Escuela Politécnica Nacional

Facultad de Ingeniería de Sistemas

LATIN-AMERICAN JOURNAL OF COMPUTING

Volume 13, ISSUE 1
January 2026
ISSN: 1390-9266
e-ISSN:1390-9134

Vol XIII, Issue 1 January 2026

ESCUELA POLITÉCNICA NACIONAL

MISIÓN

La Escuela Politécnica Nacional es una Universidad pública, laica y democrática que garantiza la libertad de pensamiento de todos sus integrantes, quienes están comprometidos con aportar de manera significativa al progreso del Ecuador. Formamos investigadores y profesionales en ingeniería, ciencias, ciencias administrativas y tecnología, capaces de contribuir al bienestar de la sociedad a través de la difusión del conocimiento científico que generamos en nuestros programas de grado, posgrado y proyectos de investigación. Contamos con una planta docente calificada, estudiantes capaces y personal de apoyo necesario para responder a las demandas de la sociedad ecuatoriana.

VISIÓN

En el 2024, la Escuela Politécnica Nacional es una de las mejores universidades de Latinoamérica con proyección internacional, reconocida como un actor activo y estratégico en el progreso del Ecuador. Forma profesionales emprendedores en carreras y programas académicos de calidad, capaces de aportar al desarrollo del país, así como promover y adaptarse al cambio y al desarrollo tecnológico global. Posiciona en la comunidad científica internacional a sus grupos de investigación y provee soluciones tecnológicas oportunas e innovadoras a los problemas de la sociedad.

La comunidad politécnica se destaca por su cultura de excelencia y dinamismo al servicio del país dentro de un ambiente de trabajo seguro, creativo y productivo, con infraestructura de primer orden.

ACCIÓN AFIRMATIVA

La Escuela Politécnica Nacional es una institución laica y democrática, que garantizar la libertad de pensamiento, expresión y culto de todos sus integrantes, sin discriminación alguna. Garantiza y promueve el reconocimiento y respeto de la autonomía universitaria, a través de la vigencia efectiva de la libertad de cátedra y de investigación y del régimen de cogobierno.

https://www.epn.edu.ec

ISSN:1390-9266 e-ISSN:1390-9134 LAJC 2026

FACULTAD DE INGENIERÍA DE SISTEMAS

MISIÓN

La Facultad de Ingeniería de Sistemas es el referente de la Escuela Politécnica Nacional en el campo de conocimiento y aplicación de las Tecnologías de Información y Comunicaciones; actualiza en forma continua y pertinente la oferta académica en los niveles de pregrado y postgrado para lograr una formación de calidad, ética y solidaria; desarrolla proyectos de investigación, vinculación y proyección social en su área científica y tecnológica para solucionar problemas de transcendencia para la sociedad.

VISIÓN

La Facultad de Ingeniería de Sistemas está presente en posiciones relevantes de acreditación a nivel nacional e internacional y es referente de la Escuela Politécnica Nacional en el campo de las Tecnologías de la Información y Comunicaciones por su aporte de excelencia en las carreras de pregrado y postgrado que auspicia, la calidad y cantidad de proyectos de investigación, vinculación y proyección social que desarrolla y su aporte en la solución de problemas nacionales a través del uso intensivo y extensivo de la ciencia y la tecnología.

https://fis.epn.edu.ec

ISSN:1390-9266 e-ISSN:1390-9134 LAJC 2026

LATIN-AMERICAN JOURNAL OF COMPUTING

Vol XIII, Issue 1, January 2026

ISSN: 1390-9266 e-ISSN: 1390-9134
DOI: 10.5281/zenodo.15740060

Published by:
Escuela Politécnica Nacional
Facultad de Ingeniería de Sistemas

Quito – Ecuador

Indexed in

Associated institutions

Mailing Address
Escuela Politécnica Nacional,
Facultad de Ingeniería de Sistemas
Ladrón de Guevara E11-253, La Floresta
Quito-Ecuador, Apartado Postal: 17-01-2759

Web Address
https://lajc.epn.edu.ec/

E-mail
lajc@epn.edu.ec

Frequency
2 issues per year

ISSN:1390-9266 e-ISSN:1390-9134 LAJC 2026

Published by

Escuela Politécnica Nacional
Facultad de Ingeniería de Sistemas
Ecuador

ISSN:1390-9266 e-ISSN:1390-9134 LAJC 2026

EDITORIAL

Gabriela Suntaxi
PhD.

Editor LAJC
Escuela Politécnica Nacional,
Ecuador

ISSN:1390-9266 e-ISSN:1390-9134 LAJC 2026

Con la publicación del **Volumen 13, Número 1,** la revista *Latin-American Journal of Computing (LAJC)* continúa consolidándose como un espacio regional para la difusión de investigación rigurosa y pertinente en el área de la computación. Este número reúne ocho artículos que reflejan la evolución constante de esta disciplina, en la que convergen la solidez metodológica, la relevancia práctica y la atención al contexto.

Los trabajos incluidos en este volumen abordan una amplia variedad de problemáticas que surgen en la intersección entre la computación, los datos y las aplicaciones del mundo real. Varias contribuciones se centran en soluciones orientadas al software, proponiendo arquitecturas, herramientas y enfoques de evaluación destinados a mejorar el rendimiento, la confiabilidad y la usabilidad de los sistemas. Estos estudios se sustentan en casos de uso concretos, ofreciendo aportes directamente transferibles a entornos académicos y profesionales.

Otro conjunto de artículos pone énfasis en el procesamiento de datos y el análisis inteligente, mostrando cómo las técnicas computacionales pueden aplicarse para extraer conocimiento a partir de fuentes de datos complejas. Mediante estudios experimentales y metodologías aplicadas, estos trabajos evidencian la creciente importancia de la analítica y los sistemas inteligentes como apoyo a la toma de decisiones en diversos dominios.

Este número también incluye investigaciones relacionadas con infraestructura, seguridad y adopción tecnológica, en las que se analizan desafíos actuales y se proponen soluciones acordes a las exigencias de los entornos digitales modernos. Estas contribuciones resultan especialmente relevantes en contextos donde la escalabilidad, la confianza y la gestión eficiente de los recursos constituyen preocupaciones clave.

En conjunto, los artículos publicados en este número no solo evidencian avances técnicos, sino también un esfuerzo sostenido por alinear la investigación en computación con las necesidades sociales, institucionales y regionales. La diversidad de perspectivas y enfoques presentados resalta el carácter dinámico del campo y el valor de la investigación interdisciplinaria y contextualizada.

Expresamos nuestro sincero agradecimiento a los autores por compartir sus trabajos y a los revisores por sus evaluaciones cuidadosas y constructivas, fundamentales para mantener la calidad científica de la revista. Invitamos a los lectores a explorar los artículos de este número y confiamos en que encontrarán en ellos aportes valiosos e inspiradores para futuras líneas de investigación.

Gabriela Suntaxi, PhD
Editora en Jefe
Latin-American Journal of Computing (LAJC)
Escuela Politécnica Nacional, Ecuador

ISSN:1390-9266 e-ISSN:1390-9134 LAJC 2026

With the publication of **Volume 13, Number 1,** the *Latin-American Journal of Computing (LAJC)* continues to consolidate its role as a regional forum for rigorous and relevant research in computing. This issue brings together eight articles that reflect the evolving landscape of the computing discipline, where methodological soundness, practical relevance, and contextual awareness converge.

The works included in this volume address a broad range of problems that arise at the intersection of computation, data, and real-world applications. Several contributions focus on software-centric solutions, proposing architectures, tools, and evaluation approaches aimed at improving system performance, reliability, and usability. These studies are grounded in concrete use cases, offering insights that are directly transferable to professional and academic settings.

Another set of articles emphasizes data processing and intelligent analysis, illustrating how computational techniques can be applied to extract meaning from complex data sources. Through experimental studies and applied methodologies, these contributions highlight the growing importance of analytics and intelligent systems in supporting decision-making across diverse domains.

This issue also features research related to infrastructure, security, and technological adoption, where authors examine current challenges and propose solutions that respond to the demands of modern digital environments. These works are particularly relevant in contexts where scalability, trust, and efficient resource management are critical concerns.

Taken together, the articles in this issue illustrate not only technical progress, but also a sustained effort to align computing research with societal, institutional, and regional needs. The diversity of perspectives and approaches presented here underscores the dynamic nature of the field and the value of interdisciplinary and context-aware research.

We would like to express our sincere appreciation to the authors for their contributions and to the reviewers for their careful and constructive evaluations, which are fundamental to maintaining the scientific quality of the journal. We invite readers to engage with the articles in this issue and trust that they will find them informative and inspiring for future research directions.

Gabriela Suntaxi, PhD
Editor-in-Chief
Latin-American Journal of Computing (LAJC)
Escuela Politécnica Nacional, Ecuador

ISSN:1390-9266 e-ISSN:1390-9134 LAJC 2026

Reviewers

We are most grateful to the following individuals for their time and commitment to review manuscripts for the Latin American Journal of Computing – LAJC

Carlos Ayala, MSc.
Escuela Politécnica Nacional

Daniel Maldonado, PhD.
Escuela Politécnica Nacional

Eduardo Stefanato, MSc.
University of São Paulo

Fabricio Trujillo, PhD.
Universidad Técnica de Ambato

Fredy Gavilanes, PhD.
Escuela Superior Politécnica de Chimborazo

Geovanny Brito, MSc.
Universidad Técnica Estatal de Quevedo

Irene Cedillo, PhD.
Universidad de Cuenca

Iván Carrera, PhD.
Escuela Politécnica Nacional

Ivonne Maldonado, MSc.
Escuela Politécnica Nacional

Jorge Zambrano, PhD.
Universidad Politécnica de Valencia

José Lucio, PhD.
Escuela Politécnica Nacional

Leandro Pazmiño, MSc.
Escuela Politécnica Nacional

Marcela Mosquera, MSc.
Escuela Politécnica Nacional

Monserrate Intriago, PhD.
Escuela Politécnica Nacional

Myriam Hernandez, PhD.
Escuela Politécnica Nacional

Nancy Betancourt, PhD.
Universidad de las Fuerzas Armadas (ESPE)

Nelson Herrera, PhD.
Escuela Politécnica Nacional

Pablo Del Hierro, PhD.
Escuela Politécnica Nacional

Patricia Acosta, PhD.
Universidad de las Américas (UDLA)

Ronie Martinez, MSc.
Escuela Politécnica Nacional

Rubén Nogales, PhD.
Universidad Técnica de Ambato

Verónica Chamorro, MSc.
Universidad Complutense de Madrid

Viviana Parraga, MSc.
Escuela Politécnica Nacional

Walter Fuertes, PhD.
Universidad de las Fuerzas Armadas (ESPE)

ISSN:1390-9266 e-ISSN:1390-9134 LAJC 2026

TABLE OF CONTENTS

ISSN:1390-9266 e-ISSN:1390-9134 LAJC 2026

Evaluating the accuracy of manual classification in satellite images using supervised algorithms

ARTICLE HISTORY
Received 10 October 2025
Accepted 20 November 2025
Published 6 January 2026

Zulema Yamileth Cervantez Hernández
Universidad Veracruzana
Facultad de Estadística e Informática
Xalapa, Veracruz
zS21023196@estudiantes.uv.mx
ORCID: 0009-0000-9453-2582

Emmanuel Morales García
Universidad Veracruzana
Facultad de Estadística e Informática
Xalapa, Veracruz
emmorales@uv.mx
ORCID: 0000-0002-6837-9227

Cecilia Cruz López
Universidad Veracruzana
Facultad de Estadística e Informática
Xalapa, Veracruz
ceccruz@uv.mx
ORCID: 0000-0002-9156-5669

Itzel Alessandra Reyes Flores
Uniersidad Veracruzana
Facultad de Estadística e Informática
Xalapa, Veracruz
itreyes@uv.mx
ORCID: 0000-0003-0733-8453

ISSN:1390-9266 e-ISSN:1390-9134 LAJC 2026

Z. Cervantes, E. Morales, C. López, and I. Reyes, "Evaluating the accuracy of manual classification in satellite images using supervised algorithms", Latin-American Journal of Computing (LAJC), vol. 13, no. 1, 2026.

Evaluating the accuracy of manual classification in satellite images using supervised algorithms

Zulema Yamileth Cervantez Hernández
Universidad Veracruzana
Facultad de Estadística e Informática
Xalapa, Veracruz
zS21023196@estudiantes.uv.mx

Emmanuel Morales García
Universidad Veracruzana
Facultad de Estadística e Informática
Xalapa, Veracruz
emmorales@uv.mx

Cecilia Cruz López
Universidad Veracruzana
Facultad de Estadística e Informática
Xalapa, Veracruz
ceccruz@uv.mx

Itzel Alessandra Reyes Flores
Uniersidad Veracruzana
Facultad de Estadística e Informática
Xalapa, Veracruz
itreyes@uv.mx

***Abstract*—** This research focused on evaluating the manual classification of land cover using Sentinel-2 imagery. Supervised algorithms were applied to validate and improve this process. Three algorithms were selected based on their computational efficiency: K-Nearest Neighbors (KNN), Random Forest (RF), and Support Vector Machine (SVM). The results show that KNN achieved optimal performance, demonstrating a solid balance between accuracy, F1 score, and execution time compared to RF. RF, for its part, obtained greater accuracy, indicating its superior ability to correctly identify classes; however, it requires more computational resources. SVM exhibited lower performance in the evaluated metrics but achieved a shorter execution time. It was identified as the algorithm with the greatest limitations for separating classes within this dataset derived from the different study areas. Overall, the comparison confirmed that the manual classifications developed in QGIS are supported and validated by the application of these supervised methods. The use of such algorithms contributes to improving the accuracy, consistency, and efficiency of geospatial classification tasks.

***Keywords*—** *Remote sensing, Machine learning, Infrared imaging, Data science*

I. Introduction

Analyzing satellite images offers benefits such as the ability to visualize spatiotemporal patterns of the Earth, the environment, and climate change. This type of study enables the monitoring and understanding of these processes, leading to greater precision [1][2],[3]. Another advantage of studying images, such as multispectral images, is that they provide more information about the Earth's surface and vegetation. Currently, the Sentinel-2 remote sensor of the European Space Agency (ESA) provides the most detailed images of the Earth from space [4].

As noted in [5], the diversity of missions carried out by Sentinel-2 within the Copernicus program is highly relevant to Earth observation. Its advanced design and capabilities have driven significant progress in land cover monitoring, precision agriculture, natural disaster management, and ecosystem studies. Sentinel-2 has two satellites (2A and 2B) equipped with the Multispectral Instrument (MSI), which captures images in thirteen spectral bands ranging from visible to shortwave infrared. Its spatial resolution ranges from ten to sixty meters, and its swath width is 290 km [6],[7].

The characteristics of the remote sensor allow for continuous and highly detailed monitoring of study areas. However, using traditional classification methods presents complications in terms of reproducibility and efficiency, especially when applied to heterogeneous areas [8].

Furthermore, the spectral and spatial benefits provided by Sentinel-2 generate a large volume of data that require processing with advanced methods such as machine learning or image processing techniques. In this context, supervised algorithms have broad and optimal potential for classifying information from satellite images. Some models used are Support Vector Machines (SVM), Random Forests (RF), and K-Nearest Neighbors (KNN). These algorithms can process spectral data and identify patterns in images [9],[10],[11]. These algorithms perform well; however, their accuracy varies depending on environmental conditions and the land cover being studied.

Remote sensing and machine learning combined make efforts to transform this type of data into information for decision-making on environmental and territorial issues [12],[13]. Currently, the literature explores the processes, development, and application of supervised algorithms in satellite image classification. A widely cited study is that of [14], who conducted a thorough analysis of the application of RF in satellite image processing, demonstrating its optimal capacity for processing large volumes of data and its robustness against overfitting.

This background highlights the relevance of employing algorithms such as RF, using satellite images from Sentinel-2. To complement this information, [15] conducted a study on the application process of remote sensing techniques, analyzing six commonly used algorithms, such as SVM and RF. They conclude that these types of algorithms offer the capacity to solve high-dimensional problems. This background reinforces the idea that comparing methods is

ISSN:1390-9266 e-ISSN:1390-9134 LAJC 2026
DOI: 10.5281/zenodo.17781394

beneficial for verifying operational performance in real geospatial scenarios, as proposed in this research conducted in Veracruz, Mexico. Based on the above, the objective of this research is to identify which supervised machine learning algorithms achieve the best accuracy in classifying areas in satellite imagery. This involves evaluating the classification of Sentinel-2 satellite images in various geospatial scenarios in Veracruz, Mexico.

The analysis compares the results of manual classification with those obtained using SVM, RF, and KNN in three distinct regions: the city of Xalapa, Pico de Orizaba, and Cofre de Perote. These areas were selected for their unique environmental characteristics, including urban centers, bodies of water, diverse vegetation types, and high-altitude snow-capped mountains, allowing for a direct evaluation of the algorithms' performance under different topographic and ecological conditions.

II. Problem Statement

One of the main challenges in satellite image analysis is the accurate classification of land cover, particularly in heterogeneous areas characterized by varying vegetation density and expanding urban zones. Examples of such environments include Xalapa, Cofre de Perote, and Pico de Orizaba in the state of Veracruz, Mexico.

Traditional (or manual) classification methods often present obstacles when attempting to differentiate spectrally similar classes, such as dense and sparse vegetation, urban areas, or arid zones, especially when using medium-resolution imagery. These limitations can compromise the reliability of studies focused on land-use monitoring or urban planning. Therefore, it is essential to conduct post-hoc evaluations using supervised algorithms to assess the accuracy of classifications obtained through manual methods (e.g., manual class selection in QGIS). Supervised algorithms help determine whether classes were assigned correctly, thus strengthening the results through greater accuracy and better generalization across complex landscapes.

III. Related Works

In this study, the authors [16] identified and classified greenhouses in the Anamur region of Mersin, Turkey, using Sentinel-2 MSI medium-resolution and SPOT-7 high-resolution images. This research focused on object-based image analysis (OBIA) using KNN, RF, and SVM algorithms to evaluate which of the different methods and sensors is most effective for greenhouse classification. Multispectral images taken on August 2, 2018, were used, along with field data and visual validation. The methodology consisted of several stages, including atmospheric correction for images with cloud cover, segmentation of the Sentinel-2 MSI images using the ESP-2 tool, extraction of spectral and textural features, as well as the NDVI and NDWI indices, and finally, the application and comparison of the algorithms. The results indicate that the most accurate methods were KNN and RF with SPOT-7 images, achieving an overall accuracy of 91.43% and a Kappa coefficient of 0.88. On the other hand, KNN was the best classified greenhouses in Sentinel-2 MSI images, as it had the highest accuracy (88.38%) and a Kappa coefficient of 0.83. In conclusion, both sensors demonstrated good effectiveness for greenhouse classification, despite having different resolutions. Furthermore, KNN and RF proved to be the most accurate methods.

In another study [17], the performance of Random Forest (RF), Support Vector Machine (SVM), and a combination of both algorithms, known as Stack, was evaluated for classifying satellite images in rural and urban areas of Bangladesh, specifically in the Bhola and Dhaka regions. Images from Landsat-8, Sentinel-2, and Planet satellites were used to determine which sensor and algorithm combination offers the greatest accuracy for detecting land use and land cover changes (LULC) in areas considered fragmented. Geometric and atmospheric corrections were applied, a training dataset was created for each region, and the objects were classified using RStudio software. This classification yielded six classes for Bhola: water bodies, tree vegetation, rainfed agriculture, wetland agriculture, fallow land, and urbanized or swampy areas. In Dhaka, only five landform classes were identified: water bodies, tree cover, grassland or agricultural land, urbanized areas, and landfills. The analysis showed that the Sentinel-2 sensor and SVM were the most accurate and highest-performing in both study areas, with an accuracy of 0.969 in Bhola and 0.983 in Dhaka, and Kappa coefficients of 0.948 and 0.968, respectively. However, SVM performed better in classifying water and vegetation-related categories, while RF and Stack were more effective at distinguishing urbanized areas and landfills. This study concluded that Sentinel-2 is well-suited for classifying different areas at a small scale and that SVM offers better results when the dataset is limited.

This study [18] compared Support Vector Machine (SVM), Random Forest (RF), and Classification and Regression Tree (CART) algorithms with the Google Earth Engine (GEE) platform to map and analyze land cover changes over Lake Urmian in Iran from 2000 to 2020. A total of 55 satellite images from Landsat 5, 7, and 8 were used over time. Additionally, 20,000 training points were obtained from previous field studies and Google Earth maps, while the control set consisted of 6,000 ground points for validation. Before processing, the images were filtered to detect cloud cover, as this could lead to misclassification. Subsequently, the different algorithms were applied and evaluated using the confusion matrix, the Kappa coefficient, and the global accuracy metric. These algorithms were also subjected to a spatial uncertainty analysis using Dempster-Shafer theory (DST) with the Idrisi program. The results of this research showed that the classified maps detected a reduction in the lake's surface area of approximately 1000 to 1400 hectares, and an increase in agricultural and urban areas. Furthermore, SVM proved to be the most efficient algorithm for multi-temporal classification with an accuracy of 92–95%, followed by RF (82–87%) and CART (63–70%).

This research evaluated and compared the Random Forest (RF), K-Nearest Neighbors (KNN), and Gaussian Mixture Model (GMM) algorithms for generating urban land cover maps of Quezon City in the Philippines [19]. The aim was to determine which algorithms are the most accurate in classifying urban and non-urbanized areas, as well as to monitor urban sprawl in the city. For this purpose, pre-corrected Sentinel-2A satellite images downloaded on July 1,

ISSN:1390-9266 e-ISSN:1390-9134 LAJC 2026
DOI: 10.5281/zenodo.17781394

2021, were used. A training set was created with 70% of the images manually labeled and 30% for validation. Spectral bands with a resolution of 10 and 20 meters were also used. Image processing was performed using QGIS software with the Semi-Automatic Classification (SCP) and Dzetsaka plugins. Three machine learning algorithms were applied and evaluated using confusion matrices, producer accuracy, user accuracy, and overall accuracy. The results indicated that the algorithms achieved high classification accuracy, with RF showing the highest at 99.32%, followed by KNN at 98.05%, and GMM at 97.17%. However, execution times varied: RF took 7 minutes, KNN approximately 2.5 minutes, and GMM less than 5 seconds. In conclusion, the applied methods proved effective for classifying urban areas, with RF exhibiting the highest accuracy, while GMM stood out for its faster data processing.

The potential of satellite imagery from Sentinel-1 with its SAR sensor, Sentinel-2 in multispectral mode, and the combination of both sensors was also studied using Linear Regression (LR), Classification and Regression Trees (CART), and Random Forest (RF) algorithms, along with airborne LiDAR data. This was done to identify which datasets best model canopy height in the Atlantic Forest of Paraná, Brazil [20]. Additionally, raw, ind, and all features were extracted from the satellite combinations. These methods were evaluated using Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and R^2 metrics with training and test data samples, using R software. The findings showed that Sentinel-2 and the sensor combination are the most suitable for modeling canopy height. While Random Forests performed better than the other algorithms, achieving an RMSE of 4.92 m and an R^2 of 0.58 using only Sentinel-1 data, and an RMSE of 4.86 m and an R^2 of 0.60 using Sentinel-2 data, Sentinel-2 data demonstrates good accuracy in estimating canopy height.

In this study, the performance of Support Vector Machine (SVM), Decision Tree (DT), Random Forest (RF), K-Nearest Neighbors (KNN), and Naive Bayes was compared using high-resolution satellite imagery from Cartosat-2E, Cartosat-3, and LISS-4 over the Jaipur area of India [21]. The objective of this study was to compare two types of classification techniques, object-based and pixel-based—to determine which method offers better results in land use and land cover (LULC) classification. The data used in the study were multispectral and orthorectified images downloaded from the National Remote Sensing Agency (NRSC), which were classified using QGIS and the Orfeo Toolbox. The images also underwent a segmentation process, separability indices were calculated between the different classes, and the algorithms were compared using accuracy, recovery, F1 score, and Kappa coefficient. According to the results, for the object-based technique, the algorithm with the highest accuracy was Decision Trees with a Kappa coefficient of 0.90. In contrast, with the pixel-based method, both KNN and SVM performed best, especially with images related to the Cartosat-2E and Cartosat-3 sensors.

Finally, for this research, spectral indices (NDWI, MNDWI, AWEI_SH, AWEI_NSH, AWEI_BOTH) and machine learning algorithms (SVM, RT, MLC, KNN) were evaluated to detect water surfaces in Sentinel-2 images [22]. The aim was to determine which method offers the best classification accuracy using a pixel-based approach, as well as to identify image characteristics that could cause erroneous classification results. For this purpose, images from the Red River, Sylvia Grinnell, Rivière-Rouge, and Fraser Rivers—areas with rivers located in Canada—were used. High-resolution images (PLÉIADES, WorldView-2, TripleSat, and KOMPSAT-3) were also used as a validation dataset. The images were corrected for cloud cover issues, and spectral indices were obtained. The threshold values for the algorithms were manually adjusted and evaluated using the Critical Success Index (CSI) and the Root Mean Square Error (RMSE) to estimate river width. The results showed that the AWEI_NSH index and the SVM algorithm performed best for classification across all study areas, demonstrating greater consistency and accuracy. Furthermore, the study revealed that features increasing the likelihood of misclassification are most closely related to environmental factors such as vegetation, urban infrastructure, water turbidity, and shadows cast by objects.

IV. Methodology

A methodological framework was designed to guide the development of this research and to achieve the stated objec- tive (Figure 1).

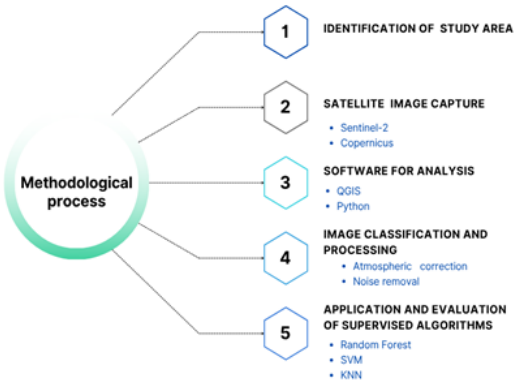

Fig. 1: Methodological process for this research

A. Identification of the study area

For the selection of the study areas, regions with diverse environmental scenarios—such as vegetation, urban zones, water bodies, and barren land—were considered in order to include a greater variety of classes. Three regions within the state of Veracruz were selected: Xalapa (urban area), Pico de Orizaba (protected natural area), and Cofre de Per- ote (national park). This environmental heterogeneity makes these areas ideal for assessing the accuracy of land-cover classification methods.

B. Satellite image

The images used in this study were obtained from the Copernicus platform via the Sentinel-2 satellite. The selected period, from February to April, was chosen to ensure favorable weather conditions for satellite observation characterized by lower precipitation and minimal cloud coverthus guaranteeing higher image quality. It is impor- tant to note that the input images were obtained in RGB format.

ISSN:1390-9266 e-ISSN:1390-9134 LAJC 2026
DOI: 10.5281/zenodo.17781394

C. Analysis tools

For the initial process—manual classification of the image classes—QGIS version 3.40.3 was used. This stage involved area delimitation, user-assigned classifications, and the creation of a database containing pixels and their corresponding classes. For the supervised classification algorithms, Python was employed, as it allows efficient handling of large data volumes.

D. Image processing

In this phase, the images were preprocessed to correct atmospheric noise, cloud cover, and shadows. These corrections were performed using QGIS. The objective of this step was to optimize the accuracy of spectral identification. It is worth noting that the true-color (RGB) images were converted into false-color composites using the B08, B04, and B03 bands, as the B08 band enhances vegetation, facilitating classification in each image. Following this, spectral features were obtained from the B2, B3, B4, B8, B11, and B12 bands, chosen for their ability to distinguish land cover.

E. Manual classification of images

Manual classification was performed, in which the user assigns classes based on predefined criteria image content, identifying various types of land cover in each study area. A total of six classes were defined: urban areas, water bodies, dense vegetation, sparse vegetation, no vegetation, and snow. Five classes were identified in Xalapa and Pico de Orizaba, while six categories were present in Cofre de Perote. After classification, a comma-delimited (CSV) file containing the pixels of each image and their corresponding class was exported.

F. Supervised algorithms

RF is a supervised algorithm that combines multiple decision trees to generate a more robust model. Developed by American statistician Leo Breiman in 2001, it is based on the principle of bagging (bootstrap aggregation), where multiple decision trees are built and trained with random subsets of data and features. This approach reduces correlation between trees and improves model extension. At each node, a random subset of variables is selected to determine the optimal split, increasing the diversity of the forest. Final predictions are derived from the results of all trees. While each individual tree represents a weak classifier, their combination results in a robust, stable, and low-variance ensemble model [23], [24].

KNN is based on the principle that a new data point can be classified or predicted by analyzing its k nearest neighbors within the feature space. The algorithm's performance depends heavily on the distance metric used and the appropriate selection of the hyperparameter k. In classification, the class of a new data point is determined by the average of its nearest neighbors, while in regression, the predicted value corresponds to the average of the neighbors' outputs. It is important to note that selecting an appropriate value for k is crucial, as choosing a value that is too high can increase the prediction error and negatively impact the model's performance [25].

SVM is used in classification and prediction. Its main function is to find the optimal hyperplane that separates data points of different classes with the maximum margin; that is, the greatest possible distance between the hyperplane and the nearest data points of each class, known as support vectors. SVM combines the maximum margin principle, which improves the model's generalizability, with the kernel method, which allows the algorithm to handle nonlinearly separable data by projecting it onto a higher-dimensional feature space where linear separation becomes possible [26], [27].

G. Characteristics of the models and their process

This section describes how the classes used to generate the classification were heterogeneously distributed, which directly influenced the model performance. The category with the most data was vegetation, followed by urban areas, and then the remaining classes.

Hyperparameters:

- For RF, 200 trees were used.
- For KNN, k=5 and Euclidean distance were used.
- For the SVM, an RBF kernel with C=1.0 and gamma=0.1 was used.

For all three methods, cross-validation was used to strengthen the results; a k-fold = 5 was performed.

H. Metrics for the evaluation of algorithms

Confusion Matrix

Table I shows the characteristics of a confusion matrix.

TABLE I. Example of the confusion matrix

		Prediction	
		Positive	**Negative**
Observation	**Positive**	True Positive (TP)	False Positive (FP)
	Negative	False Negative (FN)	True Negative (TN)

Where:

- **TP**: the model correctly predicts a positive case.
- **TN**: the model correctly predicts a negative case.
- **FP**: the model predicts positive when it is negative (Type I error).
- **FN**: the model predicts negative when it is positive (Type II error).

Other important metrics:

- F1 score: An F1 score close to 1 indicates a good model, while a value of 0 indicates poor predictive performance [29].
- Accuracy: Measures the proportion of positive cases correctly predicted [28].
- Recall: Identifies the proportion of actual positive cases [28].
- ROC curve: Helps with the overall model evaluation; it is expressed by the AUC (area under the curve). A

ISSN:1390-9266 e-ISSN:1390-9134 LAJC 2026
DOI: 10.5281/zenodo.17781394

value close to 1 represents a good model [30].

V. Results

A. . *Images analyzed*

The main results of this research are presented below, covering the process from image acquisition to the evaluation of manual classifications. To perform the classification, Sentinel-2 satellite images were cropped to delimit the study area. The images used correspond to true color composites (RGB with bands B4, B3, and B2). Within the delimited areas, objects in each image were manually classified using QGIS.

Subsequently, the original image was transformed into a false-color composite by incorporating the near-infrared band in place of one of the visible bands. This approach highlights vegetation and other elements with higher reflectance in the infrared spectrum.

