on Euclidean distance, assigning them to one of the initially defined k clusters (Bishop, 2006; Simeone, 2018; Mayer, 2020).

According to the number of clusters defined, the method works as follows. First, the k centroids are randomly determined, and each sample is assigned to a cluster based on its Euclidean distance to each centroid. Based on the distance between the samples and the centroids, the latter are recalculated until the convergence criterion is reached (Harrison, 2020).

The resulting clusters contain the samples whose sum of the squares of the distances to the corresponding centroid represents the minimum value, implying an optimal configuration for the k clusters considered (Reddy and Vinzamuri, 2014).

For some problems, the number of features can reduce the model's ability to predict the classes correctly. This can be mitigated by performing dimensionality reductions via the PCA method.

PCA converts a dataset with a given number of variables into a more concise one, in which the variables or features are contained within principal components orthogonal to each other (Davis, 2002). The method projects the original high-dimensional data to a low-dimensional subspace based on the eigenvectors of the covariance matrix (Liu and Han, 2014).

The principal components define a hyperplane that, when projected onto the data, preserves as much variance as possible, so that the selected hyperplane loses less information than the other possible hyperplanes (Géron, 2019).

The first principal component accounts for most of the variance explained, and so on until the last principal component. Thus, the method can reduce the dimensionality of the dataset while preserving the behavior of the data as much as possible.

Materials and methods

The experimental setup consisted of (i) an aluminum beam; (ii) a PZT patch; (ii) an Analog Devices AD5933 acquisition board; and (iv) a personal computer. The first step was to glue the PZT patch to the 300 x 25 x 3 mm aluminum beam (Figure 1).



Figure 1. Aluminum beam and PZT patch used in the experiments
Source: Authors

The procedure for gluing the PZT patch to the structure was of great importance. Since the impedance measurement was systemic (*i.e.*, a measurement of the electro-mechanical impedance of the structure plus the PZT patch), it was essential to ensure a thin layer that was sufficient for the adhesion of the two surfaces and was not thick enough

to induce other unwanted mechanical behaviors (Islam and Huang, 2014; Albakri and Tarazaga, 2017; Bari and Moharana, 2024).

This gluing procedure consisted of applying a thin layer of epoxy resin onto the PZT patch and mounting it onto the surface of the specimen. Next, a thin 2 mm layer of EVA foam (ethylene vinyl acetate) and a weight of 1 kg were placed to ensure the homogeneity of the cure for the glue layer.

As the technique used was a differential and not an absolute procedure, *i.e.*, the damage metrics were calculated concerning the reference integrity state, the effect of variations in the glue layer was disregarded.

Using wires, the PZT patch was connected to the AD5933 acquisition board to obtain the impedance signatures. The signal gathered from the AD5933 was read by the acquisition board's software on the personal computer.

The real part of an electrical impedance is known as the *resistance* (R), and the imaginary part is the *reactance* (X). Resistance represents the opposition to the flow of electrical current due to factors such as the resistance of conductors, circuit components, and, in this experiment, the mechanical movement of the structure. Reactance represents the opposition to the flow of current which arises from reactive elements such as capacitors (capacitive reactance) and inductors (inductive reactance). In this experiment, the reactance is capacitive because of the PZT patch. Electrical impedance is then represented as a combination of these two terms, where the real part is directly associated with the mechanical properties of the studied specimen.

In this work, the resistance was employed since it is the most sensitive to the presence of damage and is less affected by environmental conditions (Na and Lee, 2013; Hamzeloo et al., 2020; Meher et al., 2022).

The acquisition system was set to cover a frequency interval varying from 60 to 70.2 kHz in steps of 20 Hz, resulting in 511 points in the frequency domain. The selected frequency interval was determined through trial and error, as it exhibited a high degree of sensitivity to the damage being assessed. Additionally, the interval range was selected to accommodate the resolution limitations of the AD5933 acquisition board.

The experiment involved four scenarios: one with the beam in its original state (baseline) and three others with damage. The damage was simulated by attaching a 10 x 10 x 10 mm magnet to the beam's surface in three different positions. The first damage position (Damage 1) was far from the PZT patch, whereas, in the subsequent scenarios (Damage 2 and Damage 3), the magnet was gradually brought closer to the PZT patch. For each scenario, 30 samples were collected, based on an average of 20 measurements in order to reduce random variation.

The collected data were analyzed in Microsoft's Visual Studio Code using the Python language (version 3.8.17). The datasets for each condition (Healthy State, Damage 1, Damage 2, and Damage 3) were imported using the Scipy package (Virtanen et al., 2020) and consisted of two matrices. One dataset was related to the frequency data, which was presented as a one-row, 511-column vector, making it a vector of the frequency values, and the other

corresponded to the real-part values of the impedance, with 30 rows and 511 columns, where each row represented a sample. Each column corresponded to a point within the frequency range in both datasets.

From the imported data set, the median of the resistance values along the frequency domain was calculated for the baseline condition and the three damage states, aiming to analyze the damage-related variations in the impedance signatures. All numerical calculations were performed using the Numpy package (Harris et al., 2020), and plotting was performed using Matplotlib (Caswell et al., 2023).

In the sequence, all the resistance matrices, each with 30 rows and 511 columns, were merged into a single one by concatenating rows, resulting in a 120-row, 511-column matrix. The first 30 rows were attributed to the healthy condition, the next 30 to Damage 1, and so on. The resistance values were then normalized to an interval between 0 and 1 to avoid undesirable effects caused by outliers once these signatures had been characterized by the presence of peaks, which could lead to significant differences in the magnitude of the values.

A one-row, 511-column vector of labels was created for each condition. The number 0 was assigned to the pristine condition, number 1 to Damage 1, and so on. The vectors containing the labels were merged using the same process employed for the real-part matrices of the impedance, resulting in a 4-row, 511-column matrix.

The resulting dataset was split into two parts, 70,0 % for training the model and 30 % for testing, in order to verify the models' ability to correctly predict the classes, which were all previously known: one baseline condition and three different damage levels. The dataset was partitioned using the Scikit-learn package (Pedregosa et al., 2011).

The training dataset was used to obtain predictive models based on the KM, DT, and RF algorithms of the Scikit-learn package (Pedregosa et al., 2011). In this step, the training data were based on the raw signature dataset of 511 variables.

In determining the KM model, the number of clusters was an important parameter. Based on the Elbow method, this number was set to 4, and it was selected as the first point from which there were no significant changes in the within-cluster sum of squares (WCSS).

The models based on the DT and RF methods were created using entropy, *i.e.*, Equation (1), as a criterion for defining the nodes, as well as based on the $p(X)$ value, which is a fraction of examples belonging to a given class. Entropy is a measure of the purity of a group, wherein low values indicate high homogeneity or a high level of purity, and higher values indicate a lower level of purity.

$$\text{Entropy} = - \sum p(X) \log p(X) \quad (1)$$

The ability of the models to correctly predict the classes was determined by using the test dataset and submitting these values to the model. Accuracy, *i.e.*, Equation (2), was used as a parameter to quantify the capability of each model.

$$\text{Accuracy}(y, \hat{y}) = \frac{1}{n} \sum_{i=0}^{n-1} 1(\hat{y}_i = y_i) \quad (2)$$

Accuracy corresponds to the mean of the correctly predicted classes, where y is the actual class, and \hat{y} is the predicted class (Géron, 2019; Scikit-Learn Team, 2023). This verification is performed along the test vector, which refers to the number of samples evaluated (n). The total number of times that the true values are equal to the corresponding predicted ones is divided by the total number of items in the test vector, resulting in an accuracy value.

The number of variables was decreased using PCA to evaluate the effect of reducing the dimensionality of the dataset on the models' ability to predict the classes correctly.

The PCA was selected over other dimensionality reduction techniques such as multidimensional scaling (MDS), uniform manifold approximation and projection (UMAP), and t-distributed stochastic neighbor embedding (t-SNE), given its simplicity and computational processing capabilities when dealing with data in high dimensions. Additionally, the PCA is less susceptible to the influence of outliers.

Moreover, the PCA method exhibits superior computational efficiency (it is a low complexity, covariance matrix-based linear method) when compared to nonlinear approaches. While the dataset utilized in this study may not be extensive, thereby yielding similar processing times across various dimensionality reduction methodologies, it is crucial to recognize that, within an automated system operating at scale and encompassing diverse structural configurations and sensor arrays over extended durations, this efficiency advantage can yield tangible reductions in energy consumption and decision-making times.

In this sense, the original training and testing data were transformed into principal components and used in the KM, DT, and RF algorithms.

To build the KM model with the number of principal components and clusters, a routine was defined to obtain the optimal combination of parameters to maximize the model's accuracy.

Figure 2 shows the accuracy values for each tested combination. The results show that the highest accuracy value corresponds to 25 principal components and four clusters.

The execution of KM clustering – and the other methods – using the raw data and PCA information resulted in two prediction models, the first of which was based on 511 variables and the other on 25 variables (PCAs), with the latter representing a reduction of 95.1% in the number of features to be analyzed.

To allow for comparisons between the ML methods, the dataset consisting of 25 principal components was used to train the DT and RF models. Entropy was also used as a criterion to define the groups, and the parameters were the same as those used in the other models.

The resulting models were tested using the test set (30,0% of the data). Accuracy was also used as a parameter to quantify the models' ability to correctly predict the classes, and the

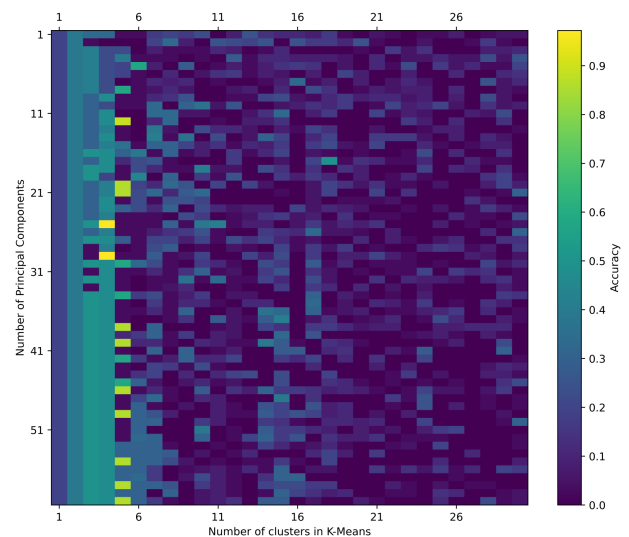


Figure 2. Matrix chart of the accuracy values as a function of the number of principal components and clusters

Source: Authors

results were used to make comparisons between the studied approaches.

Results and discussion

The analysis of the predictions generated by the KM models (Figure 3) reveals a notable discrepancy in performance. Specifically, the KM model relying on raw data exhibits many inaccuracies, as depicted in Figure 3a. This discrepancy is particularly pronounced for the first and second damage scenarios, wherein all predictions were erroneously classified. Conversely, accurate predictions were achieved for the baseline condition and the third damage scenario.

Figure 3 also shows that the same portion of the third damage scenario was misclassified as the second class, which may be due to the characteristics of the database itself – probably the number of samples.

The PCA method's improvement in the classification ability of the KM algorithm can also be observed in the WCSS results (Figure 4). Considering the four clusters used, the WCSS for the PCA-based approach was 48,7% lower than that attributed to the raw data-based KM model.

WCSS values indicate how compact clusters are, wherein high values reflect a high variability within the cluster. In this sense, the lower values obtained by the model based on principal components reflect a better classification.

The results obtained with the DT method (Figure 5) show that the model based on the raw signature database yielded predictions that agreed with the corresponding true values (Figure 5a). Misclassification occurred in the second and third damage scenarios, with only small proportions of incorrect predictions.

Furthermore, misclassification occurred mainly between consecutive damage scenarios, probably due to the small differences between impedance signatures. The exception was the 2.78% of the third damage scenario that was misclassified as Damage 1.

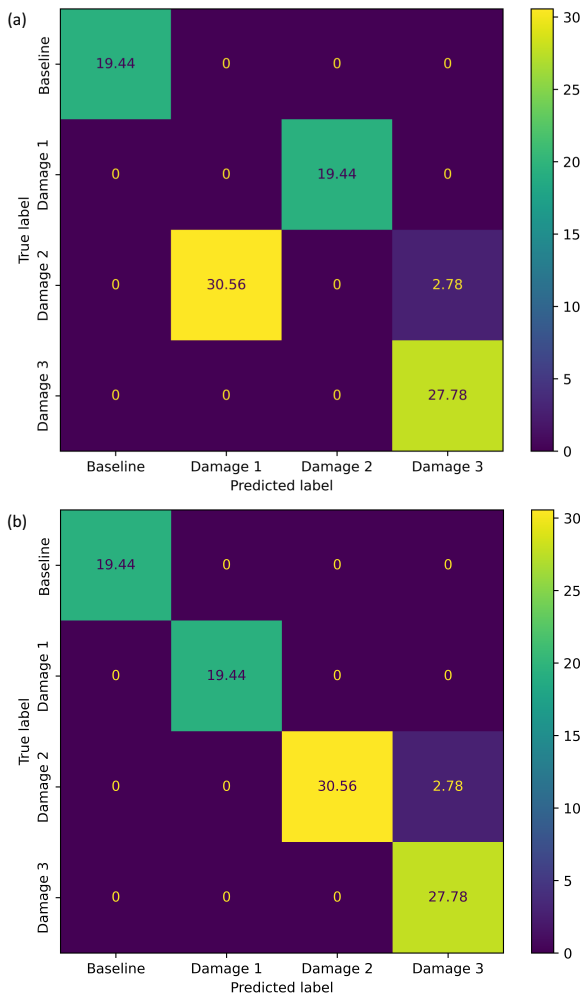


Figure 3. Confusion matrix for the KM clustering models based on raw data (a) and PCA information (b)
Source: Authors

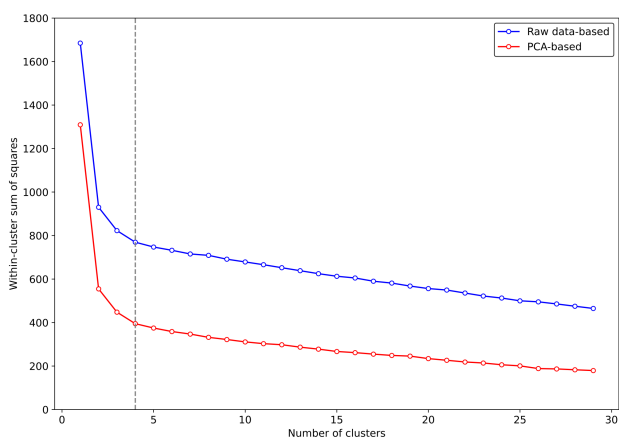


Figure 4. Plot of the WCSS obtained for the KM method using raw data and PCA information
Source: Authors

The results obtained via the DT model based on PCA information (Figure 5b) showed that all predictions matched the corresponding accurate labels, achieving a hit rate of 100%. Therefore, this model could differentiate all damaged scenarios based only on impedance signatures.

When comparing the raw data and the PCA-based results of the DT method, only a slight improvement was observed

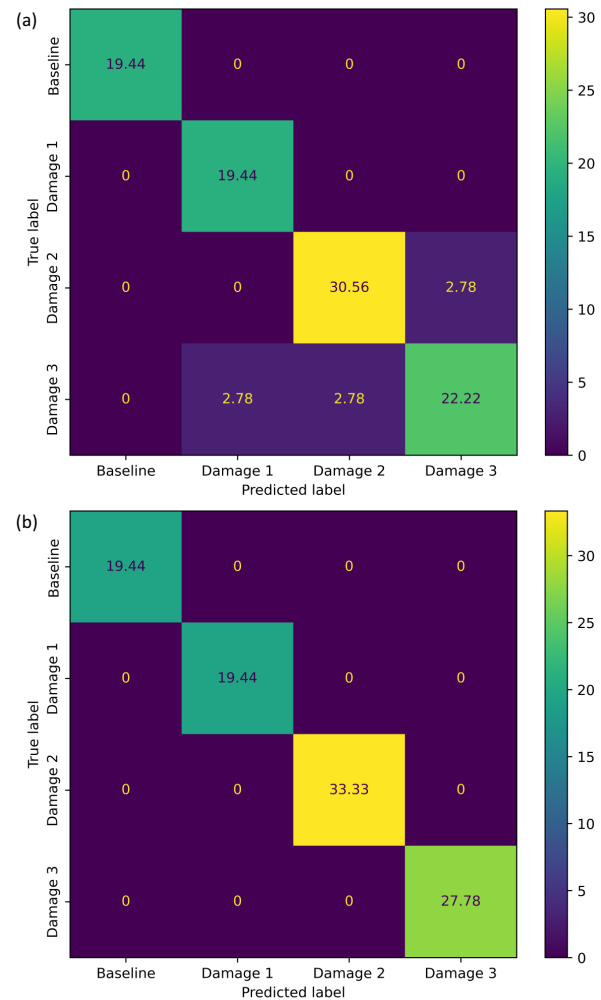


Figure 5. Confusion matrix for the DT models based on raw data (a) and PCA information (b)
Source: Authors

on one occasion. Even in the model based on the raw data, it was possible to obtain reliable results, with little deviations. This indicates that the DT method is a reliable and robust option when classifying structural conditions based on impedance signatures.

The confusion matrices obtained from the RF method are shown in Figure 6. The results show that the model based on raw data yielded predictions with minor deviations from the actual labels (Figure 6a), with only 2.78% of the third damage scenario misclassified.

The results of the PCA-based RF model (Figure 6b) show a slightly higher deviation than that based on raw data, which may be due to the fact that the method is characterized as an ensemble. This feature can be better explored when dealing with large data.

However, the misclassification reported by the PCA-based RF model was only 2.78% higher than the results based on raw data, so the former does not represent significant changes concerning the latter. In this sense, the model based on PCA information can provide reliable results.

The accuracy obtained by each model (Figure 7) shows that KM clustering was the most sensitive to the dimensionality reduction of the dataset. In this case, reducing the number

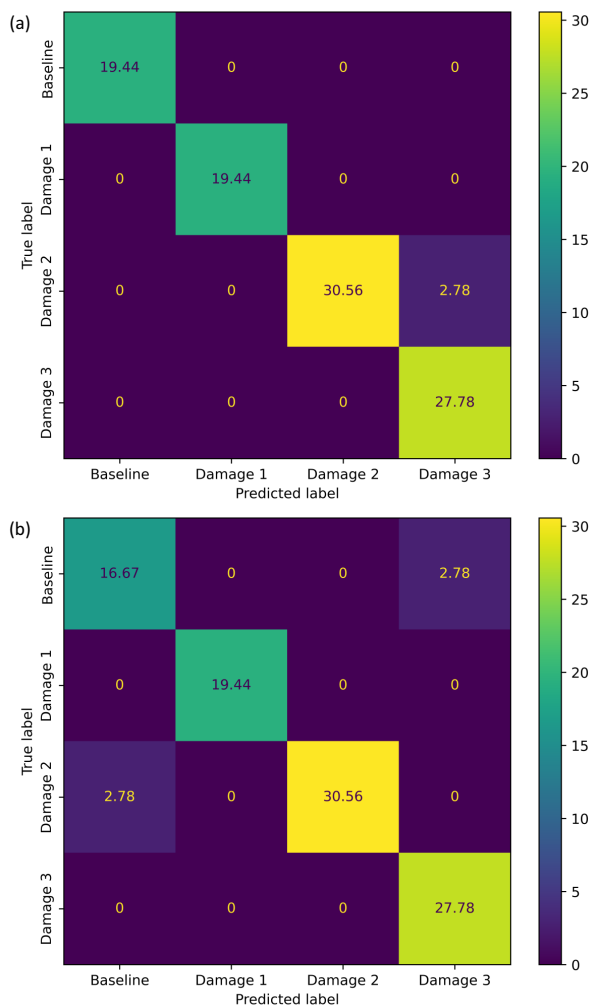


Figure 6. Confusion matrix for the RF models based on raw data (a) and PCA information (b)

Source: Authors

of variables from 511 to 25 principal components resulted in a 105.9% increase in accuracy.

On the other hand, the DT method reported an improvement of only 9.1% when comparing the PCA-based results with the raw data-based ones, and the RF method showed a 2.9% decrease in accuracy when using PCA information. This behavior can be explained by the characteristics of the RF method, which can be advantageous in situations with a large number of variables.

When it came to using raw data for model training, the RF method yielded the best results and the KM model the worst. When the PCA data set was used, the DT method performed the best, correctly classifying all classes.

The KM method utilizing PCA compression outperformed the probability neural network (PNN) (Na, 2021), which achieved an accuracy rate of 94.4%. Furthermore, the KM approach exhibited comparable outcomes to the CNN technique (de Rezende et al., 2020), which reported a minimum accuracy rate of 97%.

The DT method showed the highest level of accuracy when using PCA information. It outperformed the PNN (Na, 2021) and CNN (de Rezende et al., 2021; Du et al., 2023) approaches and was comparable to other cutting-edge ML techniques for damage detection (Alazzawi and Wang, 2021;

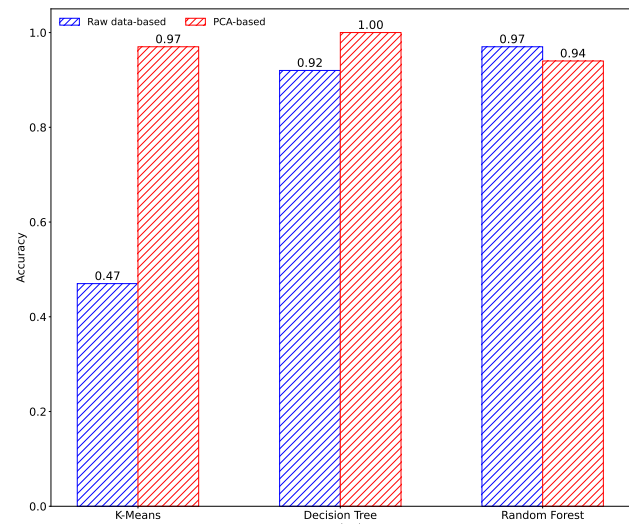


Figure 7. Accuracy results for each method in the two scenarios considered

Source: Authors

Jiang et al. 2021a; Ai and Cheng, 2023). Additionally, the DT method utilizing PCA compression yielded superior results compared to the DT-based approach presented by Jiang et al. (2021a), making it a noteworthy strategy for detecting damage using EMI data.

The results obtained with the RF approach were either comparable to or better than those obtained from the PNN method (Na, 2021). Additionally, they closely matched the outcomes of other related studies (de Rezende et al., 2020; Jiang et al., 2021a). Therefore, the RF method is a viable means to detect damage accurately. Moreover, it employs a single-step process and does not necessitate PCA compression to enhance the model's capabilities.

The results obtained herein show that reducing the number of variables generally entails an increase in accuracy, except for the RF method, without invalidating the use of the model.

The evaluation shows that the DT and RF methods are the most appropriate for damage detection in SHM applications using only a dataset of impedance signatures. When using PCA information, the former showed the highest accuracy, and the latter was the most stable, with acceptable accuracy values.