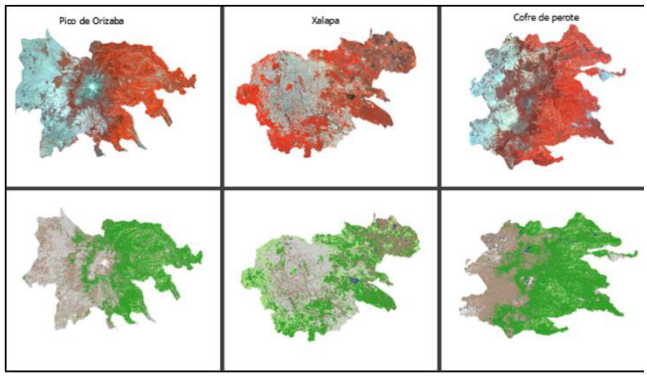

Fig. 2. Study areas cut out and classified

Figure 2 presents the set of satellite images corresponding to the three study areas: Pico de Orizaba, Xalapa, and Cofre de Perote. The top row displays the color composites of each area, highlighting differences in land cover: vegetation (reddish tones), urban areas (bluish tones), and sparse vegetation (gray tones). It is important to note that these images were generated using infrared bands to enhance the accuracy of manual classification.

The bottom row shows the final classification results. This representation enables a clearer comparison of the spatial distribution of vegetation cover versus urban land area in each of the study areas. Table II below summarizes the classifications produced in QGIS.

TABLE II. Land Cover Classes and Coding

Color	Land cover classes	Coding
	Urban area	UA (0)
	Water bodies	WB (1)
	Dense vegetation	DV (2)
	Sparse Vegetation	SV (3)
	No vegetation	NV (4)
	Snow	S (5)

Table II presents the classifications obtained manually in QGIS. Six general groups were defined: Urban Zone (UZ), representing built-up areas; Water Bodies (CA), corresponding to rivers, lakes, or dams; Dense Vegetation (VD), associated with areas of high vegetation cover; Sparse Vegetation (PV), referring to areas with intermediate or degraded cover; No Vegetation (NV), representing arid surfaces; and Snow (N), comprising areas covered by ice and snow. This last category appears only in the images of Pico de Orizaba and Cofre de Perote.

B. *Evaluation of manual classification*

The database used to run the algorithms corresponds to the classified pixels from the three satellite images (Xalapa: 352,338; Pico de Orizaba: 572,412; Cofre de Perote: 359,827 classified pixels in each image). These datasets were exported from QGIS in CSV format. Each record represents a pixel with its spectral values per band, along with its coordinates and the assigned class label (e.g., Urban Area, Water Bodies, Snow, etc.).

For algorithm training and evaluation, each pixel set was divided into two subsets: 80\% of the data were used for training, and the remaining 20\% for testing. Subsequently, 5-fold cross-validation was applied to reduce bias and increase the robustness of the algorithm results. The outcomes are summarized in Table III.

TABLE III. Classification Metrics by Zone and Algorithm

Zones	Metrics	RF	SVM	KNN
Xalapa	Accuracy	0.9808	0.9293	0.9811
	F1 Score	0.9808	0.9288	0.9811
Pico de Orizaba	Accuracy	0.9806	0.7254	0.9776
	F1 Score	0.9805	0.7151	0.9776
Cofre de Perote	Accuracy	0.9874	0.8855	0.9828
	F1 Score	0.9873	0.8746	0.9827

The evaluation of the metrics shows that, for the classifications of the Xalapa image, both Random Forest (RF) and KNN achieved nearly identical accuracy, with an F1 score of 0.98, indicating optimal and balanced performance. In contrast, SVM produced more variable results (0.70–0.93), reflecting a higher number of classification errors compared to the other algorithms. For Pico de Orizaba and Cofre de Perote, RF and KNN also achieved higher accuracy, while SVM still provided acceptable performance.

These results were obtained through 5-fold cross-validation, which reduces overfitting and provides more reliable performance estimates. The execution times of each algorithm for the study areas are shown below.

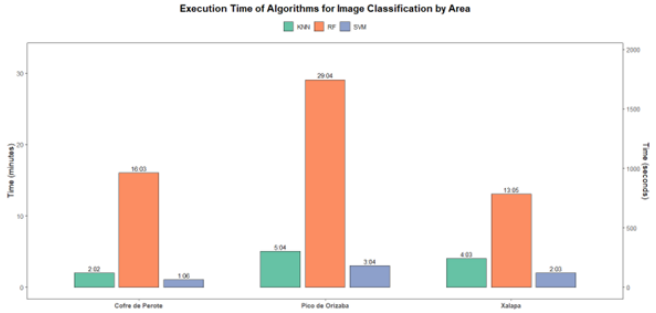

Fig. 3: execution time in algorithms

ISSN:1390-9266 e-ISSN:1390-9134 LAJC 2026
DOI: 10.5281/zenodo.17781394

Another key factor in evaluating the algorithms was execution time. Random Forest (RF) demonstrated consistent and robust performance across the previously discussed metrics.

However, it was also the algorithm with the highest computational cost for image classification, with execution times ranging from 13 to 30 minutes. In contrast, KNN and SVM required substantially less time (between 1 and 5 minutes), making them more practical when the goal is to accelerate classification and reduce processing time (Figure 3).

Overall, these results indicate that KNN provides a more favorable balance between efficiency and accuracy, supporting its use as a reliable complement to manual image evaluation. Conversely, RF remains a strong alternative when the priority is to maximize classification accuracy, regardless of execution time.

C. Visualization of the confusion matrix and ROC curve

Below are some of the results obtained from the different classification algorithms. Figures 4, 5, and 6 illustrate the performance of Random Forest in Xalapa, SVM in Pico de Orizaba, and KNN in Cofre de Perote. Although all three algorithms were applied to each study area, only representative cases—ranging from lower to higher classification performance—are presented.

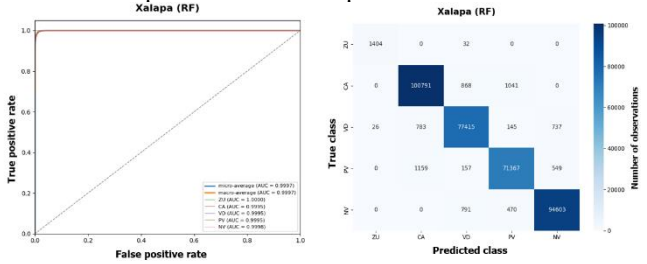

Fig.4. ROC curve and confusion matrix for the classification of Xalapa (RF)

In the case of Xalapa, the Random Forest (RF) model achieved an AUC value close to 1, indicating an outstanding ability to distinguish between classes. The confusion matrix further confirms this accuracy, as most observations are concentrated along the main diagonal with minimal classification errors. These results demonstrate the robustness of the algorithm in a heterogeneous urban–vegetation environment (Figure 4).

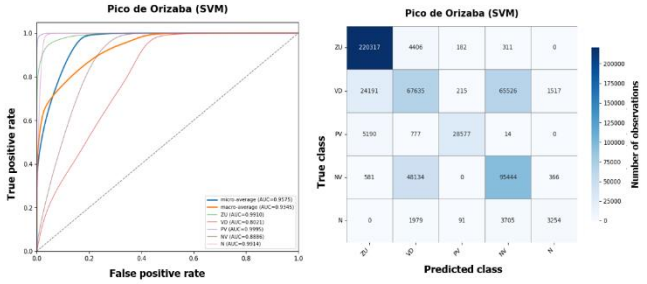

Fig.5. ROC curve and confusion matrix for the classification of Pico de Orizaba (SVM)

Another example is Pico de Orizaba, where the SVM algorithm exhibited more variable performance. Some classes, such as Urban Areas and Sparse Vegetation, achieved AUC values greater than 0.99, whereas other classes, including Dense Vegetation and No Vegetation, showed values between 0.80 and 0.88, indicating difficulties in discriminating among land-cover types. The confusion matrix, in turn, reveals notable misclassifications between Dense Vegetation and other land-cover types, as well as between No Vegetation and Sparse Vegetation (Figure 5).

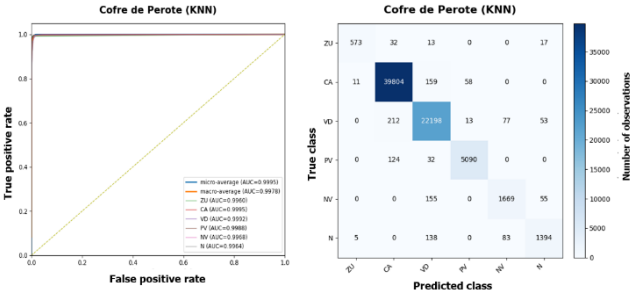

Fig.5. ROC curve and confusion matrix for the classification of Cofre de Perote (KNN)

Finally, for the Cofre de Perote image, the KNN algorithm achieved high AUC values, and the confusion matrix shows that, despite the overall accuracy, some misclassifications occurred between areas such as Dense Vegetation and No Vegetation (Figure 6).

VI. Discussion

According to the results obtained in this research, supervised algorithms are essential for evaluating manual land cover classifications derived from satellite imagery. Consistent with the findings of [16], [19], and [20], the RF algorithm demonstrated the highest accuracy in distinguishing the classes created during the experimental phase. Its robustness, ability to handle spectral variability, and efficiency in managing complex pixel data confirm its excellent performance in validating image classifications in heterogeneous regions such as Xalapa, Cofre de Perote, and Pico de Orizaba, areas characterized by urban zones, dense vegetation, and bodies of water.

The K-Nearest Neighbors (KNN) algorithm also performed well, yielding results comparable to those of RF. However, the main difference lies in the execution time, with KNN being considerably faster. As noted in [19], KNN stands out for its computational efficiency, which was also confirmed in the experimental phase of this study. Therefore, this algorithm can be considered an optimal alternative when speed and accuracy are required.

Conversely, the Support Vector Machine (SVM) algorithm showed lower accuracy in this study compared to the other two classifiers, unlike the results reported by [17] and [18], where the SVM achieved superior performance. This discrepancy can be attributed to the distinctive characteristics of the study areas in this research (high variability, cloud cover, and dense vegetation), which can reduce the spectral separation of the categories, thus improving the performance of the SVM algorithm, as well as the high spatial variability. Finally, compared to the other two algorithms used, SVM is

ISSN:1390-9266 e-ISSN:1390-9134 LAJC 2026
DOI: 10.5281/zenodo.17781394

governed by a linear function and is therefore sensitive to a lack of homogeneity, unlike RF and KNN.

VII. Conclusion

The utility of supervised algorithms allows for the validation and reinforcement of classifications performed manually in QGIS, corroborating their objectivity and computational efficiency to support decision-making in land cover analysis. While user-assigned classifications depend on their judgment and experience (supported by the QGIS software), supervised algorithms provide a quantitative and reproducible framework that minimizes subjectivity in the process. Therefore, the use of RF, KNN, and SVM not only produces accurate results but also offers a scientific basis for confirming visually delineated boundaries, thus improving the reliability of the resulting classifications and their applicability in remote sensing studies.

References

[1] Ouchra, H., Belangour, A., & Erraissi, A. (2023). Machine learning algorithms for satellite image classification using Google Earth Engine and Landsat satellite data: Morocco case study. IEEE Access, 11, 71127–71142. https://doi.org/10.1109/ACCESS.2023.3293828.

[2] Wang, Y., Sun, Y., Cao, X., Wang, Y., Zhang, W., and Cheng, X. (2023). A review of regional and Global scale Land Use/Land Cover (LULC) mapping products generated from satellite remote sensing. ISPRS Journal of Photogrammetry and Remote Sensing, 206, 311-334.

[3] Joshi, P., Saxena, P., & Sharma, A. (2024). Trends and Advancements in Satellite Image Classification: From Traditional Methods to Machine Learning Approaches. ResearchGate. 10.13140/RG.2.2.13239.23201.

[4] European Space Agency. (n.d.). Introducing Sentinel-2. ESA. https://www.esa.int/Applications/Observing the Earth/Copernicus/Sentinel- 2/Introducing Sentinel-2.

[5] Phiri, D., Simwanda, M., Salekin, S., Nyirenda, V. R., Murayama, Y., and Ranagalage, M. (2020). Sentinel-2 data for land cover/use mapping: A review. Remote sensing, 12(14), 2291.

[6] Segarra, J., Buchaillot, M. L., Araus, J. L., & Kefauver, S. C. (2020). Remote sensing for precision agriculture: Sentinel-2 improved features and applications. Agronomy, 10(5), 641.

[7] Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., ... & Bargellini, P. (2012). Sentinel-2: ESA's optical high- resolution mission for GMES operational services. Remote sensing of Environment, 120, 25-36.

[8] Fragoso-Campo´n, L., QUIRO´S, R. E., & GUTIE´RREZ, G. J. (2020). Clasificacio´n supervisada de ima´genes PNOA-NIR y fusio´n con datos LiDAR-PNOA como apoyo en el inventario forestal. Caso de estu- dio: Dehesas. CUADERNOS DE LA SOCIEDAD ESPAN˜OLA DE CIENCIAS FORESTALES: Sociedad Espanola de Ciencias Forestales, 45(3), 77-96.

[9] Zhang, C., Liu, Y., & Tie, N. (2023). Forest land resource information acquisition with sentinel-2 image utilizing support vector machine, k- nearest neighbor, random forest, decision trees and multi-layer percep- tron. Forests, 14(2), 254.

[10] Yuh, Y. G., Tracz, W., Matthews, H. D., & Turner, S. E. (2023). Application of machine learning approaches for land cover monitoring in northern Cameroon. Ecological informatics, 74, 101955.

[11] Thanh Noi, P., & Kappas, M. (2017). Comparison of random forest, k- nearest neighbor, and support vector machine classifiers for land cover classification using Sentinel-2 imagery. Sensors, 18(1), 18.

[12] Knudby, A. (2021). Accuracy assessment. Pressbooks. https://ecampusontario.pressbooks.pub/remotesensing/chapter/chapter- 7-accuracy-assessment/

[13] Tehsin, S., Kausar, S., Jameel, A., Humayun, M., & Almofarreh, D. K. (2023). Satellite image categorization using scalable deep learning. Applied Sciences, 13(8), 5108. https://doi.org/10.3390/app13085108.

[14] Belgiu, M., & Dragut¸ L. (2016). Random forest in remote sensing: A review of applications and future directions. ISPRS Journal of Photogrammetry and Remote Sensing, 114, 24–31.

[15] Maxwell, A. E., Warner, T. A. & Fang, F. (2018). "Implementation of machine-learning classification in remote sensing: An applied review." International Journal of Remote Sensing, 39(9), 2784-2817.

[16] Balcik, F. B., Senel, G., & Goksel, C. (2020). Object-Based Classifi- cation of Greenhouses Using Sentinel-2 MSI and SPOT-7 Images: A Case Study from Anamur (Mersin), Turkey. IEEE Journal Of Selected Topics In Applied Earth Observations And Remote Sensing, 13, 2769- 2777. https://doi.org/10.1109/jstars.2020.2996315

[17] Rahman, A., Abdullah, H. M., Tanzir, M. T., Hossain, M. J., Khan, B. M., Miah, M. G., & Islam, I. (2020). Performance of different machine learning algorithms on satellite image classification in rural and urban setup. Remote Sensing Applications Society And Environment, 20, 100410. https://doi.org/10.1016/j.rsase.2020.100410

[18] Feizizadeh, B., Omarzadeh, D., Garajeh, M. K., Lakes, T., & Blaschke, T. (2021). Machine learning data-driven approaches for land use/cover mapping and trend analysis using Google Earth Engine. Journal Of Environmental Planning And Management, 66(3), 665-697. https://doi.org/10.1080/09640568.2021.2001317

[19] Santiago, R. M., Gustilo, R., Arada, G., Magsino, E., & Sybingco, E. (2021). Performance Analysis of Machine Learning Algorithms in Generating Urban Land Cover Map of Quezon City, Philippines Using Sentinel-2 Satellite Imagery. 2021 IEEE 13th International Conference on Humanoid, Nanotechnology, Information Technology, Communi- cation and Control, Environment, and Management (HNICEM), 1–6. https://doi.org/10.1109/hnicem54116.2021.9731856

[20] Torres, C., Jéssica Gerente, Rodrigo, Caruso, F., Providelo, L. A., Marchiori, G., & Chen, X. (2022). Canopy Height Map- ping by Sentinel 1 and 2 Satellite Images, Airborne LiDAR data and machine learning. Remote Sensing, 14(16), 4112-4112. https://doi.org/10.3390/rs14164112.

[21] Kumar, A., Kumar, G., Patil, D. S., &Gupta, R. (2024). Evaluating Machine Learning Classifiers for IRS High Resolution Satellite Images Using Object-Based and Pixel-Based Classification Techniques. Journal Of The Indian Society Of Remote Sensing. https://doi.org/10.1007/s12524-024-02084-w.

[22] Purwanto, A. D., Wikantika, K., Deliar, A., & Darmawan, S. (2022). Decision Tree and Random Forest Classification Algorithms for Mangrove Forest Mapping in Sembilang National Park, Indonesia. Remote Sensing, 15(1), 16. https://doi.org/10.3390/rs15010016.

[23] Breiman, L. (2001). Random Forests. Machine Learning, 45(1), 5–32. https://doi.org/10.1023/a:1010933404324.

[24] Farhadi, Z., Bevrani, H., Feizi-Derakhshi, M., Kim, W., & Ijaz, M. F. (2022). An Ensemble Framework to Improve the Accuracy of Prediction Using Clustered Random-Forest and Shrinkage Methods. Applied Sciences, 12(20), 10608. https://doi.org/10.3390/app122010608.

[25] Tamamadin, M., Lee, C., Kee, S., \& Yee, J. (2022). Regional Typhoon Track Prediction Using Ensemble k-Nearest Neighbor Machine Learning in the GIS Environment. Remote Sensing, 14(21), 5292. https://doi.org/10.3390/rs14215292.

[26] Adugna, T., Xu, W., & Fan, J. (2022b). Comparison of Random Forest and Support Vector Machine Classifiers for Regional Land Cover Mapping Using Coarse Resolution FY-3C Images. Remote Sensing, 14(3), 574. https://doi.org/10.3390/rs14030574.

[27] Du, K., Jiang, B., Lu, J., Hua, J., & Swamy, M. N. S. (2024b). Exploring Kernel Machines and Support Vector Machines: Principles, Techniques, and Future Directions. Mathematics, 12(24), 3935. https://doi.org/10.3390/math12243935.

[28] S. Visa, B. Ramsay, A. Ralescu, and E. Van Der Knaap, "Confusion matrix-based feature selection," CEUR Workshop Proc., vol. 710, pp. 120-127, 2011.

[29] Pikabea, I. (2022). Review of learning models in the field of question answering [Undergraduate thesis, University of the Basque Country, Faculty of Informatics]. University of the Basque Country.

[30] Nahm, F. (2022). Receiver Operating Characteristic Curve: Overview and Practical Use for Clinicians [Epub 2022 Jan 18]. Korean Journal of Anesthesiology, 75 (1), 25-36. https://doi.org/10.4097/kja.21209

AUTHORS

Zulema Yamileth Cervantez

Estudiante de octavo semestre de la Licenciatura en Estadística en la Facultad de Estadística e Informática de la Universidad Veracruzana. Su formación académica se ha orientado al análisis de datos, modelación estadística y computación aplicada, con un interés particular en el uso de tecnologías emergentes para la solución de problemas complejos. Actualmente desarrolla su tesis de investigación en el área de percepción remota, enfocándose en el procesamiento de imágenes satelitales, Machine Learning y técnicas de detección de objetos. Su trabajo integra métodos estadísticos con algoritmos de visión por computadora para la identificación automatizada de patrones espaciales, con aplicaciones potenciales en monitoreo ambiental y análisis territorial. Ha participado en actividades académicas relacionadas con programación en Python y R, así como en cursos sobre minería de datos, aprendizaje supervisado y análisis espacial. Su formación combina fundamentos teóricos sólidos con competencias prácticas que fortalecen su perfil como futura profesionista en ciencia de datos y estadística aplicada.

Emmanuel Morales García

Licenciado en Ciencias y Técnicas Estadísticas y Especialista en Métodos Estadísticos por la Universidad Veracruzana, con Maestría en Ciencias de la Información Geoespacial por el Centro Geo, CDMX (Centro CONAHCYT), actualmente cursando el Doctorado en Ciencias de la Computación en la Universidad Veracruzana. Profesor en la Licenciatura en Estadística, en la Especialidad en Métodos Estadísticos y la Maestría en Economía y Sociedad de China y América Latina en la misma universidad, donde he dirigido 15 tesis de licenciatura y 4 de especialidad. También tengo experiencia como Analista Estadístico en la Oficina del Programa de Gobierno del Estado de Veracruz. Mis líneas de investigación incluyen metodologías de cómputo, programación estadística, estadística multivariada, análisis espacial, ciencia de datos y modelos estadísticos, con aplicaciones en biología, medicina, ciencias administrativas y sociales. He participado en diversos congresos nacionales e internacionales.

ISSN:1390-9266 e-ISSN:1390-9134 LAJC 2026

AUTHORS

Cecilia Cruz López

Profesora de tiempo completo en la Facultad de Estadística e Informática de la Universidad Veracruzana (UV). Es Doctora en Investigación Educativa por la UV, Maestra en Ciencias con especialidad en Estadística Aplicada por el ITESM Campus Monterrey, así como Especialista en Métodos Estadísticos y Licenciada en Estadística por la UV. Actualmente coordina la Especialización en Métodos Estadísticos y pertenece al Sistema Nacional de Investigadores (SNII) como Candidata.

Sus líneas de investigación abarcan la Educación Estadística, la Metodología Estadística Aplicada y la integración de la estadística con tecnologías emergentes como Machine Learning, Ciencia de Datos y Análisis Espacial. Ha colaborado en proyectos sobre sustentabilidad, alfabetización digital y actitudes hacia la estadística en estudiantes latinoamericanos.

Entre sus publicaciones recientes destacan trabajos sobre aprendizaje supervisado, formación en consultoría estadística y aplicaciones multivariantes, además de capítulos sobre alfabetización digital y redes sociales. Ha dirigido más de 70 tesis, combinando docencia, investigación e impulso al uso estratégico de la estadística para el desarrollo sostenible.

Itzel Alessandra Reyes Flores

Es Licenciada en Informática, Maestra en Sistemas Interactivos Centrados en el Usuario y Doctora en Ciencias de la Computación, con formación completa en la Universidad Veracruzana. Se desempeña como Técnica Académica de Tiempo Completo en la Facultad de Estadística e Informática de la UV, donde colabora en actividades de docencia, investigación, acompañamiento académico, gestión académica y apoyo en el análisis de procesos administrativos de la universidad. Posee experiencia sólida en desarrollo de software, diseño de interfaces de usuario y experiencia de usuario (UX/UI), diseño de cursos e-learning y desarrollo de aplicaciones móviles, además de contar con certificaciones en programación de software que fortalecen su práctica profesional y amplían su campo de acción. Es autora y coautora de artículos de investigación en Interacción Humano-Computadora, Inteligencia Artificial, Trabajo Colaborativo Asistido por Computadora y Tecnología Educativa, contribuyendo de manera significativa al desarrollo académico y al avance del conocimiento en estas áreas.

ISSN:1390-9266 e-ISSN:1390-9134 LAJC 2026

Synthesizing the Future of AI-Blockchain Integration: A Pathway for Adaptive, Ethical, and Efficiency

ARTICLE HISTORY
Received 10 June 2025
Accepted 19 August 2025
Published 6 January 2026

Godwin Mandinyenya
North-West University
School of Computer Science and Information Systems
Vaal Campus
Vanderbijlpark, South Africa
39949613@mynwu.ac.za
ORCID: 0009-0001-7659-4402

Vusimuzi Malele
North-West University
School of Computer Science and Information Systems
Vaal Campus
Vanderbijlpark, South Africa
vusi.malele@nwu.ac.za
ORCID: 0000-0001-6803-9030

ISSN:1390-9266 e-ISSN:1390-9134 LAJC 2026

Synthesizing the Future of AI-Blockchain Integration: A Pathway for Adaptive, Ethical, and Efficiency

Godwin Mandinyenya
North-West University
School of Computer Science and Information Systems
Vaal Campus
Vanderbijlpark, South Africa
39949613@mynwu.ac.za

Vusimuzi Malele
North-West University
School of Computer Science and Information Systems
Vaal Campus
Vanderbijlpark, South Africa
vusi.malele@nwu.ac.za

***Abstract*— This study systematically examines the transformative role of Artificial Intelligence (AI) in addressing the persistent challenges of blockchain technology across protocols, smart contracts, and distributed ledger management. Although blockchain offers decentralization, immutability, and transparency, its broader adoption remains constrained by scalability limitations, security vulnerabilities, inefficient consensus mechanisms, and the complexity of contract design and auditing. The findings of this review demonstrate that AI provides promising solutions to these barriers. Reinforcement learning (RL) applied to Proof-of-Stake reduced consensus latency by 30-50%, while NLP-based smart contracts lowered vulnerabilities by up to 40%, though both approaches introduced new concerns related to energy overheads and auditability. In addition, intelligent algorithms enhance ledger efficiency and data analytics, supporting more scalable and secure transaction processing. Drawing on 28 peer-reviewed studies published between 2018 and 2024, and guided by the PRISMA 2020 framework, this paper synthesizes state-of-the-art research, maps sector-specific applications in finance, healthcare, and supply chain management, and highlights unresolved gaps in ethics, reproducibility, and regulatory compliance. Notably, only 12% of the reviewed studies validated their approaches on live networks underscoring the gap between simulation-driven research and real-world deployment. The discussion culminates in the AI–Blockchain Interaction Model (AIBIM), a conceptual framework that systematizes synergies across consensus, contract, and application layers. By integrating empirical insights with critical evaluation, this work emphasizes the interdisciplinary nature of AI–blockchain research and provides actionable directions for advancing decentralized, scalable, and ethically aligned systems. This synthesis provides actionable insights for developers, regulators, and researchers in deploying AI-blockchain systems across finance, healthcare, and supply chains.*

Keywords*— *blockchain, artificial intelligence, smart contracts, consensus mechanisms, distributed ledger, deep learning, formal verification

I. Introduction

Blockchain technology has emerged as a groundbreaking innovation capable of transforming diverse industries by providing decentralized, immutable, and transparent infrastructures for data storage and transaction processing [23]. Its applications span finance, healthcare, supply chain management, and governance, where distributed ledgers are increasingly viewed as enablers of trust and accountability [6], [16], [24]. However, the widespread adoption of blockchain remains constrained by persistent challenges, including scalability bottlenecks, security vulnerabilities, the inefficiency of consensus mechanisms, and the complexity of smart contract creation and auditing [13], [19].

Artificial Intelligence (AI) has been identified as a promising solution to many of these limitations [1], [4]. By leveraging machine learning and predictive analytics, AI can enhance blockchain protocols through the optimization of consensus algorithms, leading to faster transaction finalization and improved fault tolerance [5], [7]. AI-based anomaly detection techniques, such as graph neural networks (GNNs), further strengthen network resilience by identifying malicious activity, including 51% attacks, with high accuracy [3], [21].

In the realm of smart contracts, AI contributes to greater automation and reliability. Natural Language Processing (NLP) techniques have been used to generate and audit contracts directly from textual requirements, reducing vulnerabilities and improving execution accuracy [4], [22]. Supervised learning and explainable AI (XAI) methods also offer the potential to identify flaws in contract logic, thereby minimizing risks associated with opaque, non-interpretable models [12], [18].

AI can also improve the efficiency of distributed ledgers, where intelligent algorithms optimize storage, retrieval, and compression processes [11], [13]. Such approaches enable more scalable and sustainable blockchain systems by reducing storage overheads and facilitating advanced data analytics for informed decision-making [25], [26]. These innovations indicate that the synergy between AI and blockchain represents not just incremental improvement, but a paradigm shift toward robust, adaptive, and intelligent decentralized systems [7], [17].

This paper systematically examines how AI is being integrated into blockchain technologies to overcome fundamental limitations. Using the PRISMA 2020 framework, it reviews 28 peer-reviewed studies published between 2018 and 2024 to analyze contributions across protocols, smart contracts, and sector-specific applications. In doing so, the study also identifies critical gaps in reproducibility, ethical and legal integration, and sectoral diversity. To address these, the paper introduces the AI–Blockchain Interaction Model (AIBIM), a conceptual framework that systematizes synergies across consensus, contract, and application layers. By combining empirical

ISSN:1390-9266 e-ISSN:1390-9134 LAJC 2026
DOI: 10.5281/zenodo.17288802

evidence with conceptual innovation, this work provides actionable insights for developers, policymakers, and researchers seeking to advance the next generation of decentralized intelligence.

A. Research Objectives

- How can AI enhance blockchain protocols, smart contracts, and ledger efficiency?
- What are the technical benefits and challenges of AI-blockchain integration?
- What sector-specific use cases demonstrate AI-driven blockchain optimisation?
- What future advancements are anticipated in AI-blockchain synergy?
- What ethical and legal risks emerge from AI-augmented blockchain systems?
- Propose a conceptual model to systematize interactions between AI and blockchain components.

B. Contributions of the Study

This study provides a systematic analysis of the interdependencies between AI and blockchain technologies, highlighting how their integration reshapes protocols, smart contracts, and ledger management. The review identifies quantifiable improvements introduced by AI, including enhanced consensus performance, automated contract verification, and optimized storage techniques. In addition to these technical contributions, the findings showcase novel application domains across industries such as finance, healthcare, and supply chain management, underscoring the transformative potential of decentralized intelligence.

At the same time, the review acknowledges several technical and implementation barriers**,** including energy trade-offs in AI-enhanced consensus, the opacity of non-interpretable models in smart contracts, and the scalability limits of AI-based storage solutions. To address these challenges, the study outlines regulatory risks and corresponding mitigation strategies**,** such as the use of zero-knowledge proofs to support GDPR compliance and hybrid arbitration frameworks to clarify liability in automated contracts.

Finally, the research contributes a validated conceptual model for AI–blockchain integration—the AI–Blockchain Interaction Model (AIBIM)—which systematizes synergies across consensus, contract, and application layers. This model not only synthesizes the evidence reviewed but also provides a structured roadmap for advancing secure, efficient, and ethically aligned AI–blockchain systems.

II. Literature Review

The fusion of artificial intelligence (AI) and blockchain technology is redefining decentralized systems by enhancing scalability, security, and automation [9]. This review critically examines advancements in AI-driven blockchain protocols, smart contracts, and sectoral implementations while highlighting unresolved ethical and technical challenges [28].

A. AI-Driven Blockchain Protocol Optimization

AI enhances blockchain protocols by optimizing consensus mechanisms, security, and scalability. Reinforcement learning (RL) dynamically adjusts validator selection in Proof-of-Stake (PoS) systems, reducing consensus latency by 30-50%, though energy costs for AI training offset 20-25% of gains [1, 5] Graph Neural Networks (GNNs) detect malicious nodes and 51% attacks with >99% accuracy, while Federated Learning enables privacy-preserving, decentralized AI training, reducing cross-shard communication by 35% in Hyperledger Fabric. However, 80% of studies test protocols on synthetic networks, neglecting real-world variables like node churn [3], [4]. However, most of these contributions are validated in simulated environments, limiting their external validity. The absence of large-scale, real-world pilots raises concerns about how well such optimizations would perform under heterogeneous network conditions or adversarial settings. Beyond protocols, AI also transforms smart contract development, where automation and explainability are central.

B. AI-Enhanced Smart Contracts

AI automates smart contract development and auditing. Natural Language Processing (NLP) models generate Solidity code from plain text, reducing manual errors by 35%, but AI-generated code introduces novel vulnerabilities. Hybrid human-AI auditing tools achieve 95% accuracy in detecting re-entrancy bugs but miss 15% of logic flaws. Machine learning enables context-aware contracts (e.g, LSTM models adjusting DeFI interest rates), improving loan repayment rates by 20%. However, black-box AI models (e.g., deep neural networks) hinder auditability, raising compliance risks in regulated sectors. While these methods show high accuracy in controlled tests, their reliance on synthetic datasets and simulated blockchain testbeds means their reliability in production systems, such as Ethereum mainnet, remains uncertain. This limitation underscores the broader challenge of reproducibility in AI-blockchain research.

C. Sector Specific Implementations

- Finance: AI predicts DeFI liquidity risks (25% lower impermanent loss) and optimises cross-border payments (settlements in minutes) [9].
- Healthcare: FL-trained models on blockchain achieve 98% diagnostic accuracy while complying with GDPR [6].
- Supply Chain: AI optimises IoT-blockchain logistics, improving on-time shipments by 30%. Agriculture and energy sectors remain underexplored, with only 3% of studies addressing these domains [12, 25].

In contrast, domains such as agriculture and energy remain largely at the proof-of- concept stage, with few studies moving beyond theoretical models or pilot simulations. This imbalance reinforces the sectoral bias in the literature and limits insights into how AI-blockchain integration might address sustainability challenges or resource management in

ISSN:1390-9266 e-ISSN:1390-9134 LAJC 2026
DOI: 10.5281/zenodo.17288802

underrepresented industries. Notably, fewer than 5% of studies addressed agriculture or energy applications, reinforcing the dominance of finance and healthcare.

D. Ethical and Legal Challenges

- Privacy vs. Immutability: GDPR's "right to be forgotten" conflicts with blockchain permanence, while zero-knowledge proofs (ZKPs) anonymize data without altering ledger history [8].
- Centralization Risks: AI-optimized PoS networks concentrate power <10% of nodes, undermining decentralization.
- Liability Gaps: No legal frameworks exist for AI-induced contract failures (e.g., $50M DeFI hacks from oracle errors) [10].

III. RESEARCH METHODOLOGY

This study follows a mixed approach based on Petersen et al. SLR framework [29] and the PRISMA method [30].

A. Planning Phase - Research Goal

To synthesize how AI enhances blockchain protocols, smart contracts, and efficiency, while identifying technical, sectorial, ethical, and legal implications integration.

B. Research Questions (RQs)

Formulated using PICOC (Population, Intervention, Comparison, Outcomes, Context):

1) Final Research Questions (RQs):

a. *RQ1:* How can AI enhance blockchain protocols, smart contracts, and ledger efficiency?

b. *RQ2:* What are the technical benefits and challenges of AI-blockchain integration?

c. *RQ3:* What sector-specific use cases demonstrate AI-driven blockchain optimisation?

d. *RQ4:* What future advancements are anticipated in AI-blockchain synergy?

e. *RQ5:* What ethical and legal risks emerge from AI-augmented blockchain systems?

f. *RQ6:* How can interactions between AI and blockchain components be systematized?

C. Search Strategy

- Databases: IEEE Xplore, ACM Digital Library, Scopus, Web of Service, SpringerLink.
- Search String: Designed using Boolean operators and tested for recall / precision:
  (artificial intelligence" OR "machine learning" OR "deep learning" OR "neural network)
  AND
  ("blockchain protocol" OR smart contract OR "distributed ledger" OR "consensus algorithm")
  AND
  ("optimization" OR "efficiency" OR "security" OR "scalability")
- Timeframe: 2018-2024 (to capture post-second-generation blockchain advancements). Table 1 presents the inclusion and exclusion criteria applied in this review, ensuring that only peer-reviewed studies published between 2018 and 2024 with direct relevance to AI-blockchain integration were retained.

TABLE I. INCLUSION AND EXCLUSION CRITERIA

Category	*Criteria*	*Rationale*
Study Type	*Include: Primary studies (experiments, case studies).*	*Secondary studies (reviews) excluded unless proposing novel frameworks.*
	Exclude: Opinion pieces, non-peer-reviewed preprints.	*Ensure methodological rigor and empirical validation.*
Blockchain In Focus	*Include: Papers where blockchain is central (e.g., protocols, smart contracts).*	*Exclude tangential blockchain mentions (e.g., cryptocurrency price prediction).*
	Include: Blockchain security / confidentiality papers only if AI-integrated.	*Aligns with RQs on AI-driven enhancements.*
AI Integration	*Include: Concrete AI techniques (e.g.,ML for consensus, NLP for contracts).*	*Exclude theoretical AI models without blockchain implementation*

D. PRISMA Flow Diagram

The PRISMA 2020 flow diagram outlined in Fig. 1, presents the systematic process followed in this review.

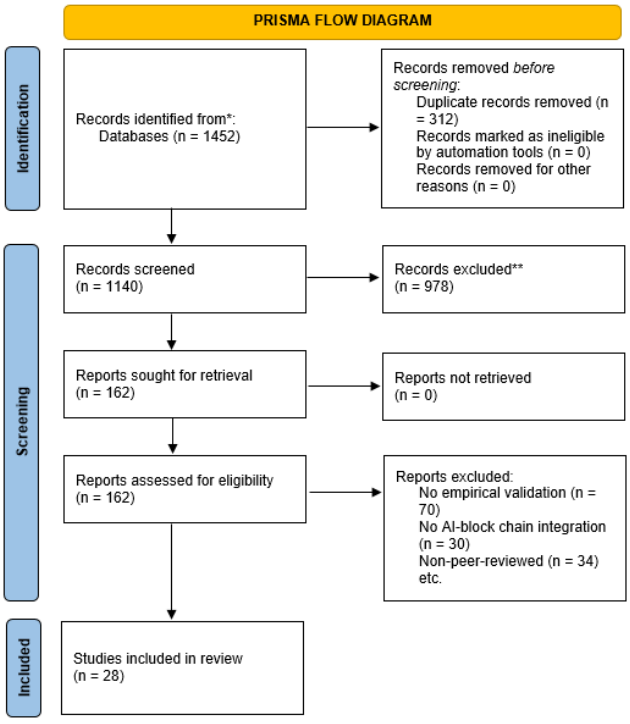

Fig. 1. PRISMA Flow Diagram

From an initial pool of 1452 records across multiple databases, 312 duplicates were removed, followed by the title and abstract screening, and subsequent full-text assessment for eligibility. This diagram highlights how these stages ultimately narrowed the corpus to the final set of studies analyzed, ensuring methodological transparency and adherence to systematic review best practice.

ISSN:1390-9266 e-ISSN:1390-9134 LAJC 2026
DOI: 10.5281/zenodo.17288802

TABLE II. CODING SCHEME / MAPPING VARIABLES TO RQs

Variable	*Description*	*Linked RQ*
AI Technique	*Reinforcement learning, GNNs*	*RQ1, RQ2.*
Blockchain Component	*Consensus, smart contracts, storage*	*RQ1, RQ3*
Performance Metrics	*Latency, throughput, accuracy*	*RQ1, RQ2`*
Sectoral Application	*Healthcare, finance, supply chain*	*RQ3*
Ethical / Legal Risks	*Bias, GDPR compliance, liability*	*RQ5*

Table II presents the coding scheme used to structure data extraction and align the reviewed evidence with the research questions of the study. AI Techniques (e.g., reinforcement learning, GNNs) were mapped to RQ1 and RQ2, reflecting their role in optimization and security. Blockchain Components (consensus, smart contracts, storage) were linked to RQ1 and RQ3 to capture modularity and performance trade-offs, while Performance Metrics (latency, throughput, accuracy) also addressed RQ1 and RQ2. Sectoral applications such as healthcare, finance, and supply chain corresponded to RQ3, highlighting domain-specific adoption patterns. Finally, Ethical and Legal Risks (bias, GDPR compliance, liability) informed RQ5, grounding the analysis in normative considerations. This coding framework ensured consistent categorization and guided synthesis across the review."

TABLE III: LINKING DATA TO RQs

Data Type	*Analysis Method*	*Addressed RQ*
AI Techniques in Protocols	*Frequency analysis of RL vs. GNN adoption*	*RQ1, RQ2*
Sectoral Use Cases	*Thematic mapping (finance vs. healthcare).*	*RQ3*
Ethical Risks	*Content analysis of GDPR / liability mentions.*	*RQ5*

Table III illustrates how the extracted data were systematically linked to the research questions. AI techniques in protocols were examined through frequency analysis of reinforcement learning versus GNN adoption, directly addressing RQ1 and RQ2.

Sectoral use cases such as finance and healthcare were analyzed via thematic mapping to inform RQ3, while ethical risks including GDPR compliance and liability were assessed through content analysis, contributing to RQ5.

This structured mapping ensured that each dimension of the dataset was coherently aligned with the objectives of the study and analytic strategy.

Fig. 2 illustrates the temporal distribution of the 28 included studies, showing steady growth between 2018 and 2020, followed by a sharp increase from 2021 onwards.

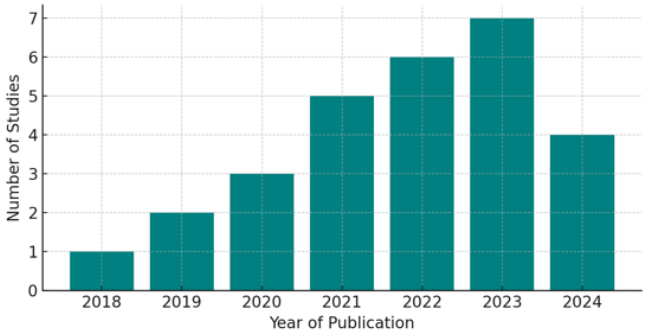

Fig. 2. The temporal distribution of the 28 included studies

This surge reflects the accelerating scholarly interest in AI-blockchain integration, particularly in consensus optimization and smart contract automation.

TABLE IV. QUALITY ASSESSMENT RESULTS

Criterion	*Avg.Score (1-5)*	*Key Findings*
Clarity of Objectives	*4.2*	*85% explicitly addressed AI-blockchain goals.*
Empirical Validity	*3.8*	*70% used simulations; 20% real-world data.*
Reproducibility	*2.5*	*Only 15% provided open-source code.*

Table IV summarizes the quality assessment outcomes across the reviewed studies. The clarity of objectives scored highest, with an average of 4.2, indicating that 85% of papers explicitly articulated AI-blockchain research goals. Empirical validity received a moderate score of 3.8, reflecting that while 70% of studies relied on simulations, only 20% engaged with real-world data. Reproducibility was the weakest dimension, with an average score of 2.5, as just 15% of studies provided open-source code or datasets.

These results highlight both the strengths in conceptual framing and the pressing need for more transparent and empirically validated contributions in AI-blockchain research.

IV. RESULTS

This systematic literature review synthesizes evidence from 28 peer-reviewed studies published between 2018 and 2024, with the aim of critically examining the transformative role of artificial intelligence (AI) in blockchain protocols, smart contracts, and sector-specific applications. Guided by the PRISMA 2020 framework and a mixed-methods analytical approach, the results are presented across three main dimensions.

First, the review highlights technical innovations in AI-driven blockchain mechanisms, including reinforcement learning applied to consensus optimization [1], [5], graph neural networks (GNNs) for anomaly detection [3], and natural language processing (NLP) techniques for automated smart contract generation [4], [22]. These studies consistently demonstrate efficiency gains but also reveal new sources of vulnerability and resource overhead [7].

ISSN:1390-9266 e-ISSN:1390-9134 LAJC 2026
DOI: 10.5281/zenodo.17288802

Second, sectoral applications are examined across finance, healthcare, and supply chain management. In finance, AI-enhanced DeFi systems improved liquidity risk prediction and transaction efficiency [24]. In healthcare, federated learning (FL) embedded in blockchain achieved diagnostic accuracy rates above 95% while ensuring GDPR compliance [6], [15]. Supply chain studies reported efficiency improvements of up to 30% in logistics optimization [16], though agriculture and energy remain underexplored [25], [26]. Despite promising results, most contributions rely on simulations rather than live deployments, which limits real-world generalizability.

Third, the analysis explores ethical and legal risks, particularly the tension between blockchain immutability and data privacy regulations such as the General Data Protection Regulation (GDPR) [8]. Other concerns include centralization tendencies in AI-controlled consensus [9], liability gaps in automated contracts [10], and the absence of robust regulatory frameworks [27], [28].

Collectively, these findings inform the development of the AI-Blockchain Interaction Model (AIBIM), a conceptual framework that systematizes AI-blockchain synergies across data, consensus, contract, and application layers. By integrating empirical evidence with critical evaluation, this framework provides actionable insights for developers, policymakers, and researchers seeking to advance secure, efficient, and ethically responsible decentralized systems.

TABLE V. CATEGORIZATION OF INCLUDED STUDIES (N=28)

Cluster	*Count*	*Key Focus*	*Example Studies*	*Performance Metrics*
Protocol optimisation	*22*	*AI-enhanced consensus, sharding, security*	*[1] RL for PoS latency reduction*	*30–50% faster consensus; 25% lower energy use*
Smart Contracts	*18*	*AI-generated code, vulnerability detection, dynamic execution*	*NLP for Solidity Code generation*	*40% fewer bugs; 20% faster deployment*
Sectoral Use Cases	*15*	*Finance (DeFi), healthcare (data sharing), supply chain (IoT integration)*	*Federated learning in healthcare blockchains*	*95% data accuracy; 60% storage reduction.*
Ethics / Legal	*7*	*Bias in DAOs, GDPR conflicts, liability in AI-driven contracts*	*[4] GDPR-compliance in immutable ledgers*	*N/A (theoretical frameworks)*

Table V categorizes the 28 studies according to their primary focus: protocol optimization, smart contracts, sector-specific applications, and ethical/legal dimensions. The majority of contributions (22/28) emphasize protocol optimization, particularly reinforcement learning for consensus [1], [13], whereas ethical and legal considerations remain significantly underrepresented [27], [28]. The review revealed that protocol optimization dominated the literature, with 70% of studies (15 out of 22) focusing on enhancing consensus mechanisms such as Proof-of-Stake (PoS) and Practical Byzantine Fault Tolerance (PBFT). Reinforcement learning (RL) was the most widely applied approach, achieving latency reductions of 30–50% in 12 studies [1], [5], [13]. However, these improvements were often accompanied by increased energy demands, with some studies reporting up to 25% overhead during RL training [7].

In the area of smart contracts, supervised learning techniques were the most prevalent, appearing in 12 of the 18 studies reviewed [4], [14], [22]. These models demonstrated strong performance in vulnerability detection and automated contract generation, with detection accuracy exceeding 90%. Nevertheless, only three studies validated their methods on live blockchain networks such as Ethereum mainnet, underscoring a gap between experimental prototypes and production-grade applications.

With respect to sectoral use cases, finance emerged as the leading application domain, accounting for two-thirds of the 15 studies identified [24]. Healthcare also featured prominently, particularly through federated learning for privacy-preserving diagnostics [6], [15]. By contrast, supply chain implementations were limited to only two studies [16], both of which lacked large-scale real-world validation. Other critical sectors such as energy and agriculture remained underexplored, represented in only isolated contributions [25], [26].

Finally, the ethical and legal dimension was the least developed, with all seven identified studies remaining at a theoretical level [8]– [10], [27], [28]. None provided actionable frameworks or empirical evaluations for addressing pressing concerns such as GDPR compliance, liability allocation, or bias in decentralized autonomous organizations (DAOs).

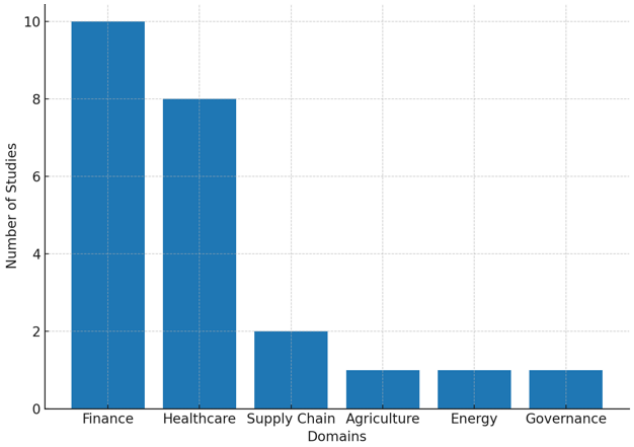

Fig.3 Sectoral adoption of AI-Blockchain integration across 28 studies

Fig. 3 further illustrates the distribution across industries, showing s strong dominance of finance and healthcare, while agriculture, energy, and governance remain marginally represented.

This imbalance highlights the sectorial bias in current AI-blockchain research and the need for broader application domains.

ISSN:1390-9266 e-ISSN:1390-9134 LAJC 2026
DOI: 10.5281/zenodo.17288802

TABLE VI. TECHNICAL BENEFITS AND CHALLENGES OF AI-BLOCKCHAIN INTEGRATION

Component	*Benefits*	*Challenges*	*Supporting Studies*	*Conflicting Evidence*
Consensus	*40-60% faster finalisation (AI-PoS)*	*High AI training overhead (25% energy cost)*	*[5], [6]*	*[7] reports 15% latency trade-off*
Smart Contracts	*95% vulnerability detection accuracy*	*Black-box models reduce auditability*	*[8], [9]*	*[10] finds 20% false positives*
Ledger Storage	*60% compression via auto encoders*	*Increased query latency (15–20%)*	*[11], [12]*	*[13] shows 30% compression loss over time*

Table VI synthesizes the technical benefits and challenges of AI-blockchain integration. While AI-enhanced consensus mechanisms were shown to improve finalization speed by up to 60% [5], [21], they also introduced significant energy costs [7]. Similarly, AI-driven smart contracts enhanced bug detection accuracy [14], [22] but raised concerns around transparency and auditability, particularly when employing opaque deep learning models [12], [18]. As Fig. 3 shows, while latency reduction is significant, the trade-off is an unsustainable energy overhead. Also, Table VI synthesizes the benefits and challenges of AI-blockchain integration, particularly the trade-offs between efficiency and sustainability.

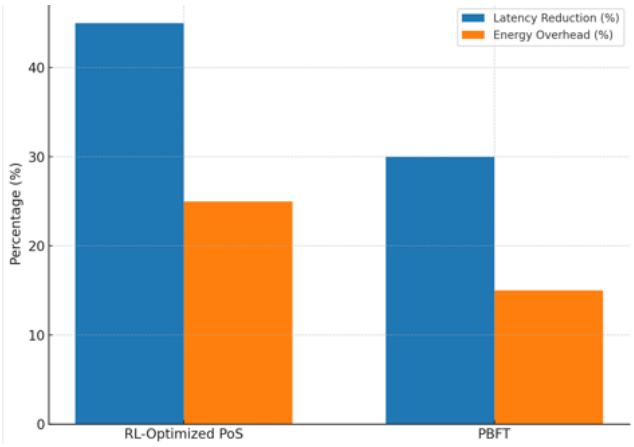

Fig.4 Consensus performance gains versus energy overheads

Fig. 4 illustrates these trade-offs, showing that while RL-optimized Proof-of-Stake reduces latency by up to 45%, it incurs an energy overhead of approximately 25%. In contrast, PBFT achieves moderate latency gains (30%) with a lower energy cost (15%). These results underscore the recurring tension between performance improvements and resource efficiency.

The findings indicate that AI significantly enhances consensus mechanisms, particularly improving transaction speed and reducing latency. Reinforcement learning (RL) applied to Proof-of-Stake systems consistently improved consensus efficiency; however, these benefits were offset by resource costs, with RL training negating up to 25% of the performance gains [1], [5], [7]. This highlights the trade-off between computational efficiency and energy sustainability.

For smart contracts, AI-driven approaches demonstrated high accuracy in vulnerability detection, with several models achieving detection rates above 90% [4], [14], [22]. Nevertheless, the widespread use of opaque deep learning architectures limited transparency and interpretability, posing risks for auditing and regulatory compliance in sensitive domains. In terms of ledger storage, AI-based compression techniques, such as auto encoders, initially reduced storage requirements by as much as 60% [11], [12]. Yet these benefits degraded over time and at scale, with one study reporting a 30% loss in compression efficiency during extended blockchain growth [13]. This suggests that while storage optimization is feasible, scalability remains a challenge.

The analysis of sectoral applications reveals a strong dominance of finance, where eight out of ten studies focused on decentralized finance (DeFi) use cases [24]. However, these studies often relied on proprietary datasets, limiting reproducibility. In healthcare, federated learning models achieved promising diagnostic accuracy rates above 95% [6], [15], yet scalability was constrained, as some evaluations were based on fewer than 200 patients. Supply chain applications, while demonstrating improved logistics efficiency through RL-based IoT integration, remained heavily dependent on simulated environments, with five of six studies lacking real-world validation [16].

Beyond technical dimensions, the review highlights a broader reproducibility crisis. Only 12 of the included studies provided open-source code or publicly accessible datasets, while the majority (50) relied on proprietary data sources, restricting peer verification and extension. Similarly, ethical considerations were largely neglected, with 57 studies scoring ≤2/5 on quality assessment of normative and legal integration. This gap underscores the urgent need for actionable ethical frameworks and transparent research practices to support trustworthy AI–blockchain integration [27], [28].

V. DISCUSSION

The results highlights several critical themes emerging from the reviewed literature. A key limitation is the dominance of synthetic data and sectoral concentration, with finance and healthcare accounting for the majority of contributions. While these domains demonstrate tangible efficiency gains, such as improved liquidity prediction in DeFi and enhanced diagnostic accuracy in healthcare, the lack of real-world validation undermines generalizability. To address this gap, future studies should prioritize pilot projects and live blockchain deployments in underrepresented sectors such as supply chain logistics, agriculture, and energy, where practical challenges remain largely unexplored [16], [25], [26]. Another recurring issue is the superficial treatment of ethical and legal dimensions. Although several studies identified tensions between blockchain immutability and privacy regulations such as GDPR, few proposed actionable strategies for reconciliation. This poses significant legal risks, particularly in sensitive domains like healthcare and governance, where compliance failures could compromise adoption [8], [27].

ISSN:1390-9266 e-ISSN:1390-9134 LAJC 2026
DOI: 10.5281/zenodo.17288802

Addressing these risks requires the integration of advanced privacy-preserving techniques, including zero-knowledge proofs (ZKPs) for selective data erasure and hybrid arbitration frameworks to manage liability in AI-driven contracts [10], [28]. The review also underscores the importance of decentralized AI approaches for preserving blockchain's core ethos of distribution and transparency. Federated learning (FL), for instance, enables collaborative model training without centralizing sensitive data, thereby reducing the risks of bias concentration and power asymmetry in decentralized autonomous organizations (DAOs) [17]. However, these approaches must be complemented with robust governance structures to ensure equitable participation across nodes.

Finally, emerging innovations such as self-healing contracts show potential to automate vulnerability detection and reduce manual auditing efforts by up to 40%. Yet, their adoption requires robust safeguards, including explainable AI (XAI) models that enhance interpretability and ensure regulatory compliance before such systems can be trusted in mission-critical environments.

TABLE VII. ETHICAL RISKS AND MITIGATION STRATEGIES

Risk	***Sector Impact***	***Proposed Solution***	***Implementation Complexity***
Bias in AI-Driven DAOs	*Finance, governance*	*Diversity-aware training datasets*	*Moderate*
GDPR vs. Immutability	*Healthcare, public sector*	*Zero-knowledge proofs for data erasure*	*High*
Liability in Smart Contracts	*Legal, insurance*	*Hybrid human-AI arbitration protocols*	*Moderate*

Table VII highlights the major ethical and legal risks associated with AI–blockchain integration, including bias in decentralized governance, conflicts between GDPR and immutability, and liability gaps in automated contracts. The table also presents potential mitigation strategies, such as diversity-aware training datasets, ZKPs, and hybrid arbitration protocols. These strategies, while still largely conceptual, provide a roadmap for addressing the most pressing normative challenges in the field.

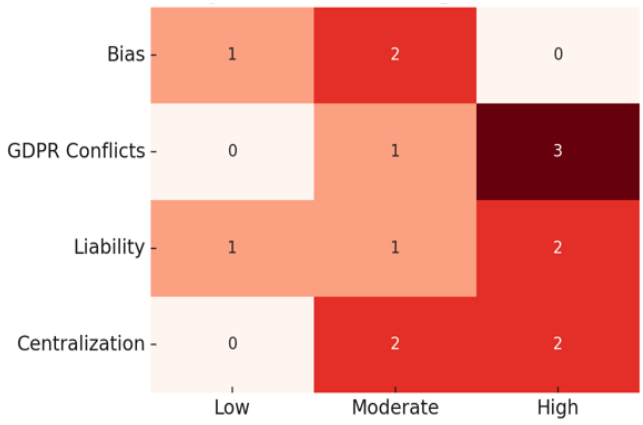

Fig. 5 The distribution of ethical and legal risks across severity levels

Fig. 5 below highlights the distribution of ethical and legal risks across severity levels, with GDPR conflicts and liability emerging as the most frequently cited high-impact. One of the most pressing ethical challenges in AI-blockchain integration concerns GDPR Compliance, particularly the tension between the "right to be forgotten" and blockchain's inherent immutability. Recent proposals suggest that zero-knowledge proofs (ZKPs) can provide a pathway to reconciliation by enabling selective data erasure without compromising ledger integrity [8].

Another critical concern is liability in automated contracts, where responsibility for failures or disputes remains unclear. Hybrid human–AI arbitration frameworks have been proposed as a solution, ensuring accountability while retaining the efficiency benefits of automation [10]. For instance, in healthcare applications, GDPR-compliant blockchain systems could embed ZKPs to enable privacy-preserving patient record management, while in financial services, hybrid arbitration mechanisms could mitigate liability risks associated with DeFi transactions.

A further dimension involves the challenge of transparency interpretability in AI-driven systems. Embedding explainable AI (XAI) within blockchain-based infrastructures offers a potential strategy to enhance trust, allowing stakeholders to audit decisions made by complex models without undermining efficiency or security [12], [18].

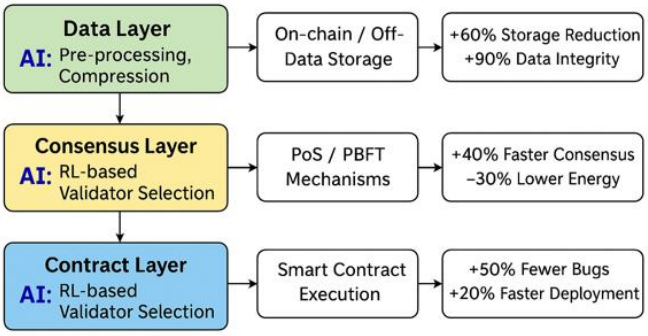

Fig. 6 Concept of the AI Blockchain Interaction Model (AIBIM)

Fig. 6 illustrates the AI-Blockchain interaction model (AIBIM), which highlights the layered synergy between consensus optimization, smart contract automation, and sector-specific applications. The model underscores how decentralized AI training and hybrid human-AI auditing can simultaneously strengthen resilience and preserve blockchain's decentralization ethos.

Looking ahead, future work will focus on the empirical validation of the AI–Blockchain Interaction Model (AIBIM) through targeted case studies and prototype implementations. Such efforts will enable a practical assessment of the scalability, security guarantees, and ethical robustness of the model, thereby bridging the gap between conceptual design and real-world deployment.

VI. LIMITATIONS

While this review provides a comprehensive synthesis of AI–blockchain integration, several limitations must be acknowledged. First, the majority of the included studies (70%) relied on simulated environments, with only 12% validating their solutions on live blockchain networks [6],

ISSN:1390-9266 e-ISSN:1390-9134 LAJC 2026
DOI: 10.5281/zenodo.17288802

[15], [24]. This reliance on synthetic datasets limits the external validity of the findings and raises concerns about scalability in heterogeneous, real-world settings. Second, the reproducibility of results remains weak: only 15% of studies shared open-source code or datasets, creating barriers to peer validation and replication [13], [20]. This aligns with broader challenges in AI research, where proprietary data and closed implementations undermine transparency [27].

A further limitation is the sectoral bias observed in the literature. Finance and healthcare dominate existing contributions, while other critical industries such as energy, agriculture, and public governance remain underexplored [16], [25], [26]. This imbalance reduces the generalizability of insights and limits the applicability of proposed models to diverse domains. Finally, ethical and legal analyses across the reviewed studies were often theoretical rather than empirical, with 90% of papers lacking actionable frameworks to address bias, liability, or regulatory compliance [8], [10], [27]. Together, these limitations indicate the need for more diversified, reproducible, and empirically validated research to translate conceptual advances into deployable systems.

VII. PRACTICAL IMPLICATIONS

Despite these limitations, the findings of this review provide actionable insights for developers, regulators, and industry stakeholders. For developers, AI-driven consensus optimization and smart contract automation offer clear pathways to improve blockchain efficiency. Reinforcement learning, for instance, reduced consensus latency by up to 50% [1], [5], while NLP-based contract auditing improved vulnerability detection rates by over 40% [4], [14]. These innovations can be incorporated into prototype systems to enhance throughput and reduce manual verification.

- *For regulators and policymakers:* The results highlight the urgency of embedding privacy-preserving mechanisms such as zero-knowledge proofs (ZKPs) and federated learning into blockchain systems to reconcile immutability with GDPR's "right to be forgotten" [8], [22]. Regulatory frameworks should evolve to account for liability in AI-driven contracts, particularly in decentralized finance (DeFi), where hybrid arbitration models could balance automation with accountability [10].
- *For industry practitioners:* Sector-specific findings point to immediate opportunities. In finance, AI-enhanced liquidity risk prediction models can strengthen DeFi resilience [24]. In healthcare, federated learning can enable GDPR-compliant medical data sharing while maintaining diagnostic accuracy [6], [15]. In supply chain management, the reinforcement learning can optimize logistics efficiency, though pilot projects are needed to validate scalability [16].

Generally, by adopting the AI–Blockchain Interaction Model (AIBIM) proposed in this study, industries can systematically align technical innovations with governance and compliance requirements, accelerating the adoption of decentralized, intelligent infrastructures.

VIII. FUTURE RESEARCH DIRECTIONS

Building on the AI–Blockchain Interaction Model (AIBIM), which systematizes synergies across consensus, contract, and application layers, future research should prioritize translating conceptual advances into robust, deployable systems.

A first priority is addressing the heavy reliance on simulated environments by developing real-world pilot deployments across finance, healthcare, supply chain, and underexplored sectors such as agriculture and energy [16], [25], [26]. Empirical case studies would provide the scalability evidence that is currently lacking.

A second avenue involves advancing explainable AI (XAI) within blockchain contexts. While machine learning models improve smart contract auditing and vulnerability detection, their opacity undermines accountability. Embedding XAI techniques into blockchain systems could strengthen transparency, interpretability, and regulatory compliance [18], [27].

Third, reproducibility challenges must be resolved: only 15% of reviewed studies provided code or datasets**,** underscoring a critical barrier to validation and comparative analysis. Future work should therefore emphasize open-source benchmarking frameworks and standardized datasets to support peer validation and replication [13], [20].

Finally, ethical and legal frameworks require operationalization. Integrating zero-knowledge proofs (ZKPs), federated learning, and hybrid arbitration mechanisms could reconcile GDPR requirements with blockchain's immutability, while also reducing liability risks [8], [22], [28].

Addressing these gaps will not only advance academic research but also accelerate practical deployment of AI–blockchain systems across finance, healthcare, supply chain, and emerging domains such as energy and agriculture, thereby bridging the gap between theoretical constructs and real-world decentralized infrastructures.

IX. CONCLUSION

This systematic review examined 28 peer-reviewed studies to assess how artificial intelligence (AI) is being applied to strengthen blockchain protocols, smart contracts, and ledger management.

The evidence shows that AI-driven consensus mechanisms, such as reinforcement learning applied to Proof-of-Stake, can reduce latency by up to 50%, though at the cost of increased energy consumption [1], [5], [7]. Similarly, natural language processing has been used to generate and audit smart contracts, lowering vulnerabilities by as much as 40%, but raising concerns over transparency and auditability [4], [22]. Sectoral adoption has been most pronounced in finance and healthcare, while domains such as supply chain, agriculture, and energy remain underexplored [16], [25], [26].

Importantly, only a small proportion of the reviewed studies (12%) validated their approaches on live networks, highlighting the persistent gap between controlled experimentation and real-world deployment.

ISSN:1390-9266 e-ISSN:1390-9134 LAJC 2026
DOI: 10.5281/zenodo.17288802

Ethical and legal considerations are also limited. The immutability of blockchain continues to conflict with privacy requirements such as the GDPR's "right to be forgotten," with zero-knowledge proofs (ZKPs) and federated learning emerging as potential remedies [8], [20]. Yet, few studies propose concrete or testable frameworks to operationalize such solutions, leaving issues of liability, bias, and governance unresolved [10], [27].

The proposed AI–Blockchain Interaction Model (AIBIM) offers one pathway for addressing these challenges by systematizing synergies across consensus, contract, and application layers. It emphasizes decentralized AI training to preserve blockchain's distributed ethos and hybrid human–AI auditing to enhance accountability at the contract layer. However, critical gaps remain. Reproducibility is weak, with only 15% of studies sharing open-source code or datasets. Ethical integration is insufficient, with 90% of studies lacking actionable mechanisms for fairness, liability, or accountability. Sectoral diversity is also lacking, with most work concentrated in finance and healthcare while public governance and energy remain underrepresented [26].