Conclusions

Identifying the presence of damage is a critical task in SHM systems. When using the EMI technique, the information derived from the impedance signature aids in detecting damage. When considering real-world applications, it is necessary to use adaptable techniques to provide reliable results in general situations.

This work used a dataset with impedance signatures of four classes in the KM, DT, and RF methods. In addition, to study the effect of reducing dataset dimensionality, PCA was applied, resulting in two scenarios for each method (i.e., raw data- and PCA-based).

The results show that, in general, reducing the number of variables in the model through PCA caused an increase in the accuracy of the models, with an exception attributed to

the RF model. However, the decrease was not pronounced in this case, and the accuracy obtained was acceptable.

An analysis of the results showed the DT and RF methods to be the most suitable methods for damage detection based only on an impedance signature dataset. The former exhibited the highest accuracy value and good stability, resulting in a close and acceptable accuracy. On the other hand, the KM method managed to provide reliable results only by applying PCA to reduce dimensionality.

In addition, the approaches based on the DT and RF methods yielded results comparable to or better than other state-of-the-art solutions for damage detection using EMI.

According to the results obtained, applying supervised or unsupervised ML techniques for damage detection using EMI yields promising outcomes. Based only on a dataset of impedance signatures, an accurate classification was possible. Furthermore, by using the DT method, it was possible to achieve an accuracy of 100%.

The methods used herein are attractive options when it comes to damage detection using EMI data. Furthermore, the studied approaches do not require any pre-processing step before establishing the models.

In the experimental procedures, damage of the same severity was induced at varying distances from the PZT patch. Other works conducted by the authors have explored scenarios involving defects of different severity levels at the same distance from the patch. This proposal only considered the first scenario to check for feasibility, but different ML model responses for discerning different damage levels can be further explored.

Additionally, future research may focus on considering scenarios with different temperatures to enhance the use of unsupervised techniques for damage detection, bringing the approaches used herein closer to shop-floor industrial applications.

Acknowledgements

P. E. C. Pereira would like to express his gratitude to the Federal University of Catalão for the provision of the license to pursue doctoral studies at the Federal University of Uberlândia.

CRedit author statement

P. E. C. Pereira: formal analysis, methodology, writing (original draft). S. W. F. de Rezende: data curation, writing (review and editing). B. P. Barella: data curation. J. R. V. de Moura Jr.: conceptualization, methodology, writing (review and editing). R. M. Finzi Neto: supervision, writing (review and editing).

Conflicts of interest

The authors declare no competing interests relevant to the content of this article.

References

- Ai, D., and Cheng, J. (2023). A deep learning approach for electromechanical impedance-based concrete structural damage quantification using a two-dimensional convolutional neural network. *Mechanical Systems and Signal Processing*, 183, 109634. <https://doi.org/10.1016/j.ymssp.2022.109634>
- Ai, D., Mo, F., Yang, F., and Zhu, H. (2022). Electromechanical impedance-based concrete structural damage detection using principal component analysis incorporated with a neural network. *Journal of Intelligent Material Systems and Structures*, 33(17), 2241–2256. <https://doi.org/10.1177/1045389X221077440>
- Alazzawi, O., and Wang, D. (2021). Damage identification using the PZT impedance signals and residual learning algorithm. *Journal of Civil Structural Health Monitoring*, 11, 1225–1238. <https://doi.org/10.1007/s13349-021-00505-9>
- Albakri, M. I., and Tarazaga, P. A. (2017). Dynamic analysis of a piezoelectric augmented beam system with adhesive bonding layer effects. *Journal of Intelligent Materials Systems and Structures*, 28(2), 178–194. <https://doi.org/10.1177/1045389X16648426>
- Alelyani, S., Tang, J., and Liu, H. (2014). Feature Selection for Clustering: A Review. In C. C. Aggarwal and C. K. Reddy (Eds.), *Data Clustering: Algorithms and Applications* (pp. 29–60). CRC Press.
- Bari, Sk. A., and Moharana, S. (2024). A novel approach to alloy-based bonding for piezoelectric sensor integrated in impedance-based structural health monitoring. *Journal of Asian Architecture and Building Engineering*, 23(3), 898–920. <https://doi.org/10.1080/13467581.2023.2257279>
- Barros, B., Conde, B., Cabaleiro, M., and Riveiro, B. (2023). Design and testing of a decision tree algorithm for early failure detection in steel truss bridges. *Engineering Structures*, 289, 116243. <https://doi.org/10.1016/j.engstruct.2023.116243>
- Bergmayr, T., Holl, S., Kralovec, C., and Schagerl, M. (2023). Local residual random forest classifier for strain-based damage detection and localization in aerospace sandwich structures. *Composite Structures*, 304(1), 116331. <https://doi.org/10.1016/j.compstruct.2022.116331>
- Bishop, C. M. (2006). *Pattern recognition and machine learning*. Springer.
- Biswal, T., and Parida, S. K. (2022). A novel high impedance fault detection in the micro-grid system by the summation of accumulated difference of residual voltage method and fault event classification using discrete wavelet transforms and a decision tree approach. *Electric Power Systems Research*, 209, 108042. <https://doi.org/10.1016/j.epsr.2022.108042>
- Caswell, T. A., Sales de Andrade, E., Lee, A., Droettboom, M., Hoffmann, T., Klymak, J., Hunter, J., Firing, E., Stansby, D., Varoquaux, N., Nielsen, J. H., Gustafsson, O., Root, B., May, R., Sundén, K., Elson, P., Seppanen, J. K., Lee, J.-J., Dale, D., ... Silvester, S. (2023). *matplotlib/matplotlib: REL: v3.7.3 (v3.7.3)*. Zenodo. <https://doi.org/10.5281/zenodo.8336761>
- Chen, D., Montano, V., Huo, L., and Song, G. (2020). Depth detection of subsurface voids in concrete-filled steel tubular (CFST) structure using percussion

- and decision tree. *Measurement*, 163, 107869. <https://doi.org/10.1016/j.measurement.2020.107869>
- Chen, W., Zhang, S., Li, R., and Shahabi, H. (2018). Performance evaluation of the GIS-based data mining techniques of best-first decision tree, random forest, and naive Bayes tree for landslide susceptibility modeling. *Science of The Total Environment*, 644, 1006–1018. <https://doi.org/10.1016/j.scitotenv.2018.06.389>
- Chencho, J., Li, H., Hao, R., Wang, L., and Li, L. (2021). Development and application of random forest technique for element level structural damage quantification. *Structural Control and Health Monitoring*, 28(3), e2678. <https://doi.org/10.1002/stc.2678>
- Davis, J. C. (2002). *Statistics and Data Analysis in Geology* (3rd ed.). John Wiley & Sons.
- de Rezende, S. W. F., de Moura Jr., J. R. V., Finzi Neto, R. M., Gallo, C. A., and Steffen Jr., V. (2020). Convolutional neural network and impedance-based SHM applied to damage detection. *Engineering Research Express*, 2(3), Article 035031. <https://doi.org/10.1088/2631-8695/abb568>
- Dinh, T. P., Pham-Quoc, C., Thinh, T. N., Nguyen, B. K. D., and Kha, P. C. (2023). A flexible and efficient FPGA-based random forest architecture for IoT applications. *Internet of Things*, 22, 100813. <https://doi.org/10.1016/j.iot.2023.100813>
- Djemana, M., Hrairi, M., and Al Jeroudi, Y. (2017). Using Electromechanical impedance and extreme learning machine to detect and locate damage in structures. *Journal of Nondestructive Evaluation*, 36, 39. <https://doi.org/10.1007/s10921-017-0417-5>
- Du, F., Wu, S., Xu, C., Yang, Z., and Su, Z. (2023). Electromechanical Impedance Temperature Compensation and Bolt Loosening Monitoring Based on Modified Unet and Multitask Learning. *IEEE Sensors Journal*, 23(5), 4556–4567. <https://doi.org/10.1109/JSEN.2021.3132943>
- Fan, X., and Li, J. (2020). Damage identification in plate structures using sparse regularization based electromechanical impedance technique. *Sensors*, 20(24), 7069. <https://doi.org/10.3390/s20247069>
- Fletcher, S., and Md. Islam, Z. (2019). Decision tree classification with differential privacy: A survey. *ACM Computing Surveys*, 52(4), 83. <https://doi.org/10.1145/3337064>
- Géron, A. (2019). *Mãos à obra aprendizado de máquina com Scikit-Learn and TensorFlow: Conceitos, ferramentas e técnicas para a construção de sistemas inteligentes*. Alta Books.
- Giurgiutiu, V. (2014). *Structural health monitoring with piezoelectric wafer active sensors* (2nd ed.). Academic Press. <https://doi.org/10.1016/C2013-0-00155-7>
- Gonçalves, D. R., de Moura Jr., J. R. V., Pereira, P. E. C., Mendes, M. V. A., and Diniz-Pinto, H. S. (2021). Indicator kriging for damage position prediction by the use of electromechanical impedance-based structural health monitoring. *CR Mécanique*, 349(2), 225–240. <https://doi.org/10.5802/crmeca.81>
- Hamzeloo, S. R., Barzegar, M., and Mohsenzadeh, M. (2020). Damage detection of L-shaped beam structure with a crack by electromechanical impedance response: Analytical approach and experimental validation. *Journal of Nondestructive Evaluation*, 39(2), 47. <https://doi.org/10.1007/s10921-020-00692-3>
- Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M. H., Brett, M., Haldane, A., del Río, J. F., Wiebe, M., Peterson, P., ... Oliphant, T. E. (2020). Array programming with NumPy. *Nature*, 585, 357–362. <https://doi.org/10.1038/s41586-020-2649-2>
- Harrison, M. (2020). *Machine learning: Guia de referência rápida*. Novatec Editora.
- Hong, H., Liu, J., Bui, D. T., Pradhan, B., Acharya, T. D., Pham, B. T., Zhu, A.-X., Chen, W., and Ahmad, B. B. (2018). Landslide susceptibility mapping using J48 Decision Tree with AdaBoost, bagging and Rotation Forest ensembles in the Guangchang area (China). *CATENA*, 163, 399–413. <https://doi.org/10.1016/j.catena.2018.01.005>
- Islam, M. M., and Huang, H. (2014). Understanding the effects of adhesive layer on the electromechanical impedance (EMI) of bonded piezoelectric wafer transducer. *Smart Materials and Structures*, 23(12), 125037. <https://doi.org/10.1088/0964-1726/23/12/125037>
- Jegadeeshwaran, R., and Sugumaran, V. (2013). Comparative study of decision tree classifier and best first tree classifier for fault diagnosis of automobile hydraulic brake system using statistical features. *Measurement*, 46(9), 3247–3260. <https://doi.org/10.1016/j.measurement.2013.04.068>
- Jiang, X., Zhang, X., Tang, T., and Zhang, Y. (2021c). Electromechanical impedance-based self-diagnosis of piezoelectric smart structure using principal component analysis and LibSVM. *Scientific Reports*, 11, 11345. <https://doi.org/10.1038/s41598-021-90567-y>
- Jiang, X., Zhang, X., and Zhang, Y. (2021a). Establishment and optimization of sensor fault identification model based on classification and regression tree and particle swarm optimization. *Materials Research Express*, 8(8), 085703. <https://doi.org/10.1088/2053-1591/ac1cae>
- Jiang, X., Zhang, X., and Zhang, Y. (2021b). Piezoelectric active sensor self-diagnosis for electromechanical impedance monitoring using K-means clustering analysis and artificial neural network. *Shock and Vibration*, 2021, 5574898. <https://doi.org/10.1155/2021/5574898>
- Kim, J., and Wang, K.-W. (2019). Electromechanical impedance-based damage identification enhancement using bistable and adaptive piezoelectric circuitry. *Structural Health Monitoring*, 18(4), 1268–1281. <https://doi.org/10.1177/1475921718794202>
- Kim, S. Y., and Upneja, A. (2014). Predicting restaurant financial distress using decision tree and AdaBoosted decision tree models. *Economic Modelling*, 36, 354–362. <https://doi.org/10.1016/j.econmod.2013.10.005>

- Koza, J. R. (1990). Concept formation and decision tree induction using the genetic programming paradigm. In H.-P. Schwefel and R. Manner (Eds.), *PPSN I: Proceedings of the 1st Workshop on Parallel Problem Solving from Nature* (pp. 124–128). Springer-Verlag.
- Li, H., Ai, D., Zhu, H., and Luo, H. (2021). Integrated electromechanical impedance technique with convolutional neural network for concrete structural damage quantification under varied temperatures. *Mechanical Systems and Signal Processing*, 152, 107467. <https://doi.org/10.1016/j.ymssp.2020.107467>
- Lim, D. K., Mustapha, K. B., and Pagwiwoko, C. P. (2021). Delamination detection in composite plates using random forests. *Composites Structures*, 278, 114676. <https://doi.org/10.1016/j.compstruct.2021.114676>
- Liu, J., and Han, J. (2014). Spectral clustering. In C. C. Aggarwal and C. K. Reddy (Eds.), *Data Clustering: Algorithms and Applications* (pp. 177–200). CRC Press.
- Liu, R., Li, S., Zhang, G., and Jin, W. (2021). Depth detection of void defect in sandwich-structured immersed tunnel using elastic wave and decision tree. *Construction and Building Materials*, 305, 124756. <https://doi.org/10.1016/j.conbuildmat.2021.124756>
- Loyola-González, O., Ramírez-Sáyago, E., and Medina-Pérez, M. A. (2023). Towards improving decision tree induction by combining split evaluation measures. *Knowledge-Based Systems*, 277, 110832. <https://doi.org/10.1016/j.knosys.2023.110832>
- Malinowski, P. H., Wandowski, T., and Singh, S. K. (2021). Employing principal component analysis for assessment of damage in GFRP composites using electromechanical impedance. *Composite Structures*, 266, 113820. <https://doi.org/10.1016/j.compstruct.2021.113820>
- Martowicz, A., and Rosiek, M. (2013). Electromechanical impedance method. In T. Stepinski, T. Uhl, and W. Staszewski (Eds.), *Advanced Structural Damage Detection: From Theory to Engineering Applications* (ch. 6, pp. 141–176). John Wiley & Sons. <https://doi.org/10.1002/9781118536148.ch6>
- Mayer, C. (2020). *Python one-liners: Write concise, eloquent Python like a professional*. No Starch Press, Inc.
- Meher, U., Mishra, S. K., and Sunny, M. R. (2022). Impedance-based looseness detection of bolted joints using artificial neural network: An experimental study. *Structural Control and Health Monitoring*, 29(10), e3049. <https://doi.org/10.1002/stc.3049>
- Na, S., and Lee, H. K. (2013). A multi-sensing electromechanical impedance method for non-destructive evaluation of metallic structures. *Smart Materials and Structures*, 22(9), 095011. <https://doi.org/10.1088/0964-1726/22/9/095011>
- Na, W. S. (2021). Bolt loosening detection using impedance-based non-destructive method and probabilistic neural network technique with minimal training data. *Engineering Structures*, 226, 111228. <https://doi.org/10.1016/j.engstruct.2020.111228>
- Ning, F., Cheng, Z., Meng, D., and Wei, J. (2021). A framework combining acoustic features extraction method and random forest algorithm for gas pipeline leak detection and classification. *Applied Acoustics*, 182, 108255. <https://doi.org/10.1016/j.apacoust.2021.108255>
- Nomellini, Q. S. S., da Silva, J. V., Gallo, C. A., Finzi Neto, R. M., Tsuruta, K. M., and de Moura Jr., J. R. V. (2020). Non-parametric inference applied to damage detection in the electromechanical impedance-based health monitoring. *International Journal of Advanced Engineering Research and Science*, 7(9), 73–79. <https://doi.org/10.22161/ijaers.79.9>
- Oliveira, M. A. de, Monteiro, A. V., and Vieira Filho, J. (2018a). A new structural health monitoring strategy based on PZT sensors and convolutional neural network. *Sensors*, 18(9), 2955. <https://doi.org/10.3390/s18092955>
- Parida, L., Moharana, S., and Giri, S. K. (2023). Machine learning approach for predicting impedance signatures of construction steel structures in various tensile pull action. *Materials Today: Proceedings*. <https://doi.org/10.1016/j.matpr.2023.03.741>
- Park, G., Sohn, H., Farrar, C. R., and Inman, D. J. (2003). Overview of piezoelectric impedance-based health monitoring and path forward. *Shock and Vibration Digest*, 35(6), 451–463. <https://doi.org/10.1177/05831024030356001>
- Park, S., Lee, J.-J., Yun, C.-B., and Inman, D. J. (2008). Electro-mechanical impedance-based wireless structural health monitoring using PCA-data compression and K-means clustering algorithms. *Journal of Intelligent Material Systems and Structures*, 19(4), 509–520. <https://doi.org/10.1177/1045389X07077400>
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830
- Perera, R., Torres, L., Ruiz, A., Barris, C., and Baena, M. (2019). An EMI-based clustering for structural health monitoring of NSM FRP strengthening systems. *Sensors*, 19(17), 3775. <https://doi.org/10.3390/s19173775>
- Reddy, C. K., and Vinzamuri, B. (2014). A survey of partitional and hierarchical clustering algorithms. In C. C. Aggarwal and C. K. Reddy (Eds.), *Data Clustering: Algorithms and Applications* (pp. 87–110). CRC Press
- Rokach, L., and Maimon, O. (2009). Classification trees. In L. Rokach and O. Maimon (Eds.), *Data Mining and Knowledge Discovery Handbook* (pp. 165–192). Springer. https://doi.org/10.1007/978-0-387-09823-4_9
- Scikit-Learn Team (2023). *Model selection and evaluation*. https://scikit-learn.org/stable/model_selection.html
- Simeone, O. (2018). A brief introduction to machine learning for engineers. *Foundations and Trends® in Signal Processing*, 12(3-4), 200–431. <https://doi.org/10.1561/2000000102>
- Sun, F. P., Chaudhry, Z., Liang, C., and Rogers, C. A. (1995). Truss structure integrity identification

using PZT sensor-actuator. *Journal of Intelligent Materials Systems and Structures*, 6(1), 134–139. <https://doi.org/10.1177/1045389X9500600117>

- Tang, X., Gu, X., Rao, L., and Lu, J. (2021). A single fault detection method of gearbox based on random forest hybrid classifier and improved Dempster-Shafer information fusion. *Computers and Electrical Engineering*, 92, 107101. <https://doi.org/10.1016/j.compeleceng.2021.107101>
- Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., ... SciPy 1.0 Contributors. (2020). SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nature Methods*, 17, 261–272. <https://doi.org/10.1038/s41592-019-0686-2>
- Wang, L., Yuan, B., Xu, Z., and Sun, Q. (2022). Synchronous detection of bolts looseness position and degree based on fusing electro-mechanical impedance. *Mechanical Systems and Signal Processing*, 174, 109068. <https://doi.org/10.1016/j.ymssp.2022.109068>
- Yan, Q., Liao, X., Zhang, C., Zhang, Y., Luo, S., and Zhang, D. (2022). Intelligent monitoring and assessment on early-age hydration and setting of cement mortar through an EMI-integrated neural network. *Measurement*, 203, 111984. <https://doi.org/10.1016/j.measurement.2022.111984>
- Zhou, L., Chen, S.-X., Ni, Y.-Q., and Choy, A. W.-H. (2021). EMI-GCN: A hybrid model for real-time monitoring of multiple bolt looseness using electromechanical impedance and graph convolutional networks. *Smart Materials and Structures*, 30(3), 035032. <https://doi.org/10.1088/1361-665X/abe292>

An Intelligent System-Based Strategic Plan for a Humanoid Robot Playing the Game of Dominoes

Plan estratégico basado en sistemas inteligentes para el juego de dominó por parte de un robot humanoide

Álex Medina¹, Daniela Charris², Mauricio Pardo³, and Christian G. Quintero M.⁴

ABSTRACT

The application of intelligent systems in humanoid robots provides research and development alternatives, as is the case with human-robot interaction. This paper focuses on the design and implementation of an intelligent system in the NAO robot to plan and execute moves in the board game known as *dominoes*. This system uses the NAO robot's vision to determine the distribution of tiles on the board, as well as those available in hand. The appropriate moves are planned using a computational intelligence technique, and a kinematics model executes them. The results show that the vision system has an average error of 5.62%, in addition to 3.37% for the decision-making system and 7.87% for the kinematics of the robot. This leads to the NAO robot being capable of making successful plays through the proposed system, with an average effectiveness of 83.15%.

Keywords: computer vision, decision tree, dominoes, board games, forward kinematics, human robot interaction, image processing, intelligent system, NAO robot, robotics

RESUMEN

La aplicación de sistemas inteligentes en robots humanoides brinda alternativas de investigación y desarrollo, como es el caso de la interacción humano-robot. Este trabajo se enfoca en el diseño e implementación de un sistema inteligente en el robot NAO para planificar y ejecutar movimientos en el juego de mesa conocido como *dominó*. Este sistema utiliza el sistema de visión del robot NAO para determinar la distribución de fichas en el tablero y de las disponibles en la mano. Los movimientos adecuados se calculan mediante una técnica de inteligencia computacional, y un modelo de cinemática los ejecuta. Los resultados muestran que el sistema de visión tiene un error promedio del 5.62 %, así como del 3.37 % para el sistema de decisión y de 7.87 % para la cinemática del robot. Esto lleva a que, a través del sistema propuesto, el robot NAO sea capaz de realizar jugadas exitosas con una efectividad promedio del 83.15 %.

Palabras clave: visión por computadora, árbol de decisión, dominó, juegos de mesa, cinemática inversa, interacción humano-robot, procesamiento de imágenes, sistema inteligente, robot NAO, robótica

Received: May 12th 2023

Accepted: September 1st 2024

Introduction

Humanoid robots' incorporation of intelligent systems creates new opportunities for research and development in fields such as human-robot interaction (HRI). HRI has recently been employed in a variety of sectors, including education [1, 2, 3], sports [4], entertainment [5], business and industry [7, 8], and healthcare [9, 10, 11], among others.

For a friendly HRI, three fundamental components are necessary: environment detection (vision system), planning (decision system), and movement actions (execution system). In particular, the NAO robot's vision system has been used for environment exploration through object detection [12, 13], object tracking [14], image segmentation [15], visual navigation [16], and transfer learning [17]. Regarding its decision system, the NAO robot has been used for sentiment analysis [18, 19] and to support psychologists during treatments [20, 21, 22]. As for the execution system, kinematic methods have been studied to perform specific tasks, such as imitating human arm movements [23, 24] and walking [25, 26], among others.

More specifically, robots playing board games have increasingly gained attention for their multiple benefits. Robots can play with people, enhancing interaction enjoyment and social engagement [27]. However, humans

may find it challenging to compete against robots in immersive entertainment settings. The cited study examines how players' gaming experiences vary depending on whether a human or an embodied robot serve as the game guide. In addition, people who do not compete with other humans can gain social experiences via robots, and students can learn programming and other STEM skills by playing board games with robots, as indicated by [28], whose study evaluates how robots affect students' perceptions of learning and their motivation to learn.