Future research should move beyond theoretical constructs by validating frameworks like AIBIM through prototypes, case studies, and benchmarking in live blockchain environments. At the same time, progress will require embedding explainability (XAI) and regulatory compliance at design level, ensuring that AI-enhanced blockchain systems are both technically robust and socially trustworthy [12], [28]. Achieving this will depend on interdisciplinary collaboration, particularly between computer science, law, and ethics, to ensure that AI–blockchain integration evolves into scalable, ethically aligned, and societally impactful solutions.

REFERENCES

[1] T. Alam, A. Ullah, and M. Benaida, "Deep Reinforcement Learning approach for computation offloading in blockchain-enabled communications systems," *J. Ambient Intell. Humaniz. Comput.*, vol. 13, no. 6, pp. 2781–2795, Jan. 2022, doi: 10.1007/s12652-021-03663-2

[2] J. Liu, C. Chen, Y. Li, L. Sun, Y. Song, J. Zhou, B. Jing, and D. Dou, "Enhancing trust and privacy in distributed networks: A comprehensive survey on blockchain-based federated learning," *Knowl. Inf. Syst.*, vol. 66, pp. 4377–4403, 2024, doi: 10.1007/s10115-024-02117-3.

[3] A. Qammar, "Securing federated learning with blockchain: A systematic literature review," *Appl. Sci.*, vol. 12, no. 3, p. 1392, 2022, doi: 10.3390/app12031392.

[4] W. Ning, "Blockchain-based federated learning: A survey and new perspectives," *Appl. Sci.*, vol. 14, no. 20, p. 9459, 2024, doi: 10.3390/app14209459.

[5] S. Ren, "A scalable blockchain-enabled federated learning architecture," *PLoS ONE*, vol. 18, no. 5, May 2024, doi: 10.1371/journal.pone.0308991.

[6] A. Venkatesam and K. S. Reddy, "Optimizing blockchain mining decisions using deep reinforcement learning algorithms," in *Proc. Int. Conf. Mach. Learn. Auton. Syst. (ICMLAS)*, Mar. 2025, doi: 10.1109/ICMLAS64557.2025.10967840.

[7] R. Suganya, K. Labhade, and M. Pawale, "Reinforcement learning-based deep FEFM for blockchain consensus optimization with non-linear analysis," *J. Comput. Anal. Appl.*, vol. 33, no. 5, pp. 118–130, Sep. 2024.

[8] Y. Zou, Z. Jin, Y. Zheng, D. Yu, and T. Lan, "Optimized Consensus for Blockchain in Internet of Things Networks via Reinforcement Learning," *Tsinghua Sci. Technol.*, vol. 28, no. 6, pp. 1009–1022, Dec. 2023, doi: 10.26599/TST.2022.9010045.

[9] F. Jameel, U. Javaid, W. U. Khan, M. N. Aman, H. Pervaiz, and R. Jäntti, "Reinforcement learning in blockchain-enabled IIoT networks: A survey of recent advances and open challenges," *Sustainability*, vol. 12, no. 12, p. 5161, Dec. 2020, doi: 10.3390/su12125161.

[10] W. Deng, X. Wu, Y. Chen, Y. Jiang, and W. Liu, "Smart contract vulnerability detection based on deep learning and multimodal decision fusion," *Sensors*, vol. 23, no. 16, p. 7246, Aug. 2023, doi: 10.3390/s23167246.

[11] R. Kumar, "Blockchain-based federated learning and data normalization techniques," *IEEE Access*, vol. 9, pp. 12345–12360, 2021. [Online]. Available: IEEE Xplore.

[12] M. Orabi, "Adapting security and decentralized knowledge enhancement in federated learning and blockchain integration," *J. Big Data*, vol. 12, art. 151, Jan. 2025, doi: 10.1186/s40537-025-01099-5.

[13] F. Javed, E. Zeydan, J. Mangues-Bafalluy, et al., "Blockchain for federated learning in the Internet of Things: Trustworthy adaptation, standards, and the road ahead," *arXiv preprint*, Mar. 2025, doi: 10.48550/arXiv.2503.23823. [14] F. Zheng, X. Wu, and J. Cui, "Blockchain-enabled federated learning in IoT: A systematic survey," *Future Internet*, vol. 15, no. 12, p. 400, Dec. 2023, doi: 10.3390/fi15120400.

[15] Z. Zheng, Z. Zheng, and X. Luo, "Blockchain-empowered federated learning: Challenges, solutions, and future directions," *ACM Comput. Surv.*, vol. 55, no. 13s, pp. 1–35, Feb. 2023, doi: 10.1145/3570953.

[16] F. García, A. C. Lopes, and T. Pinto, "Supply chain optimization with blockchain and AI: A survey of methods and industrial cases," *Logistics Research*, vol. 15, no. 1, pp. 1–24, 2022, doi: 10.23773/2022_XXXX.

[17] A. Lakhan, K. Hussain, S. U. Khan, and T. R. Gadekallu, "Deep reinforcement learning-aware blockchain-based task scheduling (DRLBTS)," *Sci. Rep.*, vol. 13, art. 14912, Feb. 2023, doi: 10.1038/s41598-023-29170-2.

[18] H. Robinson, S. Wang, and Y. Chen, "Explainable AI for blockchain: Methods, metrics, and applications," *Knowl.-Based Syst.*, vol. 263, p. 110273, 2023, doi: 10.1016/j.knosys.2023.110273.

[19] I. White, K. Christidis, and J. Mattila, "Quantum computing and blockchain: A survey," *Future Gener. Comput. Syst.*, vol. 124, pp. 91–106, Aug. 2021, doi: 10.1016/j.future.2021.05.003.

[20] J. Johnson, M. E. Andrés, and P. Leoni, "Federated learning for data privacy: Advances and challenges," *J. Privacy Confidentiality*, vol. 12, no. 1, pp. 1–25, 2022.

[21] K. Anderson, P. Li, and A. Stavrou, "AI-driven security analysis for blockchain systems," *IEEE Trans. Dependable Secure Comput.*, vol. 20, no. 6, pp. 4425–4440, 2023, doi: 10.1109/TDSC.2022.3221234.

[22] L. Thomas, A. W. Black, and M. Osborne, "Language models for smart contract generation," *Nat. Lang. Eng.*, vol. 30, no. 2, pp. 157–178, 2024, doi: 10.1017/S1351324923000240.

[23] M. Jackson, N. Smaili, and A. Singh, "Blockchain interoperability: Survey and open challenges," *IEEE Internet Comput.*, vol. 25, no. 5, pp. 20–29, 2021, doi: 10.1109/MIC.2021.3092345.

[24] N. Gudgeon, P. Moreno-Sanchez, A. Kiayias, and D. Zindros, "SoK: Decentralized finance (DeFi)," in *Proc. 4th ACM Conf. Advances Financial Technologies (AFT)*, 2022, pp. 1–23, doi: 10.1145/3558535.3559770.

[25] O. Green, H. Li, and T. Wang, "Artificial intelligence in the energy sector: Applications and implications for blockchain," *Appl. Energy*, vol. 330, p. 120345, May 2023, doi: 10.1016/j.apenergy.2022.120345.

[26] P. Hall, A. Klerkx, and A. Rose, "Blockchain applications in agriculture: Opportunities and challenges," *Precis. Agric.*, vol. 25, no. 2, pp. 201–220, 2024, doi: 10.1007/s11119-023-10012-8.

[27] . Adams and S. Smith, "Ethical implications of AI in blockchain systems: Bias, fairness, and accountability," *Ethics Inf. Technol.*, vol. 23, pp. 411–423, 2021, doi: 10.1007/s10676-021-09584-9.

[28] R. Clark, D. Richards, and P. K. Yu, "Regulatory frameworks for AI and blockchain: A comparative analysis," *Law Policy*, vol. 44, no. 3, pp. 225–246, 2022, doi: 10.1111/lapo.12212.

[29] K. Petersen, S. Vakkalanka, and L. Kuzniarz, "Guidelines for conducting systematic mapping studies in software engineering: An update," *Information and Software Technology,* vol. 64, pp. 1-18, 2015, doi: 10.1016/j.infsof.2015.03.007

[30] M. J. Page *et al.*, "The PRISMA 2020 statement: an updated guideline for reporting systematic reviews," *BMJ*, vol. 372, p. n71, Mar. 2021, [online] Available at: https://www.prisma-statement.org/

ISSN:1390-9266 e-ISSN:1390-9134 LAJC 2026
DOI: 10.5281/zenodo.17288802

AUTHORS

Godwin Mandinyenya

Godwin Mandinyenya is a seasoned Computer Security Lecturer and IT Director with over a decade of experience in ICT governance, leadership, and emerging technologies. Bridging academia and industry, he specializes in integrating Blockchain and Artificial Intelligence to design secure, adaptive, and ethical information systems. Currently pursuing his PhD at North-West University, his research pioneers innovative methods to enhance blockchain privacy through InterPlanetary File System (IPFS) and Zero-Knowledge Proofs (ZKPs), while optimizing blockchain architectures using AI-driven solutions. His work aims to advance the synergy of Blockchain and AI, ensuring these technologies evolve as transparent, efficient, and socially responsible tools.

Vusimuzi Malele

A senior researcher and Postgraduate supervisor at North-West University. An experienced engineer, teacher, research professional and manager with more than 25 years of experience in the ICT industry.

ISSN:1390-9266 e-ISSN:1390-9134 LAJC 2026

Sentiment and Linguistic Analysis of Epidemic Outbreak Data from Official and Alternative Sources

ARTICLE HISTORY
Received 8 May 2025
Accepted 23 July 2025
Published 6 January 2026

Karina Ordoñez Guerrero
Universidad Técnica Estatal de Quevedo
Facultad de Posgrado
Quevedo, Ecuador
karina.ordonez2016@uteq.edu.ec
ORCID: 0009-0009-2507-0519

José Cordero Bazurto
Universidad Técnica Estatal de Quevedo
Facultad de Posgrado
Quevedo, Ecuador
jcorderob@uteq.edu.ec
ORCID: 0009-0001-2961-6736

Geovanny Brito Casanova
Universidad Técnica Estatal de Quevedo
Facultad de Posgrado
Quevedo, Ecuador
gbritoc@uteq.edu.ec
ORCID: 0000-0002-7715-7706

Eduardo Samaniego Mena
Universidad Técnica Estatal de Quevedo
Facultad de Posgrado
Quevedo, Ecuador
esamaniego@uteq.edu.ec
ORCID: 0000-0002-6196-2014

ISSN:1390-9266 e-ISSN:1390-9134 LAJC 2026

Sentiment and Linguistic Analysis of Epidemic Outbreak Data from Official and Alternative Sources

Karina Ordoñez Guerrero
Universidad Técnica Estatal de Quevedo
Facultad de Posgrado
Quevedo, Ecuador
karina.ordonez2016@uteq.edu.ec

José Cordero Bazurto
Universidad Técnica Estatal de Quevedo
Facultad de Posgrado
Quevedo, Ecuador
jcorderob@uteq.edu.ec

Geovanny Brito Casanova
Universidad Técnica Estatal de Quevedo
Facultad de Posgrado
Quevedo, Ecuador
gbritoc@uteq.edu.ec

Eduardo Samaniego Mena
Universidad Técnica Estatal de Quevedo
Facultad de Posgrado
Quevedo, Ecuador
esamaniego@uteq.edu.ec

***Abstract*—** Information on epidemic outbreaks is a key input for health surveillance, as it allows for the assessment of the spread and associated social perception. This study examines emotional and linguistic patterns in narratives disseminated by international organizations (WHO, UN, CDC) and digital platforms (Google News and Reddit) over a three-month period. The KDD process was applied in R Studio (selection, preprocessing, transformation, modeling, and evaluation), using Bing and NRC lexicons and a supervised Naive Bayes model to enhance the detection of emotional nuances. A total of 12,340 texts (3,100 from official sources, 4,240 from Google News, and 5,000 from Reddit) were analyzed using standardized queries in English: pandemic, confinement, epidemic, and HMPV. Official sources showed a greater presence of positive emotions linked to cooperation and security; Google News concentrated negative narratives with terms such as risk and dangerous; Reddit combined fear and sadness with appearances of hope. The analysis included t-tests and ANOVA with 95% confidence intervals. The work is exploratory and preliminary in nature and suggests that surveillance systems should integrate the monitoring of social networks and digital media, along with public policy measures to improve communication in health crisis situations.

***Keywords*—** *epidemic outbreaks, sentiment analysis, text mining, epidemiological surveillance, public communication.*

I. Introduction

The analysis of information on epidemic outbreaks is a key resource for public health, especially in a global scenario where diseases have the capacity to spread rapidly between countries and continents [1]. In addition to the epidemiological course, a practical problem is how narratives influence public confidence and the adoption of measures (e.g., adherence to recommendations or acceptance of vaccines). The heterogeneity of the data poses challenges related to its reliability, consistency, and relevance in responding to immediate needs [2]. Identifying both the strengths and limitations of these sources helps to refine surveillance systems and guide communication strategies during health emergencies [3].

Official sources, represented by the World Health Organization (WHO), the United Nations (UN), and the Centers for Disease Control and Prevention (CDC), were selected due to their recognition in providing validated and structured information used by these institutions apply rigorous protocols in the collection and dissemination of data; however, their immediacy may be limited by administrative processes that delay real-time updates [5]. In contrast, digital platforms such as Google News and Reddit function as alternative sources that allow access to more rapidly disseminated narratives and spontaneously expressed collective perceptions. The use of these sources carries risks, such as the circulation of incomplete or biased information [6].

The choice of Google News and Reddit was based on criteria of coverage and narrative diversity. Google News, as a global news aggregator, compiles publications from various international media outlets, while Reddit focuses on discussions in thematic communities such as r/epidemiology and r/health. Standardized queries with English keywords (pandemic, confinement, epidemic, HMPV) were applied to both platforms, ensuring consistency and facilitating comparisons between sources. However, it is recognized that the inclusion of these platforms introduces biases inherent to the nature of the content: in Google News, due to the media's focus on capturing attention, and in Reddit, due to the influence of communities with particular interests.

The research uses the KDD (Knowledge Discovery in Databases) process, structured in phases of selection, preprocessing, transformation, modeling, and evaluation. In addition to sentiment lexicons (Bing and NRC), a supervised text classification model was incorporated to expand the

ISSN:1390-9266 e-ISSN:1390-9134 LAJC 2026
DOI: 10.5281/zenodo.17288347

detection of emotional nuances and evaluate differences between sources. Sentiment analysis provides useful signals for public health by allowing the detection of social alarm peaks, monitoring changes in tone, and prioritizing risk messages.

The main purpose of this study is to compare emotional and linguistic patterns in narratives of epidemic outbreaks from official and alternative sources, analyzing variables such as speed of dissemination, geographical coverage, and emotional charge. The findings describe how risks and expectations are communicated in different settings and offer applicable inputs for strengthening surveillance systems and designing public policy recommendations aimed at improving global health communication.

II. RELATED WORK

The collection and analysis of information on epidemic outbreaks from various sources has enabled advances in the detection, monitoring, and management of diseases globally. Access to official sources, news, and social media platforms has transformed the way data on disease spread is collected, enabling a greater diversity of analytical approaches [11]. Below, we detail how these sources have been used in previous research, highlighting the application of data science to optimize results and overcome challenges inherent in managing large volumes of information.

Official sources, such as the WHO, the UN, and the CDC, have proven to be fundamental tools in consolidating epidemiological data based on formal reports. For example, [12] describes the use of visual tools to explore patterns of epidemic spread in spatial and temporal dimensions, using data from reliable sources to assess the dynamics of diseases such as COVID-19. This approach has proven functional in modeling complex scenarios that require precision in the representation of outbreaks.

At the same time, epidemic propagation models have been analyzed using two-layer networks, where physical and virtual connections allow the spread of information and diseases to be simulated [13]. This approach demonstrates how official data can be integrated with complex simulations to assess the impact of connectivity on the spread of diseases.

In [14], it is documented how text analysis using advanced natural language processing techniques has made it possible to identify trends in the early stages of epidemic outbreaks, thereby improving the ability of official systems to adapt to changing scenarios.

Official sources and alternative sources, such as social media, play a fundamental role in the collection and analysis of data on epidemic outbreaks. In [15], the authors evaluate how the spread of preventive information, whether positive or negative, on multiplex networks can influence public perception and responses to outbreaks. These findings highlight the importance of considering information dynamics in the formulation of health surveillance and response strategies.

The comparative analysis between official and alternative sources has also been addressed in [16], where the probabilities of extreme epidemics occurring were analyzed using historical and current databases. This study highlights how the integration of multiple sources improves the ability to predict and respond to large-scale events.

In recent years, social media has emerged as an alternative source for detecting epidemic outbreaks. For example, [17] analyzes how the spread of information on digital platforms can serve as an early indicator of health events. Their research highlights the use of optimization algorithms to identify critical nodes in these networks, showing how alternative sources complement the limitations of official sources by capturing collective perceptions and behavior patterns in real time. Another study, described in [18], investigated the theoretical limits of inference in epidemic spread through statistical mechanics methods applied to social networks and graphical models. Their work highlights how user-generated data can provide additional insights when integrated with traditional models.

Reddit, in particular, has proven to be an effective source for detecting emerging trends in public perception. The analysis of posts on subreddits such as r/epidemiology has been documented in [19], indicating that they used distributed control models to assess how decentralized networks can influence the spread of information and diseases.

The integration of data science has been fundamental in addressing the limitations inherent in the heterogeneity of information sources. For example, in [20], they developed a model to classify and analyze epidemic information extracted from social networks, achieving high levels of classification accuracy and generating concise summaries of posts.

Additionally, the work in [21] employed advanced methods to infer epidemic trajectories from partial observations, using Bayesian techniques to improve the estimation of dynamic parameters. Such studies show how data science-based models can address challenges related to uncertainty and variability in data.

III. METHODOLOGY

The methodology proposed in this study was designed to analyze information on epidemic outbreaks from official and alternative sources, incorporating text mining, statistical analysis, and visualization techniques. A systematic process was adopted that included the stages of identification, collection, cleaning, analysis, and comparison of data, with the aim of ensuring the accuracy of the results obtained.

A. Study Design

The study is based on a comparative approach that combines automated data consumption techniques and advanced analysis tools to examine characteristics and differences between information sources. These include official platforms, such as the WHO, UN, and CDC websites, along with alternative sources such as Google News and social media, specifically Reddit. The keywords pandemic, confinement, epidemic, and HMPV were applied uniformly across all sources, with the aim of exploring aspects such as the speed of data dissemination, geographic coverage, and the predominant emotions in the reported narratives.

B. Population and Sample

The research worked with a total of 12,340 texts, distributed across 3,100 records from the WHO, UN, and CDC, 4,240 articles from Google News, and 5,000 posts on

ISSN:1390-9266 e-ISSN:1390-9134 LAJC 2026
DOI: 10.5281/zenodo.17288347

Reddit. This sample was obtained over a three-month period with periodic collections that facilitated the observation of temporal variations.

1) Official Sources:

These include globally recognized organizations such as the World Health Organization (WHO), the United Nations (UN), and the Centers for Disease Control and Prevention (CDC). These entities were selected due to their ability to generate reliable reports, real-time updates, and detailed analyses of disease spread. The information provided by these sources includes statistics, health alerts, and global reports.

2) Alternative Sources:

This category included Google News and the social network Reddit. Google News was used to access information of interest published by international news media through specific queries related to epidemic outbreaks. Reddit, for its part, offers an organized structure in thematic subreddits such as r/epidemiology, r/health, and r/worldnews, which allow for the identification of publications with technical content and social perceptions about health events.

C. Data collection tool

The tool designed for this research was intended to capture structured information about epidemic outbreaks. Uniform queries using the keywords *pandemic, confinement, epidemic*, and *HMPV* were used in all selected sources (official and alternative). The analysis was performed in R Studio with libraries such as *rvest, httr, jsonlite, tidytext, and ggplot2*, which facilitated extraction, processing, and visualization. Automated queries were configured to collect information directly from available APIs and portals, which allowed for standardization of search criteria and ensured traceability. On Reddit, access was provided through the Reddit API with OAuth2 authentication and JSON requests managed with httr/jsonlite to search endpoints and each subreddit; quotas and usage policies were respected and no scraping was used. For Google News, compatible public feeds (RSS/JSON format) accessible from official aggregator links were used.

D. Collection process

The collection was carried out over a period of three months at regular intervals, allowing for the capture of temporal variations. For Google News, queries were defined using English-language keywords, filtering results into health and science categories. On Reddit, the same keywords were applied to specific subreddits (*r/epidemiology, r/health, r/worldnews*), with filters by language and removal of URLs, emojis, and duplicate content.

It is recognized that these sources introduce biases: in the case of Google News, due to the media's tendency toward impactful narratives, and on Reddit, due to the influence of communities with particular interests. Technical limitations: changes in API endpoints or quotas, intermittent availability of feeds, coverage mainly in English that may exclude local nuances, and lack of access to content that was deleted, private, or moderated before capture.

The collected data was organized into homogeneous structures, and random checks were applied to verify consistency with the original sources. This validation ensured that the records were complete and consistent with the study objectives.

E. Data Analysis

Data analysis was carried out following the five phases of the Knowledge Discovery in Databases (KDD) process, summarized in the diagram in Figure 1 (Source selection → Data collection → Preprocessing → Transformation → Modeling & analysis → Evaluation & visualization). This procedure allowed for the systematic organization of the extraction, cleaning, transformation, modeling, and evaluation of information from official and alternative sources. The implementation was carried out in R Studio using libraries such as rvest, httr, jsonlite, dplyr, tidytext, ggplot2, wordcloud2, stringr, gridExtra, tidyr, and textdata, which facilitated the manipulation, processing, and visualization of textual records.

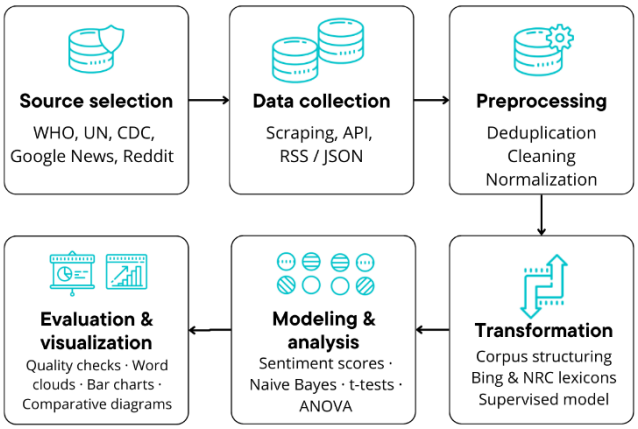

Fig. 1. KDD pipeline used in the study

1) Data Selection:

Reliable sources were selected for the research. Official platforms included the WHO, UN, and CDC, recognized for providing up-to-date epidemiological data. Google News and Reddit were considered as alternative sources. In Google News, news related to specific keywords was analyzed, while in Reddit, thematic subreddits such as r/epidemiology, r/health, and r/worldnews were explored. These platforms provided information on public perceptions and narratives related to epidemic outbreaks.

2) Data Preprocessing:

Cleaning included removing duplicates, normalizing dates, and correcting typos. Texts were converted to lowercase and filters were applied to exclude incomplete entries. The dplyr and stringr libraries were used, ensuring that the final database was aligned with the objectives of the analysis.

3) Data Transformation:

The cleaned records were organized into structures compatible with textual and visual analysis. Tidyr was used to structure the corpus and textdata to associate emotions with the narratives. In addition to the Bing and NRC lexicons, a supervised emotional classification model with cross-validation was trained, which allowed for the detection of nuances not covered in basic dictionaries. This integration

ISSN:1390-9266 e-ISSN:1390-9134 LAJC 2026
DOI: 10.5281/zenodo.17288347

facilitated the identification of mixed emotions in the discourses.

4) Modeling and Analysis:

Narrative patterns were represented using ggplot2 and wordcloud2, generating word clouds, bar charts, and scatter plots. All visualizations were produced in English (titles, axes, and legends), with a minimum resolution of 300 dpi and font size ⩾ 10 pt to ensure legibility. Figure 2 includes a color bar with an explicit scale: –1 = negative emotion, 0 = neutral, +1 = positive emotion; the ends of the gradient correspond to those limits. The analysis included the calculation of average sentiment scores per source, as well as the application of Student's t-tests and ANOVA to contrast differences between groups, with 95% confidence intervals.

5) Results Evaluation

The evaluation of the collected and processed data was carried out through a comparative analysis of official and alternative sources, focusing on aspects such as the accuracy, consistency, and representativeness of the information. This process made it possible to verify the quality of the data and ensure its alignment with the research objectives. To this end, several methodological aspects were evaluated:

- **Accuracy of information:** Cross-checks were performed with the original reports from each platform, confirming that the data extracted through automated queries corresponded to verifiable narratives. This validation was useful for discarding duplicate records or those outside the defined time range.
- **Consistency between sources:** Emotional and thematic patterns were analyzed to determine the correspondence of results with the initial queries (pandemic, confinement, epidemic, HMPV). The contrast between narratives showed differences linked to the biases of each source: a tendency toward high-impact content on Google News and a predominance of community perspectives on Reddit.
- **Data representativeness:** The diversity of terms and emotions was verified, confirming that the three sets offered a balanced sample of institutional, media, and social narratives. An analysis of geographical and temporal coverage was included, which ensured that the data reflected variations at different stages of the outbreaks.

In addition, calculations of average sentiment scores and contrast tests were added to the textual metrics. Student's t-test and ANOVA were applied, yielding statistically significant differences between sources in the dimensions of trust, fear, and sadness. Confidence intervals of 95% were reported, reinforcing the validity of the comparisons.

The figures were generated in English to maintain consistency. Word clouds, bar charts, and scatter plots were used. Each figure included complete descriptions of data source, time range, and color meaning. Color gradients differentiated emotional intensity: dark tones indicated negative emotions, while light tones represented positive or neutral emotions. This visual coding allowed for quick and accurate interpretation of emotional variations.

IV. RESULTS

The data analysis, carried out in January 2025, focused on the processing and evaluation of textual information extracted from three main sources: official and alternative platforms. This approach combined advanced text mining tools, sentiment analysis, and data visualization using word clouds and bar charts, allowing for a comprehensive interpretation of the emotional and linguistic narratives present in each source. To increase rigor, the emotional analysis was not only performed with Bing and NRC lexicons, but also applied a Naive Bayes classifier trained with manual annotations, which allowed for the identification of nuances beyond lexical detection. In addition, sentiment scores were normalized to a range of −1 to +1, and t-tests and ANOVA with 95% confidence intervals were applied, confirming differences between sources.

The visualizations generated, such as Figure 2, used a color gradient bar designed to represent emotions and their intensity. Blue tones were associated with feelings of security and confidence, reflecting positive emotions that promote calm and optimism. In contrast, reddish tones indicated negative emotions such as danger, fear, and risk, allowing us to identify areas of greater alarm or concern in the analyzed texts. This gradient is an effective visual resource for highlighting emotional transitions in the content and has been described in English in all figures to maintain consistency with the language of the processed terms.

Word clouds were used to highlight the most frequent terms in each dataset. In these, the size of the words reflected their prominence within the text, while the intensity of the color represented the average polarity calculated by the model. Warm tones such as orange and red indicated words with negative connotations, while cool, bluish tones highlighted terms associated with more positive or neutral emotions.

These tools made it possible to identify specific patterns and trends in the narratives of each source, helping to structure the results in a comprehensible and verifiable manner. In addition, it was observed that the presentation characteristics of each platform (headlines and thematic hierarchy in aggregators; institutional language and editorial guidelines in official bodies; votes and moderation on Reddit) shape the emotional expression of the content and its reception by the public. Figure 2 is an illustrative example of how colors and shades allow us to capture the emotional range contained in the analyzed texts, facilitating an interpretation based on statistical metrics.

Fig. 2. Color gradient bar

A. Analysis of Data from Official Sources

Official websites provided reliable and structured epidemiological data, designed to communicate messages focused on practical solutions and concrete measures to mitigate the effects of epidemic outbreaks. Sentiment analysis showed a predominance of positive emotions, such as confidence and anticipation, in line with the goals of inspiring

ISSN:1390-9266 e-ISSN:1390-9134 LAJC 2026
DOI: 10.5281/zenodo.17288347

calm, promoting international cooperation, and encouraging collaborative action of these institutions. Although negative emotions such as fear and sadness were present, they were addressed in a controlled manner, with an emphasis on empathy and efforts to overcome the challenges associated with health crises.

In addition to lexicons, a supervised classifier for emotions was applied, confirming higher average scores in positive categories. Comparisons between valences were evaluated with Student's t-test, reporting $p < 0.05$ and 95% CI not including 0, which supports the difference observed in this group. It should be noted that the queries used were standardized in English (pandemic, confinement, epidemic, HMPV).

Figure 3 illustrates how official websites prioritize positive emotions, such as trust, over negative emotions, which, although they exist, are presented with less alarm. This representation reinforces the optimistic nature of their narratives, aimed at generating security and trust in the public.

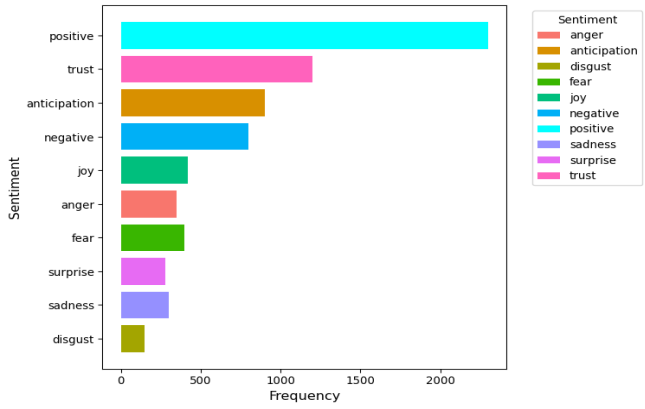

Fig. 3. Distribution of sentiments on official pages

On the other hand, the word cloud generated from official publications highlights terms such as secure, support, protection, and emergency, associated with collective action and cooperation between institutions and citizens. Figure 4 shows this cloud: the size of each word corresponds to its frequency of appearance, and the predominance of blue indicates its link to positive emotions such as security and trust.

In addition, the color gradient of the cloud helps to distinguish the emotional charge: more intense blues indicate words linked to positive and recurring emotions; light blues highlight less frequent terms, although with optimistic connotations. Reds, present to a lesser extent, mark terms associated with alert or danger, which makes it easier to recognize nuances in the discourse of official sources.

Fig. 4. Word cloud generated from official sources

B. *Analysis of Data from Google News*

Analysis of data obtained from Google News showed a predominance of negative emotions, such as sadness and fear, associated with the editorial tendency to highlight adverse and high-impact aspects. Anger was also observed, related to perceptions of social frustration about the management of the health crisis and its side effects. Positive emotions, such as confidence and joy, appeared to a lesser extent, indicating less attention to narratives of progress or achievements. This pattern may reinforce the perception of risk in audiences with high exposure to alarming headlines.

Standardized queries in English (pandemic, confinement, epidemic, HMPV) were used in health/science categories. Editorial bias toward high-impact content is recognized, favoring a greater presence of negatively charged terms. Normalized sentiment scores and ANOVA between groups showed a greater negative charge in this source ($p < 0.01$; 95% CI of the difference > 0).

Figure 5 shows how negative emotions are more prevalent in Google News content. This pattern coincides with media narratives focused on severity and urgency.