¹Mechatronics Engineering, Universidad de las Fuerzas Armadas, Ecuador. MSc Electronics Engineering, Universidad del Norte, Colombia. Affiliation: Member of the Robotics and Intelligent Systems Research Group, Universidad del Norte, Colombia. E-mail: medinafa@uninorte.edu.co

²Electronics Engineering, Universidad del Norte, Colombia. MSc Electronics Engineering, Universidad del Norte, Colombia. Affiliation: Member of the Robotics and Intelligent Systems Research Group, Universidad del Norte, Colombia. E-mail: dmcharris@uninorte.edu.co

³Electronics Engineering, Universidad del Norte, Colombia. MSc and PhD in Electrical Engineering, Georgia Institute of Technology, USA. Affiliation: Assistant professor and member of the Robotics and Intelligent Systems Research Group, Universidad del Norte, Colombia. E-mail: mpardo@uninorte.edu.co

⁴Electronics Engineering, Universidad Industrial de Santander, Colombia. MSc and PhD in Information Technologies, Universidad de Girona, Spain. Affiliation: Associate professor and member of the Robotics and Intelligent Systems Research Group, Universidad del Norte, Colombia. E-mail: christianq@uninorte.edu.co



Attribution 4.0 International (CC BY 4.0) Share - Adapt

As for teleoperation roles, robots can deliver haptic feedback in addition to visual information, which enhances the user experience [29]. Moreover, robotic board games can boost creative capacity [30]. In particular, humanoid robots are used for research in different types of games, where they play against a person or are required to play for themselves. Several humanoid robots have been used to play chess by processing board and tile images [31, 32, 33]. In [34], a NAO robot was employed to this effect. Another NAO robot capable of playing Simon Says was presented in [35].

Current advances show the potential of HRI in board games for a variety of purposes, including education, entertainment, and research. The goal of robots playing board games is to make the experience more engaging and challenging for humans. Recent research has focused on developing algorithms that can learn to play high-level board games. Machine learning techniques such as evolutionary computation, neural networks, and reinforcement learning can be used by robots to improve their gameplay over time [36]. In particular, reinforcement learning techniques in robotics can be developed and tested using robots, enabling the system to complete tasks that would normally need a sophisticated decision-making algorithm [37].

Robots can learn to play board games utilizing a variety of techniques, including programming instruction, machine learning, and cognitive human-robot interfaces [38]. Yet, how well they work depends on the objectives and the study goals. In short, the effectiveness of robots in playing board games is determined by a variety of factors such as the complexity of the game, the robot's level of autonomy, and the quality of the algorithms used. Robots may have an advantage in terms of strategy and decision-making in certain scenarios, where they may use machine learning and reinforcement learning techniques to discover complex behaviors and strategies in real time. Humans may still have an advantage in situations where creativity, intuition, and social interaction are important decision-making factors. Robots may be incapable of adapting to unexpected moves or strategies employed by human opponents, and they may be unable to comprehend the social and emotional aspects of gameplay that humans value. However, by using intelligent techniques, robots can learn and improve their gameplay. In this sense, one of the main challenges in designing robots for board games is to ensure that the robot understands the game rules and can make appropriate decisions. This calls for the creation of algorithms capable of interpreting the game state and determining the best move to make. Another challenge is creating a way for the robot to interact with physical game pieces that is intuitive and natural for human players. This could entail creating specialized sensors or grippers that can handle various types of game pieces. Another challenge is to make the robot socially acceptable to human players. This includes considering aspects such as the robot's appearance, behavior, and level of autonomy. A robot that is overly competitive or aggressive, for example, may be off-putting to human players, whereas a robot that is overly passive may not provide a satisfying challenge. In this sense, the robot's design can influence how humans perceive and interact with it during gameplay. The ultimate challenge is to design the robot so that it can adapt to different types of board games, which calls for the development of algorithms and hardware

capable of handling a wide range of game dynamics and rules.

To summarize, designing robots for board games involves a variety of challenges related to game understanding, physical interaction, social acceptability, and adaptability. Overcoming these challenges necessitates a multidisciplinary approach that combines expertise in robotics, artificial intelligence, HRI, and game design. For example, [36] describes a real-time vision-based robotic system that plays Carrom against a skilled human opponent, and [39] presents an automated laser alignment and image processing-based robotic Carrom player. The above references demonstrate the potential for robots to play board games, but there are no widely known examples of successful robots designed specifically for this purpose. In this vein, this paper aims to contribute to the field of HRI entertainment, specifically board games. Dominoes was chosen for this study because it is a traditional game from Central America and the Caribbean. This study examines whether a humanoid robot is capable of interacting with people and planning and executing its plays based on game conditions (tiles on the table and those in hand) via an intelligent system-based strategic plan.

This paper is structured as follows. First, an explanation of the game of dominoes is presented in the *Materials and method* section. Afterwards, the proposed approach is divided into vision, decision, and execution systems, outlining each aspect. Then, the complete system is presented and demonstrated via a user interface. Finally, the *Results and discussion* section describes the tests carried out to validate the system's performance.

Materials and method

Dominoes board game

The most common game of dominoes consists of 28 tiles, each with two numbers represented by pips and blanks, as shown in Fig. 1. There are seven tiles of each number, going from zero to six. In the most popular game mode, four players participate in either pairs or score modes. In this paper, a three-player game variant is explored in order to implement individual game strategies, as the robot's vision of the available tiles in hand covers a maximum of nine.

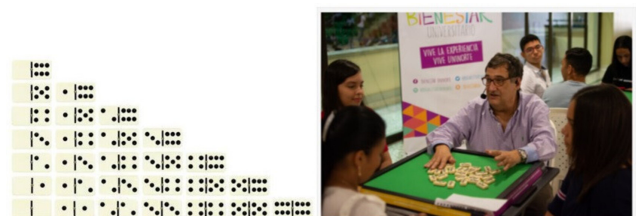


Figure 1. Dominoes game

Source: Authors

For the three-player variant, the blank-blank tile is discarded, so that nine tiles correspond to each player. The player who draws the double-six tile starts the game. However, if it is a recurring game, the winner of the last set starts a new one with the double-six tile, but, if they do not have it, they can start playing with the highest double tile they have. If the player does not have a double tile, then the tile with the

most pips comes out. The starting tile is placed face-up at the center of the board. The next player must play to match the blanks/pips on either side of the starting tile.

In the following rounds, the players take turns repeating the procedure. As the game unfolds, there will be two available ends to place a playing tile. However, if a player does not have a tile that matches the blanks/pips at either end, they pass the playing turn to the next player. The game ends when a player manages to place all their tiles or when the game is closed (no player can place a tile on the board). The player with the least pip count wins the set [40].

Game theory deals with the study of games like dominoes, classifying them as *deterministic* or *non-deterministic*. The former are those for which it is possible to generate a winning strategy. A sequence of steps or *strategy* is said to be winning because it increases the chance of winning for the player who applies it, regardless of the circumstances. Non-deterministic games, in turn, are those for which there is no winning strategy. In the case of dominoes, the presence of *randomness* implies that it is not possible to determine a strategy that always wins. For this research, we selected a non-deterministic game, since the aim is for the players to interact with the robot, winning or losing the games in which they participate.

Proposed approach

This paper describes an intelligent system as a strategic plan that allows a NAO robot to play the game of dominoes. The proposed system is divided into three subsystems: image processing (vision), computational intelligence (decision-making), and robot kinematics (execution). The components of each subsystem are presented in Fig. 2.

The image processing subsystem captures what the robot perceives of its surroundings with its built-in camera. Through this process, the most relevant variables of the game are obtained, such as the tiles on the table and those available in hand. These data are then entered into the computer system. In the computational intelligence subsystem, a decision tree-based algorithm is executed, so that the robot can decide on the tile to be played. This decision tree considers several game strategies collected from people with medium to high experience in the game. Finally, the robot kinematics subsystem provides the most viable way to execute a play, taking the selected tile from those available in hand and placing it correctly on the board. For this work, direct kinematics was proposed and then experimentally verified by a model that determined its viability, as it minimized the error when choosing a tile.

The vision system includes two cameras to separately observe the tiles available in hand and those on the board. The resulting data are then entered into the image processing block to obtain the game variables.

As observed in Fig. 2, there is an interconnection block, *i.e.*, game variables acquisition, that continues into the decision system. It was devised using a voice recognition algorithm, and its role is to let the NAO robot know when its turn to play has begun.

Once the game variables have been acquired, the decision system processes the tiles available on the table and the ones in hand, using an intelligent system that provides the final

decision, *i.e.*, the tile to be played. Once said tile has been selected, the execution system takes the decision system output as input to the direct kinematics model of the NAO robot. The forward kinematics direct one of the robot's arms to pick the selected tile and place it on the board, which is followed by a voice command.

Finally, the main system returns to the game variables acquisition block, where the NAO robot waits for the voice command to resume the game.

Game scenario

For the proposed system, we decided that at least three players would play the game, as the NAO robot's field of vision can cover up to nine tiles, given the tile set selected for testing. Then, the tile and board colors were established. In this case, white tiles with black pips and a blackboard were selected, so that there would be a better contrast for the robot's vision and image processing. Similarly, the game area was delimited, defining the location of the tiles in hand for the NAO robot and the gaming space to place the tiles on the board. The tiles in hand had to be located in two planes (Fig. 3), so that the NAO robot could visualize them correctly and they were within reach of its arms.

Vision system

To process the images of the tiles in hand, an algorithm was developed, the partial results of which are shown in Fig. 4. As observed, the objective is to detect all the tiles together with their corresponding pips, leading to the identification of the associated number.

To this effect, the Edge Detection algorithm provides information about object boundaries, and it is employed for analysis and filtering applications. Its goal is to drastically reduce the amount of data to be processed while preserving structural information. This algorithm's working principle detects sudden color intensity changes and enhances them, highlighting the resulting image edges. It is very effective in accentuating contrast and detecting isolated spots or small details.

One of the algorithms most commonly used in this field is Canny, which looks for the maximum gradient along an edge. The result is a border/non-border binary threshold-adjustable image, as shown in Fig. 4 (adaptive binarization). In addition, the Segmentation algorithm subdivides an image into its constituent regions or objects, so that the pixels in said regions have similar properties or attributes, *e.g.*, gray level, contrast, or texture.

The segmentation process is entrusted with evaluating each pixel in the image and deciding whether it contains characteristics of interest.

The Hough transform allows detecting curves in an image and is based on the search for geometric characteristics such as lines, triangles, and circular objects, among others. The Hough transform is one of the most widely used model-based segmentation techniques due to its robustness against noise and its behavior in the presence of holes in the object's border. When applying the Hough transform to an image, it is first necessary to obtain a binary image of the pixels that are part of the object boundary using threshold-based segmentation. The goal of this technique is to find points in

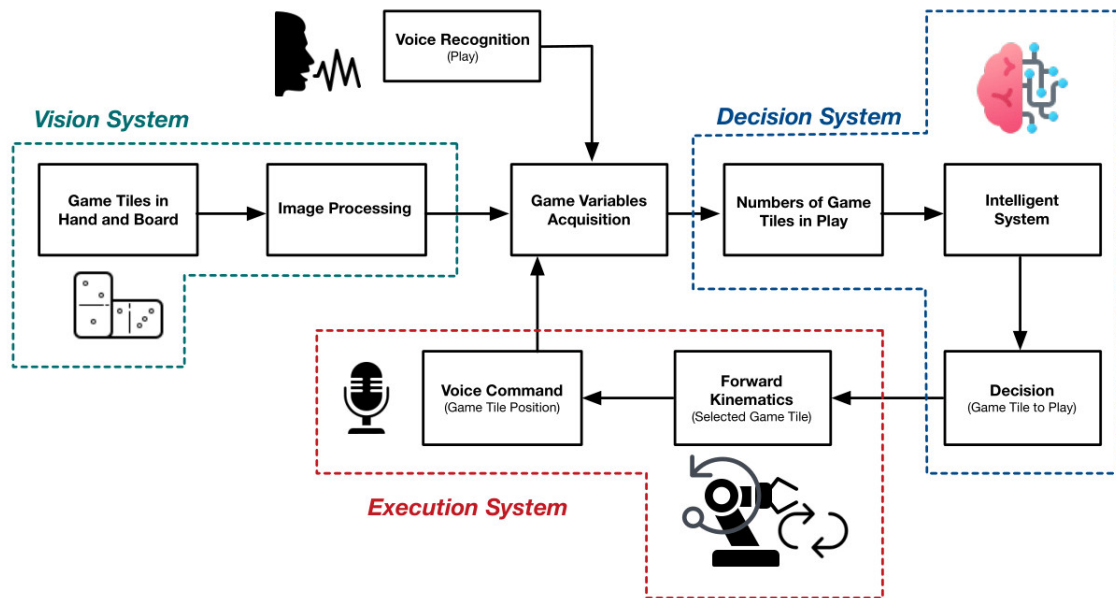


Figure 2. Diagram of the proposed approach
Source: Authors



Figure 3. Game scenario for the proposed system
Source: Authors

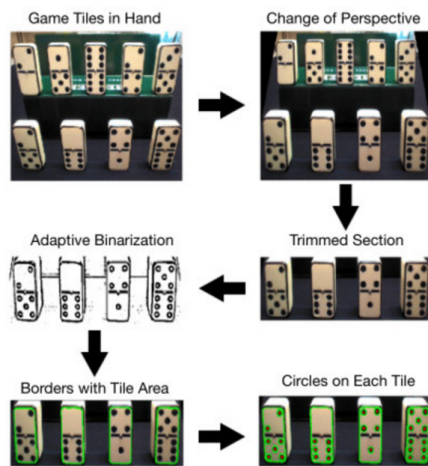


Figure 4. Algorithm for the detection of the game tiles in hand
Source: Authors

the image that are aligned [41]. Fig. 4 shows the use of this transform to find circles in an image during the final step of the detection process.

To process the images of the tiles placed on the board, a similar algorithm is employed, with the difference that there was no change of perspective regarding the image to be processed. The partial results of board processing are shown in Fig. 5.

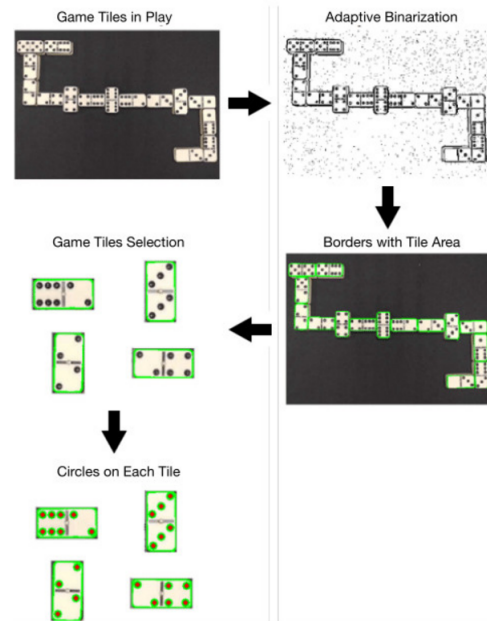


Figure 5. Algorithm for the detection of the game tiles placed on the board
Source: Authors

Once all the tiles on the board have been identified, the system identifies the ones at the end and their number. To this effect, four axes are assigned (Fig. 6).

Then, the axes of each tile are compared against each other. When two axes have the same location within the image, it means that two tiles are joined. Therefore, if a tile has only a single axis shared with another tile, it means that it is a tile at the end of the current game. A combination of the above-mentioned axes is made to determine the tile side that indicates which end has to be played. For example, if the tile is vertical, the axes Y2 and X1 and X2 and Y1 are combined. If one of these combinations contains the shared axis, it is discarded. Therefore, the opposite side is chosen, whose number will be obtained (Fig. 7).

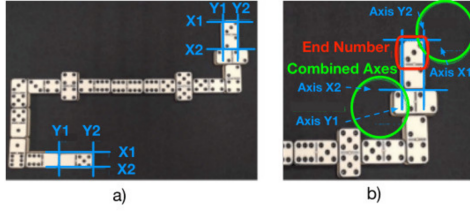


Figure 6. a) Axis assignment of both ends for the game tiles in play, b) end number example

Source: Authors

Decision system

A decision tree takes an object or situation as input (which is described through a set of attributes) and returns a decision, i.e., the expected value of the output. The input attributes can be discrete or continuous. Thus, the decision tree develops a sequence of conditionals to reach a decision. Each internal node of the tree corresponds to a conditional for the value of one of the properties, and the branches that come out of the node are labeled with the possible values of said property. Each leaf node of the tree represents the value that must be returned upon reaching it.

For the intelligent system, the decision tree shown in Fig. 7 was used. In this tree, each node contains the main techniques or strategies according to their importance or priority in the game. These strategies have been compiled through research and collaboration from experienced dominoes players. Each of them is listed below, in order of priority.

Tile counting: The most important strategy is counting how many tiles of the same number are in hand and how many have already been placed on the board.

Blocking: Once the tile counting strategy has been executed, if the player has five or more tiles of the same number (adding both tiles in hand and on the board), the game should be played so that both ends match that number, as shown in Fig. 8.

Keeping ahead: After counting all the tiles in hand and those already placed on the board, if the player has six or seven tiles of the same number, the game should be played only at the other end, always keeping the tiles with that number until the end (if possible), as shown in Fig. 9.

One of each: When the player only has one option for playing on either side, they should play by placing the tile whose number has come out the most during the game.

Doubles first: Once the player has counted the tiles and determined that it is not possible to make a blocking move, they should get rid of the doubles first.

Best tile: If the player does not have doubles to place, the next strategy is called *the best tile*. To choose such a tile, the player first selects the number that has been placed the most on the board. Then, since a tile has two numbers, the player must choose the one that complements the already selected number with the one that exhibits the next majority of the tiles played, as shown in Fig. 10.

The first nodes of the decision tree correspond to the initial and basic plays of the game. A basic play is when a player only has one tile with a number that matches one of the

ends of the tiles on the board. There are also situations when a player has two or more tiles that can be played. In that case, the doubles first or best tile rules apply. Once the tile to be played has been selected using this decision tree, the information is passed on for the execution system to physically perform the corresponding moves.

Execution system

To take the tiles and place them on the board, the robot uses its left arm. The joints involved in the movement are shown in Fig. 11. Here, the forward kinematic problem is used to find a transformation matrix that relates the coordinate system linked to the body in motion to a reference coordinate system. Homogeneous transformation matrices are used to achieve this representation, as shown in Eqs. (1) and (2), which includes translation and orientation operations. To this effect, a 4 x 4 matrix transforms a vector expressed in homogeneous coordinates from one coordinate system to another [42].

$$T = \begin{bmatrix} \text{rotation matrix} & \text{position vector} \\ f_{ix3} & \text{scaled} \end{bmatrix} = \begin{bmatrix} n_x & s_x & a_x & p_x \\ n_y & s_y & a_y & p_y \\ n_z & s_z & a_z & p_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (1)$$

$$T = \begin{bmatrix} n & s & a & p \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (2)$$

The vectors n , s , and a are unit orthogonal vectors, and p is a vector that describes the position x , y , and z of the origin of the current system with respect to the reference (Fig. 12).

Within the execution system, a forward kinematics model was created for the NAO robot's arm. With this model, the robot can take the tile selected for each move.

Axes must be assigned for each joint of the arm in order to establish this model, for which they must first be identified. Then, all the links that make up the robot's arm are listed, and the z -axis is assigned depending on the rotation of each joint. Afterwards, the x - and y -axes are located while following the right-hand rule. Once the axes have been properly assigned, the Denavit-Hartenberg (DH) parameters are obtained. These are the main basis for developing the forward kinematics model. The identification (numbering) or allocation of axle systems is first performed for each joint θ_i to obtain said parameters.

After assigning the axes, the DH parameters are determined as follows:

- θ_i : Angle around the Z_{i-1} axis, from the X_{i-1} axis to the X_i axis.
- d_i : Distance along the Z_{i-1} axis, from the origin of the $i - 1$ system to the X_i axis.
- a_i : Distance along the X_i axis, from the Z_{i-1} axis to the Z_i axis.
- α_i : Angle around the X_i axis, from the Z_{i-1} axis to the Z_i axis.

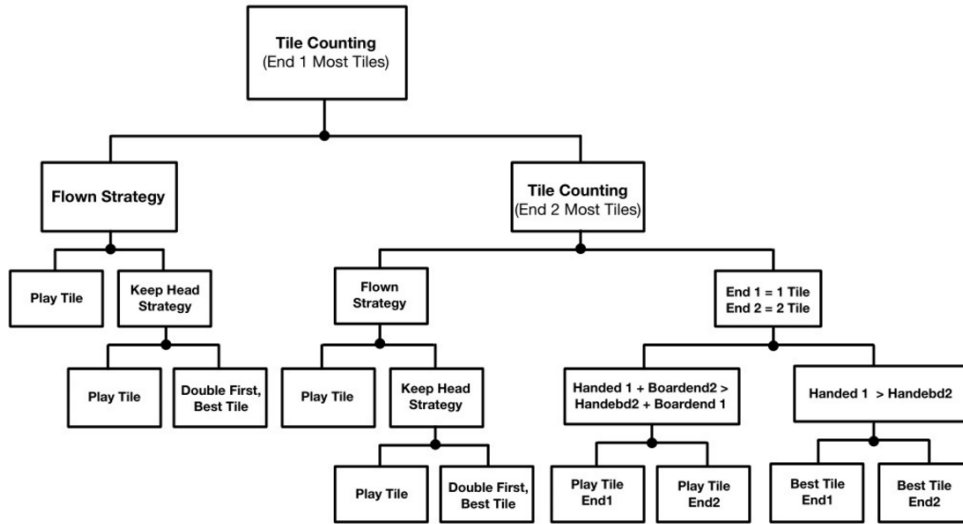


Figure 7. Decision tree for the intelligent game system
Source: Authors



Figure 8. Blocking play
Source: Authors



Figure 9. Keep-ahead play
Source: Authors

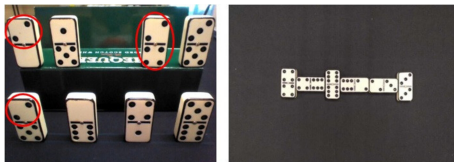


Figure 10. Best tile play
Source: Authors

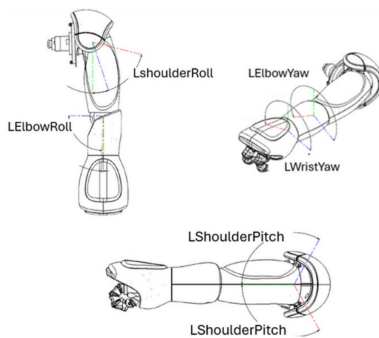


Figure 11. Joints used for the NAO robot's movement
Source: Authors

Based on the information in Table 1, the transformation matrix for each joint is obtained using the general matrix presented in Eq. (3).