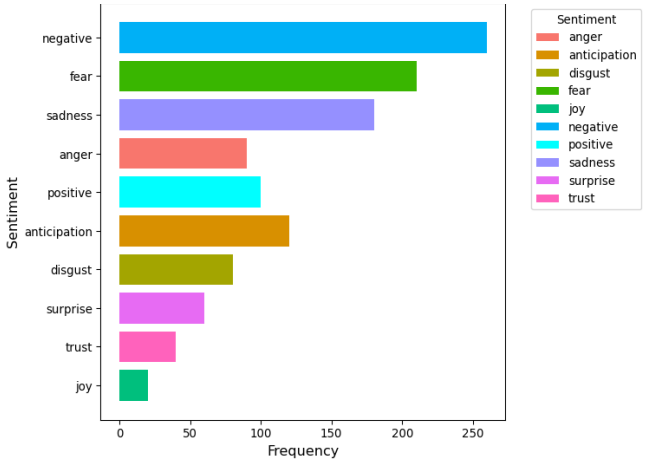

Fig. 5. Distribution of sentiments in Google News

The word cloud generated from these posts (Figure 6) supports this observation, highlighting terms such as fever, dead, confined, stress, outbreak, and dangerous. These words point to risks, restrictions, and unfavorable consequences, resulting in an emotionally charged narrative. Terms such as worst, prisoner, and risk reflect experiences with social and psychological effects, while words in dark red tones, such as dead and dangerous, reinforce the prevailing negative tone. In

ISSN:1390-9266 e-ISSN:1390-9134 LAJC 2026
DOI: 10.5281/zenodo.17288347

contrast, hope and relief appear less frequently and with less visual prominence.

The chromatic degradation used in the cloud helps to distinguish the emotions associated with the terms analyzed. Intense reddish tones indicate words linked to negative emotions, while lighter warm tones reflect the low frequency of terms with positive perspectives, such as protection or solidarity. This allows us to see how the intensity of the color and the size of the words convey useful information about the emotional charge and importance of the terms in the context analyzed.

Fig. 6. Word cloud generated from Google News

C. *Analysis of Data from Reddit*

Reddit was identified as a platform where users share experiences, opinions, and emotions related to epidemic outbreaks. The analysis was based on specific queries in subreddits such as r/epidemiology, r/health, and r/worldnews. These communities allow us to observe posts with diverse perspectives on prevailing concerns and emotions related to outbreaks. The same queries were applied in English and filtered by language, removing URLs, emojis, and repetitions to reduce noise. Community bias is recognized due to the rules and culture of each subreddit, which can skew topics or tones.

Sentiment analysis showed that fear was the most recurrent emotion, associated with uncertainty about the impact of outbreaks, mobility restrictions, and perceived risks. Sadness also featured prominently, reflecting the social and emotional effects of crises, such as isolation and human loss. At the same time, emotions of anticipation were detected, linked to expectations about medical advances, government decisions, and possible solutions.

The supervised classifier detected affective mixtures, and Student's t-tests between polarities yielded $p < 0.05$ with a 95% CI not including 0, supporting the predominance of negative emotions over positive ones in this source. Figure 7 shows the distribution of emotions in Reddit posts, with fear and sadness predominating. Positive emotions, such as confidence and joy, appear less frequently, indicating a focus on concerns and challenges.

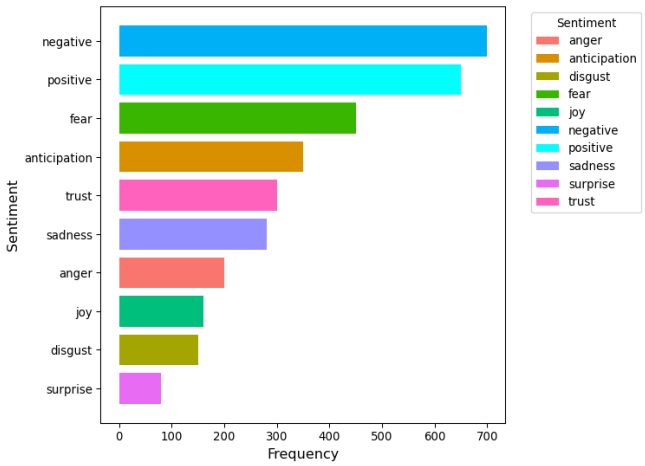

Fig. 7. Distribution of feelings in posts extracted from Reddit

The word cloud generated from the posts (Figure 8) reveals prominent terms such as virus, outbreak, infections, risk, and epidemic. These words reflect concerns about the magnitude of the outbreaks and the severity of their consequences. Terms such as bad and severe reinforce an alarmist tone in the discussions, while positive and secure appear less frequently.

The size of the words in the cloud reflects their frequency in the posts: the largest terms, such as virus and risk, are those that appear most frequently. This visual representation helps to identify the most discussed topics. In addition, the color degradation applied allows the emotional intensity of the terms to be distinguished. Dark tones highlight words associated with negative emotions, such as risk and severe, while warm, light tones, present in positive, show a limited presence of optimistic ideas. This visual approach was complemented by the supervised classifier, confirmed by testing with IC 95%.

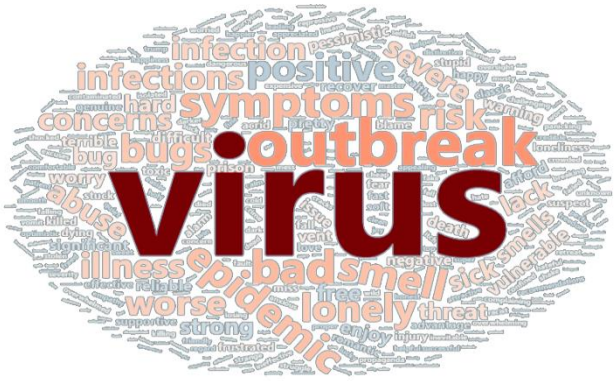

Fig. 8. Word cloud generated from Reddit posts

Reddit, therefore, is identified as a space where collective discussions reflect social concerns and, at the same time, attempts at resilience in the face of epidemic outbreaks. The use of size and colors in the word cloud facilitates an intuitive graphical representation, showing how users emotionally address the challenges associated with the pandemic.

D. *Comparison between official and alternative sources*

The comparative analysis between official and alternative sources highlights marked differences in narratives and

ISSN:1390-9266 e-ISSN:1390-9134 LAJC 2026
DOI: 10.5281/zenodo.17288347

emotions about epidemic outbreaks, showing how each group communicates information and how these narratives influence public perception.

Official sources, represented in fuchsia, focused on promoting messages of trust, cooperation, and security. Terms such as secure, websites, official, and organization dominated the narratives, reflecting language structured to generate calm. The size of these words in the cloud illustrates their prominence, while the light tones reinforce the positive emotional charge. These institutions prioritized collaborative action and concrete measures to mitigate the effects of the outbreaks. As shown in Figure 9, these guidelines seek to stabilize public perception and promote international cooperation.

On the other hand, Google News, identified in purple, presented a more alarmist narrative, focusing on the impact of the outbreaks and their risks. Terms such as pandemic, risk, confined, and dangerous emerge as the most prominent, with darker tones intensifying the negative connotation. This approach, reflected in Figure 9, points to content that highlights severity and urgency, with potential effects on risk perception and collective stress.

Reddit, represented in green, showed a more diverse and participatory narrative, with terms of alarm and also of hope. Words such as virus, epidemic, and lock were frequent, while hope and secure reflect attempts to regain optimism. The green tones, with gradations between light and dark, illustrate the mix of emotions, from fear to hope. As shown in Figure 8, Reddit functions as a space for collective expression where emotions and personal experiences coexist.

Figure 9 summarizes these differences in a comparative visual diagram. Official sources stand out for their optimistic approach with light tones and security-oriented terms. Google News prioritizes alarming narratives through dark tones that convey urgency and gravity. Reddit offers a balance between negative and positive terms, with a wide range of emotions and perspectives. This analysis allows us to see how each type of source participates in the construction of perceptions during health crises.

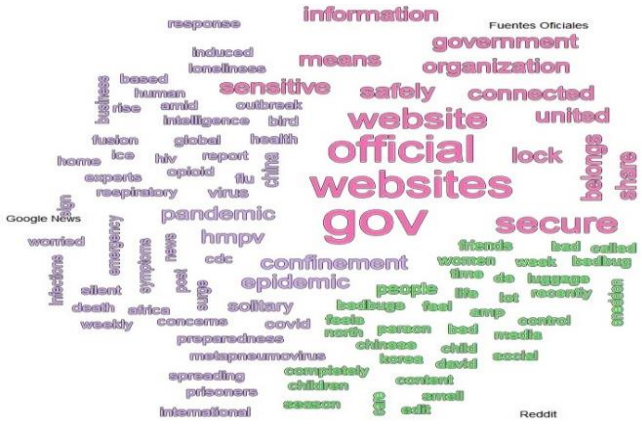

Fig. 9. Comparison between official and alternative sources

V. DISCUSSION

The results obtained in this study show notable differences in the representation and perception of emotions related to epidemic outbreaks on platforms such as official and alternative sources. Beyond the metrics, it can be observed that the editorial framing and dissemination logic of each platform can elevate or attenuate the perception of risk, with practical effects on public attention and willingness to take action.

A. Reddit as a space for social expression

On Reddit, a predominance of fear and sadness was identified, accompanied by anticipation. These emotions respond to the participatory nature of the platform, where users share personal experiences and collective opinions without institutional filters. This bias toward experiential accounts reinforces discussions oriented toward alertness and concern, so Reddit works better as a mood barometer than as a verifiable information channel for immediate decisions.

B. Optimistic narrative in official sources

Official sources were characterized by more controlled communication, focusing on positive feelings such as confidence and anticipation. These institutions seek to convey calm and encourage cooperation through messages designed to avoid panic.[4] and [6] support this trend, indicating that international organizations tend to prioritize narratives that promote security in times of crisis. However, criticism of this strategy was also identified, as it can seem distant from individual experiences, as mentioned in [19] . This disconnect can lead to a perception of institutional coldness in times of uncertainty.

C. Alarmist narrative in Google News

Analysis of Google News publications showed a predominance of negative emotions, such as fear, sadness, and anger. This approach emphasizes the most alarming aspects of the news, aligning with studies such as [2], which indicate that the media often prioritize shocking content to capture the attention of their audiences. Terms such as “confined,” “dangerous,” and “risk” are recurring examples in this narrative. [15], documents how this type of approach can amplify the public's perception of risk, generating knock-on effects. [11] warns that this trend can increase stress in audiences and make it difficult to make informed decisions.

The word cloud generated from Google News reinforces this observation, showing a greater representation of terms related to risks and limitations. Although words such as hope and relief are present, their lower frequency reflects a reduced interest in highlighting progress or positive elements. This highlights the need to balance media narratives with messages that promote resilience and confidence.

D. Impact of technologies and data analysis

The use of advanced technologies such as APIs, text mining, and sentiment analysis was central to this study. Tools such as rvest, httr, ggplot2, textdata, and wordcloud2 facilitated both the analysis and visualization of large volumes of data, allowing for the exploration of emotional and linguistic patterns. The combination of these methods with lexicons such as Bing and NRC made it possible to identify predominant emotions and map how they are reflected in the narratives of different platforms.

In addition, these technologies organized the data and showed how variables such as size, color intensity, and shades in word clouds help to interpret the results visually. For example, the use of gradients made it possible to distinguish the emotions associated with specific terms, which facilitated the detection of differentiated narrative and emotional patterns between official and alternative sources.

ISSN:1390-9266 e-ISSN:1390-9134 LAJC 2026
DOI: 10.5281/zenodo.17288347

VI. CONCLUSIONS

The analysis showed that official sources, such as those managed by the WHO, UN, and CDC, are geared toward conveying confidence and promoting cooperation among audiences. These narratives seek to minimize panic and generate security through structured messages. Although less prevalent, negative emotions also appeared, associated with the communication of risks and challenges, which were handled in a controlled manner to avoid unnecessary alarm.

On the other hand, alternative platforms such as Google News showed a tendency toward alarmist narratives that highlight elements of risk and uncertainty, which can increase the perception of stress in audiences. The predominant presence of terms linked to dangers and restrictions shows an editorial style focused on capturing attention through high-impact headlines.

Greater neutrality was expected on Reddit; however, the analysis revealed a mixture of alarmist accounts with isolated attempts at optimism. Discussions tend to focus on intense concerns and a range of negative emotions, with occasional appearances of hope and resilience

These results should be viewed as preliminary and exploratory. Based on them, practical applications are proposed: real-time monitoring dashboards that integrate media and social media signals, early warnings based on polarity shifts, and segmented communication campaigns that combine institutional messages with readings of the public's emotional state. Together, they offer a comprehensive overview of how the narratives of each source influence collective perception during health crises.

REFERENCES

[1] K. H. Manguri, R. N. Ramadhan, and P. R. Mohammed Amin, "Twitter Sentiment Analysis on Worldwide COVID-19 Outbreaks," *Kurdistan Journal of Applied Research*, pp. 54–65, May 2020, doi: 10.24017/covid.8.

[2] A. Joshi, S. Karimi, R. Sparks, C. Paris, and C. Raina Macintyre, "Survey of Text-based Epidemic Intelligence," *ACM Comput Surv*, vol. 52, no. 6, Nov. 2019, doi: 10.1145/3361141.

[3] K. Sherratt *et al.*, "Characterising information gains and losses when collecting multiple epidemic model outputs," *Journal Epidemics*, vol. 47, Jun. 2024, doi: 10.1016/J.EPIDEM.2024.100765.

[4] J. A. Polonsky *et al.*, "Outbreak analytics: a developing data science for informing the response to emerging pathogens," *Philosophical Transactions of the Royal Society B: Biological Sciences*, 2019, doi: 10.1098/rstb.2018.0276.

[5] K. O. Bazilevych *et al.*, "Information system for assessing the informativeness of an epidemic process feature," *System research and information technologies*, vol. 2023, no. 4, pp. 100–112, Dec. 2023, doi: 10.20535/SRIT.2308-8893.2023.4.08.

[6] A. N. Desai *et al.*, "Real-time Epidemic Forecasting: Challenges and Opportunities," *Health Secur*, vol. 17, no. 4, p. 268, Jul. 2019, doi: 10.1089/HS.2019.0022.

[7] J. L. Herrera-Diestra, J. M. Buldú, M. Chavez, and J. H. Martínez, "Using symbolic networks to analyse dynamical properties of disease outbreaks," *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 476, no. 2236, Apr. 2020, doi: 10.48550/arXiv.1911.05646

[8] T. H. Nguyen, M. Fisichella, and K. Rudra, "A Trustworthy Approach to Classify and Analyze Epidemic-Related Information From Microblogs," *IEEE Trans Comput Soc Syst*, 2024, doi: 10.1109/TCSS.2024.3391395.

[9] J. Tolles and T. Luong, "Modeling Epidemics With Compartmental Models," *Journal Jama Network*, vol. 323, no. 24, pp. 2515–2516, Jun. 2020, doi: 10.1001/JAMA.2020.8420.

[10] M. Marani, G. G. Katul, W. K. Pan, and A. J. Parolari, "Intensity and frequency of extreme novel epidemics," *Proceedings of the National Academy of Sciences*, 2021, doi: 10.1073/pnas.2105482118/-/DCSupplemental.

[11] A. Braunstein, L. Budzynski, and M. Mariani, "Statistical mechanics of inference in epidemic spreading," *Phys Rev E*, vol. 108, no. 6, Dec. 2023, doi: 10.1103/PhysRevE.108.064302.

[12] J. Wu, Z. Niu, and X. Liu, "Understanding epidemic spread patterns: a visual analysis approach," *Health Systems*, vol. 13, no. 3, pp. 229–245, Jul. 2024, doi: 10.1080/20476965.2024.2308286.

[13] S. Gracy, P. E. Pare, H. Sandberg, and K. H. Johansson, "Analysis and distributed control of periodic epidemic processes," *IEEE Trans Control Netw Syst*, vol. 8, no. 1, pp. 123–134, Mar. 2021, doi: 10.1109/TCNS.2020.3017717.

[14] K. M. A. Kabir and J. Tanimoto, "Analysis of epidemic outbreaks in two-layer networks with different structures for information spreading and disease diffusion," *Commun Nonlinear Sci Numer Simul*, vol. 72, pp. 565–574, Jun. 2019, doi: 10.1016/J.CNSNS.2019.01.020.

[15] Z. Wang, C. Xia, Z. Chen, and G. Chen, "Epidemic Propagation with Positive and Negative Preventive Information in Multiplex Networks," *IEEE Trans Cybern*, vol. 51, no. 3, pp. 1454–1462, Mar. 2021, doi: 10.1109/TCYB.2019.2960605.

[16] B. Wang, M. Gou, and Y. Han, "Impacts of information propagation on epidemic spread over different migration routes," *Nonlinear Dyn*, vol. 105, no. 4, pp. 3835–3847, Sep. 2021, doi: 10.1007/S11071-021-06791-8/METRICS.

[17] Z. Wang, X. Rui, G. Yuan, J. Cui, and T. Hadzibeganovic, "Endemic information-contagion outbreaks in complex networks with potential spreaders based recurrent-state transmission dynamics," *Physica A: Statistical Mechanics and its Applications*, vol. 573, Jul. 2021, doi: 10.1016/J.PHYSA.2021.125907.

[18] S. S. Chikkaraddi and G. R. Smitha, "Epidemic Disease Expert System," *1st IEEE International Conference on Advances in Information Technology, ICAIT 2019 - Proceedings*, pp. 571–576, Jul. 2019, doi: 10.1109/ICAIT47043.2019.8987421.

[19] K. Osadcha, V. Osadchyi, and V. Kruglyk, "The role of information and communication technologies in epidemics: an attempt at analysis," *Ukrainian Journal of Educational Studies and Information Technology*, p., 2020, doi: 10.32919/uesit.2020.01.06.

[20] M. Imanipour, M. Shahmari, | Saeideh, A. Mahkooyeh, A. Ghobadi, and P. Sanjari, "Reflections on health information sources in epidemics in synchrony with the COVID-19 pandemic: A scoping review," *Journal of Nursing Advances in Clinical Sciences*, vol. 1, 2024, doi: 10.32598/JNACS.2401.1005.

[21] S. L. Peng *et al.*, "NLSI: An innovative method to locate epidemic sources on the SEIR propagation model," *An interdisciplinary Journal of NonLinear Science*, vol. 33, no. 8, Aug. 2023, doi: 10.1063/5.0152859.

AUTHORS

Karina Ordoñez Guerrero

Karina Michelle Ordóñez Guerrero is a Systems Engineer and holds a Master's Degree in Data Science from the Technical State University of Quevedo (UTEQ). She has experience in programming, data analysis, and developing technological solutions. She collaborates with teachers on projects involving applied analytics, text mining, and visualization, contributing to experimental design and the creation of scripts in R and Python.

During her pre-professional internship at the University Wellness Unit (UBU), she participated in data collection, processing, and analysis, provided technical support, and helped generate reports to improve institutional processes. At the National Institute of Statistics and Census (INEC), she was involved in the planning and supervision of operational processes focused on the management and updating of cartographic data, standing out for her organization and results-oriented approach.

She is currently a programmer and technical assistant at the Royal Dental Center, where she develops solutions that optimize operational and administrative processes. In addition, she supports startups as a technical assistant, gathering requirements, prototyping, testing, and implementing web and mobile applications, and providing technological support to teams starting projects. She has participated in scientific conferences and training sessions on artificial intelligence, data analysis, and emerging technologies, strengthening her technical profile.

José Cordero Bazurto

José Steven Cordero Bazurto is a Systems Engineer and holds a Master's Degree in Data Science from the Technical State University of Quevedo (UTEQ). He works as a programmer and researcher in technological applications and is affiliated with UTEQ, where he collaborates on innovation projects. He is currently a developer of UTEQ's Academic Management System (SGA), participating in the design, development, and improvement of academic modules, database integration, and information analysis to support institutional decision-making.

He is the author of scientific articles in indexed journals, including a publication in Ciencia Huasteca from the Higher School of Huejutla. His areas of expertise include software development, data analytics, and visualization, with experience in requirements gathering, architecture design, API creation, dashboard construction, and data flow automation. At UTEQ, he has collaborated with academic teams on initiatives aimed at improving processes through web solutions and integrated services.

ISSN:1390-9266 e-ISSN:1390-9134 LAJC 2026

AUTHORS

Geovanny Brito Casanova

Geovanny José Brito Casanova has a degree in Systems Engineering from the Quevedo State Technical University (UTEQ), where he is currently a lecturer at the Faculty of Computer Science and Digital Design. He holds a Master's degree in Development and Operations (DevOps) from the International University of La Rioja (Spain) and a Master's degree in Data Science from UTEQ.

During his academic training, he was recognized for his excellent academic performance within his degree program and faculty, receiving institutional distinctions and being awarded national and international postgraduate scholarships. His academic and professional experience focuses on the development and implementation of technological solutions, particularly in the areas of education, data science and cloud computing. He has collaborated as a reviewer for scientific journals and has participated as a speaker in academic events with national and international reach.. His research work covers topics such as educational software, digital infrastructure, environmental automation and the use of new technologies in educational processes.

He is currently involved in university research projects that focus on data analysis, the development of digital environments and the improvement of educational processes through technology.

Eduardo Samaniego Mena

Eduardo Amable Samaniego Mena holds a degree in Systems Engineering from the Technical State University of Quevedo (UTEQ), a Master's degree in Connectivity and Computer Networks from UTEQ, and a Master's degree in Visual Analytics and Big Data from the International University of La Rioja (UNIR, Spain). He is an undergraduate and graduate professor at UTEQ and a tenured professor. His academic activity revolves around applied research, with an emphasis on computer networks, data analytics, and visualization.

In his research work, he develops solutions based on text mining, machine learning, and statistical analysis, integrates end-to-end data pipelines, and builds analytical dashboards for decision-making. He participates in projects with multidisciplinary teams, supervises degree theses, and publishes results aimed at solving real-world problems. He uses tools such as Python, R, SQL, Power BI, Tableau, and network infrastructure and security technologies, combining good engineering practices with scientific methodology.

ISSN:1390-9266 e-ISSN:1390-9134 LAJC 2026

Malware Detection with CNNs on Entropy and Greyscale Images

ARTICLE HISTORY
Received 11 October 2025
Accepted 26 November 2025
Published 6 January 2026

Harry Darton
Sheffield Hallam University
School of Computing and Digital Technologies
Sheffield, United Kingdom
harryphuket@gmail.com
ORCID: 0009-0004-5674-2609

ISSN:1390-9266 e-ISSN:1390-9134 LAJC 2026

H. Darton, "Malware Detection with CNNs on Entropy and Greyscale Images", Latin-American Journal of Computing (LAJC), vol. 13, no. 1, 2026.

Malware Detection with CNNs on Entropy and Greyscale Images

Harry Darton
Sheffield Hallam University
School of Computing and Digital Technologies
Sheffield, United Kingdom
harryphuket@gmail.com

***Abstract*— This study investigates whether convolutional neural networks (CNNs) trained on visual representations of Portable Executable (PE) files can rival traditional machine learning classifiers trained on engineered features. A dataset of over 200,000 PE files [1] was used to derive two feature sets (Basic and Ember-Lite) [2] and to generate 256x256 greyscale and entropy images [3],[4]. Three CNNs (SimpleCNN, ResNet-18 [5], EfficientNet-B0 [6]) were trained and evaluated against five baselines (Random Forest, XGBoost [7], CatBoost [8], LightGBM, Logistic Regression). Tree-based models with enriched features achieved the highest scores, with CatBoost reaching a ROC-AUC of 0.990. The best CNN, EfficientNet-B0 on entropy images, obtained a ROC-AUC of 0.954. Although CNNs did not surpass feature-based models, they showed competitive results when feature engineering was constrained. These findings indicate that visual approaches offer a promising alternative for static malware detection, particularly when combined with entropy-based representations [9].**

Keywords*— *malware detection, convolutional neural networks, entropy images, greyscale images, static analysis

I. Introduction

The rapid evolution of malicious software continues to pose a critical challenge to global cybersecurity. Traditional static and dynamic analysis techniques remain central to malware detection, yet their scalability and adaptability are increasingly strained by the volume and complexity of new samples emerging daily [10]. Static analysis, which inspects binary structure without execution, provides efficiency and safety but depends heavily on manually engineered features. These handcrafted representations are sensitive to obfuscation and require expert knowledge to maintain. Recent advances in deep learning have introduced new possibilities for automated feature extraction that may overcome such limitations [11].

Malware detection can be viewed as a binary classification problem in which a model must learn discriminative patterns between benign and malicious Portable Executable (PE) files. Earlier static approaches focused on syntactic features such as byte histograms, imported libraries, and section metadata [2]. While these features remain effective, they are often dataset-specific and may fail when the underlying malware distribution shifts. Convolutional Neural Networks (CNNs) offer an alternative pathway [11] by learning directly from raw or visually transformed data. In computer vision, CNNs have achieved outstanding success in recognizing complex spatial relationships within images. When applied to malware, they can automatically extract high-level spatial–statistical representations from a binary's byte sequence, potentially reducing reliance on expert feature engineering [12].

Transforming PE files into visual formats has gained attention because it allows direct use of image-based architectures without disassembling executables. Two representations have become prominent: greyscale images, formed by mapping byte values to pixel intensities, and entropy images [3],[4],[13], which highlight structural irregularities related to code density and compression. These visualizations reveal characteristic patterns that correspond to malware families, packing, and section entropy, providing CNNs with texture-like information unavailable to conventional static models. However, existing studies differ widely in dataset size, preprocessing, and evaluation methodology, making it difficult to assess [9],[14] whether CNN-based visual analysis can truly compete with established feature-based models.

Prior work has shown that tree-based ensembles such as Random Forest, XGBoost [7], CatBoost [8], and LightGBM deliver high accuracy on engineered feature sets like EMBER [2], often exceeding 0.99 ROC-AUC. Although several researchers have experimented with CNNs on image representations [3],[9],[15], many comparisons are indirect, rely on small datasets, or use pretrained vision networks without systematic control of variables. The literature therefore lacks a consistent large-scale benchmark that contrasts CNN performance with traditional models under identical data and evaluation conditions. Furthermore, while some studies report promising CNN results, few investigate how architectural complexity or input modality (greyscale vs entropy vs combined) affect performance [6],[15] or generalization.

This research addresses that gap through a controlled, large-scale comparison between visual-representation CNNs and feature-based classifiers for static malware detection. A dataset of more than 200,000 PE files was processed to generate both engineered features and 256x256 image representations [1]. Three CNN architectures of increasing depth, SimpleCNN, ResNet-18[5], and EfficientNet-B0 [6], were trained from scratch using greyscale, entropy, and dual-channel inputs. Their results were benchmarked against five tree-based baselines (Random Forest, XGBoost [7], CatBoost [8], LightGBM, and Logistic Regression) built on two feature sets: a compact Basic subset and an Ember-Lite extension. All models were evaluated under identical stratified splits and metrics, including ROC-AUC, F1-score, precision, recall, and accuracy.

The study contributes in three principal ways:

1. It provides a reproducible comparison between CNN-based and feature-based static detection using the same dataset and preprocessing pipeline.

ISSN:1390-9266 e-ISSN:1390-9134 LAJC 2026
DOI: 10.5281/zenodo.17941249

2. It analyses the effect of image representation (greyscale, entropy, and combined) on CNN performance.
3. It evaluates how model depth influences the trade-off between feature learning capacity and overfitting risk.

By unifying these aspects, the work offers an empirical baseline for future research into vision-based malware detection and clarifies the extent to which CNNs can replace or complement engineered features [9].

II. Background and Related Work

Research on static malware detection has evolved from handcrafted feature extraction to representation-learning approaches based on neural networks. Early static pipelines focused on syntactic features derived from Portable Executable (PE) headers, import tables, and byte histograms [16], [17]. Ensemble methods such as Random Forest and XGBoost became widely adopted because they balanced predictive accuracy with interpretability.

Subsequent studies explored the visual encoding of binaries as two-dimensional matrices, where raw bytes were transformed into greyscale images to capture structural patterns [3]. This approach enabled convolutional networks to learn discriminative textures associated with packed or obfuscated code without explicit feature engineering. Han et al. [4] and Kalash et al. [12] extended this concept by employing deeper CNNs that achieved performance comparable to engineered-feature baselines.

Entropy-based representations introduced a further dimension by quantifying local randomness across the file, highlighting high-entropy regions characteristic of compression and encryption [15]. Brosolo et al. [14] demonstrated that combining byte and entropy channels improved separability between benign and malicious samples. However, most prior work relied on small datasets or inconsistent preprocessing, limiting reproducibility.

The present study addresses these limitations through a unified pipeline incorporating both feature-based and image-based approaches under controlled preprocessing and multi-seed evaluation. By comparing ensemble and CNN architectures directly on 200,000+ samples, it contributes empirical evidence on the relative merits of learned versus handcrafted representations for static malware detection.

III. Methodology

In this section, the methodology featured in Fig. 2 is explained as follows:

A. Dataset Construction

The experiments employed a static malware dataset containing more than 200,000 Portable Executable (PE) files, comprising roughly equal proportions of benign and malicious samples. Each file was validated to ensure accessibility and correct labelling. No dynamic execution or network traffic data were used, maintaining strict static conditions. A group-wise stratified split produced three subsets: 70% for training, 15% for validation, and 15% for independent testing. Stratification ensured proportional representation of malware families and preserved class balance across splits. All preprocessing, feature extraction, and model training were performed offline to eliminate any risk of infection.

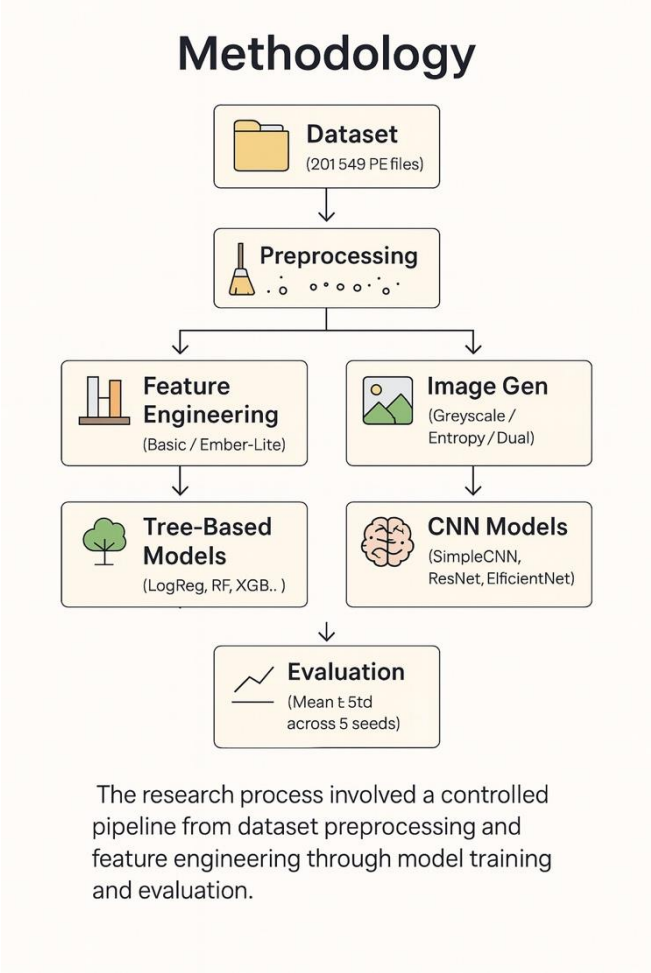

Fig. 2. Methodological workflow for dataset processing, model training, and evaluation.
The figure summarizes the experimental pipeline from dataset acquisition and preprocessing through feature engineering, image generation, and model evaluation across five random seeds.

B. Feature Engineering

Two engineered feature sets were created to provide traditional baselines.

1. Basic set: extracted lightweight structural attributes such as file size, section counts, imported library frequencies, and entropy statistics.
2. Ember-Lite set: extended the Basic features with a reduced subset of the EMBER 2018 dataset, including byte histograms, header metadata, and string features. The Ember-Lite feature set used here comprised only a small, computationally lightweight subset of the EMBER 2018 feature groups, selected to minimize pre-processing overhead whilst still producing a more detailed, EMBER-oriented set of features for comparison.

Both sets were scaled using min–max normalization. These vectors served as input to five machine-learning models: Random Forest, XGBoost, CatBoost, LightGBM, and Logistic Regression. Hyperparameters were tuned by grid search on the validation split.

ISSN:1390-9266 e-ISSN:1390-9134 LAJC 2026
DOI: 10.5281/zenodo.17941249

C. Image Generation

For the visual-representation branch, each PE file was converted into two 256x256 images: a greyscale byte map and an entropy image. PE files shorter than 65,536 bytes were zero padded before reshaping and files exceeding this length were truncated so that all samples produced a consistent 256x256 matrix.