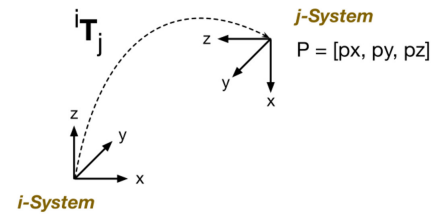


Figure 12. Reference coordinate system
Source: Authors

Table 1. DH parameters

Joints	Parameter			
	a	α	d	θ
LShoulderPitch	0	$-\pi/2$	0	θ_1
LShoulderRoll	0	$\pi/2$	0	$\theta_2 - \pi/2$
LElbowYaw	0	$-\pi/2$	d_3	θ_3
LElbowRoll	0	$\pi/2$	0	θ_4
LWristYaw	0	$-\pi/2$	d_5	θ_5

Source: Authors

$$\begin{bmatrix} \cos \theta_i & -\sin \theta_i \cos \alpha_i & \sin \theta_i \sin \alpha_i & \alpha_i \cos \theta_i \\ \sin \theta_i & \cos \theta_i \cos \alpha_i & \cos \theta_i \sin \alpha_i & \alpha_i \sin \theta_i \\ 0 & \sin \alpha_i & \cos \alpha_i & d_i \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (3)$$

Then, Eq. (4) is applied to obtain the general transformation matrix of the entire kinematics model.

$$T_{\text{end of vector}}^{1st \text{ link}} = T_1^0 + T_2^1 + T_3^2 + T_4^3 + T_5^4 + T_6^5. \quad (4)$$

The resulting matrix, shown in Eq. (5), contains the rotation and translation components of the NAO robot's arm with respect to the lower part of its torso (Fig. 13). The system outputs correspond to P_x , P_y , and P_z .

$$T_n^0 = \begin{bmatrix} n_x & 0_x & a_x & P_x \\ n_y & 0_y & a_y & P_y \\ n_z & 0_z & a_z & P_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (5)$$

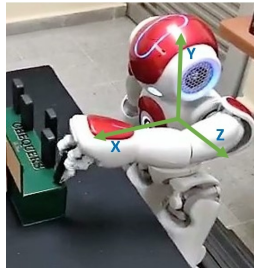


Figure 13. Reference coordinate system on the NAO robot's arm
Source: Authors

For this specific case, the Eqs. for each coordinate axis are as follows.

For the x -coordinate:

$$P_x = -d_3 \cos(\theta_1) \cos(\theta_2) + d_5 [(-\sin(\theta_1) \sin(\theta_3) + \sin(\theta_2) \cos(\theta_1) \cos(\theta_3)) \sin(\theta_4) - \cos(\theta_1) \cos(\theta_2) \cos(\theta_4)]. \quad (6)$$

For the y -coordinate:

$$P_y = -d_3 \sin(\theta_1) \cos(\theta_2) + d_5 [(\sin(\theta_1) \sin(\theta_2) \cos(\theta_3) + \sin(\theta_3) \cos(\theta_1) \sin(\theta_4)) - \sin(\theta_1) \cos(\theta_2) \cos(\theta_4)]. \quad (7)$$

For the z -coordinate:

$$P_z = d_3 \sin(\theta_2) + d_5 (\sin(\theta_2) \cos(\theta_4) + \sin(\theta_4) \cos(\theta_2) \cos(\theta_3)). \quad (8)$$

Complete system and user interface

Once all the systems had been developed, they were put together in a single program using the Naoqi (Python) framework. Naoqi was selected because it provides all the libraries that allow access to the elements of the robot and its language recognition and processing functions. The latter are very important, as voice commands indicate the start and the end of the NAO robot's turn. The complete system takes less than 10 seconds to decide on and perform a play.

An interface to visualize the joint-system process of each play was devised using the Python Tkinter tool and an external computer. Through it, the visualization, processing, and decision system outputs could be observed (Fig. 14). The execution of the play, *i.e.*, the movements made to reach a tile, were visualized using the Coreographe software, which is part of the NAO robot's tools.

The interface was designed with two modes of visualization: (i) the process can be observed step by step, and (ii) the operation of each subsystem can be continuously visualized in an actual dominoes game.

Results and discussion

The testing setup comprised nine games of dominoes in a three-player variant (two humans, one robot), where the NAO robot interacted in a certain number of moves. If the move performed by the robot was incorrect, the conditions of the game before it were restored, and the interaction was repeated. For testing, the NAO robot was meant to interact in the most challenging situations, which is why the human players knew all the tiles in hand.

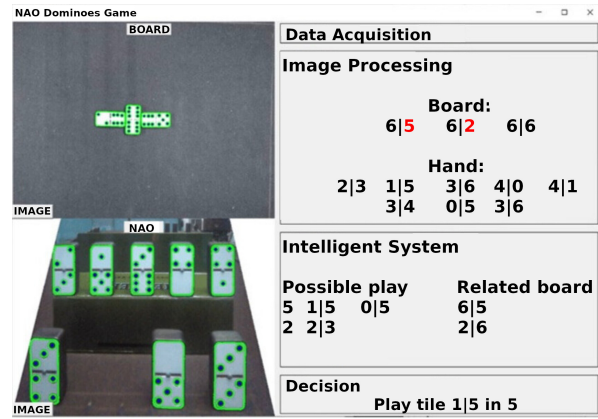


Figure 14. Complete system interface
Source: Authors

This section presents the data collected during the nine games played, classifying the errors reported by the vision, decision, and execution systems and counting the number of successful and unsuccessful plays for the complete system. A confidence interval of 90% was used in all cases.

Vision system

Table 2 lists the number of times that a play was incorrect due to bad image processing regarding the game tiles on the board (B) and in hand (H), *i.e.*, when the tile was incorrectly detected, resulting in an invalid move. Note that failures in the vision system were attributed to the image processing algorithms used. This table also shows the calculated percent error for each game and a total accumulated error after the nine games.

The number of times that there was an error when viewing the tiles on the board and in hand (X) was 5, and the total number of moves (n) was 89. Therefore, the value of the proportion (p) was 0,0562. Knowing the value of $Z_{0.95}$, an interval of 0.0562 ± 0.0402 was obtained.

Table 2. Results obtained for the vision, execution, and decision systems

Game	Moves	Error Vision			Execution		Decision	
		B	H	%	#	%	#	%
1	10	0	0	0	1	10	0	0
2	10	0	0	0	1	10	0	0
3	11	1	0	9.09	0	0	1	9.09
4	9	0	1	11.11	0	0	1	11.11
5	9	0	0	0	0	0	1	11.11
6	11	1	0	9.09	1	9.09	0	0
7	10	1	0	10	2	20	0	0
8	10	1	0	10	1	10	0	0
9	9	0	0	0	1	11.11	0	0
Total	89	4	1	5.62	7	7.87	3	3.37

Source: Authors

According to the results obtained, it can be said that the probability that the robot's vision system makes a mistake during a play is between 1.6 and 9.6%, with 90% confidence. Depending on the development of a full three-players game, the robot makes between 9 and 11 moves, which indicates that the robot can make vision-related mistakes at most once (if any). This error may be due to external factors, such as variations in the external lighting of the game scenario.

Decision system

For this system, the errors listed not only correspond to erroneous moves, but also to failures in selecting the most suitable tile at a given moment in the game. Table 2 shows the calculated percent error for each game and a cumulative total after the nine games.

The number of times that there was an error related to the intelligent system's decision on the tile to be played (X) was 3, and n had the same value: 89. Therefore, the value of the proportion (p) was 0.0337. Knowing the value of $Z_{0.95}$, an interval of 0.0337 ± 0.0315 was obtained.

According to the results, it can be said that the probability that the robot makes a decision-related mistake during a play is between 0.2 and 6.5%, with 90% confidence.

This subsystem's error probability is very low, which indicates that errors only occur in very specific cases. As previously described, since tile selection is entrusted to a decision tree that contains hierarchically organized game strategies, the robot may make an incorrect decision if the optimal play does not follow the hierarchy established with the help of highly experienced players.

Execution system

For this system, the errors listed correspond to the times that the selected tile was not taken or was incorrectly placed on the board. Table 2 shows the calculated percent error for each game and a cumulative total after the nine games.

The number of times that the robot made a mistake while taking the selected tile (X) was 7, and n was 89. Hence, the value of the proportion (p) was 0.0787. Knowing the value of $Z_{0.95}$, an interval of 0.0787 ± 0.0469 was obtained.

According to the results, it can be stated that the probability that the robot will not take the selected tile due to its kinematics model is between 3.2 and 12.6%, with 90% confidence.

Complete system

For the complete system, all the plays, both successful and unsuccessful, were counted, and the total performance of the system was calculated. Table 3 shows the results obtained by calculating the robot's efficiency in each game, as well as a cumulative total after the nine games.

Table 3. Results obtained for the complete system

Game	Moves	Move		% Efficiency
		Successful	Unsuccessful	
1	10	9	1	90
2	10	9	1	90
3	11	9	2	81.82
4	9	7	2	77.78
5	9	8	1	88.89
6	11	9	2	81.82
7	10	7	3	70
8	10	8	2	80
9	9	8	1	88.89
Total	89	74	15	83.15

Source: Authors

For the complete system, the number of times that the robot made successful plays when interacting with the board game

(X) was 74, and the total number of moves (n) was 89. Therefore, the value of the proportion (p) was 0.8315. As with the subsystems, knowing the value of $Z_{0.95}$, an interval of 0.8315 ± 0.0653 was obtained.

According to the results, it can be said that the probability that the robot makes a successful move during a game of dominoes is between 76.6 and 89.7%, with 90% confidence.

The correct operation of the complete system is contingent upon an absence of errors in its subsystems. Therefore, correctly processing the captured images, deciding on the optimal tile to be played, and properly taking and delivering the tile are a necessity. In this vein, the probability that the robot makes a successful move each time it interacts with the game is approximately eight to nine times out of 10.

The findings of this study are consistent with previous research on HRI in games. The proposed approach aligns other studies that have employed user-centered robot designs to improve the performance of HRI systems [7]. For example, the results of refining the vision system [13, 14, 15], improving the robot's decision-making [19, 18], and experimenting with different players and humanoid robots [23, 24]. Overall, this study contributes to the growing body of research on HRI entertainment (specifically, board games) and highlights the potential for humanoid robots to interact with humans in various settings [27, 28]. According to the experimental results, further research is needed to refine the system and improve its performance in real-world settings while considering the complexity of the game, the robot's level of autonomy, and the quality of the algorithms used.

Conclusions

This paper provides an overview of research on the use of humanoid robots in a variety of fields, including education, therapy, and entertainment. One of the most important contributions of this study is the description of an intelligent system for playing dominoes with a NAO robot. The system includes image processing algorithms for detecting and identifying tiles on the board and in hand, a decision tree algorithm for determining the best move, and a forward kinematics model for physically moving the robot's arm to perform the selected move. The system also includes a user interface for process visualization and monitoring.

The proposed system successfully enables a humanoid robot to interact with people by executing plays in a game of dominoes. The case studies considered include actual game situations, in which the robot correctly interacts with other players based on conditions provided as the game progresses. Thanks to the proposed approach, the robot can select the most appropriate tile, which entails a friendlier HRI. The proper operation of the entire system is dependent on the absence of errors in its subsystems. Thus, correctly processing the captured images, determining the most tile to be played, and properly taking and placing tiles on the board are required.

According to the findings and the analysis presented in this paper, the vision system has an average error of 5.62%, the decision system reports 3.37%, and the value for the execution system is 7.87%. When all the components function properly, the NAO robot can complete successful plays at an 83% rate. Each HRI-enabled system was validated and comprised various subsystems that executed particular

functions based on the application's requirements, with each subsystem's error being minimized to enhance overall system performance.

The vision system of a robot can be affected by various factors such as changes in external lighting conditions, player hand occlusions, or quick movements, which can result in image processing errors. Additionally, hardware or software limitations can further impair the robot's ability to accurately interpret and respond to visual cues, which can increase the likelihood of errors.

The decision system helps to make optimal plays with its structured decision tree framework, but deviations from predefined strategies are sometimes necessary due to unforeseen game states or unexpected player actions. Relying solely on historical data or predefined rules may overlook nuanced situational factors, leading to suboptimal decisions. To maximize success, strategies must be adapted accordingly.

The execution system is vulnerable to mechanical issues, such as imprecise motor control or workspace obstructions, that could compromise its reliability. Furthermore, communication delays between subsystems may hinder the timely execution of planned actions, leading to suboptimal gameplay performance. It is crucial to consider these factors for an optimal system performance.

After analyzing various factors, it is clear that the NAO robot's ability to play dominoes is dependent on addressing a number of technical challenges and environmental variables. By thoroughly assessing these factors and their impact on subsystem functionality, we can develop specific strategies to improve the robot's overall reliability and reduce errors during gameplay.

Finally, this study demonstrates the potential of humanoid robots in promoting and improving HRI in various settings. The development of intelligent systems that can interact with people and plan and execute their moves in board games represents a significant advancement in this field of study. The proposed method demonstrates that the robot can successfully play a board game with people by combining image processing, decision-making, and physical movement systems, making the HRI more effective. Further research could improve the system's performance by refining the vision system, experimenting with different players and humanoid robots, and exploring other intelligent algorithms.

CRedit author statement

Conceptualization: Alex Medina and Christian G. Quintero M.. Data curation: Alex Medina, Daniela Charris, and Christian G. Quintero M.. Formal analysis: Alex Medina, Daniela Charris, Mauricio Pardo, and Christian G. Quintero M.. Funding acquisition: Alex Medina and Christian G. Quintero M.. Investigation: Alex Medina, Daniela Charris, Mauricio Pardo, and Christian G. Quintero M.. Methodology: Mauricio Pardo and Christian G. Quintero M.. Project administration: Christian G. Quintero M.. Resources: Mauricio Pardo. Supervision: Mauricio Pardo and Christian G. Quintero M.. Validation: Daniela Charris and Mauricio Pardo. Visualization: Mauricio Pardo. Writing - original draft: Alex Medina, Daniela Charris, Mauricio Pardo, and Christian G. Quintero M.. Writing - review and editing:

Alex Medina, Daniela Charris, Mauricio Pardo, and Christian G. Quintero M..

References

- [1] J. Barnes, S. FakhrHosseini, E. Vasey, J. Ryan, C. Park, and M. Jeon, "Promoting STEAM education with child-robot musical theater," in *2019 14th ACM/IEEE Int. Conf. Human-Robot Interact. (HRI)*, 2019, pp. 366. <https://doi.org/10.1109/HRI.2019.8673311>
- [2] W. Budiharto, A. D. Cahyani, P. C. B. Rumondor, and D. Suhartono, "EduRobot: intelligent humanoid robot with natural interaction for education and entertainment," *Procedia Comp. Sci.*, vol. 116, pp. 564-570, 2017. <https://doi.org/10.1016/j.procs.2017.10.064>
- [3] M. Jeon et al., "Robot opera: A modularized afterschool program for STEAM education at local elementary school," in *2017 14th Int. Conf. Ubiquitous Robots Ambient Intel. (URAI)*, 2017, pp. 935-936. <https://doi.org/10.1109/URAI.2017.7992869>
- [4] Z. Bi, Q. Yang, M. Zhang, A. Goupil, and L. Feng, "Accurate football detection and localization for NAO robot with the improved HOG-SVM approach," in *2018 Chinese Autom. Cong. (CAC)*, 2018, pp. 567-571. <https://doi.org/10.1109/CAC.2018.8623727>
- [5] E. Knox and K. Watanabe, "AIBO robot mortuary rites in the Japanese cultural context," in *2018 IEEE/RSJ Int. Conf. Intel. Robots Syst. (IROS)*, 2018, pp. 2020-2025. <https://doi.org/10.1109/IROS.2018.8594066>
- [6] K. Chin, C. Wu, and Z. Hong, "A humanoid robot as a teaching assistant for primary education," *2011 5th Int. Conf. Genetic Evolut. Comp.*, 2011, pp. 21-24. <https://doi.org/10.1109/ICGEC.2011.13>
- [7] T. Asfour et al., "ARMAR-6: A collaborative humanoid robot for industrial environments," in *2018 IEEE-RAS 18th Int. Conf. Humanoid Robots (Humanoids)*, 2018, pp. 447-454. <https://doi.org/10.1109/HUMANOIDS.2018.8624966>
- [8] Z. Barakeh, S. Alkork, A. Karar, S. Said, and T. Beyrouthy, "Pepper humanoid robot as a service robot: A customer approach," in *2019 3rd Int. Conf. Bio-Eng. Smart Tech. (BioSMART)*, 2019, pp. 1-4. <https://doi.org/10.1109/BIOSMART.2019.8734250>
- [9] O. Ajani, H. Obasekore, B. Kang, and M. Rammohan, "Robotic assistance in radiology: A COVID-19 scenario," *IEEE Access*, vol. 11, pp. 49785-49793, 2023. <https://doi.org/10.1109/ACCESS.2023.3277526>
- [10] M. Mohan and K. Kuchenbecker, "A design tool for therapeutic social-physical human-robot interactions," in *2019 14th ACM/IEEE Int. Conf. Human-Robot Interact. (HRI)*, 2019, pp. 727-729.
- [11] B. Schadenberg, "Predictability in human-robot interactions for autistic children," in *2019 14th ACM/IEEE Int. Conf. Human-Robot Interact. (HRI)*, 2019, pp. 748-750. <https://doi.org/10.1109/HRI.2019.8673135>
- [12] F. Sun, C. Wang, T. Zheng, and H. Liu, "An improved object detection method based on NAO robot," in *2023 IEEE 3rd Int. Conf. Power*

- Elec. Comp. App. (ICPECA)*, 2023, pp. 1184-1188. <https://doi.org/10.1109/ICPECA56706.2023.10076034>
- [13] H. Zhu, H. Yi, R. Chellali, and L. Feng, "Object recognition and localization algorithm based on NAO robot," in *2018 27th IEEE Int. Symp. Robot Human Interac. Comm. (RO-MAN)*, 2018, pp. 483-486. <https://doi.org/10.1109/ROMAN.2018.8525656>
- [14] L. Juang and J. Zhang, "Visual tracking control of humanoid robot," *IEEE Access*, vol. 7, pp. 29213-29222, 2019. <https://doi.org/10.1109/ACCESS.2019.2901009>
- [15] Q. Yan, S. Li, C. Liu, M. Liu, and Q. Chen, "RoboSeg: Real-time semantic segmentation on computationally constrained robots," *IEEE Trans. Syst. Man Cyber. Syst.*, vol. 52, pp. 1567-1577, 2022. <https://ieeexplore.ieee.org/document/9259011>
- [16] K. Lobos-Tsunekawa, F. Leiva, and J. Solar, "Visual navigation for biped humanoid robots using deep reinforcement learning," *IEEE Robotics Automat. Lett.*, vol. 3, pp. 3247-3254, 2018. <https://doi.org/10.1109/LRA.2018.2851148>
- [17] E. Ovalle-Magallanes *et al.*, "Transfer learning for humanoid robot appearance-based localization in a visual map," *IEEE Access*, vol. 9, pp. 6868-6877, 2021. <https://doi.org/10.1109/ACCESS.2020.3048936>
- [18] S. Goenaga, L. Navarro, C. Quintero M., and M. Pardo, "Imitating human emotions with a NAO robot as interviewer playing the role of vocational tutor," *Electronics*, vol. 9, art. 971, 2020. <https://doi.org/10.3390/electronics9060971>
- [19] S. Díaz, J. Shaik, J. Santofimio, and C. Quintero M., "Intelligent execution of behaviors in a NAO robot exposed to audiovisual stimulus," in *2018 IEEE 2nd Colombian Conf. Robotics Automat. (CCRA)*, 2018, pp. 1-6. <https://doi.org/10.1109/CCRA.2018.8588149>
- [20] H. Cao *et al.*, "Robot-assisted joint attention: A comparative study between children with autism spectrum disorder and typically developing children in interaction with NAO," *IEEE Access*, vol. 8, pp. 1-1, Dec. 2020. <https://doi.org/10.1109/ACCESS.2020.3044483>
- [21] M. Feidakis, I. Gkolompia, A. Marnelaki, K. Marathaki, S. Emmanouilidou, and E. Agrianiti, "NAO robot, an educational assistant in training, educational and therapeutic sessions," in *2023 IEEE Global Eng. Educ. Conf. (EDUCON)*, 2023, pp. 1-6. <https://doi.org/10.1109/EDUCON54358.2023.10125229>
- [22] X. Yang, M. Shyu, H. Yu, S. Sun, N. Yin, and W. Chen, "Integrating image and textual information in human-robot interactions for children with autism spectrum disorder," *IEEE Trans. Multimedia*, vol. 21, no. 3, pp. 746-759, 2019. <https://doi.org/10.1109/TMM.2018.2865828>
- [23] Y. Wei, "A comprehensive approach to the generation of human-like arm movements on robot NAO," *IEEE Access*, vol. 8, pp. 172869-172881, 2020. <https://doi.org/10.1109/ACCESS.2020.3025532>
- [24] R. Zhai, S. Wen, J. Zhu, and G. Guo, "Trajectory planning of NAO robot arm based on target recognition," in *2017 Int. Conf. Adv. Mechatronic Syst. (ICAMEchS)*, 2017, pp. 403-407. <https://doi.org/10.1109/ICAMEchS.2017.8316507>
- [25] S. Piperakis, M. Koskinopoulou, and P. Trahanias, "Nonlinear state estimation for humanoid robot walking," *IEEE Robotics Automat. Lett.*, vol. 3, no. 4, pp. 3347-3354, 2018. <https://doi.org/10.1109/LRA.2018.2852788>
- [26] Y. Wang, X. Xue, and B. Chen, "Matsuoka's CPG with desired rhythmic signals for adaptive walking of humanoid robots," *IEEE Trans. Cyber.*, vol. 50, pp. 613-626, 2020. <https://doi.org/10.1109/TCYB.2018.2870145>
- [27] T. Lin, S. Ng, and S. Sebo, "Benefits of an interactive robot character in immersive puzzle games," in *2022 31st IEEE Int. Conf. Robot Human Interac. Comm. (RO-MAN)*, 2022, pp. 37-44. <https://doi.org/10.1109/RO-MAN53752.2022.9900828>
- [28] T. Inoue, F. Jimenez, T. Haruta, and M. Oonuki, "Effect of impression on learners during partnered robots learning programming while playing board games against each other," in *2022 Joint 12th Int. Conf. Soft Comp. Intel. Syst. and 23rd Int. Symp. Adv. Intel. Syst. (SCIS&ISIS)*, 2022, pp. 1-4. <https://ieeexplore.ieee.org/document/10001985>
- [29] V. Moya, E. Slawiński, V. Mut, and B. Wagner, "Intercontinental bilateral-by-phases teleoperation of a humanoid robot," *IEEE Latin Amer. Trans.*, vol. 20, pp. 64-72, May 2021. <https://doi.org/10.1109/TLA.2022.9662174>
- [30] M. Mercier and T. Lubart, "The effects of board games on creative potential," *J. Creat. Behavior*, vol. 55, Feb. 2021. <https://doi.org/10.1002/jocb.494>
- [31] P. Kołosowski, A. Wolniakowski, and K. Miatliuk, "Collaborative robot system for playing chess," in *2020 Int. Conf. Mechatronic Sys. Mat. (MSM)*, 2020, pp. 1-6. <https://doi.org/10.1109/MSM49833.2020.9202398>
- [32] D. Patil, D. Fegade, A. Kadam, P. Patil, and N. Singhaniya, "A novel framework for robotic chess," in *2021 2nd Int. Smart Elec. Comm. (ICOSEC)*, 2021, pp. 1-6. <https://doi.org/10.1109/ICOSEC51865.2021.9591961>
- [33] P. Rath, N. Mahapatro, P. Nath, and R. Dash, "Autonomous chess playing robot," in *2019 28th IEEE Int. Conf. Robot Human Interac. Comm. (RO-MAN)*, 2019, pp. 1-6. <https://doi.org/10.1109/RO-MAN46459.2019.8956389>
- [34] L. Juang, "Humanoid robots play chess using visual control," *Multimedia Tools App.*, vol. 81, pp. 1-22, Jan. 2022. <https://doi.org/10.1007/s11042-021-11636-y>
- [35] C. Li, E. Imeokparia, M. Ketzner, and T. Tsahai, "Teaching the NAO robot to play a human-robot interactive game," in *2019 Int. Conf. Comp. Sci. Comp. Intel. (CSCI)*, 2019, pp. 712-715. <https://doi.org/10.1109/CSCI49370.2019.00134>
- [36] N. Karunanayake *et al.*, "Towards a smart opponent for board games: Learning beyond simulations," in *2020 IEEE Int. Conf. Syst. Man Cyber. (SMC)*, 2020, pp. 1943-1950. <https://doi.org/10.1109/SMC42975.2020.9283458>

- [37] E. Karmanova, V. Serpiva, S. Perminov, A. Fedoseev, and D. Tsetserukou, "SwarmPlay: Interactive tic-tac-toe board game with swarm of nano-UAVs driven by reinforcement learning," in *2021 30th IEEE Int. Conf. Robot Human Interac. Comm. (RO-MAN)*, 2021, pp. 1269-1274. <https://doi.org/10.48550/arXiv.2108.01593>
- [38] J. Hu, F. Zhao, J. Meng, and S. Wu, "Application of deep reinforcement learning in the board game," in *2020 IEEE Int. Conf. Info. Tech. Big Data Art. Intel. (ICIBA)*, vol. 1, 2020, pp. 809-812. <https://doi.org/10.1109/ICIBA50161.2020.9277188>
- [39] A. Raghavan, H. Kumar M. B., R. Srinivasan, S. Dey, and T. Chandar, "Automated laser alignment and image processing-based robotic carrom player," in *2021 20th Int. Conf. Adv. Robotics (ICAR)*, 2021, pp. 499-504. <https://doi.org/10.1109/ICAR53236.2021.9659332>
- [40] P. Tres, "Estudio e implementación de algoritmos de resolución del juego del dominó para un robot antropomórfico," M.S. thesis, Universidad Politécnica de Cataluña, Spain, 2014.
- [41] N. Lestriandoko and R. Sadikin, "Circle detection based on Hough transform and Mexican Hat filter," in *2016 Int. Conf. Control Infor. App. (IC3INA)*, 2016, pp. 153-157. <https://doi.org/10.1109/IC3INA.2016.7863041>
- [42] N. Kofinas, E. Orfanoudakis, and M. Lagoudakis, "Complete analytical inverse kinematics for NAO," in *2013 13th Int. Conf. Autonom. Robot Syst.*, 2013, pp. 1-6. <https://doi.org/10.1109/Robotica.2013.6623524>

Conecta2: An Android App for Communication between Disaster Survivors through an *ad hoc* Network

Conecta2: una aplicación en Android para la comunicación entre sobrevivientes de desastres a través de una red *ad hoc*

María del Pilar Salamanca¹, Elio H. Cables², and Juan Camilo Ramírez³

ABSTRACT

Natural and man-made disasters often render network infrastructure inoperative, severely hampering communication between rescue workers and survivors. While smartphone-based *ad hoc* networks are frequently proposed as a solution, the implementation of mobile applications optimized for disaster scenarios remains a challenge. This study aims to address this gap by developing Conecta2, a prototype Android application for communication in post-disaster situations. The application establishes an *ad hoc* network using Wi-Fi Direct or Bluetooth in order to enable communication between nearby smartphones without relying on regular network infrastructure. Conecta2 allows users to transmit text, photos, and audio, as well as to estimate the location of nearby devices running the same software. Experimental tests were conducted to evaluate the performance of the application in simulated disaster scenarios. Results demonstrate that Conecta2 effectively facilitates communication between smartphones in situations analogous to disaster aftermath. The ability of the application to locate survivors via their smartphones also aids search and rescue efforts. In conclusion, Conecta2 offers a viable communication alternative for post-disaster scenarios, potentially improving the effectiveness of disaster response and survivor location efforts.

Keywords: peer-to-peer communication, Wi-Fi Direct, Bluetooth, emergency communication, disaster response

RESUMEN

Los desastres naturales y provocados por el hombre a menudo dejan inoperante la infraestructura de red, lo que dificulta gravemente la comunicación entre los trabajadores de rescate y los sobrevivientes. Si bien las redes *ad hoc* basadas en teléfonos inteligentes se proponen con frecuencia como una solución, la implementación de aplicaciones móviles optimizadas para escenarios de desastre sigue siendo un desafío. Este estudio tiene como objetivo abordar esta brecha mediante el desarrollo de Conecta2, un prototipo de aplicación Android para la comunicación en situaciones posteriores a un desastre. La aplicación establece una red *ad hoc* utilizando Wi-Fi Direct o Bluetooth, en aras de permitir la comunicación entre teléfonos inteligentes cercanos sin depender de la infraestructura de red regular. Conecta2 permite a los usuarios transmitir texto, fotos y audio, así como estimar la ubicación de dispositivos cercanos que ejecutan el mismo *software*. Se realizaron pruebas experimentales para evaluar el rendimiento de la aplicación en escenarios de desastre simulados. Los resultados demuestran que Conecta2 facilita eficazmente la comunicación entre teléfonos inteligentes en situaciones análogas a las posteriores a un desastre. La capacidad de la aplicación para localizar sobrevivientes a través de sus teléfonos inteligentes también ayuda en los esfuerzos de búsqueda y rescate. En conclusión, Conecta2 ofrece una alternativa de comunicación viable para escenarios post-desastre, mejorando potencialmente la efectividad de la respuesta al desastre y los esfuerzos de localización de sobrevivientes.

Palabras clave: comunicación *peer-to-peer*, Wi-Fi Direct, Bluetooth, comunicación de emergencia, gestión de desastres

Received: August 21st 2023

Accepted: December 5th 2024

Introduction

Reduced network connectivity due to infrastructure damage caused by natural disasters hinders communication between victims and rescue crew members. The main driver of this issue is the static nature of the cellular network, which makes communication efforts vulnerable to this type of emergency. Examples of this include the Turkey-Syria earthquake in 2023, with many survivors found by rescue services several days after the incident [1, 2]; the massive destruction of infrastructure reported in the aftermath of the 2011 earthquake in Japan [3, 4]; the cellular networks rendered inoperative by massive floods during Hurricane Sandy in 2012 [5]; and the network collapse observed after the Haiti earthquake in 2010 [6]. Among the most crucial issues faced by disaster response efforts, communication constraints have been reported, a problem that is especially critical during the first 72 hours following the catastrophe, as

the probability of finding survivors decreases considerably thereafter [7, 8].

Smartphone-based *ad hoc* networks are commonly explored as a solution to this problem [8, 9, 10, 11], given the ubiquity and extended use of these mobile devices among the general public [12]. Consequently various paradigms have emerged for the configuration of such *ad hoc* networks, with varying degrees of applicability depending on the scenario

¹PhD in Engineering, Universidad de los Andes, Colombia. Affiliation: Associate professor Facultad de Ingeniería de Sistemas, Universidad Antonio Nariño, Colombia. E-mail: mpsalamanca@uan.edu.co

²PhD in Informatics, Universidad de Granada, Spain. Affiliation: Professor Facultad de Ingeniería de Sistemas, Universidad Antonio Nariño, Colombia. E-mail: ehcables@uan.edu.co

³PhD in Computer Science, University of Sheffield, United Kingdom. Affiliation: Associate professor, Facultad de Ingeniería de Sistemas, Universidad Antonio Nariño, Colombia. E-mail: juan.ramirez@uan.edu.co



Attribution 4.0 International (CC BY 4.0) Share - Adapt

[13, 14]. Even though Bluetooth was initially used for this purpose, the use of Wi-Fi Direct technology has gained momentum in the implementation of *ad hoc* networks for disaster scenarios, as this standard has become the *de facto* technology for communication between modern devices [13, 14]. Furthermore, when using this standard, power consumption derived from data transfer is notoriously lower than that of the cellular network [8], which makes Wi-Fi Direct more desirable in emergency situations with seriously limited power supply. This approach also has challenges, however, given that both Wi-Fi Direct and Bluetooth only enable communication between devices in close proximity of each other, with the coverage of the latter being reportedly shorter.

Regardless of the technology in use, interoperability, *i.e.*, any software's ability to operate correctly in devices from different manufacturers, is a crucial ingredient for any integral application of *ad hoc* networks in disaster scenarios. Nevertheless, this feature remains limited, given the existing variety of paradigms for these networks. Furthermore, research on this topic has mainly focused on simulated scenarios, rather than on real implementations. Moreover, traditional approaches to implementing these networks frequently rely on gaining *root* access to the device, which most manufacturers disable by default due to security concerns, making solutions non-portable [15, 16].

This article presents *Conecta2*, a prototype application for Android that allows for communication between the devices where it is installed, in the absence of network infrastructure. The application offers both *group* and *one-to-one* communication through the automated configuration of an emergency *ad hoc* network, using either Wi-Fi Direct or Bluetooth. *Conecta2* also includes a module to locate nearby smartphones, provides a snapshot of the current layout of such devices, and it is characterized by an easy-to-use user interface, guaranteeing that a user without technical skills can employ it without inconvenience. *Conecta2* has been successfully tested in devices with Android versions 10.0 and 11.0 and should nevertheless run smoothly in smartphones with versions 5.0 and above.

The contributions of this article are as follows:

- A prototype application for Android named *Conecta2*, which allows any survivor to collaborate with or request aid from other connected users (survivors or rescue personnel). The most remarkable features of *Conecta2* are its simple user interface and the possibility of using either Wi-Fi Direct or Bluetooth to connect with nearby smartphones.
- A performance evaluation of the application modules, aimed at characterizing the coverage of the networks deployed with each communication technology. Such characterization is based on field measurements taken from real scenarios with and without obstacles.

This paper is organized as follows. The next section summarizes recent works on mobile applications based on the analyzed technologies, which have been designed to allow survivors to communicate in the aftermath of a disaster. Then, the three modules of *Conecta2* are described: *Group communication* (based on Bluetooth technology), *One-to-one communication* (based on Wi-Fi

Direct), and *Device location* (based on Wi-Fi Direct and Global Positioning System (GPS) technology). Thereafter, the results (based on field measurements) regarding the network performance of *Conecta2* are presented, as well as the conclusions of this work.

Related work

The use of Wi-Fi Direct and Bluetooth for the automated configuration of *ad hoc* networks, including their use in disaster scenarios, is not new and is featured in several proposals. However, most of the scientific literature in this regard focuses on the design of new protocols, topologies, paradigms, and technologies, while significantly fewer studies aim specifically at the development of mobile applications as a solution to the problem of communication between survivors following a natural disaster. Furthermore, the use of Bluetooth is remarkably less frequent than that of Wi-Fi Direct, possibly due to the fact that the coverage of the latter has been reported to greatly exceed that of the former. In the literature, the performance of proposals with both technologies is generally evaluated in terms of various metrics. However, the most critical and frequently considered are the message delivery rate and the associated energy consumption efficiency. The former is clearly fundamental, since it directly addresses the main problem at hand, and the latter is also crucial, given that survivors' access to power sources following a natural disaster is likely to be seriously limited due to infrastructure damage. This evaluation is generally conducted through simulations or field experiments, and sometimes using a combination of the two. The use of Bluetooth in the development of mobile apps for emergency *ad hoc* networks is notably lower than that of Wi-Fi Direct, especially in the literature published within the last five years (as of October 2024), probably due to the former's reportedly reduced range, which makes it less attractive in this context. This may be the reason why the most recently published studies investigating the applicability of mobile *ad hoc* networks using Bluetooth in disaster situations generally do so in combination with Wi-Fi Direct or other technologies [15, 16, 14, 13].

Even though there is a diversity of messaging applications for Android that offer offline communication through either Wi-Fi Direct or Bluetooth (or both) without the need for conventional network coverage, many of these have been designed to serve purposes other than aiding the location and rescue of survivors after a natural disaster. Applications such as Bridgefy* and Serval Mesh have been developed as alternatives to conventional messaging services, *e.g.*, WhatsApp and Facebook Messenger, when Internet connectivity is unavailable. Other applications have been designed primarily to address privacy concerns, especially those encountered by individuals facing persecution, such as journalists and activists. One example of such applications is Briar, which encrypts and syncs messages between users via Wi-Fi Direct and Bluetooth, as well as through the Tor network[§] when an Internet connection is available, in order to deliver messages to a wider range. It deliberately lacks

*Bridgefy (<https://bridgefy.me>).

†Serval Mesh (<https://www.servalproject.org/>).

‡Briar (<https://briarproject.org>).

§Tor (<https://www.torproject.org>).

a central server, delegating all authentication and storage tasks to each device running the software.

BeWare is an Android application developed to establish an emergency *ad hoc* network between smartphones using both Wi-Fi Direct and Wi-Fi legacy, the latter for multi-hop communication between groups formed by the former [17]. An evaluation was conducted in an indoor space using mobile nodes, each corresponding to one model from various manufacturers, *i.e.*, Samsung Galaxy S7, Samsung Galaxy S5, Xperia Z2, and Huawei P8. Performance was evaluated using Android Profiler[†] for measuring both network and energy efficiency, Android Debug Bridge[‡] to assess packet loss, and Android Studio IDE^{**} to measure the time required by the app to form/update groups and deliver messages between nodes. These experiments show, among other results, that group forming times increase with the number of nodes involved, and that the most critical performance metrics achieved by BeWare exceed those of previous related solutions found in the literature. Nevertheless, the authors acknowledge that further experimental work in more realistic scenarios is needed, particularly with the inclusion of obstacles, in order to more thoroughly assess the usability of the app in real life situations.

In addition to the above-presented approaches, various proposals can be found in the literature which seek to improve the protocols and algorithms involved in the automated configuration of *ad hoc* networks in a variety of situations, not only disaster scenarios, although they do not necessarily contemplate the development of a mobile app designed for the average user. [18], for instance, propose a novel method for the configuration of a multi-hop *ad hoc* network that first divides the disaster area into hexagonal virtual cells and then enables the formation of various Wi-Fi Direct groups in each cell. Each group exists for a limited period of time, after which the cells are reallocated and new groups are formed, thus allowing each node to change its membership from one group to another and relay information from the former to the latter. A preliminary evaluation was conducted by the authors *via* simulations, showcasing the system's effective information dissemination. Yet, the authors acknowledge that further experimental work is required, especially involving the evaluation of the information diffusion model and its performance in a realistic scenario. It is worth adding that this system is preliminary and not based on a single mobile app that an average user can use without technical expertise. In the same vein, COPE is a recently proposed scheme (not a mobile app) designed to provide an emergency communications channel for disaster survivors through the integration of various network technologies, including Wi-Fi Direct and Bluetooth, each representing a layer in a multi-tier architecture. In this scheme, devices are grouped by tier and thus by their corresponding coverage range and energy consumption. Additionally, the scheme contemplates the use of drones to complement communication efforts by scanning the area and relaying messages among nodes that are mutually out of range [19]. This approach was evaluated *via* simulations, showing effective message transmission between nodes in various disaster scenarios through self-organizing routing by the drones while optimizing energy consumption by the connected devices. Each simulation was executed 10 times, and the performance metrics



Figure 1. Architecture of Conecta2

Source: Authors

(namely energy management efficiency and message relay efficacy) were averaged over all repetitions. No field experiments were conducted. Despite these positive results, this approach does not involve a single app that can be installed and used by an average user, and it relies strongly on the use of drones in order to achieve optimal communication.

Other initiatives using either Wi-Fi Direct or Bluetooth include the use of drones for the configuration of the *ad hoc* network [20], novel protocols [21], methods for enhanced connectivity [17], new algorithms for improved energy management, new mobility models, the use of Bluetooth Mesh to enable communication with household devices [22], and other similar proposals aimed mostly at optimizing Wi-Fi Direct or Bluetooth applications [23, 24]. In the case of Bluetooth, similar attempts have been made at establishing *ad hoc* networks for disaster scenarios, albeit without the development of a mobile app that includes the use of novel protocols and topologies using Bluetooth Low Energy [25].

Description of Conecta2

Conecta2 is the first prototype of an *ad hoc* network-based emergency messaging application for Android smartphones that comprises three modules: a group communication module, a one-to-one communication module, and a module to locate nearby smartphones, which was developed using the architecture displayed in Fig. 1 and publicly accessible community libraries^{§§}. The first time Conecta2 is opened, the application asks the user for a device name that will identify the smartphone in the network. Once the user enters the requested information, the application switches to the main screen, illustrated in Fig. 2, where the user may choose one of the three available modules. These are described below.

Module 1: group communication

This Bluetooth-based module allows for group communication with nearby devices, *i.e.*, any message sent by a user will be delivered to all the users connected to the same group. When a user taps on the *Group Communication* button (Fig. 2), the application shows two options: *Create Group* and *Find an Existing Group* (Fig. 3a). If the first option is selected, the application creates a new Bluetooth group

[†]Android Profiler (<https://developer.android.com/studio/profile/android-profiler>).

[‡]ADB Shell (<https://adbshell.com/>).

^{**}Android Studio IDE (<https://developer.android.com/studio>).

^{††}BlueChat (<https://github.com/alexkang/blue-chat>).

^{‡‡}Direct Chat (<https://github.com/0xSG/DirectChat>).

^{§§}android-wifi-direct-chat-app (<https://github.com/blackchalk/android-wifi-direct-chat-app>).

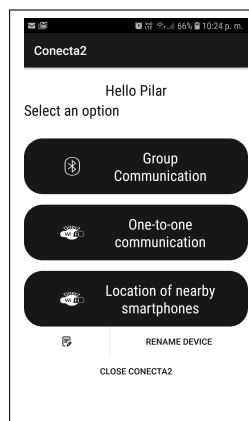
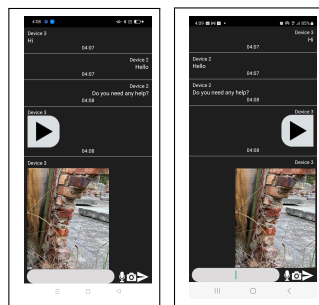


Figure 2. Main screen of Conecta2

Source: Authors



(a) Device 2 chat screen. **(b)** Device 3 chat screen.

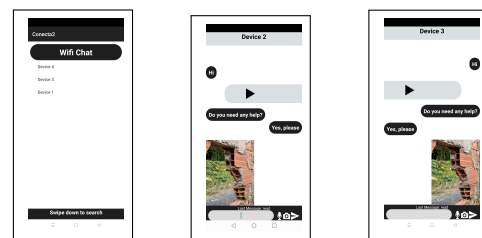
Figure 3. An example conversation between two devices, named Device 2 and Device 3, through the Bluetooth-based group communication module of Conecta2

Source: Authors

and waits for nearby users to join. If the second option is selected instead, the application starts searching for an existing group; when it is found, the joining smartphone issues a request to pair with any other device already in the group. The incoming smartphone joins the group if the pairing request is accepted by both devices. When no group is not found, the application informs the user that there are no nearby groups and that it will return to the main menu of the Bluetooth-based module (Fig. 3a), so that the user can create a new group. Whether the user decides to create a group or join an existing one, the application switches to the chat screen shown in Fig. 3b. The messages aligned to the right correspond to those sent from the current smartphone, while the received messages are aligned to the left. In this screen, a user can send text messages by typing in the blank space located in the bottom line. In addition, the user can tap on the microphone icon to send voice messages or touch the camera icon to send photos to the group. Even though the current prototype's group communication module was developed using only Bluetooth for simplicity in its implementation, future versions are planned which also incorporate Wi-Fi Direct.

Module 2: one-to-one communication

This Wi-Fi Direct-based module was conceived to only allow for communication between any two users, in contrast to group communication. The module is activated when the *One-to-One Communication* button is tapped in the app's



(a) List of nearby devices.

(b) Device 2's chat screen.

(c) Device 3's chat screen.

Figure 4. An example conversation between two devices, named Device 2 and Device 3, through the Wi-Fi Direct-based one-to-one communication module of Conecta2. When the module is opened, a list of nearby devices is shown, from which the user can choose any to communicate with.

Source: Authors

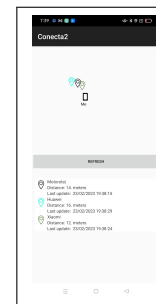


Figure 5. Device location module

Source: Authors

main screen (Fig. 2). In the first screen of the module, shown in Fig. 4a, a list of nearby devices can be obtained by sliding the finger from top to bottom, after which the user can tap on the name of any of the devices listed in order to start a conversation.

Upon selecting a device from the list, the application will show a connection request message in the screen of the selected device. If this incoming connection is accepted, one device will take the role of GO and the other will be the client. Then, the application on both devices will switch to the Wi-Fi Direct-based chat, as illustrated in Fig. 4b. Text and audio messages, as well as photos, can be sent from this screen, similarly to the group communication chat. Nonetheless, this screen has an additional feature: the sending user receives confirmation of whether a sent message was delivered to and read by the receiving user.

Module 3: location of nearby smartphones

This module allows the user to search for nearby smartphones where the application is installed and estimate their location through Android's Network Service Discovery (NSD), advantageously bypassing the need for pairing between devices. Several seconds after tapping on the *Location of Nearby Smartphones* button of the main screen (Fig. 2), an image similar to that shown in Fig. 5 is displayed. The module shows the current smartphone (identified by the word *Me*) as well as nearby devices on the map, based on geographical coordinates exchanged via Wi-Fi Direct. The layout of nearby devices reflects their location at a given time, and it can be refreshed by touching the *Refresh List* button. Since this module needs to start the GPS to read

the coordinates and send them to nearby devices, energy consumption is increased while it is active, so it should be used carefully to avoid battery drain.

Performance evaluation

The performance of Conecta2 was experimentally evaluated in various scenarios, both with and without obstacles, in order to determine the maximum network coverage achieved when using Wi-Fi Direct or Bluetooth, *i.e.*, the group and one-to-one communication modules, respectively. All these experiments were conducted with the conditions listed below.

- Data transmission attempts were repeatedly made between two Android devices running Conecta2 as the distance between them increased by either 10 m (without obstacles) or 5 m (with obstacles), until no communication was possible (*i.e.*, when data were not transmitted or when the devices failed to establish a piconet or a P2P group), or until 100 m were reached.
- At each distance, 10 texts were sent (with 20 characters each), then 10 audios (5 s each), and, finally, 10 photos.
- At each distance, all transmission attempts were conducted first through the group communication module (*i.e.*, Wi-Fi Direct) and then through the one-to-one communication module (*i.e.*, Bluetooth).
- The number of successfully delivered messages, out of the 10 attempts at each distance and with each module, was recorded.