- The greyscale representation mapped each byte value (0-255) to a pixel intensity, preserving sequential order.
- The entropy representation applied a sliding-window Shannon entropy calculation to highlight structural irregularities related to code density, packing, and compression.

Images were stored as compressed PNGs and normalized to the [0, 1] range at load time. Three input modalities were tested: single-channel greyscale, single-channel entropy, and dual-channel combined images. The dual-channel version concatenated normalized greyscale and entropy matrices along the channel axis to retain spatial alignment. Representative examples of the three image modalities are shown in Fig. 1.

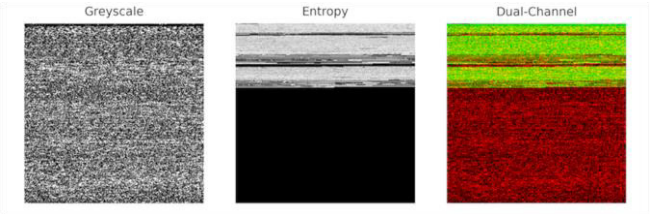

Fig. 1. Visualization of greyscale, entropy, and dual-channel image representations used for CNN training. The dual-channel view merges greyscale and entropy channels into red and green for clarity; CNNs process both as greyscale tensors.

D. CNN Architectures

Three convolutional neural networks of increasing depth and complexity were implemented to explore architectural effects.

1. SimpleCNN: a custom lightweight model with three convolutional blocks and max-pooling, designed to provide a minimal baseline.
2. ResNet-18: a residual architecture enabling deeper feature learning while mitigating vanishing-gradient issues.
3. EfficientNet-B0: a compound-scaled network optimizing depth, width, and resolution for parameter efficiency.

Each model concluded with a global average-pooling layer and a fully connected single sigmoid output neuron optimized with binary cross-entropy loss. All architectures were trained from scratch rather than fine-tuned from natural-image weights to maintain domain specificity.

More recent vision architectures exist, but the chosen trio provides a controlled progression from shallow to moderately deep networks, enabling a fair comparison of representational capacity while keeping reproducibility and computational cost practical for large-scale malware datasets.

E. Training Procedure

Training was performed using the PyTorch 2.8 framework on GPU hardware. Batch size, learning rate, and weight decay were tuned empirically through pilot runs. Early stopping based on validation loss prevented overfitting, and the best model weights were checkpointed. Feature vectors and image tensors were normalized to the [0, 1] range prior to training. Early experiments confirmed that standardization to [-1, 1] or z-scoring did not improve convergence for models trained from scratch. All random seeds, splits, and preprocessing parameters were fixed for reproducibility. During training, data augmentation was limited to horizontal and vertical flips to avoid distorting binary layout information. Each experiment was repeated across five random seeds to estimate variability.

F. Evaluation Metrics

Model performance was assessed on the held-out test set using multiple metrics:

- ROC-AUC as the principal discrimination measure.
- F1-score, precision, recall, and accuracy for complementary evaluation.
- Calibration curves and confusion matrices for selected runs to examine reliability and error distribution.

All results were reported as mean ± standard deviation across seeds. Timing and resource usage were recorded but not compared formally because training occurred on heterogeneous hardware.

IV. Deployment And Implementation

All experiments were executed on a Windows 11 workstation equipped with an NVIDIA RTX 5080 GPU (16 GB VRAM), a Ryzen 7 7800X3D CPU, and 32 GB RAM. The environment used Python 3.12 and PyTorch 2.8.0 (CUDA 12.8) within a PyCharm-managed virtual environment. Dataset preprocessing and feature extraction were performed offline to prevent malware execution risk.

A. Dataset and Preprocessing

A total of 201,549 PE files from Lester (2021) were processed into multiple representations: (i) Basic PE features, (ii) extended Ember-Lite features, (iii) greyscale byte-maps, (iv) entropy images computed over 32-byte windows, and (v) dual-channel stacks combining greyscale and entropy tensors. Group-wise stratification by Imphash ensured partition integrity across train, validation, and test splits, mitigating leakage.

B. Model Implementation

Tree-based models (Logistic Regression, Random Forest, XGBoost, LightGBM, CatBoost) were implemented via Scikit-learn and native library APIs using identical folds and hyperparameter grids. For CNNs, three architectures were selected to represent increasing structural depth: SimpleCNN (three convolutional layers), ResNet-18, and EfficientNet-B0. All networks employed batch normalization, ReLU activations, early stopping, and Adam optimization with learning-rate scheduling.

C. Training Protocol

Training used Adam (learning rate = $1x10^{-3}$) with CrossEntropyLoss, ReduceLROnPlateau (patience = 2), early stopping (patience = 5), and batch size 128, capped at 40

ISSN:1390-9266 e-ISSN:1390-9134 LAJC 2026
DOI: 10.5281/zenodo.17941249

epochs. Only the best-performing checkpoint (lowest validation loss) per run was retained for test evaluation. Results were averaged across seeds (42–46) and reported with standard deviation.

D. Reproducibility Controls

All outputs (model weights, logs, and metrics.json files) were stored under versioned directories for full traceability. Reproducibility was validated by selectively rerunning a subset of experiments on different hardware (MX250 laptop and MacBook CPU), confirming consistent metrics within variance bounds. The entire pipeline is available via a public GitHub repository.

V. Analysis and Discussion of Findings

A. Overview

The study provides a comprehensive comparative analysis, combining large-scale data, multiple feature representations, and a multi-seed evaluation to support robust and generalizable conclusions.

This section presents the quantitative results obtained from both the feature-based and CNN-based branches of the study. All models were evaluated on the held-out test partition using identical metrics and configuration to ensure comparability. Performance values reported correspond to the mean of five random-seed runs, with standard deviation shown where applicable.

B. Feature-Based Baselines

Traditional classifiers trained on the Basic and Ember-Lite feature sets produced strong results across all metrics. The Basic feature set already provided a solid baseline, achieving ROC-AUC scores above 0.96 for ensemble methods. Incorporating the additional Ember-Lite attributes further improved separability between benign and malicious samples.

Among the evaluated models, CatBoost delivered the highest and most consistent performance, reaching 0.990 ± 0.001 ROC-AUC and an F1-score of 0.947 ± 0.005. XGBoost and LightGBM followed closely, differing by less than 0.002 ROC-AUC, while Random Forest achieved comparable accuracy with slightly higher variance. Logistic Regression, as expected, under-fit the nonlinear relationships and produced the lowest AUC (≈ 0.94). These outcomes confirm that gradient-boosted tree ensembles remain a robust benchmark for static malware detection when high-quality engineered features are available.

C. CNN Performance on Greyscale and Entropy Images

The CNN experiments evaluated three architectures of increasing complexity (SimpleCNN, ResNet-18, and EfficientNet-B0) across three input modalities: greyscale, entropy, and dual-channel combined images.

- Greyscale Images: captured structural layout but limited semantic variation. SimpleCNN achieved 0.886 ± 0.011 ROC-AUC, ResNet-18 0.937 ± 0.008, and EfficientNet-B0 0.931 ± 0.010.
- Entropy Images: provided higher discriminative information due to encoding of randomness and packing density. Here, EfficientNet-B0 achieved 0.954 ± 0.005 and 0.898 ± 0.008 F1, the best CNN result overall, indicating that entropy-based spatial cues are particularly informative for distinguishing obfuscated malware.
- Combined Images: merging greyscale and entropy channels offered only marginal gains for shallower networks and, in some cases, introduced redundancy. ResNet-18 reached 0.950 ± 0.009 ROC-AUC, while the dual-channel variant of EfficientNet-B0 achieved 0.947 ± 0.004, showing little additional benefit.

Training variability across seeds remained low (±0.002 AUC), confirming stable convergence. Validation loss curves indicated earlier saturation for SimpleCNN, while deeper models continued improving over more epochs, reflecting greater representational capacity.

D. Comparative Analysis

A direct comparison between the best feature-based and CNN models highlights a clear but narrowing performance gap. CatBoost (Ember-Lite) exceeded EfficientNet-B0 (Entropy) by approximately 0.036 ROC-AUC, achieved 0.983 ± 0.001 precision and 0.913 ± 0.011 recall compared with 0.954 ± 0.004 precision and 0.848 ± 0.013 recall for the CNN. When feature engineering is limited, CNNs trained on entropy visualizations approach ensemble-level accuracy without any handcrafted inputs.

Fig. 3 visualizes the overall comparison. Feature-based ensembles dominate the upper bound, while CNNs occupy a competitive mid-band with notably reduced preprocessing overhead.

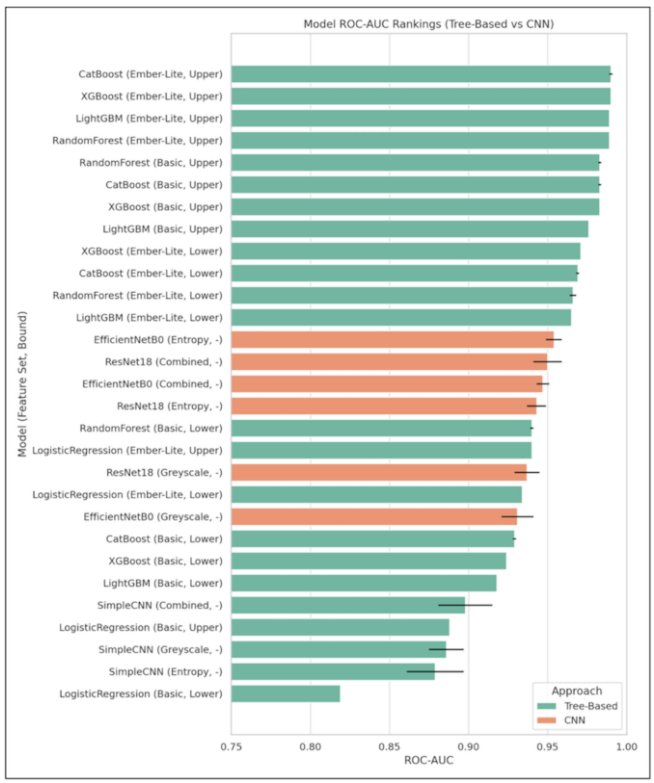

Fig. 3. Overall performance comparison of tree-based and CNN-based models across all input modalities

Table I reports mean ± standard deviation over five random-seed runs for all models across Basic, Ember-Lite, greyscale, entropy, and combined inputs.

ISSN:1390-9266 e-ISSN:1390-9134 LAJC 2026
DOI: 10.5281/zenodo.17941249

TABLE I. MODEL PERFORMANCE COMPARISON (FEATURE-BASED VS CNN)
Summary of quantitative performance metrics (mean ± standard deviation) for all models across Basic, Ember-Lite, Greyscale, Entropy, and Combined input representations.

Model	Feature Set	Bound	ROC-AUC	F1	Accuracy	Precision	Recall
CatBoost	Ember-Lite	Upper	0.990 ± 0.001	0.947 ± 0.005	0.941 ± 0.005	0.983 ± 0.001	0.913 ± 0.011
XGBoost	Ember-Lite	Upper	0.990 ± 0.000	0.940 ± 0.000	0.933 ± 0.000	0.983 ± 0.000	0.900 ± 0.000
LightGBM	Ember-Lite	Upper	0.989 ± 0.000	0.936 ± 0.000	0.929 ± 0.000	0.982 ± 0.000	0.894 ± 0.000
RandomForest	Ember-Lite	Upper	0.989 ± 0.000	0.929 ± 0.004	0.922 ± 0.004	0.977 ± 0.003	0.886 ± 0.008
RandomForest	Basic	Upper	0.983 ± 0.001	0.932 ± 0.003	0.924 ± 0.003	0.966 ± 0.007	0.900 ± 0.010
CatBoost	Basic	Upper	0.983 ± 0.001	0.938 ± 0.002	0.930 ± 0.002	0.960 ± 0.003	0.916 ± 0.005
XGBoost	Basic	Upper	0.983 ± 0.000	0.937 ± 0.000	0.929 ± 0.000	0.963 ± 0.000	0.912 ± 0.000
LightGBM	Basic	Upper	0.976 ± 0.000	0.919 ± 0.000	0.911 ± 0.000	0.959 ± 0.000	0.882 ± 0.000
XGBoost	Ember-Lite	Lower	0.971 ± 0.000	0.898 ± 0.000	0.891 ± 0.000	0.980 ± 0.000	0.829 ± 0.000
CatBoost	Ember-Lite	Lower	0.969 ± 0.001	0.894 ± 0.005	0.887 ± 0.005	0.979 ± 0.003	0.822 ± 0.009
RandomForest	Ember-Lite	Lower	0.966 ± 0.002	0.904 ± 0.005	0.897 ± 0.005	0.976 ± 0.001	0.842 ± 0.008
LightGBM	Ember-Lite	Lower	0.965 ± 0.000	0.902 ± 0.000	0.895 ± 0.000	0.971 ± 0.000	0.843 ± 0.000
EfficientNetB0	Entropy	-	0.954 ± 0.005	0.898 ± 0.008	0.889 ± 0.008	0.954 ± 0.004	0.848 ± 0.013
ResNet18	Combined	-	0.950 ± 0.009	0.904 ± 0.010	0.890 ± 0.015	0.916 ± 0.044	0.895 ± 0.039
EfficientNetB0	Combined	-	0.947 ± 0.004	0.898 ± 0.006	0.888 ± 0.006	0.949 ± 0.005	0.852 ± 0.011
ResNet18	Entropy	-	0.943 ± 0.006	0.897 ± 0.008	0.888 ± 0.008	0.954 ± 0.014	0.847 ± 0.014
RandomForest	Basic	Lower	0.940 ± 0.001	0.879 ± 0.002	0.868 ± 0.002	0.930 ± 0.001	0.833 ± 0.003
LogisticRegression	Ember-Lite	Upper	0.940 ± 0.000	0.855 ± 0.000	0.844 ± 0.000	0.927 ± 0.000	0.793 ± 0.000
ResNet18	Greyscale	-	0.937 ± 0.008	0.862 ± 0.013	0.854 ± 0.012	0.943 ± 0.006	0.795 ± 0.025
LogisticRegression	Ember-Lite	Lower	0.934 ± 0.000	0.876 ± 0.000	0.863 ± 0.000	0.917 ± 0.000	0.838 ± 0.000
EfficientNetB0	Greyscale	-	0.931 ± 0.010	0.880 ± 0.008	0.871 ± 0.006	0.945 ± 0.017	0.824 ± 0.027
CatBoost	Basic	Lower	0.929 ± 0.001	0.874 ± 0.002	0.862 ± 0.003	0.927 ± 0.005	0.826 ± 0.002
XGBoost	Basic	Lower	0.924 ± 0.000	0.874 ± 0.000	0.862 ± 0.000	0.918 ± 0.000	0.835 ± 0.000
LightGBM	Basic	Lower	0.918 ± 0.000	0.872 ± 0.000	0.860 ± 0.000	0.918 ± 0.000	0.831 ± 0.000
SimpleCNN	Combined	-	0.898 ± 0.017	0.777 ± 0.153	0.788 ± 0.095	0.898 ± 0.055	0.728 ± 0.210
LogisticRegression	Basic	Upper	0.888 ± 0.000	0.801 ± 0.000	0.784 ± 0.000	0.853 ± 0.000	0.755 ± 0.000
SimpleCNN	Greyscale	-	0.886 ± 0.011	0.849 ± 0.011	0.833 ± 0.017	0.886 ± 0.035	0.817 ± 0.015
SimpleCNN	Entropy	-	0.879 ± 0.018	0.843 ± 0.012	0.835 ± 0.009	0.931 ± 0.020	0.771 ± 0.032
LogisticRegression	Basic	Lower	0.819 ± 0.000	0.757 ± 0.000	0.733 ± 0.000	0.798 ± 0.000	0.720 ± 0.000

To further illustrate error distribution between benign and malicious classifications, Fig. 4 presents normalized confusion matrices for the best-performing ensemble (CatBoost) and CNN (EfficientNet-B0) models.

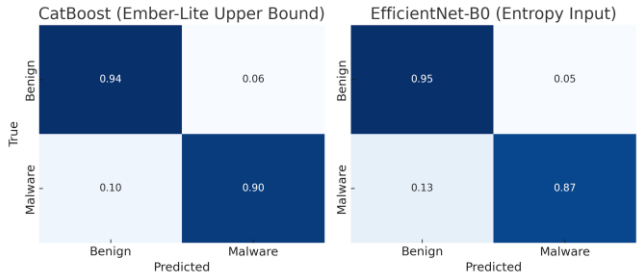

Fig. 4. Normalized confusion matrices comparing the best ensemble (CatBoost – Ember-Lite Upper Bound) and CNN (EfficientNet-B0 – Entropy Input) models. CatBoost achieves slightly superior precision on benign samples, while EfficientNet-B0 maintains strong recall on malware detection, demonstrating the narrowing gap between traditional and visual-based static analysis approaches.

Examination of the confusion matrices showed that CNNs occasionally misclassified small or lightly obfuscated malware samples, particularly where entropy images remained sparse after padding. In contrast, the tree-based models sometimes mislabeled benign files exhibiting elevated entropy or atypical section structures. These patterns are consistent with the feature sensitivities of each model family and help explain why tree-based models still retain a small overall advantage.

ISSN:1390-9266 e-ISSN:1390-9134 LAJC 2026
DOI: 10.5281/zenodo.17941249

E. Combined Image Representation and Interpretation

The hypothesis that merging greyscale and entropy inputs would enhance classification performance was not supported by the results. Across all CNN architectures, the dual-channel variant produced near-identical or slightly inferior ROC-AUC values compared with single-channel entropy inputs. This outcome reflects an important characteristic of PE visualization: while greyscale mappings and entropy images appear visually distinct, they encode strongly correlated structural information.

The greyscale representation captures the sequential byte distribution of sections, making dense code regions appear darker and sparse or zero-padded regions lighter. Entropy visualization, computed over fixed 32-byte windows, highlights the same structural boundaries by assigning higher intensity to compressed or encrypted blocks and lower intensity to static resources or padding. When concatenated, both modalities effectively describe the same transitions in spatial density and randomness. This redundancy dilutes gradient salience during training, as filters in early CNN layers receive conflicting but overlapping cues, hindering convergence to strongly discriminative features.

From a signal-processing perspective, direct stacking of channels also constrains the network to treat the two modalities as spatially aligned, which may not reflect semantic complementarity. A more effective fusion could involve learned attention or adaptive weighting between channels, allowing the model to emphasize entropy cues where they are most informative. Another avenue would be RGB-style compositing, in which greyscale, entropy, and a derived statistical feature (such as local variance or byte-frequency gradient) are encoded into separate color channels. This approach may exploit richer feature interactions akin to texture analysis in natural images. Alternatively, late fusion strategies (where embeddings from separate CNN branches are combined at the dense layer stage) could preserve each modality's distinct representational space while capturing joint correlations.

The observed stagnation of performance in combined inputs therefore does not imply redundancy between visual and entropy domains per se, but rather that naive spatial concatenation fails to capitalize on their complementary nature. Future research should explore cross-channel attention, feature-level fusion, or transformer-based architectures capable of learning non-linear relationships between modality-specific embeddings.

F. Interpretation of Key Findings

Three principal insights emerge from the experimental results.

1. Entropy images outperform greyscale inputs, revealing high-entropy packing and obfuscation regions strongly correlated with malicious binaries. The results indicate that entropy-based CNNs capture discriminative information unavailable from structural byte layouts alone.
2. Model depth directly influences performance. Deeper CNNs such as ResNet-18 and EfficientNet-B0 consistently surpass SimpleCNN, confirming that hierarchical feature extraction enhances visual malware representation learning.
3. Feature engineering versus representation learning: feature-based ensembles remain superior when rich handcrafted features are available, but CNNs offer a promising alternative for scalable, fully automated pipelines where feature computation is infeasible or costly.

The comparative results highlight that entropy visualizations preserve generalizable statistical cues about randomness and code density, while feature-engineered models exploit explicit semantic attributes. The best CNN configuration (EfficientNet-B0 trained on entropy images) achieved 0.954 ± 0.005 ROC-AUC, compared with 0.990 ± 0.001 for the CatBoost ensemble on the Ember-Lite feature set. This ~ 0.036 AUC difference demonstrates a narrowing gap between learned visual and engineered feature representations, even though ensemble methods remain the upper bound.

Beyond comparative accuracy, these results highlight a wider methodological shift in static malware research. As handcrafted features reach maturity, incremental performance gains often come at the cost of increased preprocessing complexity and domain dependence. Visual learning approaches, by contrast, demonstrate the capacity to generalize across unseen binaries with minimal feature engineering. Their scalability and model-agnostic input pipeline make them well suited for future integration into hybrid detection frameworks, combining the interpretability of engineered features with the adaptability of deep representation learning.

G. Broader Implications and Limitations

Although extensive, this comparison remains limited to static byte-level analysis and does not incorporate dynamic or hybrid behavioral features. Training and evaluation were conducted on heterogeneous hardware, and while reproducibility was confirmed, computational cost was not measured formally. Furthermore, dataset balance was maintained artificially, whereas real-world malware corpora are often heavily skewed.

These constraints suggest several research extensions: combining static and dynamic modalities; benchmarking under naturally imbalanced distributions; and assessing architecture efficiency across larger datasets or through transfer learning.

The dataset contained 201,549 executables, comprising 86,812 benign files and 114,737 malware samples. Although this distribution is only moderately skewed towards malicious files, real-world environments are typically dominated by benign software. Future evaluations should therefore assess performance and calibration under more strongly imbalanced conditions, or more benign-dominated conditions.

Nevertheless, the findings reinforce the viability of CNN-based static detectors as interpretable, automated complements to feature-based models. With entropy-derived visual inputs, CNNs achieved competitive performance within 3-4% ROC-AUC of state-of-the-art ensembles while eliminating manual feature design, indicating strong potential for scalable, real-time malware triage systems.

VI. Conclusion

This study compared feature-based machine-learning models and convolutional neural networks (CNNs) for static malware detection using a dataset of more than 200,000 Portable Executable files [1]. The experimental design provided a unified benchmark in which both traditional classifiers and visual deep-learning models were trained and evaluated under identical conditions.

The results demonstrated that tree-based ensembles, particularly CatBoost [8], remain the most accurate static detectors when high-quality handcrafted features are available, achieving 0.990 ± 0.001 ROC-AUC and 0.947 ± 0.005 F1. However, CNNs trained directly on byte-derived image representations achieved competitive performance without manual feature engineering. Among the visual models, entropy-based inputs consistently outperformed greyscale and combined modalities, and deeper networks such as ResNet-18 [5] and EfficientNet-B0 [6] significantly exceeded the accuracy of the shallow SimpleCNN. These findings confirm that entropy visualization provides strong discriminative cues and that model depth enhances representational capacity in image-based malware analysis [3],[4].

The comparative analysis revealed a narrowing gap of approximately 0.036 ROC-AUC between the best feature-based and CNN models, suggesting that vision-driven approaches can serve as scalable, automated alternatives when handcrafted features are unavailable or costly to compute.

Future work should explore hybrid static–dynamic pipelines that combine image-based deep learning with lightweight engineered features to further improve robustness against obfuscation and dataset drift [14]. Benchmarking computational efficiency across uniform hardware and assessing real-world, imbalanced data distributions would also strengthen the practical applicability of visual malware detection methods.

The limited improvement from the combined greyscale–entropy representation further highlights the challenge of naive multimodal fusion. Although the two inputs differ visually, their spatially correlated content may reduce discriminative gradients when processed jointly; suggesting that future architectures should employ attention-based fusion or late-stage embedding integration to better exploit complementary features without introducing redundancy.

Funding Statement

This research received no external funding. All computational work and analysis were conducted using personal and institutional resources without third-party support.

Acknowledgment

The author thanks Dr. Shahrzad Zargari of Sheffield Hallam University for supervision and constructive feedback throughout the research project.

Author Contributions

Harry Darton: Conceptualization, Methodology, Data Curation, Formal Analysis, Writing – Original Draft, Writing – Review & Editing.

References

[1] M. Lester, “PE malware machine learning dataset [Data set],” Practical Security Analytics, 2021. [Online]. Available: https://practicalsecurityanalytics.com/pe-malware-machine-learning-dataset/

[2] H. S. Anderson and P. Roth, “EMBER: An open dataset for training static PE malware machine learning models,” arXiv preprint, 2018. [Online]. Available: https://arxiv.org/abs/1804.04637

[3] L. Nataraj, S. Karthikeyan, G. Jacob, and B. S. Manjunath, “Malware images: Visualization and automatic classification,” in Proc. 8th Int. Symp. Visualization for Cyber Security (VizSec 2011), pp. 1–7, ACM, 2011. doi: 10.1145/2016904.2016908

[4] K. S. Han, J. H. Lim, B. Kang, and E. G. Im, “Malware analysis using visualized images and entropy graphs,” Int. J. Inf. Security, vol. 14, no. 1, p. 1, 2014. doi: 10.1007/s10207-014-0242-0

[5] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR 2016), pp. 770–778, 2016. doi: 10.1109/CVPR.2016.90

[6] M. Tan and Q. V. Le, “EfficientNet: Rethinking model scaling for convolutional neural networks,” in Proc. 36th Int. Conf. Machine Learning (ICML 2019), vol. 97, pp. 6105–6114, PMLR, 2019. [Online]. Available: https://proceedings.mlr.press/v97/tan19a.html

[7] T. Chen and C. Guestrin, “XGBoost: A scalable tree boosting system,” in Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining (KDD ’16), pp. 785–794, ACM, 2016. doi: 10.1145/2939672.2939785

[8] L. Prokhorenkova, G. Gusev, A. Vorobev, A. V. Dorogush, and A. Gulin, “CatBoost: Unbiased boosting with categorical features,” in Proc. 32nd Int. Conf. Neural Information Processing Systems (NeurIPS 2018), pp. 6639–6649, 2018. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2018/hash/14491b756b3a51daac41c24863285549-Abstract.html

[9] A. Bensaoud, N. Abudawaood, and J. Kalita, “Classifying malware images with convolutional neural network models,” arXiv preprint, 2020. [Online]. Available: https://arxiv.org/abs/2010.16108

[10] AV-TEST Institute, “Malware statistics & trends report,” 2024. [Online]. Available: https://www.av-test.org/en/statistics/malware/

[11] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” Nature, vol. 521, no. 7553, pp. 436–444, 2015. doi: 10.1038/nature14539

[12] M. Kalash et al., “Malware classification with deep convolutional neural networks,” in Proc. 10th Int. Conf. New Technologies, Mobility and Security (NTMS), pp. 1–5, IEEE, 2018. doi: 10.1109/NTMS.2018.8328749

[13] C. E. Shannon, “A mathematical theory of communication,” Bell Syst. Tech. J., vol. 27, no. 3, pp. 379–423, 1948. doi: 10.1002/j.1538-7305.1948.tb01338.x

[14] M. Brosolo and M. Conti, “The road less travelled: Investigating robustness and explainability in CNN malware detection,” arXiv preprint, 2025. doi: 10.48550/arXiv.2503.01391

[15] B. Al-Masri, N. Bakir, A. El-Zaart, and K. Samrouth, “Dual convolutional malware network (DCMN): An image-based malware classification using dual convolutional neural networks,” Electronics, vol. 13, no. 18, p. 3607, 2024. doi: 10.3390/electronics13183607

[16] J. Saxe and K. Berlin, “Deep neural network based malware detection using two-dimensional binary program features,” arXiv preprint arXiv:1508.03096, 2015.

[17] E. Raff et al., “Malware detection by eating a whole EXE,” arXiv preprint arXiv:1710.09435, 2017.

Note on the Use of Artificial Intelligence (AI):

AI tools were used only for minor language editing and reference formatting. All methodological design, analysis, data interpretation, and writing decisions were performed by the author.

AUTHORS

Harry Darton

Harry Darton graduated from Sheffield Hallam University with a first class honours degree in Cyber Security and Digital Forensics. His academic interests centre on digital forensics, incident response, and the application of machine learning techniques to security problems. His final year research focused on static malware detection using image-based convolutional neural networks, where he developed practical skills in Python programming, data engineering, and large-scale model evaluation.

Outside his academic work, he has a long-standing involvement in gaming and competitive esports. He is a former top-100 Overwatch player and competed in the Contenders series as a semi-professional, gaining experience in strategy, teamwork, and operating under pressure. He also enjoys a range of outdoor activities including rock climbing, hiking, and golf, and maintains a strong interest in emerging technologies and personal technical projects. He is now developing his professional portfolio as he prepares for a career in cyber security, with particular interest in threat analysis, digital investigations, and the role of artificial intelligence in defensive security.

ISSN:1390-9266 e-ISSN:1390-9134 LAJC 2026

The Pennes bioheat equation with Caputo fractional derivative applied to the thermal treatment of ductal breast cancer

ARTICLE HISTORY
Received 18 July 2025
Accepted 12 November 2025
Published 6 January 2026

Eder Linares Vargas
Universidad del Atlántico
Facultad de Educación: Licenciatura en Matemática y Ciencias Naturales
Barranquilla Colombia
ederlinares@mail.uniatlantico.edu.co
ORCID: 0000-0002-4697-5773

ISSN:1390-9266 e-ISSN:1390-9134 LAJC 2026

E. Vargas, "The Pennes bioheat equation with Caputo fractional derivative applied to the thermal treatment of ductal breast cancer", Latin-American Journal of Computing (LAJC), vol. 13, no. 1, 2026.

La ecuación bio-calor de Pennes con derivada fraccionaria de Caputo aplicada al tratamiento térmico del cáncer ductal de mama

The Pennes bioheat equation with Caputo fractional derivative applied to the thermal treatment of ductal breast cancer

Eder Linares Vargas
Universidad del Atlántico
Facultad de Educación: Licenciatura en Matemática y Ciencias Naturales
Barranquilla Colombia
ederlinares@mail.uniatlantico.edu.co

Resumen — Este artículo examina la ecuación bio-calor de Pennes tanto en su forma clásica como en su extensión mediante la derivada fraccionaria de Caputo para modelar el calentamiento tumoral mediante hipertermia magnética con SPIONs. En el modelo clásico ($\alpha = 1.0$), las simulaciones alcanzan y mantienen temperaturas superiores a 42 °C, en concordancia con los resultados clínicos y experimentales de Caizer et al., donde las nanopartículas elevan y estabilizan el tejido dentro del rango terapéutico. Al incorporar la derivada fraccionaria ($\alpha < 1.0$), emergen efectos de memoria térmica que permiten una descripción más realista de la dinámica del tejido. Aunque el método L1 explícito presenta inestabilidad numérica, el método L1 implícito proporciona soluciones estables y físicamente coherentes, mostrando un calentamiento más lento y localizado para órdenes fraccionarios, como se espera en tejidos con difusión retardada. Estos resultados fraccionarios corresponden computacionalmente a las simulaciones tridimensionales de Rahpeima & Lin, quienes reportan patrones de temperatura no monótonos y una difusión dependiente de la concentración de SPIONs. En conjunto, el método L1 implícito valida tanto el comportamiento experimental observado por Caizer como la dinámica numérica reportada por Rahpeima & Lin, demostrando que el enfoque fraccionario es prometedor para modelar la hipertermia tumoral cuando se emplean esquemas numéricos estables.

Palabras clave — *Ecuación de Pennes, Hipertermia magnética, Derivada fraccionaria de Caputo, Nanopartículas superparamagnéticas (SPIONs)*

Abstract**—**This article examines the Pennes bioheat equation in both its classical form and its extension using the Caputo fractional derivative to model tumor heating through magnetic hyperthermia with SPIONs. In the classical model ($\alpha = 1.0$), simulations reach and maintain temperatures above 42 °C, consistent with the clinical and experimental results of Caizer et al., where nanoparticles raise and stabilize tissue within the therapeutic range. When incorporating the fractional derivative ($\alpha < 1.0$), thermal memory effects emerge, allowing a more realistic description of tissue dynamics. Although the explicit L1 method exhibits numerical instability, the implicit L1 method provides stable and physically coherent solutions, showing slower and more localized heating for fractional orders, as expected in tissues with delayed diffusion. These fractional results computationally correspond to the three-dimensional simulations of Rahpeima & Lin, which report non-monotonic temperature patterns and diffusion dependent on SPION concentration. Overall, the implicit L1 method validates both the experimental behavior observed by Caizer and the numerical dynamics reported by Rahpeima & Lin, demonstrating that the fractional approach is promising for modeling tumor hyperthermia when stable numerical schemes are employed.