The experiments without obstacles, carried out at the location shown in Fig. 6 and using smartphones with the characteristics listed in Table 1, can be summarized as follows.

Two users, each one with a smartphone running Conecta2, were placed one in front of the other, without objects between them to ensure a direct line of sight. The initial separation between the users was 10 m, and this distance was gradually increased in 10 m intervals. In the first set of experiments, either a piconet or a P2P group was formed when the devices were 10 m apart. Then, after group formation, the distance between smartphones was increased as previously described. At each distance, text and voice messages, as well as photographs, were exchanged between devices. The results obtained for the one-to-one and group communication modules are listed in Tables 2 and 3, respectively, the latter of which includes measurements at a maximum separation of 50 m between the two devices, since communication was not possible at longer distances. At various distances between the two smartphones, *e.g.*, at 30, 60, 90, and 100 m, communication with both modules was occasionally lost while conducting the tests and needed to be reestablished in order to continue the experiment. In one exceptional case, at 30 m, one of the photos appeared to have been partially delivered to the receiving device. More specifically, communication through the group module was lost and unrecoverable at distances longer than 50 m, and an analogous situation was observed with the one-to-one module when the devices were more



Figure 6. Location where the experiments without obstacles were conducted

Source: Authors

than 100 m apart. Nevertheless, both modules exhibited high message transmission rates at all the distances shown in Tables 2 and 3. At a maximum distance of 100 m, all text and audio messages, as well as 80% of the photos, were successfully transmitted via the one-to-one communication module. On the other hand, at a maximum distance of 50 m, all text and audio messages, as well as 60% of the photos, were successfully transmitted via the group communication module. However, in both cases, at greater distances, group formation and pairing were impossible. Therefore, only the distances shown in Table 2 and Table 3 are reported.

On the other hand, the experiments with obstacles were carried out inside a building with the layout illustrated in Fig. 7. The same smartphones were used, with the devices placed in different rooms, separated by brick walls and metal doors, while increasing the distance between them in 5 m intervals. The results obtained for the one-to-one and the group communication modules are listed in Tables 4 and 5, respectively. During these experiments, the communication between the two devices was sporadically lost and needed to be restarted in order to continue. In some cases, messages were received in an order that was different from that in which they had been originally sent. These experiments made it evident that the presence of obstacles, specifically brick walls, notably attenuate the signal. It was observed that, when sending text messages, communication through Wi-Fi Direct was possible with a separation of up to 35 m between sender and receiver, with losses of 60%, whereas Bluetooth allowed for successful transmission at distances of up to 45 m and without data loss. However, this is different when larger messages are sent. Audio message transmission starts to fail at distances equal to or longer than 25 m when using Bluetooth (80% loss rate), while Wi-Fi Direct is successful at distances of up to 35 m (40% loss rate). Greater distances result in loss rates of 100% in both modes. The situation becomes more critical with photo message transmission, which was observed to be limited to 15 m while using Bluetooth and 35 m while using Wi-Fi Direct, exhibiting a 60% loss rate in both cases. Even though we expected the delivery of photos and audio messages through Bluetooth to be more challenging due to its reported lower transmission rate in comparison with Wi-Fi Direct, it is noteworthy that Bluetooth was observed to have a greater range when sending small text messages. All transmission attempts at distances longer than those listed in Tables 4 and 5 were unsuccessful.

Discussion

Conecta2, an Android application designed for post-disaster communication, presents a compelling case for utilizing *ad hoc* networks to bridge communication gaps when

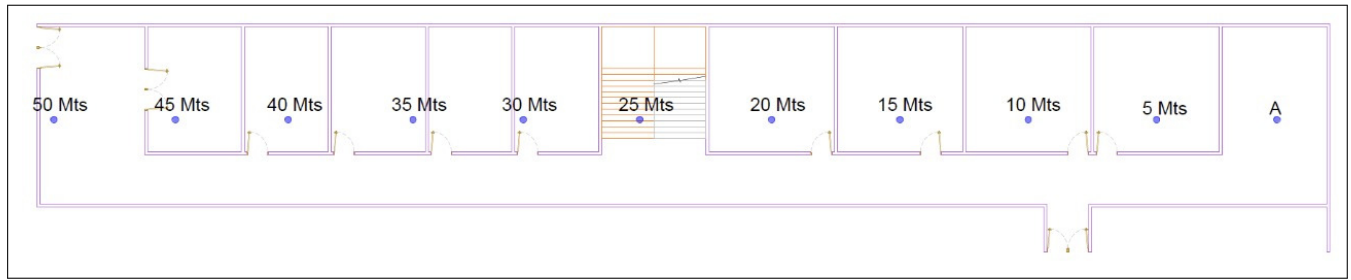


Figure 7. Layout of the building where the experiments with obstacles were conducted. A single floor of the building was used, with 12 adjacent spaces or rooms separated by brick walls. One device was placed at point A, and transmission tests were conducted with a second device in the other rooms, at the locations indicated by the other points.

Source: Authors

Table 1. Specifications of the smartphones used in the performance evaluation experiments

Source: Authors

Brand	Model	Android version	Bluetooth version
Xiaomi	Redmi Note 9	10	4.2
Oppo	A16	11	4.2

Table 2. Successful message delivery rates (with 0 being the lowest value and 1 the highest) between two Android devices running Conecta2 through the one-to-one communication module (i.e., Wi-Fi Direct) for each data type at various distances without obstacles

Source: Authors

Distance between devices (m)	10	20	30	40	50	60	70	80	90	100
Text (20 characters)	1	1	1	1	1	1	1	1	1	1
Audio (5 s)	1	1	1	0.9	1	0.9	1	1	1	1
Photos	1	1	1	1	1	1	1	1	1	0.8

Table 3. Successful message delivery rates (with 0 being the lowest value and 1 the highest) between two Android devices running Conecta2 through the group communication module (i.e., Bluetooth) for each data type at various distances without obstacles

Source: Authors

Distance between devices (m)	10	20	30	40	50
Text (20 characters)	1	1	1	1	1
Audio (5 s)	1	1	1	0.9	1
Photos	1	1	0.9	1	0.6

Table 4. Successful message delivery rates (with 0 being the lowest value and 1 the highest) between two Android devices running Conecta2 through the one-to-one Communication module (i.e., Wi-Fi Direct) for each data type at various distances with obstacles

Source: Authors

Distance between devices (m)	5	10	15	20	25	30	35
Text (20 characters)	1	1	1	0.9	0.7	0.3	0,4
Audio (5 s)	0.9	1	1	1	1	0.9	0,6
Photos	1	1	1	0.8	0.9	0.7	0.4

Table 5. Successful message delivery rates (with 0 being the lowest value and 1 the highest) between two Android devices running Conecta2 through the group communication module (i.e., Bluetooth) for each data type at various distances with obstacles

Source: Authors

Distance between devices (m)	5	10	15	20	25	30	35	40	45
Text (20 characters)	1	1	1	1	1	0.7	1	1	1
Audio (5 s)	1	1	1	1	0.2	0	0	0	0
Photos	1	1	1	0	0	0	0	0	0

traditional infrastructure fails. This study's findings resonate with recent research, highlighting the increasing importance of alternative communication solutions in disaster scenarios. For instance, the work by [9] underscores the potential of LoRa-based device-to-device communication in crisis scenarios, aligning with Conecta2's goal of enabling communication without relying on conventional network infrastructure. Similarly, the research presented by [10] on building dynamic multi-hop Wi-Fi Direct networks for Android smartphones complements this study's exploration of Wi-Fi Direct technology for emergency communication.

Conecta2's emphasis on a user-friendly interface directly addresses a critical gap identified in the related literature. While many studies focus on optimizing protocols and algorithms for *ad hoc* networks, Conecta2 prioritizes usability and accessibility for individuals without technical expertise. This focus on user experience aligns with the findings of [11], who emphasize the importance of effective application distribution systems for internet-less communication during disasters.

Furthermore, Conecta2's device location module offers a unique functionality that is often absent in other *ad hoc* communication solutions: the ability to estimate the location of nearby devices provides a crucial advantage in search and rescue efforts, directly contributing to the efficient location and assistance of survivors. This feature sets Conecta2 apart from applications like Bridgefy, Serval Mesh, and Briar, which primarily focus on privacy and general offline messaging, rather than on post-disaster communication and survivor location.

The comprehensive performance evaluation conducted in this study provides valuable insights into the practical limitations and capabilities of Wi-Fi Direct and Bluetooth in disaster-like conditions. The findings contribute to a deeper understanding of the trade-off between range and energy efficiency, as well as of the impact of obstacles on signal strength and communication range. This research aligns with the growing body of literature on the performance of *ad hoc* networks in realistic scenarios, such as the work by [16] on routing protocols for unmanned aerial vehicle-aided vehicular *ad hoc* networks.

It is important to highlight that Conecta2's development was guided by a user-centered design approach, prioritizing the needs and limitations of disaster survivors. This approach is reflected in the application's intuitive interface and versatile functionalities, allowing users to easily connect with others, share information, and estimate the location of nearby devices. By prioritizing user experience, Conecta2 aims to empower survivors and facilitate effective communication and coordination during critical situations.

Conclusions

This research successfully addressed the critical need for effective communication solutions in post-disaster scenarios by developing and evaluating Conecta2, an Android application that facilitates communication and survivor location through *ad hoc* networks. The application's user-friendly interface, combined with its versatile use of both Wi-Fi Direct and Bluetooth, provides a robust platform for communication and coordination when traditional infrastructure is compromised.

Conecta2's ability to establish an *ad hoc* network using either Wi-Fi Direct or Bluetooth offers a crucial lifeline in disaster-stricken areas, allowing survivors and first responders to connect and coordinate rescue efforts. The device location module further enhances the application's value by providing an estimate of the location of nearby devices, facilitating search and rescue operations and potentially saving lives.

While Conecta2 demonstrates significant potential for improving disaster communication, it is essential to acknowledge its limitations. The application's performance can be affected by environmental factors such as obstacles and user density, which may impact signal strength and communication range. Future research should focus on addressing these limitations by exploring strategies for optimizing the app's performance in high-density scenarios and mitigating the impact of obstacles on *ad hoc* network performance.

The implications of Conecta2 extend beyond its technical functionalities. By offering a resilient and efficient communication pathway, the app can be seamlessly integrated into disaster response strategies, particularly in situations where traditional communication infrastructure is compromised. Disasters such as earthquakes, hurricanes, and floods often result in the disruption of cellular networks and internet connectivity, leaving survivors and first responders isolated and vulnerable. Conecta2's ability to establish an *ad hoc* network using Wi-Fi Direct or Bluetooth provides a crucial lifeline under these circumstances, enabling communication and coordination when it is needed most.

In conclusion, Conecta2 has the potential to revolutionize disaster communication by providing a reliable, user-friendly, and adaptable platform for communication and coordination. Its integration into disaster response strategies could significantly improve the efficiency of rescue efforts, ultimately leading to more lives saved and a more resilient response to disaster situations. This research underscores the potential for Conecta2 to be integrated into existing disaster response strategies, potentially saving lives and improving communication efficiency in critical situations.

Declaration of competing interest

The authors declare that they have no significant competing interests, including those financial, non-financial, professional, or personal in nature, that interfere with the full and objective presentation of the research documented herein.

Funding

The authors would like to express their gratitude to Universidad Antonio Nariño^{¶¶} for the financial support offered during the completion of this research (grant 2021021).

^{¶¶}Universidad Antonio Nariño (<http://www.uan.edu.co/>).

CRedit author statement

Salamanca: conceptualization, methodology, software.
Cables: data curation. Ramírez: writing (review and editing).






References

- [1] M. Naddaf, "Turkey-Syria earthquake: What scientists know," 2023. [Online]. Available: <https://www.nature.com/articles/d41586-023-00373-x>
- [2] F. Graham, "Daily briefing: The science underlying the Turkey-Syria earthquake," 2023. [Online]. Available: <https://www.nature.com/articles/d41586-023-00373-x>
- [3] B.-G. Son, S. Chae, and C. Kocabasoglu-Hillmer, "Catastrophic supply chain disruptions and supply network changes: A study of the 2011 Japanese earthquake," *Int. J. of Oper. Prod. Management*, vol. 41, pp. 781–804, 2021. <https://www.emerald.com/insight/content/doi/10.1108/ijopm-09-2020-0614/full/html>.
- [4] E. Maly and A. Suppasri, "The Sendai framework for disaster risk reduction at five: Lessons from the 2011 great East Japan earthquake and tsunami," *Int. J. Disaster Risk Sci.*, vol. 11, pp. 167–178, 2020. <https://doi.org/10.1007/s13753-020-00268-9>.
- [5] J. Xu and Y. Qiang, "Spatial assessment of community resilience from 2012 hurricane Sandy using nighttime light," *Rem. Sens.*, vol. 13, p. 4128, 2021. <http://dx.doi.org/10.3390/rs13204128>.
- [6] M. K. C. Arnaouti, G. Cahill, M. D. Baird, L. Mangurat, R. Harris, L. P. P. Edme, M. N. Joseph, T. Worlton, S. A. Jr et al., "Medical disaster response: A critical analysis of the 2010 Haiti earthquake," *Front. Pub. Health*, vol. 10, p. 995595, 2022. <https://doi.org/10.3389/fpubh.2022.995595>.
- [7] B. Tang, Q. Chen, X. Liu, Z. Liu, Y. Liu, J. Dong, and L. Zhang, "Rapid estimation of earthquake fatalities in China using an empirical regression method," *Int. J. Disaster Risk Red.*, vol. 41, p. 101306, 2019. <https://doi.org/10.1016/j.ijdr.2019.101306>.
- [8] X. Su, K. Ming, X. Zhang, J. Liu, and D. Lei, "Development of a targeted recommendation model for earthquake risk prevention in the whole disaster chain," *J. Inf. Process. Syst.*, vol. 17, pp. 14–27, 2021. <https://doi.org/10.3745/JIPS.04.0201>.
- [9] J. Hochst, L. Baumgartner, F. Kuntke, A. Penning, A. Sterz, and B. Freisleben, "LoRa-based device-to-device smartphone communication for crisis scenarios," in *Proc. 17th Int. Conf. Inf. Syst. Crisis Res.Manag.*, 2020. https://peasec.de/paper/2020/2020_HC3B6chstetal_LoRaDeviceSmartphoneCommunicationCrisisScenarios.pdf.
- [10] I. M. Ahmed and H. M. Ali, "Building a dynamic multi-hop wi-fi direct network for android smartphones," *Solid State Tech.*, vol. 63, pp. 2193–2205, 2020.
- [11] M. Fujimoto, S. Matsumoto, E. M. Trono, Y. Arakawa, and K. Yasumoto, "Effective application distribution system for Internet-less communication during disasters," *Sens. Mater.*, vol. 32, pp. 79–95, 2020. <https://doi.org/10.18494/SAM.2020.2614>.
- [12] StatCounter, "Mobile operating system market share worldwide," 2021. [Online]. Available: <https://gs.statcounter.com/os-market-share/mobile/worldwide>
- [13] Z. A. Younis, A. M. Abdulazeez, S. R. M. Zeebaree, R. R. Zebari, and D. Q. Zeebaree, "Mobile ad hoc network in disaster area network scenario: A review on routing protocols," *Int. J. Online Biomed. Eng.*, vol. 17, pp. 49–75, 2021. <https://doi.org/10.3991/ijoe.v17i03.16039>.
- [14] Y. Jahir, M. Atiquzzaman, H. Refai, A. Paranjothi, and P. G. LoPresti, "Routing protocols and architecture for disaster area network: A survey," *Ad Hoc Net.*, vol. 82, pp. 1–14, 2019. <https://doi.org/10.1016/j.adhoc.2018.08.005>.
- [15] G. C. Deepak, A. Ladas, Y. A. Sambo, H. Pervaiz, C. Politis, and M. A. Imran, "An overview of post-disaster emergency communication systems in the future networks," *IEEE Wireless Comm.*, vol. 26, pp. 132–139, 2019. <https://doi.org/10.1109/MWC.2019.1800467>.
- [16] R. A. Nazib and S. Moh, "Routing protocols for unmanned aerial vehicle-aided vehicular ad hoc networks: A survey," *IEEE Access*, vol. 8, pp. 77 535–77 560, 2020. <https://doi.org/10.1109/ACCESS.2020.2989790>.
- [17] R. Alnashwan and H. Mokhtar, "Disaster management system over wifi direct," 2019, pp. 1–6. <https://doi.org/10.1109/CAIS.2019.8\protect\protect\leavevmode@ifvmode\kern+.1667em\relax769\protect\protect\leavevmode@ifvmode\kern+.1667em\relax460>.
- [18] T. Furutani, Y. Kawamoto, H. Nishiyama, and N. Kato, "A novel information diffusing method with virtual cells based Wi-Fi direct in disaster area networks," 2018, pp. 1–6. <https://doi.org/10.1109/WCNC.2018.8\protect\protect\leavevmode@ifvmode\kern+.1667em\relax377\protect\protect\leavevmode@ifvmode\kern+.1667em\relax148>.
- [19] F. Mezghani and N. Mitton, "Opportunistic multi-technology cooperative scheme and uav relaying for network disaster recovery," *Information*, vol. 11, p. 37, 2020. <https://doi.org/10.3390/info11010037>.
- [20] A. Reichman and S. Wayer, "Ad-hoc network recovery after severe disaster," 2019, pp. 1–3. <https://doi.org/10.1109/COMCAS44\protect\protect\leavevmode@ifvmode\kern+.1667em\relax984.2019.8\protect\protect\leavevmode@ifvmode\kern+.1667em\relax958\protect\protect\leavevmode@ifvmode\kern+.1667em\relax195>.
- [21] S. D. Gupta, S. Choudhury, and R. Chaki, "Disaster management system using vehicular ad hoc networks," in *Advanced Computing and Systems for Security*, R. Chaki, A. Cortesi, K. Saeed, and N. Chaki, Eds. Berlin, Germany: Springer, 2019, pp. 93–107. https://doi.org/10.1007/978-981-13-3702-4_6.

- [22] F. Álvarez, L. Almon, H. Radtki, and M. Hollick, "Bluemergency: Mediating post-disaster communication systems using the Internet of Things and Bluetooth mesh," 2019. [Online]. Available: <https://arxiv.org/pdf/1909.08094>
- [23] M. Lescisin and Q. H. Mahmoud, "Ad hoc messaging infrastructure for peer-to-peer communication," *Peer-to-Peer Net. Appl.*, vol. 12, pp. 60–73, 2019. <https://doi.org/10.1007/s12083-017-0628-7>.
- [24] Y. Jung, "Community-based localized disaster response through temporary social overlay networks," *Mobile Net. Appl.*, vol. 24, pp. 1641–1653, 2019. <https://doi.org/10.1007/s11036-017-0892-z>.
- [25] A. Ito and H. Hatano, "A study on a protocol for ad hoc network based on Bluetooth low energy," in *Cognitive Infocommunications, Theory and Applications*, R. Klemmou, J. Nikodem, and P. Baranyi, Eds. Berlin, Germany: Springer, 2019, pp. 433–458. https://doi.org/10.1007/978-3-319-951667-em\protect\leavevmode@ifvmode\kern+.1667em\relax996--2_20.

Digital Transformation: Perspectives and Post-Pandemic Adaptation in the Education Sector – A Case Study

Transformación digital: percepciones y adaptación post pandemia en el sector educativo – un estudio de caso

Jhon A. García Camargo ¹, Jesús D. Niño Torres ², Julieth A. López-Castiblanco ³,
Liz K. Herrera-Quintero ⁴, and Jorge I. Bula Escobar ⁵

ABSTRACT

The digital transformation of the public sector seeks to create an effective and efficient government through technology. In the education sector, the pandemic accelerated the adaptation process and the use of new technologies, evidencing, at a global level, complex situations related to university activities and their adaptation to virtuality. This work presents a study conducted at Universidad Nacional de Colombia following the 2020 crisis, which sought to determine the perception of digital transformation within the institution. A study framed within a descriptive methodology was conducted to validate the advances and challenges that persist in relation to the digitization of the university's activities. All this was based on a non-probabilistic sample of 519 people from the academic community. The main results indicate the existence of context-related aspects that facilitate or represent barriers regarding the use of digital tools, such as age, territory, and university campus. The main issue associated with information and communication technologies is Internet access, followed by physical access to related infrastructure and the lack of skills. In addition, we observed a preference for mixed modalities by students, who indicated a desire for change in the education model. Lastly, the applied instrument is highly reliable, which is why it can be used and complemented for future studies under more favorable conditions.

Keywords: education, distance education, COVID-19, pandemic, digital transformation, digital divide

RESUMEN

La transformación digital del sector público busca crear un gobierno eficaz y eficiente a través de la tecnología. En el sector educativo, la pandemia aceleró el proceso de adaptación y el uso de nuevas tecnologías, evidenciando, a nivel global, situaciones complejas relacionadas con el quehacer universitario y su adaptación a la virtualidad. Este trabajo presenta un estudio elaborado en la Universidad Nacional de Colombia, posterior a la crisis del 2020, que buscó establecer la percepción de la transformación digital en la entidad. Se desarrolló un estudio con una metodología descriptiva, en aras de validar los avances y retos que persisten en relación con la digitalización de la universidad en sus actividades. Esto, a partir de una muestra no probabilística de 519 personas de la comunidad académica. Los principales resultados indican que existen aspectos de contexto que facilitan o representan barreras respecto al uso de herramientas digitales, tales como la edad, el territorio y la sede universitaria. El principal problema respecto a las tecnologías de la información y las comunicaciones es el acceso a Internet, seguido del acceso físico a la infraestructura relacionada y la falta de habilidades. Además, se percibió una preferencia por la modalidad mixta por parte de los jóvenes, indicando un deseo de cambio en el modelo educativo. Por último, el instrumento utilizado es altamente fiable, por lo que puede ser utilizado y complementado para futuros estudios en condiciones más favorables.

Palabras clave: educación, educación a distancia, COVID-19, pandemia, transformación digital, brecha digital

Received: January 25th, 2024

Accepted: December 12th, 2024

Introduction

The COVID-19 pandemic disrupted the daily activities of various actors in society [1]. One of the most affected sectors in terms of practices and modes of operation was education, which was strongly restricted for almost two years, during which it could not tend to its pedagogical work in the *traditional* manner [2]. In particular, the university education sector faced severe restrictions for an almost continuous period of two years [3], given the volume of people who could have been affected by the crowds.