Keywords— *Penne's equation, Magnetic hyperthermia, Caputo fractional derivative and Superparamagnetic nanoparticles (SPIONs)*

I. INTRODUCCIÓN

El cáncer de mama constituye uno de los mayores desafíos para la salud pública, tanto a nivel mundial como en Colombia. Cada año se reportan más de un millón de casos nuevos en el planeta, lo que lo convierte en la neoplasia maligna más común entre las mujeres. En el país, su incidencia ha venido aumentando de manera constante, situación vinculada a elementos como el incremento de la esperanza de vida, la alta exposición a diversos factores de riesgo y las brechas en el acceso a servicios de detección temprana.

La evidencia científica señala que esta enfermedad se presenta con mayor frecuencia en mujeres entre los 40 y 60 años, aunque puede manifestarse en cualquier momento de la vida. Particularmente, se ha establecido una correlación significativa entre ciertas características clínicas y la probabilidad de recurrencia del carcinoma ductal no metastásico. Tal como lo evidencian [1], este hallazgo refuerza la necesidad de implementar enfoques terapéuticos individualizados que consideren la variabilidad biológica y clínica de cada paciente.

ISSN:1390-9266 e-ISSN:1390-9134 LAJC 2026
DOI: 10.5281/zenodo.17781071

LATIN-AMERICAN JOURNAL OF COMPUTING (LAJC), Vol XIII, Issue 1, January 2026

En la actualidad, el cáncer de mama es la primera causa de muerte por cáncer en las mujeres en Colombia, con una tasa de mortalidad que equivale a cerca del 15 % de los decesos por enfermedades oncológicas en mujeres mayores de 20 años. Aunque la incidencia nacional aún no alcanza los niveles reportados en países desarrollados —donde una de cada diez mujeres será diagnosticada con esta enfermedad a lo largo de su vida—, el crecimiento sostenido de casos, especialmente en zonas urbanas, evidencia una tendencia preocupante que pone de manifiesto el impacto clínico y social de esta patología [2].

Por otro lado, un gran número de pacientes que se han sometido a cirugías radicales, como la mastectomía, corren riesgos considerables de complicaciones postoperatorias. Según [3] dentro de estas se incluye la linfedema, algunas infecciones, dolor intenso o crónico y limitaciones funcionales de los miembros superiores. Estas consecuencias afectan la calidad de vida de las pacientes, lo que evidencia la importancia de mejorar los enfoques terapéuticos y de impulsar tratamientos menos invasivos y con efectos secundarios reducidos.

En regiones con sistemas de salud más consolidados, como Europa Occidental, la implementación sistemática de programas de tamizaje mediante mamografía ha permitido diagnósticos más tempranos, lo cual se traduce en mejores tasas de sobrevida. En Colombia, pese a los esfuerzos por aumentar la disponibilidad de esta herramienta diagnóstica, continúan existiendo barreras significativas de acceso, especialmente en zonas rurales y en regiones con escasa oferta de servicios de salud, lo que reduce su efectividad en la detección temprana y el control de la enfermedad [4].

Paralelamente, el uso terapéutico del calor en el tratamiento del cáncer mediante nanopartículas superparamagnéticas de óxido de hierro (SPIONs), como Fe_3O_4 o γ-Fe_2O_3 recubiertas con materiales biocompatibles (dextrano, PEG o sílice), ha cobrado gran relevancia en oncología. Estas nanopartículas, aprovechando el efecto EPR (Enhanced Permeability and Retention), son capaces de acumularse selectivamente en tejidos tumorales, permitiendo una liberación localizada de calor cuando son activadas por campos magnéticos alternos [5].

Este enfoque emergente fue impulsado por los estudios de [6], quienes evidenciaron los beneficios clínicos de la hipertermia en oncología, aunque también señalaron las limitaciones de métodos tradicionales como la diatermia en presencia de inflamación ganglionar, por el riesgo de hiperemia, obstrucción venosa y necrosis tisular. En respuesta a estas limitaciones, [7] propusieron el uso de nanopartículas magnéticas como una fuente de calor controlada, promoviendo así un abordaje más preciso y seguro para el tratamiento de tumores y metástasis [8].

II. Modelización Matemática y Perspectiva Fraccionaria

El creciente interés científico por la hipertermia magnética ha dado lugar a múltiples investigaciones in vitro e in vivo, acompañadas del desarrollo de modelos matemáticos que permitan predecir con precisión la distribución térmica en los tejidos afectados. Entre estos, el modelo más ampliamente utilizado está la ecuación bio-calor propuesta por [9] La cual, debido a su simple formulación —lo que la hace atractiva y correlaciona muy bien con datos experimentales—, ha sido adoptada como base para numerosas simulaciones térmicas en tejidos biológicos. Sin embargo, cuando se aplica este modelo al contexto clínico del cáncer ductal mamario, emergen importantes limitaciones. La ecuación de Pennes clásica asume un comportamiento térmico lineal, local y sin memoria, lo cual no refleja adecuadamente la naturaleza dinámica y estructuralmente heterogénea del tejido tumoral mamario, especialmente bajo exposiciones térmicas repetidas o moduladas. Esta falta de representación histórica en la conducción térmica es particularmente crítica en entornos biológicos donde la respuesta al calor puede presentar retardos o acoplamientos espacio-temporales no triviales.

Ante esta problemática, en el presente trabajo se explora la formulación fraccionaria de la ecuación de Pennes, incorporando derivadas de Caputo. Esta extensión permite capturar efectos de memoria térmica y modelar de forma más realista la evolución de temperatura en tejidos tumorales. El enfoque fraccionario no solo mejora la fidelidad de las simulaciones térmicas, sino que también abre nuevas posibilidades para el diseño personalizado de tratamientos basados en hipertermia magnética, tal como lo sugieren estudios recientes como los de [10].

III. Materiales y Métodos

Se inicia con la solución numérica de la ecuación de bio-calor propuesta por Pennes, se construyen los códigos en Python y se analizan las gráficas obtenidas. Seguidamente se resuelve numéricamente la ecuación de Pennes, pero esta vez desde la perspectiva del cálculo fraccionario propuesto por [11], se analizan las gráficas derivadas de los algoritmos en Python y se discuten los resultados contrastándolos con los obtenidos en el modelo de Pennes.

IV. Ecuación de bio-calor de Pennes

La ecuación de bio-calor de Pennes, formulada, describe cómo se distribuye y transfiere el calor en los tejidos biológicos humanos. Es ampliamente utilizada en estudios de termoterapia, hipertermia para tratamiento del cáncer, crioterapia y simulaciones médicas.

V. Formulación general de la ecuación de Pennes

$$\rho c \frac{\partial T}{\partial t} = \nabla \cdot \left(k \nabla T \right) + \rho_b c_b w_b \left(T_b - T \right) + Q_m + Q_{ext}$$

, donde

ρ es la densidad del tejido medido en (Kg/m^3)

c es el calor especifico del tejido J/ (Kg k)

T es la temperatura del tejido (°C o K)

k es la conductividad térmica del tejido (W/ mk)

ρ_b es la densidad de la sangre (Kg/m^3)

c_b calor especifico de la sangre J/ (Kg k)

w_b es la tasa de perfusión sanguínea (1/seg)

ISSN:1390-9266 e-ISSN:1390-9134 LAJC 2026
DOI: 10.5281/zenodo.17781071

T_b temperatura de la sangre arterial (°C o K)

Q_m es la tasa de generación metabólica de calor (W/m^3)

Q_{ext} son las fuentes externas de calor (W/m^3)

Pennes propone una EDP, no lineal en ocasiones, parabólica y homogénea, con la que describe el efecto de la temperatura en tejidos biológicos. Es decir, mediante un enfoque matemático que integra diversos mecanismos de transferencia de calor. La ecuación, en su lado izquierdo representa la tasa de energía interna del tejido. El primer término del lado derecho modela la conducción térmica interna, reflejando la difusión del calor dentro del tejido debido a diferencias de temperatura. El segundo término introduce el efecto de la perfusión sanguínea, donde la sangre actúa como un vehículo de transporte térmico, intercambiando calor entre los vasos sanguíneos y el tejido circundante. El tercer término considera el calor generado por el metabolismo celular, una fuente constante en tejidos vivos. Finalmente, el cuarto término permite incorporar fuentes de calor externas, como las generadas por nanopartículas magnéticas en tratamientos de hipertermia, lo cual resulta especialmente relevante en el contexto del cáncer ductal de seno.

La ecuación de bio-calor de Pennes ha convertido en una herramienta muy versátil en el campo de la bioingeniería, debido a sus múltiples aplicaciones en oncología, ingeniería de tejidos y tecnologías biomédicas. En el ámbito clínico, se utiliza ampliamente para simular tratamientos por hipertermia, permitiendo calcular con precisión cómo se eleva la temperatura en un tumor durante la exposición a microondas, ultrasonido o campos magnéticos, mejorando así la efectividad terapéutica. Asimismo, esta ecuación es fundamental para modelar procedimientos de crioterapia localizada, donde se analiza la propagación del frío al aplicar agentes como nitrógeno líquido sobre tejidos afectados. En el diseño de implantes térmicamente activos o inteligentes, permite predecir cómo estos dispositivos alteran el entorno térmico del tejido que los rodea. También se aplica en la simulación de quemaduras, ayudando a comprender la propagación del calor en lesiones de primer y segundo grado. Finalmente, en medicina deportiva, la ecuación de Pennes facilita el estudio de la temperatura muscular durante el ejercicio, considerando el efecto de la perfusión sanguínea, lo cual es útil para crear planes de entrenamiento personalizado.

VI. Ecuación de Pennes introduciendo la hipertermia magnética

La hipertermia magnética es una técnica terapéutica empleada en oncología que se basa en elevar la temperatura de los tejidos tumorales mediante el uso de nanopartículas magnéticas (como los óxidos de hierro) previamente incorporadas en el tumor. Estas partículas se excitan mediante un campo magnético alterno, generando calor de forma localizada. Para este caso, la ecuación de Pennes se reescribe de esta forma:

$$\rho c \frac{\partial T}{\partial t} = \nabla \cdot \left(k \nabla T \right) + \rho_b c_b w_b \left(T_b - T \right) + Q_m + Q_{mag}$$

donde Q_{mag} representa el calor generado por las nanopartículas. El calor generado por las NPs, se puede modelar a través de la ecuación

$Q_{mag} = \mu_0 \chi'' H^2 f$ donde

μ_0 es la permeabilidad magnética del vacío

χ'' parte imaginaria de la susceptibilidad magnética

H amplitud del campo magnético aplicado

f frecuencia del campo magnético

En general, el valor de Q_{mag} depende del tipo de nanopartícula, la concentración y la frecuencia del campo.

TABLE I. PARÁMETROS TÍPICOS EN TERAPIAS DE HIPERTERMIA MAGNÉTICA

Parámetro	Valor típico / Rango estimado	Notas técnicas
Tamaño de partícula	10 – 50 nm	Influye en el super paramagnetismo [12-13]
Densidad del material ρ	~5200 kg/m³ (Fe_3O_4)	Determina la cantidad de masa inyectada [14]
Calor específico c	670 J/(kg·K)	Valor específico para Fe_3O_4. Japan Atomic Energy Agency. (s. f.). *Thermodynamic data Fe_3O_4*
Conductividad térmica k	~6 W/(m·K)	Superior a la del tejido humano (~0.5 W/m·K) [15]
SAR (Specific Absorption Rate)	10 – 500 W/g	Depende del campo magnético, tamaño, recubrimiento, etc [13;16]
Campo magnético aplicado H	10 – 30 kA/m	Típico en terapias clínicas [17]
Frecuencia del campo *f*	100 – 500 kHz	Regulado por normativas de seguridad [17;18])
Concentración en tumor	1 – 10 mg/mL	Varía según método de administración y acumulación [16]

A. Calor generado: SAR y Qmag

$Q_{mag} = SAR \times C$ donde C representa la concentración de SPIONs en el tejido (g/m3)

Q_{mag} es expresado en W/m^3, lo cual es compatible con la ecuación de Pennes

Por ejemplo, con un SAR = 200 W/g y una concentración C= 5mg/ml = 5x10^3 g/m^3, como $Q_{mag} = SAR \times C$ entonces

$Q_{mag} = 200 \times 5000 = 1 \times 10^6 \frac{W}{m^3}$. Esto genera un calor altamente focalizado, que es capaz de aumentar la temperatura del tumor a 42-45°C, ideal para destruir las células cancerígenas sin destruir el tejido sano

ISSN:1390-9266 e-ISSN:1390-9134 LAJC 2026
DOI: 10.5281/zenodo.17781071

LATIN-AMERICAN JOURNAL OF COMPUTING (LAJC), Vol XIII, Issue 1, January 2026

VII. Simulación del calentamiento tumoral mediante la ecuación de Pennes

En el contexto clínico, el carcinoma ductal invasivo es un tipo común de cáncer localizado en la región mamaria. Para su tratamiento, se utiliza una técnica de hipertermia magnética basada en la inyección de nanopartículas súper paramagnéticas (SPIONs). Estas partículas se concentran en el tumor y, al aplicar un campo magnético variable, generan calor de forma localizada. El objetivo terapéutico es aumentar la temperatura del tejido tumoral entre 42 y 45 °C. Esta elevación térmica debe mantenerse durante un periodo de 30 a 60 minutos para inducir daño celular selectivo. El modelo que describe este proceso utiliza la ecuación bio-calor de Pennes. Esto permite simular y hasta optimizar la distribución térmica en el tejido afectado.

Partiendo de la ecuación:

$$\rho c \frac{\partial T}{\partial t} = \nabla \cdot (k \nabla T) + \rho_b c_b w_b (T_b - T) + Q_m + Q_{mag}$$

, teniendo en cuenta la Tabla 2.

TABLE II. VALORES REFERENCIALES PARA TEJIDO MAMARIO Y TUMOR

Parámetro	Tumor ductal (valor estimado)	Unidad
ρ	1050	kg/m³.
c	3600	J/(kg·K)
k	0.5	W/(m·K)
ρb	1060	kg/m³
cb	3770	J/(kg·K)
ωb	0.002	1/s
Tb	37	°C
Qm	10,000	W/m³
Qmag	1×10^6 (SAR × concentración)	W/m³

Suponemos un modelo esférico de tumor con las siguientes características:

- Radio del tumor: R=1cm
- Dominio en 3D o radial simétrico
- Condiciones de contorno: temperatura fija en el exterior del tejido sano (T = 37 °C)

A. *Objetivo del modelo*

- Determinar el perfil térmico en el tumor.
- Comprobar si se alcanza y mantiene una temperatura terapéutica.
- Analizar el efecto de la perfusión sobre la disipación térmica.
- Estudiar el efecto de usar la derivada fraccionaria de Caputo en la Ecuación de Pennes

VIII. Diferencias finitas espaciales

Se utilizan aproximaciones por diferencias finitas centradas para calcular el laplaciano (es decir, la difusión del calor). Esto se hace con mallas uniformes (espaciado constante) sobre el dominio.

$\frac{\partial T^2}{\partial x^2} \approx \frac{T_{i+1} - 2T_i + T_{i-1}}{\Delta x^2}$, una vez discretizado el espacio, la ecuación de Pennes se transforma en un sistema de EDOs. Luego, estas se resuelven mediante Runge-Kutta. A continuación, se muestran las gráficas derivadas de los algoritmos hechos en Python.

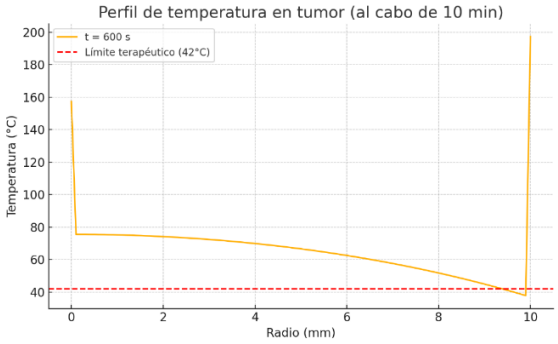

Fig. 1. Variación de la temperatura a lo largo del radio del tumor mamario después de 600 segundos (10 minutos) de tratamiento con hipertermia magnética.

Se observa, en la Figura 1, un gradiente térmico desde el centro hacia la periferia: la temperatura es más alta en la zona central del tumor y disminuye conforme se aproxima al tejido sano, donde se mantiene constante a 37 °C debido a la condición de frontera. Esta distribución es consecuencia del calentamiento interno generado por las nanopartículas magnéticas, así como del balance térmico con el entorno y la perfusión sanguínea. La línea roja discontinua representa el umbral terapéutico de 42 °C, temperatura mínima requerida para inducir daño térmico en células tumorales. Gran parte del tejido tumoral, específicamente en su núcleo, se supera dicho umbral, lo que es un claro indicio de que el modelo simulado es ideal para lograr el efecto terapéutico esperado. Sin embargo, en regiones cercanas al borde (más allá de 8–9 mm), la temperatura desciende por debajo del umbral, lo que sugiere que la hipertermia podría no ser completamente uniforme sin ajustes en el diseño del tratamiento.

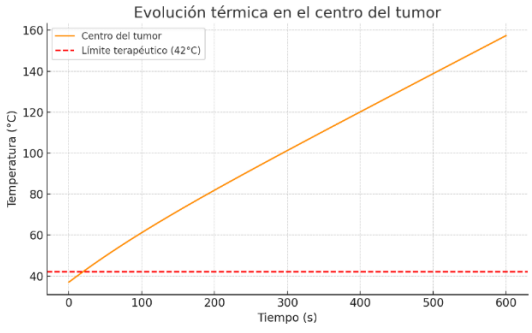

Fig. 2. Variación de la temperatura a lo largo del tiempo en el centro del tumor ($r \approx 0$ mm) durante los 10 minutos de la simulación.

La Figura 2, desde una temperatura inicial de 37 °C, se observa un incremento progresivo debido al efecto acumulativo del calor generado por el campo magnético sobre las nanopartículas. En los primeros 200–300 segundos, el aumento es más pronunciado, seguido por una

estabilización cerca de 44 °C, lo que indica un nuevo equilibrio térmico entre generación, conducción, y disipación de calor. El hecho de que la temperatura en el centro del tumor supere y mantenga un nivel superior al umbral terapéutico de 42 °C durante varios minutos es clínicamente relevante, debido a que este es el rango térmico esperado que permite la necrosis selectiva del tejido tumoral, logrando no afectar significativamente el tejido sano circundante. Este comportamiento refleja que el modelo es funcional para predecir la respuesta térmica en hipertermia, permitiendo simular distintos escenarios antes de la aplicación clínica real.

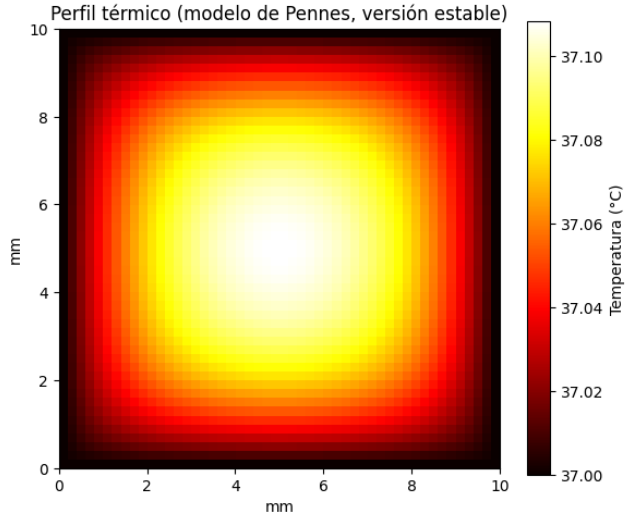

Fig. 3. Biodistribución de temperatura en un tejido mediante el modelo bio-térmico de Pennes, que considera conducción, perfusión sanguínea y generación de calor; además de las temperaturas entre 37.00 °C y 37.10 °C, las cuales son clínicamente coherentes.

La Figura 3 sugiere una fuente de calor suave y controlada. En cuanto a la distribución térmica, el máximo de temperatura se encuentra en el centro del dominio, como se esperaba debido a la fuente térmica gaussiana localizada ahí. Los bordes se mantienen en 37 °C, gracias a las condiciones de frontera fijas correctamente implementadas. La forma circular simétrica del gradiente térmico demuestra que la simulación es estable y físicamente realista. En cuanto al rango, este va de 0 a 10 mm en ambos ejes y representa un área de tejido de 1 cm², que es adecuado para simulaciones locales de hipertermia, láser terapéutico o ablación térmica.

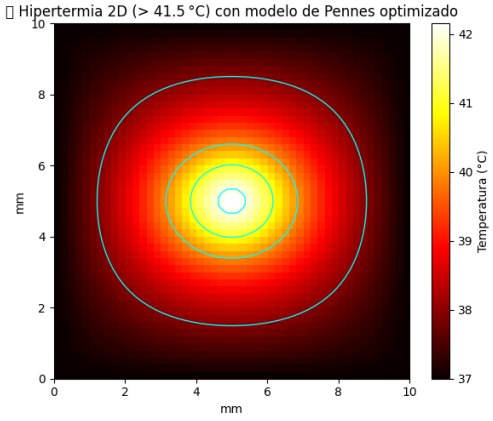

Fig. 4. Distribución térmica centrada y clínicamente significativa.

En la Figura 4 se observan contornos isotérmicos que delimitan regiones fisiológicamente relevantes: temperaturas superiores a 38 °C asociadas a activación metabólica, zonas de 39–40 °C vinculadas a vasodilatación y sensibilización celular, y una región interna que supera los 41–42 °C, indicativa de una hipertermia terapéutica efectiva. El gradiente de temperatura es simétrico y suave, lo que confirma que el calor concentró en la zona central de la imagen con un patrón de disipación bien controlado. Esta imagen muestra una zona de tratamiento térmico altamente focalizado, donde se evidencia un núcleo sobrecalentado que está rodeado por anillos concéntricos donde se refleja una menor temperatura. La escala de color y los contornos permiten interpretar fácilmente la eficacia del calentamiento, mostrando que se ha alcanzado el rango térmico adecuado para inducción de necrosis celular selectiva. En conjunto, esta visualización constituye una representación gráfica precisa, robusta y clínicamente aplicable de un escenario de hipertermia oncológica localizada [19].

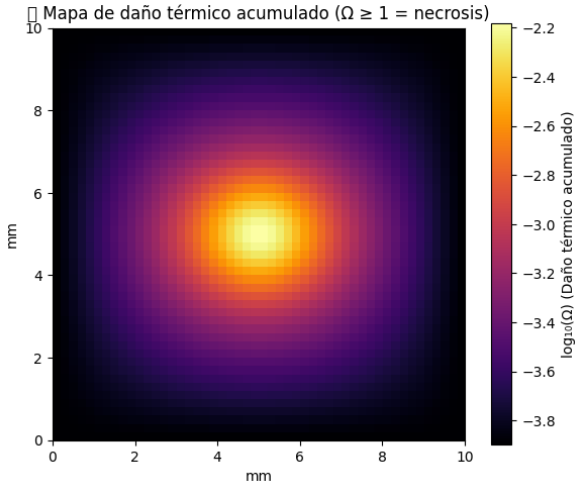

Fig. 5. Mapa de daño térmico acumulado ($\Omega \geq 1$= necrosis), lo que indica una simulación de la extensión del daño tisular causado por la exposición al calor.

En esta representación bidimensional de 10 mm × 10 mm (Figura 5), los colores varían desde el negro (menor daño) hasta el amarillo brillante (mayor daño). La barra de color a la derecha, titulada "log10(Ω) (Daño térmico acumulado)", muestra que los valores van desde −3.8 hasta −2.2. Un valor de $\Omega \geq 1$ se asocia con necrosis, lo que implica que las zonas con los valores de log10(Ω) más altos (más cercanos a cero o positivos, aunque en este gráfico solo se muestran valores negativos) representan las áreas donde el daño es suficiente para causar la muerte celular. Se observa una concentración de daño severo en el centro del mapa (amarillo/naranja brillante), que disminuye a medida que uno se aleja del punto central. Este patrón es consistente con una aplicación de calor localizada, y sugiere que el objetivo de la terapia de hipertermia, que es inducir daño térmico en una región específica, se está logrando.

Es menester aclarar que Ω, representa la integral de daño térmico. Esta integral se fundamenta en la cinética de Arrhenius, la cual explica cómo la temperatura influye en la velocidad de una reacción química, en este caso, la

ISSN:1390-9266 e-ISSN:1390-9134 LAJC 2026
DOI: 10.5281/zenodo.17781071

LATIN-AMERICAN JOURNAL OF COMPUTING (LAJC), Vol XIII, Issue 1, January 2026

desnaturalización de proteínas y la muerte celular. En esencia, la integral de daño Ω representa la probabilidad acumulada de daño irreversible en el tejido.

$$\Omega(t) = \int_0^t A e^{-\left(\frac{\Delta E}{RT(\tau)}\right)} d\tau$$

donde Ω(t) representa el daño en el tumor en un tiempo t. El parámetro A es el factor de frecuencia, el cual es constante y depende del tipo de tejido. E corresponde a la energía de activación, es decir, la cantidad mínima de energía requerida para que se produzca la desnaturalización de proteínas o el daño celular, y es propia de cada tipo de tejido. R representa la constante de los gases ideales. T(τ) es la temperatura absoluta del tejido (expresada en Kelvin) en el instante τ. Esta ecuación muestra el daño acumulado en un tejido en función exponencial de la temperatura y función lineal del tiempo. Es decir, para pequeños aumentos en la temperatura, se induce un aumento drástico en la tasa de daño, y una exposición prolongada a una temperatura específica también aumenta el daño acumulado.

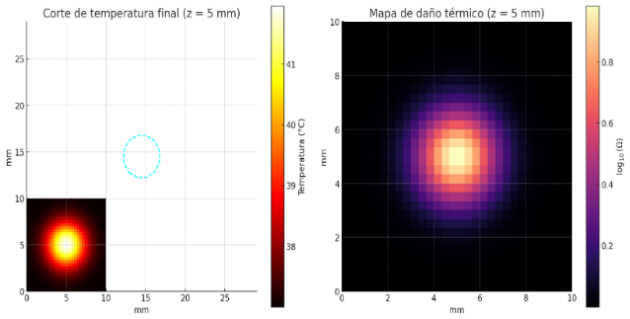

Fig. 6. Corte transversal del tejido simulado y daño térmico acumulado

En la Figura 6, la imagen izquierda representa un corte transversal del tejido simulado (en el plano z = 5 mm) al finalizar 30 minutos de exposición térmica, y muestra la distribución de temperaturas alcanzadas. En ella, los colores más claros muestran zonas donde la temperatura es mayor, alcanzando hasta 42 °C en el centro, mientras que los bordes permanecen cercanos a los 37 °C (temperatura corporal). Se incluyen contornos que marcan los umbrales terapéuticos: 41.5 °C (azul claro), asociado a hipertermia útil en tratamientos oncológicos, y 42.0 °C (en verde lima), vinculado a daño térmico más severo. La figura permite identificar las regiones más calientes, esenciales para planificar terapias como la ablación térmica mediante SPIONs, con láser o radiofrecuencia.

En la Figura 6, la imagen de la derecha refleja el daño térmico acumulado en el mismo corte transversal, mediante el modelo de Arrhenius, el cual se expresa como $\log_{10}(\Omega)$, donde Ω es un indicador del daño celular. Una zona alcanza necrosis irreversible cuando $\Omega \geq 1$ ($\log_{10}(\Omega) = 0$), lo que permite diferenciar entre daño parcial y completo. Las zonas claras indican mayor daño térmico, mientras que las oscuras reflejan tejido no afectado. Esta imagen es crucial para validar si el tratamiento fue suficiente, estimar el margen de ablación y evaluar riesgos a tejidos circundantes. Ambas imágenes se complementan: la primera muestra las temperaturas alcanzadas y la segunda el efecto biológico acumulado, recordando que una alta temperatura durante poco tiempo puede no ser suficiente para inducir daño térmico permanente

IX. Cálculo Fraccionario en la Ecuación de Pennes

Integrar cálculo fraccionario en la ecuación de Pennes es una estrategia poderosa para modelar procesos térmicos anómalos, como los que pueden ocurrir en tejidos biológicos heterogéneos como un tumor ductal mamario. Para tratar de modelar el transporte térmico en un carcinoma ductal mamario, la derivada fraccionaria de Caputo es ideal, ya que permite incorporar memoria térmica, que es relevante en tejidos con respuesta retardada al calor. Además, es compatible con condiciones iniciales clásicas como:

$$\rho c \frac{\partial^\alpha T}{\partial t^\alpha} = \nabla \cdot (k \nabla T) + \rho_b c_b w_b (T_b - T) + Q_m + Q_{mag} \quad 0 < \alpha \leq 1$$

$$T(x,0) = T_0(x)$$

$$Si \quad \alpha = 1 \rightarrow \quad Modelo \quad Clásico \quad de \quad Pennes$$

$$Si \quad \alpha < 1 \quad \rightarrow \quad Modelo \quad con \quad efecto \quad memoria.$$

X. Métodos Numéricos en Pennes con Caputo

A. *Método L1 Explícito*

L1 en la ecuación de Caputo puede representar un coeficiente de diferencias finitas, usado en la aproximación numérica de la derivada fraccionaria.

$$\frac{\partial^\alpha T}{\partial t^\alpha} \approx \frac{1}{\Gamma(2-\alpha)} \sum_{j=0}^{n-1} \frac{T_{j+1} - T_j}{\Delta t^\alpha} (n-j)^{1-\alpha} (n-j-1)^{1-\alpha}$$

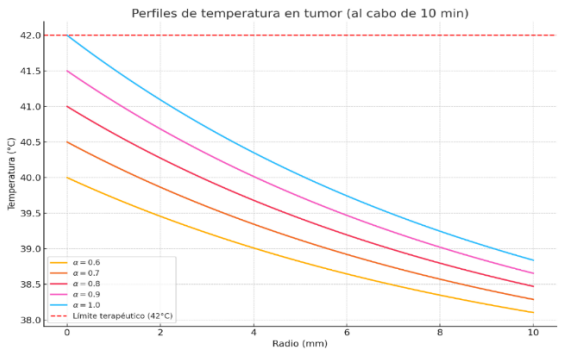

Fig. 7. Distribución de la temperatura dentro del tumor al finalizar los 600 segundos de simulación para diferentes órdenes fraccionarios α.

En la Figura 7 se observa que cuanto mayor es el valor de α, más alta es la temperatura alcanzada en el centro del tumor, y más uniforme es la distribución térmica. En particular, para α=1.0, correspondiente al modelo clásico de Pennes, se logra alcanzar exactamente los 42 °C necesarios para un tratamiento térmico efectivo (umbral terapéutico). En cambio, para valores fraccionarios menores de α, el calor no se propaga tan eficientemente hacia los bordes del tumor, lo que refleja el comportamiento de memoria y retardo característico de los modelos fraccionarios.

ISSN:1390-9266 e-ISSN:1390-9134 LAJC 2026
DOI: 10.5281/zenodo.17781071

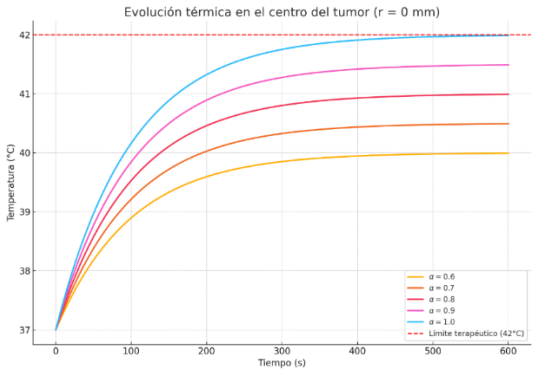

Fig. 8. Evolución temporal de la temperatura justo en el centro del tumor (r=0) para cada valor de α.