¹ Electrical engineer, Universidad Nacional de Colombia, Colombia. Master of Industrial Engineering, Universidad Nacional de Colombia. Affiliation: PhD (c) in Engineering, Universidad Nacional de Colombia, Colombia. E-mail: jhagarcia@unal.edu.co

² Statistician, Universidad Nacional de Colombia, Colombia. Affiliation: Statistician, Universidad Javeriana, Colombia. E-mail: jd.nino@javeriana.edu.co

³ Statistician, Universidad Nacional de Colombia. Affiliation: PhD student in Biotechnology, Universidad Nacional de Colombia, Bogotá, Colombia. E-mail: jula-lopezcas@unal.edu.co

⁴ Materials engineer, Universidad de Antioquia, Colombia, PhD in Materials Science, Universidad de Sevilla, Spain. Affiliation: Associate professor, Universidad Nacional de Colombia. E-mail: lkherreraq@unal.edu.co

⁵ Economist, Universidad Externado, Colombia. PhD in Sociology, Catholic University of Leuven, Belgic. Affiliation: Escuela Superior de Administración Pública, Colombia. E-mail: jibulae@unal.edu.co



Attribution 4.0 International (CC BY 4.0) Share - Adapt

This scenario revealed the need to adapt to the new conditions of isolation in order to continue guaranteeing access to higher education in Colombia, despite the socioeconomic effects derived from the pandemic [4]. Some studies in the education sector, particularly in higher education, have explored and investigated students' perceptions regarding the 'adapted' teaching-learning processes implemented during the contingency [5]. Other authors have conducted qualitative studies related to the innovative factor that triggered the digital transformation in the work of educators across 10 different universities in Latin America [6].

Universidad Nacional de Colombia (UNAL) was no stranger to this situation, which is why it had to undertake, with varying degrees of success, different actions aimed at adapting its missional functions, i.e., research, teaching, and extension [7]. The university started treading a path with considerable uncertainty regarding possible effects and results [8], which also required quick decisions to minimize aspects such as dropout, the shutdown of its academic-administrative operation, and the interruption of processes that had been proposed [9] and implemented in the institution before the pandemic [10], [11].

Under these circumstances, technological tools proved to be a necessary and even essential ally, even for those with strong positions against them due to risks related to academic quality and teamwork, among others [12]. In addition, the difficulties and challenges posed by the adaptation, i.e., focusing the entirety of university activities on digital and distance channels without adequate pedagogical preparation and technological skills due to the imminent closure [13], implied little time to plan and respond [14]. As for the digital transformation processes, it was essential to create technological (connectivity, infrastructure), cultural (adaptation to change), and human resource management (leadership and coordination) conditions in place, so that these initiatives could modify both the processes and the results obtained (efficiency, effectiveness, effectiveness).

To accelerate the digital transformation, different mechanisms and tools associated with the use of technology in teaching-learning processes were sought [15]. One of the strategies led by UNAL through the Laboratory-University Division (UNALab), with the support of the Innovation, Creativity, and New Technologies Laboratory LAB101 UNAL, had to do with understanding the use of technologies and the way in which the academic community was coping with the pandemic, as well as the actions undertaken through different strategies or activities aimed at mitigating the difficulties caused by isolation. To this effect, an instrument was built. One of its first approaches involved understanding the context and the way in which the academic community was dealing with the situation, seeking to determine a set of actions to support decision-making concerning the different needs identified in the short and medium term. Although UNAL defined a digital transformation strategy [10], [11], which was tested in a context of uncertainty, it is essential to assess and verify, through various studies, whether this strategy added value

to the institution's operations. This is particularly important given that some studies indicate that only one in four higher education institutions (HEIs) with a digital strategy perceive any 'valuable return' in terms of organizational impact or cultural change among their members [16].

For this study, two specific objectives were proposed:

- To determine the perception of the academic community (teachers, students, administrators, and contractors) regarding the use and acceptance of technologies in their educational work after the pandemic
- To validate whether there are relevant aspects associated with the context (rural, urban, municipal capitals, campus, among others) that facilitate or generate barriers around the use of digital tools and information and communication technologies (ICTs) in academic-administrative activities

Materials and methods

Instrument for measuring the perception of the digital transformation

A survey was designed to obtain information regarding the main perceptions of the members of the university community on explicit issues related to development and digital transformation in higher education during and after the pandemic. It provided an approximate look at the disruption caused by new technologies in the university scenario during the health emergency due to COVID-19. The questions were designed according to some of the main axes or themes for measuring digital processes mentioned in *Higher education data collection in Hungary and considerations for the development of indicators*, a section of the document titled *Supporting the digital transformation of higher education in Hungary*, issued by the Organization for Economic Cooperation and Development [17]. This document discusses digital infrastructure, digital skills, the environment, and digital culture. The analyzed axes and themes were modified according to the reality of Colombia, and a component with sociodemographic questions was added, allowing to disaggregate the respondents' information, thereby enriching the research on issues of population distribution and access to basic conditions for the use of new technologies.

The instrument contained two types of questions with three approaches. The first type included multiple-choice questions with a single answer, aiming to characterize the respondents according to sociodemographic aspects and the main factors of digital transformation. The latter assessed the digital infrastructure axis (e.g., Internet access). The second type included perceptual questions associated with a Likert scale-type response (i.e., a psychometric scale), where the respondent had to indicate the degree of agreement or disagreement with a statement or item. The

Likert scale was selected since this type of aggregate rating measures is relevant and common in measuring phenomena associated with feelings, attitudes, or perceptions [18], [19]. These questions primarily focused on measuring the axes of digital competences, the environment, and digital culture.

The responses to the virtual form were collected by means of snowball sampling, *i.e.*, a non-probability sampling method in which a set of selected subjects incorporates other subjects into the study [20] (in our case, this was done via e-mail). This sampling methodology was selected since, due to the characteristics and conditions of the confinement measures implemented by public health entities at the time [21], there was no guarantee of defining a relevant and appropriate random sampling framework for all the university's campuses, nor could the researchers go and survey subjects in person with all the conditions that the situation required.

Analysis and data

To analyze the data, only exploration and statistical description tools were applied, namely contingency tables; cake and doughnut diagrams; bar graphs; stacked Likert bars; a correlogram, where a polychoric correlation was used [23]; and boxplots, to which an argument of variation was applied. Since it adds a slight random alteration to the points in order to visualize those that overlap, this argument only had a visual effect and was not used for statistical tests. Regarding the consistency of the instrument, a reliability analysis was conducted, mainly using Cronbach's Alpha [24] and Guttman's Lambda 6 [25]. On the other hand, association and group comparison tests were used: Chi-Square to test for association [26], Cramér's V to determine the strength of association, and Wilcoxon Mann-Whitney with Holm adjustment to test the difference between groups [27], [28]. Additionally, a dimensionality reduction analysis was performed to measure and find associations between variables (MCA), and data mining methods were applied to identify underlying patterns in the data, *i.e.*, HCPC and K-Means [29]. However, the latter did not yield significant results, which is why they are not presented in this document.

The tools used to analyze the information were Excel and R. Organization and debugging were carried out in the former. It is important to mention that 538 answers were initially obtained, and that filtering the data resulted in 519 responses. Meanwhile, R was used for a basic visualization of the primary variables. It should be noted that, in order to make the presented figures friendlier, the viridis color palette was used, which was designed to allow for greater precision in interpreting scientific data, mainly for people with color vision deficiency. In addition, reliability analyses of the instrument were carried out, *i.e.*, tests of association and differences between groups, Likert-type response visualization (using the Likert library) [30], and association and clustering analysis.

To interpret the tests, a significance value of 0.05 was used. This means that, for a given hypothesis, it is determined whether the p-value is less than 0.05. If so, the hypothesis is rejected; otherwise, it is not rejected. The hypotheses may vary, but, for this study, the null hypothesis implied an association between variables or, failing that, a difference between groups. On the other hand, for the correlations and Cramér's V, the criteria shown in Tables I and II were used.

Table I. Criteria for interpreting Cramér's V

Effect size	Interpretation
$ES \leq 0.2$	The result is weak. Although the result is statistically significant, the fields are only weakly associated.
$0.2 < ES \leq 0.6$	The result is moderate. The fields are moderately associated.
$ES > 0.6$	The result is strong. The fields are strongly associated.

Source: Cramér's V [31]

Table II. Criteria for interpreting the correlations

Coefficient range	Interpretation
0-0.19	Very low correlation
0.2-0.39	Low correlation
0.4-0.59	Mean correlation
0.6-0.79	Strong correlation
0.8-1	Very strong correlation

Source: [32]

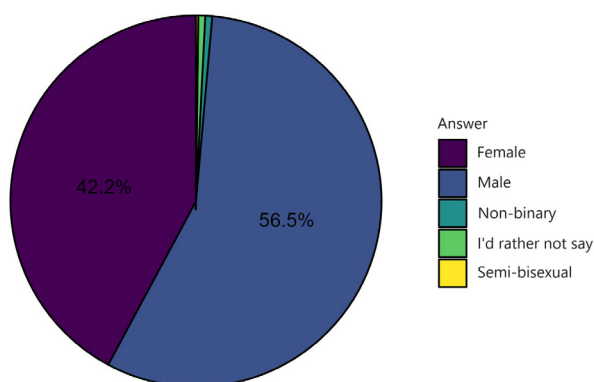
Results

The results obtained through the developed instrument are shown below. They are grouped in relation to (i) the sociodemographic aspects of the respondents, (ii) the general context of the digital transformation and the variables associated with access and connectivity, and (iii) the perception of the digital transformation in the context of work and education.

Sociodemographic aspects of the respondents

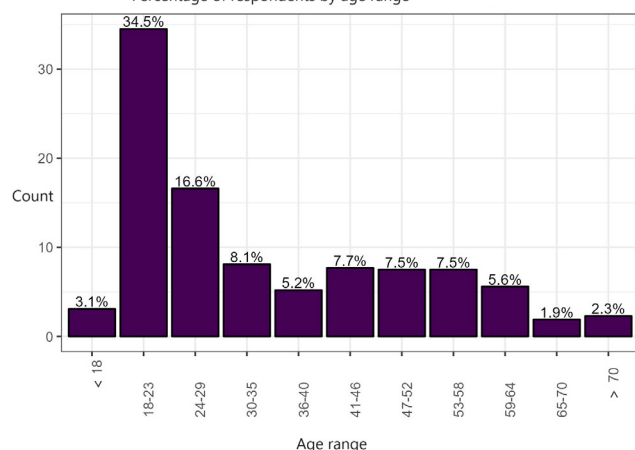
519 respondents participated in this study, out of which 56.5% identified themselves as male, 42.2% as female, 0.6% as non-binary, and 0.6% preferred not to declare. This is depicted in Fig. 1. It should be noted that there was a high participation of the age groups 18-23 and 24-29 (Fig. 2), with 34.5 and 16.6%, respectively. Meanwhile, for the age groups between the ages of 30 and 65, the average percent participation was 6.9% – for those under 18 and over 65, the value was less than 8%. However, 90.2% of the participants live in urban areas, 9.2% in rural areas, and the other 0.2% in other locations, alternating between rural and urban, municipal, and semi-rural capitals (Fig. 3). The responses are concentrated in the Bogotá and Medellín campuses, as can be seen in Fig. 4.

Percentage of respondents by gender

**Figure 1.** Sociodemographic data by gender

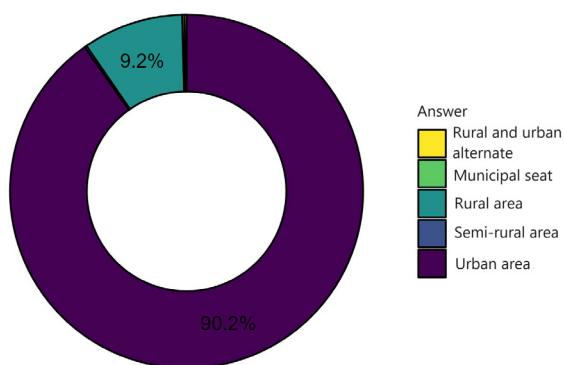
Source: Authors

Percentage of respondents by age range

**Figure 2.** Sociodemographic data by age

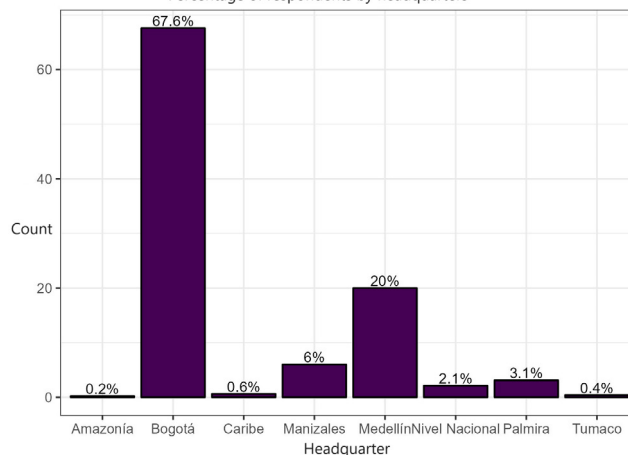
Source: Authors

Percentage of respondents by place of residence

**Figure 3.** Sociodemographic data by place of residence

Source: Authors

Percentage of respondents by headquarters

**Figure 4.** Sociodemographic data by UNAL campus

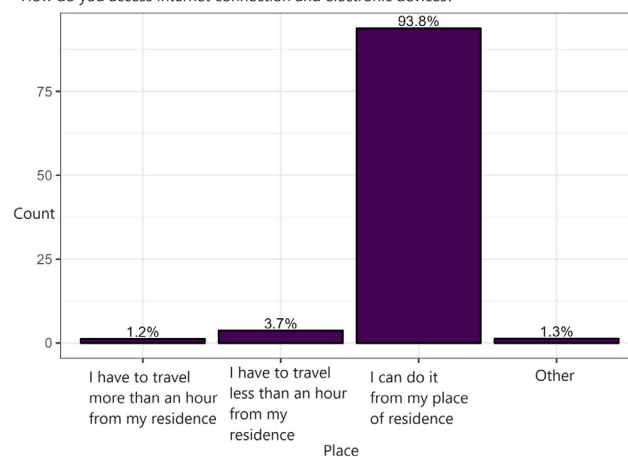
Source: Authors

General aspects of digital transformation in the university community

Within digital transformation models, digital infrastructure is highlighted as a highly relevant dimension. The university community was asked about basic aspects associated with the availability of the necessary infrastructure for their digital processes, such as access to quality electricity, a strong Internet connection, and the corresponding technological devices. For instance, Fig. 5 shows that the community was able to access the Internet mainly from home.

One of the relevant indicators to evaluate Internet quality is the contracted browsing speed. According to the survey, 34.5% of the respondents contracted more than 50 megabits per second (Mbps), 21.8% had 26-50 Mbps, 11% had 16-25 Mbps, 17.3% had 6-10 Mbps, 6.9% had 11-15 Mbps, and 8.5% reported less than 5 Mbps (Fig. 6).

How do you access internet connection and electronic devices?

**Figure 5.** Internet access data

Source: Authors

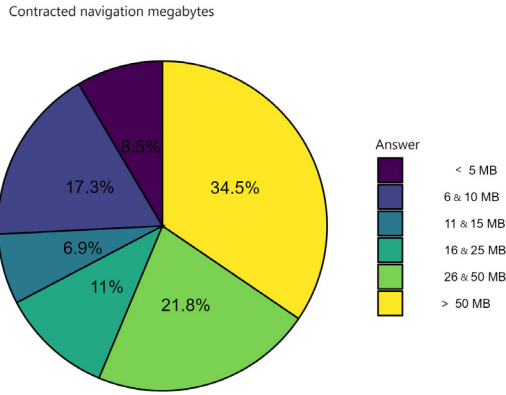


Figure 6. Type of access and capacity of the respondents' Internet service

Source: Authors

Fig. 6 indicates that the ICTs currently used by the respondents in their roles allow them to completely (73.6%) or partially (25.4%) solve problems, challenges, or difficulties. Within this group, 21.8% indicated advanced skills, 65.9% intermediate skills, and 12.3% basic skills (Fig. 8). Meanwhile, 1.0% of the respondents gave a negative answer to this question.

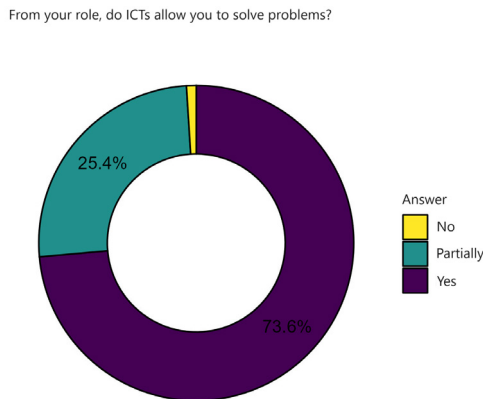


Figure 7. Appropriation of digital skills

Source: Authors

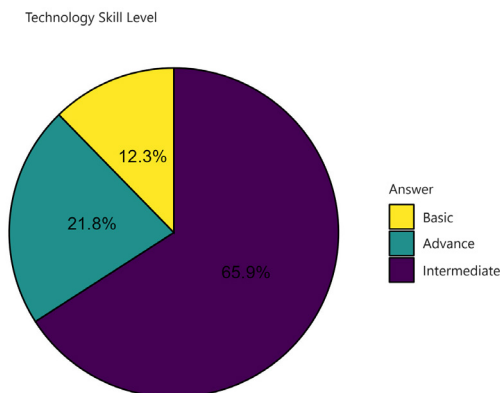


Figure 8. Digital skills level

Source: Authors

Additionally, as can be seen in the figures below, three important aspects should be highlighted. Regarding digital culture in the university community, 18.5% of the respondents indicated that the university is not a beacon of digital transformation. However, 21.6% declared the opposite. The former was partially supported by 59.9% of the respondents, who considered the university to be partially leading the digital transformation (Fig. 9).

On the other hand, Fig. 10 shows that 59.9% of the participants consider the use of digital technologies by UNAL to partially respond to the needs of the university community or society in general. 27.6% indicated that the use of technologies fully responds to the needs of individuals within and outside the institution. Only 12.5% provided a negative response to the statement.

Additionally, we asked if the processes conducted within the university have become more efficient with the use of digital technologies. 45.9% of the respondents considered this to be partially true, 43.2% provided an affirmative answer, and 11% responded negatively. The responses are summarized in Fig. 11.

Is the university a digital benchmark?

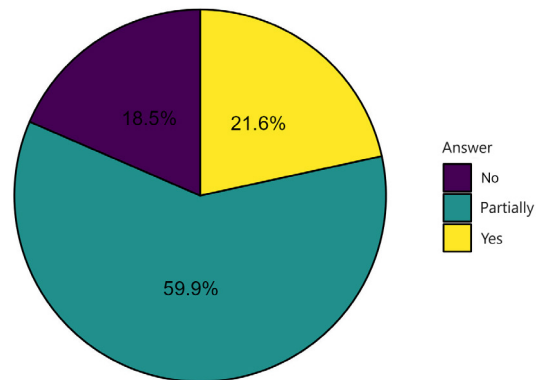


Figure 9. Perception of digital processes at the university

Source: Authors

The use of digital technologies at UNAL responds to the needs of the community?

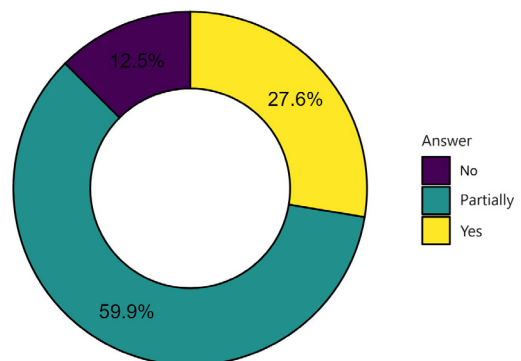


Figure 10. Perception of the response of digital technologies to the needs of the university community

Source: Authors

Have the processes at UNAL become more efficient with the use of digital technologies?

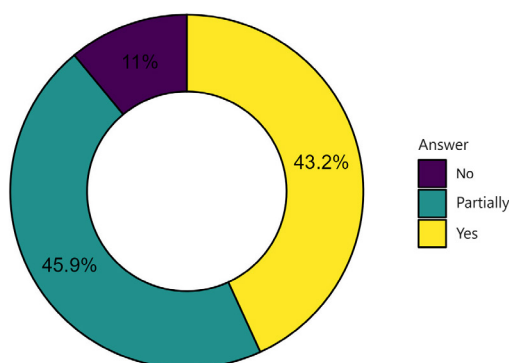


Figure 11. Perception of the use of ICTs in university processes
Source: Authors

Perception of the digital transformation in the university community

To validate whether the instrument was consistent and helped to address the questions we aimed to answer, we initially assessed it using Cronbach's Alpha. Table III shows an evaluation of the reliability of the Likert-type responses used in the second and third sections of the instrument to validate the subsequent analyses via Cronbach's Alpha and Guttman's Lambda 6.

Table III. Reliability test results

Alpha based on correlations	Standardized alpha based on correlation	Guttman Lambda 6 reliability	Average correlation between elements	Standard deviation of the total score	Median correlation between elements
0.82	0.83	0.84	0.36	0.55	0.36

Source: Authors

Table III shows high reliability or internal consistency, with a Cronbach's Alpha of 0.82 and a Guttman's Lambda 6 of 0.84. For a high level of consistency in the instruments, the values of these coefficients should be greater than 0.8 [22]. However, the average correlation between the items is 0.32, which indicates a low association between them.

To gain an initial understanding of the relationships between the variables in the instrument, an analysis of pairwise correlations was conducted. Given that the variables are ordinal and not continuous, we applied a polychoric correlation method, which provided insights beyond previous analyses focused on independently describing each question and response. By correlating variables, this method highlights the connections between the questions and the responses provided by the sample, allowing for a deeper exploration of the data's structure and implications.

The polychoric correlation shows moderate relationships between some of the variables. Considering that a moderate value is between 0.4 and 0.6, we can relate the variables associated with digital competencies or skills to most of the other variables in the figure, as well as to those related to the streamlining of processes thanks to the use of ICTs, which exhibit a medium correlation with several of the variables in the graph. In addition, there is a correlation of interest between access to ICTs and Internet access during the pandemic, as well as between the satisfaction with ICTs, the motivation to use them, and their frequency of use (Fig. 12).

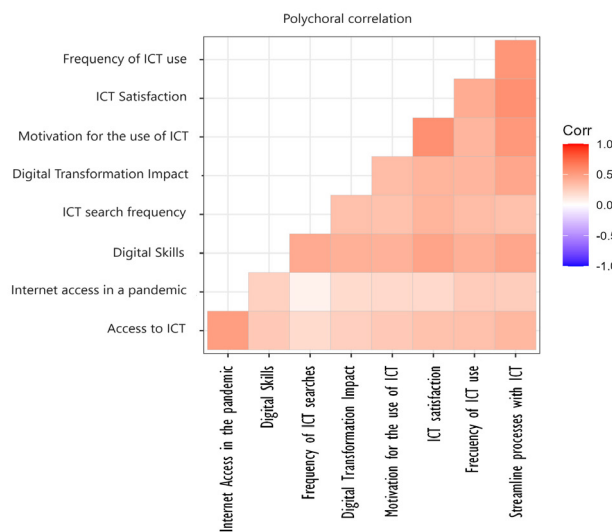


Figure 12. Polychoric correlation
Source: Authors

Fig. 13 summarizes the information obtained through five-point Likert-type responses. It should be noted that the graph indicates the proportion corresponding to each of the ratings per question and the percentage trend in descending order.

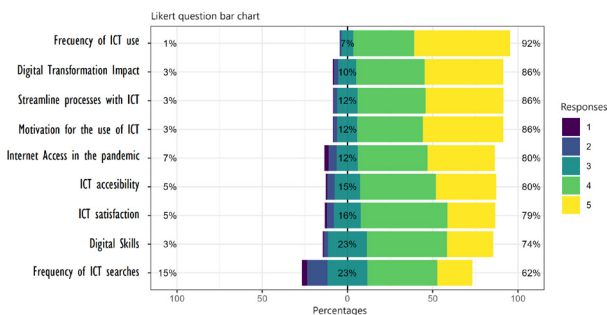


Figure 13. Bar chart for the Likert-type variables
Source: Authors

Fig. 13 shows that 92% of the people surveyed have a positive tendency to use ICTs in their activities, while 7% have a neutral position and 1% have a negative one. 86% of the participants expressed a positive preference regarding

the use of ICTs, 12% indicated a neutral position, and almost 3% did not perceive the use of these technologies to have streamlined their processes. Most of the interviewees were positively motivated to use ICTs, but not as satisfied: 79% indicated high satisfaction, 16% partial satisfaction, and 5% contrary satisfaction. However, 80% of the respondents stated that physical access to ICTs has been both easy and relatively easy, while 15% found it to be partially easy and 5% did not perceive it as easy.

Regarding digital environment and competence, Fig. 13 shows that 86% of the surveyed population have seen an impact on their daily activities. Moreover, 10% have a neutral position, and approximately 3% stated that the digital transformation has had no impact on them. Likewise, 74% rated their digital skills with using digital tools as high, while 23% indicated a central position, and 3% rated their digital skills as low.

Finally, the respondents were asked how often they searched for new ICTs to solve problems. 61% of them searched quite often, 15% did not do it so often, and 15% tended not to do it. They were also asked how difficult it was for them to access the Internet during the pandemic: 81% of the participants had difficulties with Internet access and 7% did not encounter major issues.

To analyze the differences between the categories of nominal variables with respect to the ordinal qualitative variables, the Wilcoxon test was employed, as well as the Holm method to adjust the p-value. The variables that showed significant differences between their categories are presented below, and said categories are specified.

Differences were found between genders in relation to technological skills and the motivation to use ICTs. Regarding the former, a p-value of $4.2e-07$ was found, which is lower than the level of significance ($p\text{-value} < 0.05$), between the male and female participants, i.e., there are differences between these groups. With respect to segmentation by university campus, in both Bogotá and Medellín, it can be clearly seen that males tend to rate their skills more positively than females (Fig. 14).

On the other hand, as can be seen in Fig. 15, there are differences in the territory categories with respect to Internet access during the pandemic. For the territory variable, the test yielded a p-value of 0.00089 between the urban and rural areas, indicating significant differences in the assessments provided by the participants in terms of their Internet access during the pandemic; a better perception was found in urban areas.

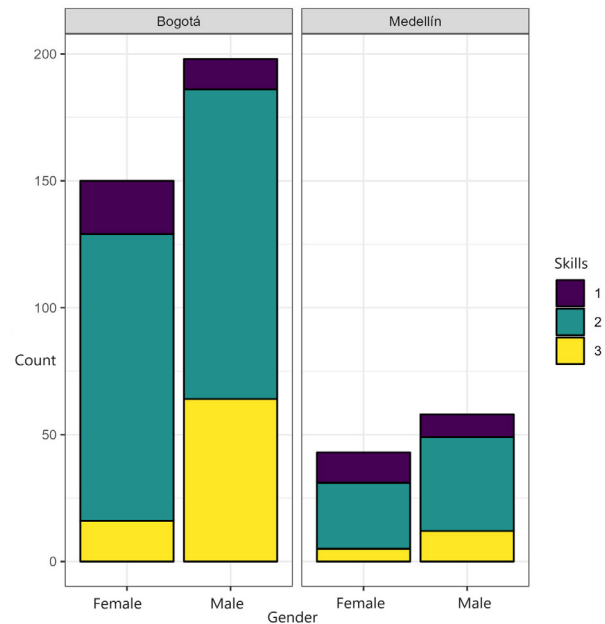


Figure 14. Bar chart for digital skills by gender and campus
Source: Authors

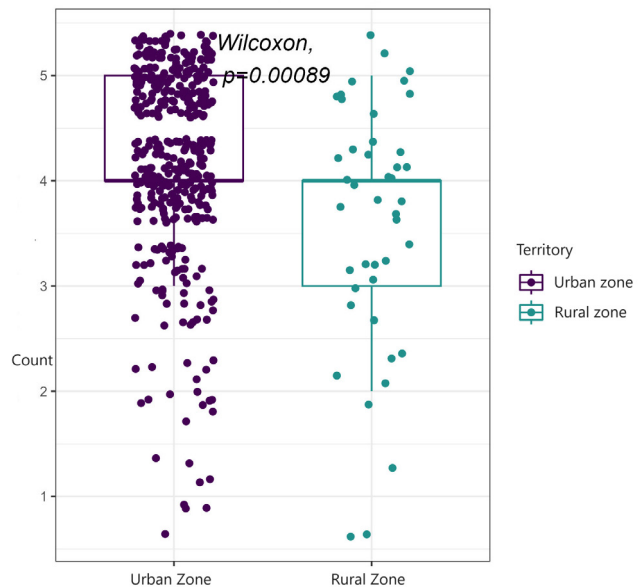


Figure 15. Internet access by territory during the pandemic
Source: Authors

Regarding segmentation by campus, Bogotá and Medellín exhibited the same behavior, but the latter showed no statistically significant difference ($p\text{-value}=0.14$) compared to the former (Fig. 16).

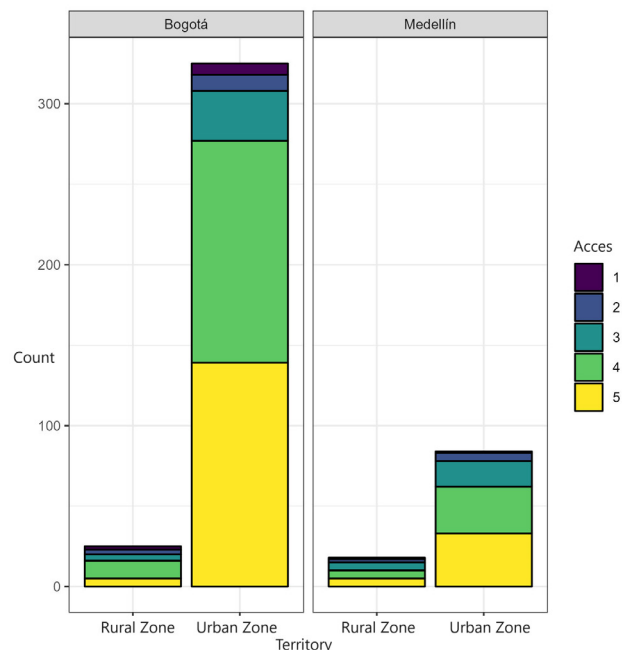


Figure 16. Internet access vs. territory

Source: Authors

Conclusions

Our instrument allowed collecting information on people from different contexts who are associated with UNAL, providing an initial idea of their knowledge and use of technology, especially among the population with access to the Internet.

According to the results, the surveyed UNAL community does not perceive the institution as a digital transformation leader in higher education. This aligns with and supports previous studies indicating that a digital transformation strategy is not commonly perceived by most HEIs to generate practical value for their development [16]. This implies, on the one hand, recognizing and better highlighting the efforts made by the institution with regard to the digital transformation, and, on the other hand, it necessitates a more accurate determination of the institution's digital maturity level, as well as the implementation of actions aimed at strengthening competencies in this area [33].

Regarding the barriers or challenges of implementing technologies in university settings, there are significant differences in male and female individuals' appropriation of and motivation to use ICTs. This finding holds great value in association with other studies suggesting that efforts for such appropriation should be focused on teacher training [6]. Our results reveal important gaps that should lead to the development of effective actions to mitigate this situation from the perspective of female students and their contexts [34].

It is also interesting that, in general, the findings indicate acceptance, knowledge, and use of ICTs in both professional and academic settings, with even a positive perception regarding the ease and speed that these tools provide in solving everyday problems. However, it is particularly noteworthy that the interest in seeking new ICT tools is relatively low. This suggests, as proposed by other authors, a need for strategies that promote interest in and knowledge of new tools or uses of ICTs in academic or professional contexts from the perspective of educational teaching, in order to foster curiosity about the use of new technologies [6], [35].

Initially, through Cronbach's Alpha, we identified a high level of reliability and consistency in the questions of the applied instrument. We observed higher participation by young people (mostly undergraduate students) aged between 18 and 23 years from the Bogotá and Medellín campuses. However, there were participants in all age ranges, from all campuses, and with different types of association with the university.

Furthermore, when analyzing the instrument's questions, associations were found between digital variables in relation to age, territory, and university campus, which implies context-related aspects that facilitate or generate barriers for the use of digital tools. Firstly, differences between territories were found with regard to Internet access during the pandemic. Secondly, differences between UNAL campuses were found in relation to the level of satisfaction expressed by participants when using ICTs, their perception of whether these tools help to solve day-to-day problems, and their opinion regarding the university as a benchmark in terms of digital transformation and the efficiency of ICTs. Thirdly, significant relationships were found between the gender variable in relation to technological skills and the motivation to use ICTs.

On the other hand, although most of the participants live in urban areas and have favorable perceptions regarding the use of ICTs, those in rural areas indicated that, despite the simplicity of physical access to digital devices, access to the Internet has not been easy. This reveals a need for better Internet distribution, accessibility, and coverage in the country, as most participants expressed it as a necessity.

In addition, the perception of the academic community regarding the use of technologies after the pandemic indicates that the main issues have to do with the lack of Internet access, followed by limitations in physical access to ICTs and lack of skills. Finally, we found that virtuality is currently used as a means of study or work, but that mixed modalities (e.g., blended learning) are significantly preferred by young people. Our tests also suggested a desire for change in the pre-pandemic educational model.

Recommendations and limitations

A more specific study on this issue could be carried out. Our sampling method allowed drawing conclusions descriptively, since it did not allow for generalization or inference. Still, it should be noted that the procedures could not have been more rigorous due to the confinement measures in effect during 2021.

In addition, three approaches were considered in this work: sociodemographic aspects, digital infrastructure, and perception of the digital transformation. These approaches were appropriate, understandable, and accessible to the respondents, in addition to covering part of the axes proposed by [17], i.e., digital infrastructure, digital skills, environment, and digital culture.

Given the need to perform snowball sampling during the pandemic and under the immediate post-quarantine conditions, the results of this work may be biased by the favorable conditions of those who were able to fill out the survey. Therefore, our results cannot be extended beyond the conditions presented by the respondents.

Additionally, we suggest carrying out a new test that allows contrasting our results, aiming to determine which variables or categories were altered or accentuated under the new conditions of work and study after UNAL resumed its 100% face-to-face classes. All this, in order to establish if teaching-learning processes continue to be facilitated or favored by the use of technology.

With the instrument applied, the information collected, and the adaptation of the Hungarian model that inspired the construction of our approach [17], longitudinal studies with several universities could be conducted, objectively comparing the progress made and the challenges present in the implementation of digital transformation strategies that add value to educational organizations [6].

In the future, more robust studies could be carried out to establish baselines of action, whose results should facilitate decision-making regarding the challenges faced by HEIs in digital transformation processes, in the use of technologies in the classroom, and in their acceptance by teachers, students, and administrators [8].

Additionally, we suggest undertaking new studies that deepen these preliminary findings, correct the biases inherent to our sampling and data collection processes, and confirm or complement the results obtained through implementations in other HEIs, enabling the generalization of the proposed indicators and the strengthening of comparative analyses [36].

Acknowledgements

The authors would like to express their gratitude to Universidad Nacional de Colombia and its National

Directorate of Digital Strategies for the support and funding provided within the framework of the UNALAB initiative titled Implementation of a technological capabilities and digital transformation observatory at UNAL.

We would also like to thank Prof. Carlos Eduardo Moreno Mantilla, whose accompaniment within the framework of the research seminar (subject 2) of the doctoral program in Engineering, Industry, and Organizations allowed for great progress in this work.

Finally, we would like to thank the students, teachers, administrators, and the rest of the university community who, through their answers, provided valuable information for this work.

CRedit author statement

J. A. García conceived the idea, proposed theoretical and empirical references to carry out this research within the framework of his doctoral thesis, and wrote a large portion of this text. **J. D. Niño** and **J. A. López** validated and analyzed the data, developed statistical models, and made important contributions to the document. **L. K. Quintero** reviewed and supported the deployment of the instrument within the academic community, provided suggestions, and helped to write this manuscript. **J. Bula** provided comments and suggestions for the development of this document.

Conflicts of interest

The authors declare no conflict of interest.

Data availability

The data collected did not include sensitive or personal data. In accordance with the objectives of this study, there was no need to collect private information that allowed distinguishing people's contributions. Additionally, the statistical treatment was done in an aggregate manner to obtain more general responses. *The data and other aspects related to this research can be consulted at the following link for its replication.*

References

- [1] Z. Budayová, M. Pavliková, A. Samed Al-Adwan, and K. Klasnja, "The impact of modern technologies on life in a pandemic situation," *J. Educ. Cult. Soc.*, vol. 13, no. 1, pp. 213–224, Jun. 2022. <https://doi.org/10.15503/jecs2022.1.213.224>
- [2] C. Son, S. Hegde, A. Smith, X. Wang, and F. Sasangohar, "Effects of COVID-19 on college students' mental health in the United States: Interview survey study," *J. Med. Internet Res.*, vol. 22, no. 9, art. e21279, Sep. 2020. <https://doi.org/10.2196/21279>

- [3] C. Giovannella, "Effect induced by the COVID-19 pandemic on students' perception about technologies and distance learning," in *Ludic, Co-design and Tools Supporting Smart Learning Ecosystems and Smart Education*, O. Mealha, M. Rehm, and T. Rebedea, Eds. Singapore: Springer, 2021, vol. 197, pp. 105–116. doi: https://doi.org/10.1007/978-981-15-7383-5_9
- [4] N. Palacios Mena and J. F. Ariza Bulla, "Socioeconomic conditions and academic performance in higher education in Colombia during the pandemic," *Qual. Higher Educ.*, vol. 29, no. 2, pp. 242–260, Oct. 2022. <https://doi.org/10.1080/13538322.2022.2088564>
- [5] M. D. Díaz Noguera, C. Hervás-Gómez, M. D. L. Á. Domínguez-González, and O. Guijarro-Cordobés, "La transformación digital en la docencia de las universidades desde la visión de los estudiantes," in *Formación en ecosistemas de aprendizaje*, E. Vázquez Cano, L. López Catalán, E. Pérez Navío, and O. Guijarro-Cordobés, Eds.. Madrid, Spain: Dykinson, S.L., 2022. <https://doi.org/10.2307/j.ctv2gz3s3t.6>
- [6] A. Deroncel-Acosta, M. L. Palacios-Núñez, and A. Toribio-López, "Digital transformation and technological innovation on higher education post-COVID-19," *Sustainability*, vol. 15, no. 3, art. 2466, Jan. 2023. <https://doi.org/10.3390/su15032466>
- [7] D. M. Osorno Alzate and Á. M. Benítez Góez, "La nueva misión de la universidad. contextualización y resultados: casos de tres universidades públicas colombianas la nueva misión de la universidad. contextualización y resultados: casos de tres universidades públicas colombianas-the new mission of...," *Panorama*, vol. 11, no. 20, pp. 82–94, Nov. 2017. [Online]. Available: <http://hdl.handle.net/10823/5426>
- [8] A. García-Holgado and F. J. García-Peñalvo, "A model for bridging the gender gap in STEM in higher education institutions," in *Women in STEM in Higher Education*, F. J. García-Peñalvo, A. García-Holgado, A. Dominguez, and J. Pascual, Eds. Singapore: Springer, 2022, pp. 1–19. https://doi.org/10.1007/978-981-19-1552-9_1
- [9] M. L. Cardona Orozco, "Capacidades digitales como base de la transformación digital. Una propuesta para la Universidad Nacional de Colombia Sede Manizales," Master's thesis, Univ. Nac. Colombia, Manizales, Colombia, 2021. [Online]. Available: <https://repositorio.unal.edu.co/handle/unal/81030>
- [10] L. K. Herrera Quintero et al., "Transformación digital: experiencias colectivas", 2023. [Online]. Available: <https://repositorio.unal.edu.co/handle/unal/83677>
- [11] L. K. Herrera Quintero, Z. F. Jacobo, and U. P. P. Carolina, Bit a bit : la transformación digital de la Universidad Nacional de Colombia. Bogotá, Colombia: Universidad Nacional de Colombia, 2024. [Online]. Available: <https://repositorio.unal.edu.co/handle/unal/86007>
- [12] R. van der Vlies, "Digital strategies in education across OECD countries", 2020. <https://doi.org/10.1787/33d-d4c26-en>
- [13] C. Hodges, S. Moore, B. Lockee, T. Trust, and A. Bond, "The difference between emergency remote teaching and online learning," *Educause Review*, 2020, [Online]. Available: <https://er.educause.edu/articles/2020/3/the-difference-between-emergency-remote-teaching-and-online-learning>
- [14] E.-J. Díez-Gutiérrez and K. Gajardo Espinoza, "Educar y evaluar durante el confinamiento en España: mundo rural y urbano," *Educ. Knowledge Soc. (EKS)*, vol. 23, art. e26312, Apr. 2022. <https://doi.org/10.14201/eks.26312>
- [15] A. Roß, A. Bikfalvi, and P. Marques, "Pandemic-accelerated digital transformation of a born digital higher education institution: Towards a customized multimode learning strategy," *Educ. Tech. Soc.*, vol. 25, no. 1, pp. 124–141, 2022, [Online]. Available: <https://www.jstor.org/stable/48647035>
- [16] A. Fernández, B. Gómez, K. Binjaku, and E. K. Meçe, "Digital transformation initiatives in higher education institutions: A multivocal literature review," *Educ. Infor. Tech.*, vol. 28, no. 10, pp. 12351–12382, Mar. 2023. <https://doi.org/10.1007/s10639-022-11544-0>
- [17] OECD, "Supporting the digital transformation of higher education in Hungary," 2021. <https://doi.org/10.1787/d30ab43f-en>
- [18] S. E. Harpe, "How to analyze Likert and other rating scale data," *Curr. Phar. Teach. Learn.*, vol. 7, no. 6, pp. 836–850, Nov. 2015, doi: <https://10.1016/j.cptl.2015.08.001>
- [19] H. Taherdoost, "What is the best response scale for survey and questionnaire design; review of different lengths of rating scale / attitude scale / Likert scale", *Int. J. Acad. Res. Manage.*, vol. 8, no. 1, pp. 1-10, Jun. 2019. <https://ssrn.com/abstract=3588604>
- [20] L. A. Goodman, "Snowball sampling," *The Annals of Mathematical Statistics*, vol. 32, no. 1, pp. 148–170, 1961. <https://doi.org/10.1214/aoms/1177705148>
- [21] Ministerio de Salud de Colombia, "Presidente Duque declara emergencia sanitaria frente a COVID-19," Minsalud, 2020. [Online]. Available: <https://minsalud.gov.co/Paginas/Presidente-Duque-declara-Emergencia-Sanitaria-frente-a-COVID-19.aspx>
- [22] M. Schweinberger, "Questionnaires and surveys: Analyses with R," 2022. [Online]. Available: https://rstudio-pubs-static.s3.amazonaws.com/1154767_5aaacfd-63244c49a04ae93e6e0b1a71.html
- [23] M. C. Richaud, "Desarrollos del análisis factorial para el estudio de ítem dicotomicos y ordinales," *Interdisciplinaria*, vol. 22, no. 2, pp. 237-251, 2005. <https://www.redalyc.org/articulo.oa?id=18022206>
- [24] L. J. Cronbach, "Coefficient alpha and the internal structure of tests," *Psychometrika*, vol. 16, no. 3, pp. 297–334, Sep. 1951. <https://doi.org/10.1007/bf02310555>
- [25] L. Guttman, "A basis for analyzing test-retest reliability," *Psychometrika*, vol. 10, no. 4, pp. 255–282, Dec. 1945. <https://doi.org/10.1007/bf02288892>
- [26] V. Bewick, L. Cheek, and J. Ball, "Statistics review 7: Correlation and regression," *Crit. Care*, vol. 7, no. 6, art. 451, 2003. <https://doi.org/10.1186/cc2401>
- [27] K. M. Ramachandran and C. P. Tsokos, *Mathematical Statistics with Applications in R*. Amsterdam, Netherlands: Elsevier Science, 2014. <https://books.google.it/books?id=5A-JOAwAAQBAJ>
- [28] S. Nørgaard Olesen and D. Giacalone, "The influence of packaging on consumers' quality perception of carrots," *J. Sens. Stud.*, vol. 33, no. 1, Jan. 2018. <https://doi.org/10.1111/joss.12310>

- [29] F. Husson, S. Le, and J. Pagès, "Exploratory Multivariate Analysis by Example Using R". Boca Raton, FL, USA: Chapman and Hall/CRC, 2017. <https://doi.org/10.1201/b21874>
- [30] J. Bryer and K. Speerschneider, "Package 'likert,'" 2022. [Online]. Available: <https://www.maths.bris.ac.uk/R/web/packages/likert/likert.pdf>
- [31] M. Kearney, "Cramér's V," Sage Encyclopedia of Communication Research Methods, 2017. [Online]. Available: <https://methods.sagepub.com/ency/edvol/the-sage-encyclopedia-of-communication-research-methods/chpt/cramer-s-v>
- [32] D. Napitupulu *et al.*, "Analysis of student satisfaction toward quality of service facility," *J. Phys. Conf. Ser.*, vol. 954, art. 012019, 2018. <https://doi.org/10.1088/1742-6596/954/1/012019>
- [33] R. Gurumurthy and D. Schatsky, "Pivoting to digital maturity: Seven capabilities central to digital transformation," Deloitte Insights, 2019, [Online]. Available: <https://www2.deloitte.com/us/en/insights/focus/digital-maturity/digital-maturity-pivot-model.html>
- [34] D. E. Flor, E. Henrique Molina da Cruz, A. T. Possebom, C. Roberto Beleti Junior, R. Hubner, and L. B. Ruiz Aylon, "MannaTeam: A case of interinstitutional collaborative learning and Education 5.0," in *2020 Int. Conf. Comp. Sci. Comp. Intel. (CSCI)*, 2020, pp. 964–970. <https://doi.org/10.1109/csci51800.2020.00179>
- [35] F. D. Guillén-Gámez, J. Ruiz-Palmero, E. Sánchez-Rivas, and E. Colomo-Magaña, "ICT resources for research: an ANOVA analysis on the digital research skills of higher education teachers comparing the areas of knowledge within each gender," *Educ. Info. Tech.*, vol. 25, no. 5, pp. 4575–4589, Apr. 2020. <https://doi.org/10.1007/s10639-020-10176-6>
- [36] G. Rodríguez-Abitia and G. Bribiesca-Correa, "Assessing Digital Transformation in Universities," *Future Internet*, vol. 13, no. 2, art. 52, Feb. 2021. <https://doi.org/10.3390/fi13020052>