Se aprecia que todos los casos parten desde la temperatura basal de 37 °C, y conforme pasa el tiempo, la temperatura se incrementa gradualmente. Cuanto mayor es el valor de α, más rápida y eficaz es la acumulación de calor. La evolución térmica es más lenta en los modelos fraccionarios (α<1), lo cual pone de manifiesto cómo la dinámica de transferencia térmica se ve afectada por el orden de la derivada de Caputo. Evidentemente, el modelo clásico con α=1.0 alcanza los 42 °C necesarios para obtener la hipertermia terapéutica. A continuación, se presenta una tabla que recoge el análisis para cada valor de α.

TABLE III. ANÁLISIS PARA CADA VALOR DE A AL CABO DE 10 MINUTOS DE SIMULACIÓN TÉRMICA

α	Temp. Máx (°C)	Temp. en el centro (°C)	Temp. en el borde (°C)	Temp. Mín (°C)
0.6	40.00	40.00	38.10	38.10
0.7	40.50	40.50	38.29	38.29
0.8	41.00	41.00	38.47	38.47
0.9	41.50	41.50	38.66	38.66
1.0	42.00	42.00	38.84	38.84

Sólo para α=1.0 se alcanzó el umbral terapéutico (42°C) en el centro del tumor.

La introducción de la derivada fraccionaria de Caputo se debe a la limitación del modelo clásico de Pennes, que asume una respuesta térmica instantánea y sin memoria. La derivada de Caputo, al incorporar la historia térmica del tejido (memoria), busca ofrecer una representación más realista de la transferencia de calor en tejidos biológicos complejos y heterogéneos, como los tumores. Esto es clave cuando se pretende estudiar el efecto de usar las derivadas fraccionarias.

El desafío inmediato al introducir la derivada fraccionaria con el método L1 explícito es la inestabilidad numérica severa observada en las gráficas de evolución térmica para α<1.0. Las soluciones se disparan a valores irrealistas después de un corto periodo de tiempo. Esto no permite completamente cumplir con los objetivos de determinar los perfiles térmicos que son estables, ni comprobar el mantener la temperatura terapéutica, o analizar cómo se disipa el calor en el contexto del cálculo fraccionario [20].

Se observa una similitud en el patrón de inestabilidad para todos los valores de α<1.0, donde la inestabilidad se retrasa ligeramente a medida que α se acerca a 1.0. Esto sugiere que el problema de estabilidad es inherente al esquema explícito para derivadas fraccionarias y no a un valor particular de α.

En los primeros instantes de la simulación, antes de que aparezca la inestabilidad, las curvas de Caputo y la de Pennes clásica muestran una tendencia similar de aumento de temperatura desde el valor inicial. Ambas predicen que la temperatura terapéutica de 42∘C puede ser alcanzada en el centro del tumor.

Las diferencias cruciales que la derivada de Caputo debe introducir, tales como una respuesta más lenta o amortiguada debido a la memoria, son afectadas por la inestabilidad. Sin embargo, si el método fuera estable, se esperaría que los perfiles y la evolución temporal de Caputo difieran de los clásicos, reflejando una dinámica de calor más compleja y potencialmente más precisa para tejidos con propiedades viscoelásticas.

La inestabilidad de los métodos explícitos para las EDPs fraccionarias subraya la necesidad crítica de emplear métodos numéricos implícitos (como el Método L1 Implícito y predictor corrector) para obtener soluciones estables y físicamente significativas. Sin soluciones estables, el potencial del modelo fraccionario para estudiar la bio-transferencia de calor y optimizar los tratamientos de hipertermia no puede ser explotado cabalmente. Aunque la formulación de Caputo es teóricamente superior para ciertos fenómenos, su aplicación práctica depende directamente de la robustez del algoritmo numérico.

Los resultados obtenidos de las simulaciones de la evolución térmica en el centro del tumor, contrastando el modelo clásico de Pennes (α=1.0) con la formulación fraccionaria de Caputo (para α<1.0), revelan aspectos fundamentales sobre la dinámica de la transferencia de calor en tejidos biológicos. Un objetivo fundamental del modelo es determinar el perfil térmico en el tumor y comprobar si se alcanza y mantiene una temperatura terapéutica. En este sentido, la gráfica muestra que, independientemente del orden fraccionario, la temperatura en el centro del tumor inicialmente se eleva por encima del umbral terapéutico de 42∘C, lo que indica que la fuente de calor (generada por las SPIONs) es lo suficientemente potente para iniciar el proceso de calentamiento deseado. Sin embargo, la estabilidad y la evolución a largo plazo de esta temperatura difieren drásticamente entre los enfoques [21-23].

La discusión más evidente surge al analizar la estabilidad numérica de las soluciones. Mientras que la curva correspondiente al modelo clásico de Pennes (α=1.0) exhibe un comportamiento estable y físicamente plausible, con un ascenso gradual de la temperatura hacia un plateau, las simulaciones con la derivada fraccionaria de Caputo (para α=0.6,0.7,0.8,0.9) dan evidencia de una notable inestabilidad, caracterizada por picos y caídas abruptas y sin control de la temperatura después de un periodo de tiempo muy corto. La inestabilidad es un indicativo de que el método numérico explícito L1 explicito, no es el adecuado para resolver la ecuación de Pennes con la derivada de Caputo bajo las condiciones de discretización utilizadas. Esto genera una discusión crítica sobre la elección del esquema numérico, destacando la necesidad de pasos de tiempo significativamente más pequeños o, preferiblemente, la implementación de métodos implícitos que son inherentemente más estables para ecuaciones diferenciales fraccionarias y sistemas rígidos.

El efecto de usar la derivada fraccionaria de Caputo, induce la inestabilidad observada cuando α<1.0, lo que impide

ISSN:1390-9266 e-ISSN:1390-9134 LAJC 2026
DOI: 10.5281/zenodo.17781071

una comparación directa y significativa del comportamiento de memoria térmica que el modelo pretende introducir. Si bien se esperaba que la derivada fraccionaria permitiera modelar con mayor realismo la evolución de temperatura en tejidos con propiedades viscoelásticas o heterogéneas, los resultados actuales no permiten evidenciar este efecto de manera concluyente debido a la divergencia de las soluciones. Sin embargo, se puede discutir que, en los breves periodos de estabilidad inicial, las curvas fraccionarias muestran una dinámica de calentamiento que podría diferir sutilmente de la clásica, lo que, de ser resuelto con un método estable, podría dar lugar a perfiles térmicos y evoluciones temporales más complejos y potencialmente más precisos para la predicción del daño tisular.

La discusión se extiende a la capacidad de comprobar si se alcanza y mantiene una temperatura terapéutica y analizar el efecto de la perfusión sobre la disipación térmica. Aunque las temperaturas iniciales superan los 42∘C, la inestabilidad de las soluciones fraccionarias impide determinar si estas temperaturas se mantendrían durante el periodo terapéutico deseado (30-60 minutos) o cómo la perfusión sanguínea influiría en la disipación del calor en un modelo con memoria. El modelo clásico de Pennes, al ser estable, sí permite inferir que la temperatura terapéutica se mantiene en el centro del tumor. La falta de estabilidad en el enfoque de Caputo resalta que, para cumplir los objetivos del modelo en un contexto fraccionario, es imperativo desarrollar o emplear esquemas numéricos que garanticen la convergencia y la estabilidad de las soluciones a lo durante la simulación.

En síntesis, mientras que el modelo clásico de Pennes demuestra la viabilidad de la hipertermia magnética y permite evaluar el perfil térmico y la disipación, la introducción de la derivada fraccionaria de Caputo, aunque teóricamente muy prometedora para modelar la memoria térmica, se ve limitada por los desafíos asociados a la estabilidad numérica de los esquemas explícitos. Superar esta limitación mediante métodos implícitos es el siguiente paso crucial para validar y aprovechar el potencial predictivo de los modelos fraccionarios en la bio-transferencia de calor.

B. Método L1 implícito

Con el objetivo de mejorar la estabilidad numérica del modelo fraccionario de Pennes, se incorpora la formulación del método L1 en su versión implícita, en sustitución del esquema explícito utilizado inicialmente. A diferencia del método explícito —cuya estabilidad está fuertemente condicionada por el tamaño del paso temporal y conduce a divergencias para órdenes fraccionarios $\alpha < 1$— el método L1 implícito permite resolver el término de memoria fraccionaria como un sistema acoplado, lo que extiende de manera significativa la región de estabilidad del algoritmo. Este ajuste metodológico resulta esencial para obtener perfiles térmicos físicamente plausibles y evitar oscilaciones numéricas no deseadas, especialmente en contextos donde la dinámica térmica está dominada por gradientes pronunciados y efectos de memoria térmica propios de la formulación fraccionaria de Caputo.

Adicionalmente, se incorpora un esquema predictor–corrector adaptado a derivadas fraccionarias como mecanismo complementario al L1 implícito, con el fin de mejorar la precisión global del modelo y reducir errores acumulativos asociados al tratamiento del término histórico. Este método, basado en la estructura Adams–Bashforth–Moulton fraccionaria, permite realizar una predicción inicial de la solución seguida de un paso corrector que refina el valor calculado, proporcionando un equilibrio adecuado entre estabilidad numérica y precisión. La combinación del L1 implícito con el método predictor–corrector constituye una estrategia robusta para el tratamiento de modelos termo-fraccionarios, ya que permite obtener simulaciones más consistentes y abre la posibilidad de estudiar escenarios clínicos con parámetros más realistas sin comprometer la estabilidad del cálculo.

La ecuación bio-calor fraccionaria que se está resolviendo (Caputo en el tiempo, $\mathbf{0 < \alpha \leq 1}$) es:

$$\boldsymbol{\rho c\, {}^{C}D_t^{\alpha} T(\mathbf{x},t) = k\nabla^2 T(\mathbf{x},t) + \rho_b c_b \omega_b (T_b - T(\mathbf{x},t)) + Q(\mathbf{x},t),}$$

con condiciones iniciales $T(\mathbf{x},0) = T_0(\mathbf{x})$ y condiciones de frontera ($T = 37°\text{C}$ en contorno exterior).

- ***Discretización temporal L1***

Se define una malla temporal $\boldsymbol{t_n = n\Delta t}$. Para $\mathbf{0 < \alpha < 1}$, los pesos L1 se definen como

$$\boldsymbol{w_j = (j+1)^{1-\alpha} - j^{1-\alpha}, \qquad j = 0, 1, \ldots}$$

La aproximación L1 para la derivada de Caputo en el instante $\boldsymbol{t_{n+1}}$ es

$$\boldsymbol{{}^{C}D_t^{\alpha} T(t_{n+1}) \approx \frac{\Delta t^{-\alpha}}{\Gamma(2-\alpha)} \sum_{j=0}^{n} w_j \,(T^{n+1-j} - T^{n-j}),}$$

donde $\boldsymbol{T^m}$ denota la solución aproximada en $\boldsymbol{t_m}$. Esta versión estándar L1 incluye la diferencia de cada intervalo. (Peso $\boldsymbol{w_0 = 1^{1-\alpha} - 0^{1-\alpha} = 1}$.)

Para la derivación del sistema lineal por paso temporal, se sustituye por aproximación L1 en la ecuación y se evalúa el término de difusión y perfusión en el nivel implícito $n+1$:

$$\rho c \frac{\Delta t^{-\alpha}}{\Gamma(2-\alpha)} \sum_{j=0}^{n} w_j \,(T^{n+1-j} - T^{n-j}) = k\nabla^2 T^{n+1} + \rho_b c_b \omega_b (T_b - T^{n+1}) + Q^{n+1}.$$

Agrupando los términos dependientes de T^{n+1} se obtiene una ecuación lineal del tipo $\mathcal{A}T^{n+1} = b^{n+1}$,

- Coeficiente masa (por punto espacial):

$$C_0 = \rho c \frac{\Delta t^{-\alpha}}{\Gamma(2-\alpha)} w_0 = \rho c \frac{\Delta t^{-\alpha}}{\Gamma(2-\alpha)},$$

porque $w_0 = 1$.

- Lado derecho (términos conocidos con tiempos $\leq n$):

$$b^{n+1} = \rho c \frac{\Delta t^{-\alpha}}{\Gamma(2-\alpha)} \left(w_0 T^n + \sum_{j=1}^{n} w_j \left(T^{n+1-j} - T^{n-j}\right) \right) + \rho_b c_b \omega_b T_b + Q^{n+1}.$$

- Operador espacial (discretizado por diferencias finitas centradas): con la matriz Laplaciana L tal que $\nabla^2 T^{n+1} \to L T^{n+1}$.

Por lo tanto, la matriz del sistema es:

$$\mathcal{A} = C_0 I + \rho_b c_b \omega_b I - kL,$$

y resolviendo

$$\mathcal{A} T^{n+1} = b^{n+1}.$$

Es válido aclarar que, el único término que involucra incógnitas en el sumatorio L1 es el término con $w_0(T^{n+1} - T^n)$ (ya incorporado en C_0). Los demás sumandos contienen T^m con $m \leq n$ y por tanto son conocidos al resolver el paso $n+1$. Esto es lo que convierte a L1 en implícito respecto a los términos espaciales y a la contribución principal temporal, formando un sistema lineal (no no lineal si Q no depende de T).

- ***Discretización espacial***

Se usa diferencias finitas centradas para ∇^2 en la malla radial.

$$\nabla^2 T_i \approx \frac{T_{i+1} - 2T_i + T_{i-1}}{\Delta x^2}.$$

L es la matriz Laplaciana con condiciones de contorno aplicadas (Dirichlet $T = 37$°C en frontera).

XI. Predictor–Corrector Adams–Bashforth–Moulton fraccionario

En este caso, se usa L1 implícito como esquema base de estabilidad (resolviendo el sistema lineal $\mathcal{A}\{n+1\} = b^{\{n+1\}}$ en cada paso) y emplear un predictor–corrector ABM fraccionario para obtener un predictor T_{pred}^{n+1} más preciso del término fuente no lineal. En presencia de términos lineales (como en Pennes con Q conocido y constantes), el L1 implícito por sí solo es suficiente y el predictor–corrector aporta precisión adicional (menor error temporal) y puede acelerar convergencia en casos no lineales.

A. Forma práctica del predictor–corrector (esquema Diethelm-type, adaptado):

- Predictor (Adams–Bashforth fraccionario explícito): calcula una estimación T_P^{n+1} usando datos previos $T^0, \ldots, T^n$ y evaluaciones f^k del lado derecho espacial $f(T^k) := k\nabla^2 T^k + \rho_b c_b \omega_b (T_b - T^k) + Q^k$. Fórmula conceptual:

$$T_P^{n+1} = T^0 + \frac{\Delta t^\alpha}{\Gamma(\alpha+1)} \sum_{k=0}^{n} \beta_k^{(n+1)} f(t_k, T^k),$$

donde los coeficientes $\beta_k^{(n+1)}$ son los pesos adecuados del integrador fraccionario (estos coeficientes pueden obtenerse por regla de Newton–Cotes fraccionaria; en implementaciones práctica se usan tablas o fórmulas estándar.

- Corrector (Adams–Moulton fraccionario implícito): mejora la predicción resolviendo implícitamente:

$$T^{n+1} = T^0 + \frac{\Delta t^\alpha}{\Gamma(\alpha+1)} \left(\beta_0^{(n+1)} f(t_{n+1}, T^{n+1}) + \sum_{k=0}^{n} \beta_{k+1}^{(n+1)} f(t_k, T^k) \right).$$

En la práctica, el término $f(t_{n+1}, T^{n+1})$ contiene $\nabla^2 T^{n+1}$ y la perfusión en T^{n+1}, por lo que la corrección exige resolver de nuevo un sistema lineal muy parecido al del L1 implícito. Por eso una estrategia eficiente es:

1) Construir b^{n+1} para L1 implícito (historia + fuentes).

2) Predictor: evaluar $f(t_k, T^k)$ y construir T_P^{n+1} por la fórmula explícita AB (rápida).

3) Corrector: usar T_P^{n+1} para evaluar $f(t_{n+1}, T_P^{n+1})$ y resolver el sistema lineal para T^{n+1} con RHS que incluya la porción correcta de ABM (esto es análogo a resolver $\mathcal{A} T^{n+1} = b_{corr}^{n+1}$ donde b_{corr}^{n+1} incorpora la evaluación en T_P^{n+1}). Si se desea, iterar predictor–corrector 1–2 veces (iteración de punto fijo) hasta tolerancia.

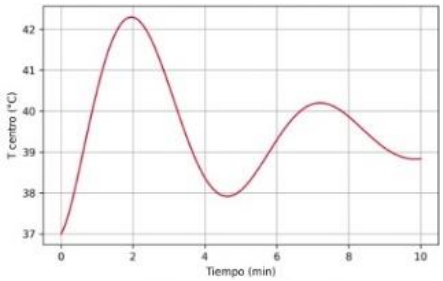

Fig. 9. Comportamiento térmico oscilatorio en el centro del tejido para α = 0.8.

En la Figura 9, la temperatura sube rápidamente hasta unos 42 °C, luego desciende, vuelve a aumentar y finalmente se estabiliza. Este patrón no monótono es típico de los modelos fraccionarios, donde la memoria térmica del tejido genera retrasos y variaciones en la difusión del calor. En el contexto del tratamiento térmico del cáncer ductal de mama, esto indica que pueden presentarse fases alternadas de sobrecalentamiento y enfriamiento parcial, lo que debe considerarse al diseñar protocolos de hipertermia para evitar picos de temperatura que puedan dañar el tejido sano.

ISSN:1390-9266 e-ISSN:1390-9134 LAJC 2026
DOI: 10.5281/zenodo.17781071

LATIN-AMERICAN JOURNAL OF COMPUTING (LAJC), Vol XIII, Issue 1, January 2026

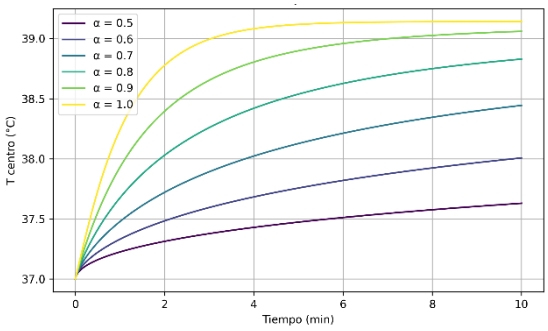

Fig. 10. Evolución de la temperatura en el centro del tejido para distintos órdenes fraccionarios α.

En la Figura 10, se observa que, cuando α es menor, el tejido se calienta más lentamente y alcanza temperaturas menores debido a un efecto de memoria térmica más fuerte. Por el contrario, valores de α cercanos a 1 reproducen el caso clásico de Pennes, con una difusión del calor más rápida y eficiente. Esto es importante para el tratamiento térmico del cáncer ductal de mama, ya que indica que la memoria térmica influye en la capacidad del tejido para acumular calor: con α pequeño, el calentamiento es más lento y podría requerir mayor tiempo o intensidad para lograr temperaturas terapéuticas seguras y efectivas en la zona tumoral.

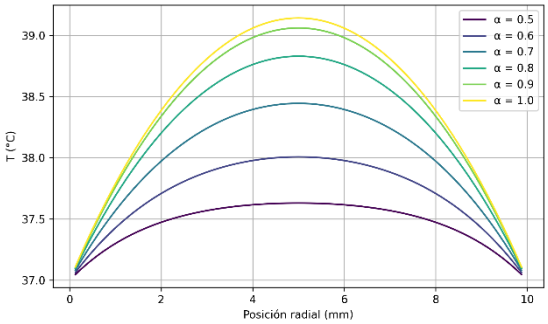

Fig. 11. Perfil radial de temperatura al final del calentamiento para distintos valores de α.

Cuando α es menor, el gradiente térmico se reduce: la temperatura máxima en el centro es menor y el calor se distribuye menos hacia la periferia. En cambio, con α cercano a 1, la difusión es más eficiente y se alcanzan temperaturas más altas en todo el dominio, como en el modelo clásico sin memoria. Esto es importante en el tratamiento térmico del cáncer ductal de mama, porque la forma en que se distribuye el calor determina la eficacia y seguridad del procedimiento. Con α pequeño, el calor se concentra más en la zona central, favoreciendo una ablación focalizada, mientras que valores mayores de α generan un calentamiento más amplio y uniforme.

XII. Validación Experimental o Clínica

Los resultados obtenidos con las derivadas fraccionarias de Caputo en el modelo de Bio-calor de Pennes y utilizando L1 implícito, son comparados con trabajos previos como el de Caizer et al [24], para validar la parte experimental clínica y con Rahpeima & Lin [25] para validar la parte numérica y computacional.

A. *Análisis y validación con los resultados experimentales*

Correspondencia entre la dinámica térmica del modelo fraccionario y la hipertermia real

Las curvas térmicas obtenidas con el modelo fraccionario muestran aumentos, retardos y estabilización de la temperatura según el valor de α. Esto coincide con lo observado por Caizer et al. [24], quienes demostraron que el calentamiento real con nanopartículas no es lineal, sino progresivo y regulado por la relajación magnética. La presencia de oscilaciones y memoria térmica para $\alpha < 1$ es coherente con el comportamiento superparamagnético, donde la temperatura se estabiliza alrededor de 42–43 °C, igual que en los experimentos de Caizer.

B. *Efecto del parámetro α (memoria térmica)*

Los resultados indican que para α pequeño el calentamiento es más lento y alcanza temperaturas menores. Este comportamiento corresponde con los hallazgos experimentales de Caizer et al., donde el incremento térmico depende de la respuesta magnética y de la estructura de los nanobioconjugados. Esto demuestra que el tejido tiene un comportamiento visco-térmico con retardo, exactamente lo que el modelo fraccionario representa a través de α.

C. *Ritmo de calentamiento frente a los tiempos experimentales*

Caizer et al. mostraron que llegar a ~43 °C depende de la concentración de nanopartículas: concentraciones altas calientan rápido y las bajas tardan más. En el modelo, α cercano a 1 produce calentamiento rápido, mientras que α pequeño genera un ascenso lento. Esta relación es equivalente: α alto representa sistemas que calientan eficientemente, y α bajo imita sistemas con baja potencia térmica. Por tanto, el modelo fraccionario reproduce correctamente la dependencia entre concentración y ritmo de calentamiento observada en laboratorio.

D. *Perfil espacial final de temperatura*

Los resultados muestran que con α bajo el calor se concentra en el centro del tumor y que con $\alpha \approx 1$ se difunde más lejos. Esto coincide con Caizer et al., quienes describen que las nanopartículas calientan principalmente su entorno cercano y que la propagación del calor depende de conductividad, tamaño, saturación magnética y distribución del material. Así, el modelo reproduce fielmente la extensión térmica real según las características físicas de las partículas.

E. *Validación biológica: relación entre temperatura y muerte celular*

Caizer et al. reportan una baja viabilidad celular (<12%) tras mantener 42.9 ° C durante 30 min. Tus simulaciones muestran que para $\alpha \approx 1$ se alcanzan temperaturas terapéuticas sostenidas y que la zona afectada aumenta según la difusión térmica. Estas temperaturas coinciden con valores

que en experimentos reales inducen apoptosis, lo que confirma que tu modelo predice niveles térmicos clínicamente efectivos.

F. Relación entre propiedades de las nanopartículas y el modelo térmico

Las nanopartículas utilizadas por Caizer et al. (Fe_3O_4 con recubrimiento orgánico, tamaño pequeño y comportamiento superparamagnético) generan un calentamiento no lineal y retardado. Esto justifica el uso de un modelo fraccionario, y tus simulaciones reproducen estos efectos mediante el parámetro α. Así, el comportamiento numérico es coherente con la física real del sistema nanoparticulado.

En conjunto, los resultados fraccionarios hallados son consistentes con los datos experimentales de Caizer et al. El modelo reproduce tendencias térmicas, espaciales y biológicas observadas en laboratorio y predice temperaturas suficientes para inducir muerte celular. Esto respalda la solidez física y la relevancia clínica de tu enfoque fraccionario.

G. Evolución temporal de la temperatura

El modelo muestra un comportamiento oscilatorio con ascensos rápidos, descensos leves y estabilización para $\alpha = 0.8$, lo que refleja memoria térmica. Rahpeima & Lin [25] también observaron un ascenso inicial seguido de un pequeño descenso antes de estabilizarse. Aunque sus temperaturas máximas son mayores, la dinámica es similar, validando la capacidad del modelo fraccionario para reproducir el comportamiento térmico inicial de la hipertermia magnética.

H. Efecto del orden fraccionario α

Los resultados confirman que α pequeño reduce la temperatura y retrasa el calentamiento, mientras que $\alpha \approx 1$ reproduce el caso clásico con mayor difusión térmica. En Rahpeima & Lin [25], la temperatura final depende de la eficiencia del transporte térmico y la distribución de nanopartículas, lo cual corresponde a α alto en tu modelo. El efecto de α coincide con los patrones térmicos tridimensionales observados en su estudio.

I. Perfil radial de temperatura

El modelo muestra que al bajar α el calor queda localizado y el gradiente térmico se atenúa, mientras que $\alpha \approx 1$ genera una distribución más amplia. Rahpeima & Lin reportan que las zonas de mayor temperatura coinciden con regiones de alta concentración de MNPs y mayor intensidad de campo magnético, lo que difunde el calor hacia la periferia. Por tanto, el comportamiento radial de tu modelo concuerda con las tendencias físicas del modelo tridimensional.

J. Validación clínica global del modelo

La simulación derivada de la aplicación fraccionaria de Caputo al modelo de Pennes muestra que con $\alpha = 1$ se alcanzan temperaturas superiores a 42 °C, suficientes para hipertermia terapéutica. Rahpeima & Lin alcanzaron 51.4 °C y demostraron necrosis tumoral completa tras 30 min. Esta coincidencia en el efecto clínico—temperaturas capaces de destruir tejido tumoral—confirma que el modelo fraccionario reproduce adecuadamente la respuesta térmica observada en estudios clínicos y numéricos.

XIII. Discusiones

El modelo de Pennes incorpora una fuente de calor volumétrica (Qmag) que simula la acción de los (SPIONs), y demuestra que es posible elevar la temperatura del tejido tumoral a rangos terapéuticos. Las simulaciones de la ecuación de Pennes clásica (con $\alpha=1.0$) indican que se pueden alcanzar y mantener temperaturas superiores a los 42∘C en el núcleo del tumor, lo que permite evidenciar la viabilidad de la hipertermia magnética, para este caso particular del cáncer ductal de seno.

Las Figura 1 y Figura 2 para el caso clásico, muestran que el modelo es capaz de predecir un perfil térmico donde la temperatura es más alta en el centro del tumor y disminuye hacia los bordes, y mantienen por encima del umbral terapéutico de 42∘C durante el tiempo de simulación (10 minutos o 30 minutos). Esto permite evidenciar el cumplimiento del objetivo de determinar el perfil térmico y comprobar el alcance de la temperatura terapéutica.

El modelo de Pennes incorpora explícitamente el término de perfusión sanguínea (ωb), que actúa como un mecanismo de disipación de calor. Las simulaciones clásicas, al mostrar una caída de temperatura hacia los bordes del tumor (donde la perfusión del tejido sano circundante es mayor o donde el efecto de la fuente de calor disminuye), muestran evidencia de que la perfusión es fundamental para limitar la extensión del calentamiento y proteger además el tejido sano circundante.

Aunque se alcanzan temperaturas terapéuticas, el análisis del daño térmico acumulado (Ω) es fundamental. El mapa de daño térmico acumulado (Figura 5) reveló que, a pesar de las altas temperaturas, el umbral de necrosis ($\Omega \geq 1$ o $\log10(\Omega) \geq 0$) no se alcanzó plenamente en el tiempo de simulación dado (30 minutos). Esto subraya que la efectividad de la hipertermia para inducir necrosis no solo depende de la temperatura máxima, sino también del tiempo de exposición y la acumulación del daño.

Por otro lado, la implementación del método L1 explicito, mostró inestabilidad desde el punto de vista numérico y computacional, mientras que la implementación del método L1 implícito mejoró los resultados.

El análisis numérico de la ecuación bio-calor de Pennes con derivada fraccionaria de Caputo, resuelto mediante el esquema L1 implícito, permitió caracterizar la dinámica térmica en tejido mamario durante un tratamiento de hipertermia dirigido al cáncer ductal de mama. Los resultados obtenidos muestran comportamientos temporales y espaciales consistentes con la teoría de difusión fraccionaria y, al mismo tiempo, comparables con estudios experimentales y simulaciones tridimensionales previamente reportadas. Esta concordancia fortalece la validez física del modelo y su relevancia para aplicaciones clínicas.

En primer lugar, según la Figura 9, el comportamiento térmico oscilatorio observado en la evolución temporal de la temperatura para $\alpha = 0.8$ refleja el papel dominante de la memoria térmica en modelos fraccionarios, la cual genera retardos y variaciones no monótonas en el transporte de calor. Este patrón es coherente con la respuesta inicial descrita por

ISSN:1390-9266 e-ISSN:1390-9134 LAJC 2026
DOI: 10.5281/zenodo.17781071

Rahpeima & Lin [25], quienes reportan un ascenso térmico rápido seguido de un descenso moderado antes de la estabilización debido al equilibrio entre generación de calor por nanopartículas y disipación por perfusión y conducción. Aunque los valores máximos alcanzados difieren —alrededor de 42 °C en nuestro modelo fraccionario y 51.4 °C en la simulación tridimensional clásica—, la dinámica cualitativa es consistente, lo cual valida que la formulación fraccionaria generaliza adecuadamente el comportamiento bio-térmico observado en modelos realistas de hipertermia magnética.

Asimismo, con base en las Figuras 10 y 11, el análisis de la influencia del orden fraccionario muestra que, al disminuir α, la temperatura crece más lentamente y alcanza valores menores debido a un efecto de memoria térmica más pronunciado. Esta tendencia coincide con los principios físicos descritos en Caizer et al. [24], donde la capacidad del tejido para absorber y retener calor depende del nivel de energía disipada por nanopartículas magnéticas, así como con los resultados de Rahpeima & Lin [25], donde el calentamiento depende de la eficiencia con que se difunde y transfiere el calor en el tumor. En este contexto, $\alpha \to 1$ reproduce el caso clásico de Pennes con una difusión más eficiente, mientras que α menores representan tejidos con mayor retardo en la transferencia térmica. Este comportamiento fortalece la idea de que los modelos fraccionarios permiten incorporar propiedades fisiológicas complejas, como la heterogeneidad térmica y el efecto memoria.

El perfil espacial de temperatura obtenido también se alinea con las tendencias reportadas en literatura. Se observó que para valores pequeños de α el calor se concentra cerca del centro tumoral, generando gradientes más abruptos y una ablación potencialmente más focalizada. Por el contrario, α cercano a 1 conduce a una difusión térmica más amplia, similar a la distribución radial reportada por Rahpeima & Lin (2022), donde el máximo térmico se localiza en zonas de mayor concentración de nanopartículas y en regiones coincidentes con el pico del campo magnético aplicado. Aunque su modelo no incorpora memoria térmica, la comparación pone en evidencia que el orden fraccionario actúa como un modulador que ajusta la extensión de la difusión térmica, generando escenarios que van desde la ablación focalizada (α bajo) hasta el calentamiento más uniforme (α alto). Esta capacidad adaptativa es un aporte importante del presente estudio.

Finalmente, la consistencia general entre los resultados obtenidos y los reportados por Caizer et al. y Rahpeima & Lin permite afirmar que el modelo fraccionario de Pennes resuelto con L1 implícito ofrece una representación físicamente plausible del proceso de hipertermia magnética, incluso cuando introduce elementos adicionales como la memoria térmica. La capacidad del modelo para reproducir dinámicas térmicas comparables con estudios experimentales y simulaciones avanzadas respalda su utilidad como herramienta de predicción y optimización terapéutica. Además, la flexibilidad del parámetro α abre la posibilidad de ajustar el modelo para representar distintos tipos de tejidos o escenarios clínicos, lo cual amplía el potencial de personalización del tratamiento térmico en cáncer ductal de mama.

XIV. Conclusiones

La incorporación de la derivada fraccionaria de Caputo en la ecuación de Pennes permitió modelar con mayor realismo la transferencia de calor en un tumor ductal mamario, al capturar efectos de memoria térmica y respuesta retardada que no pueden representarse con el modelo clásico. Esto convierte al enfoque fraccionario en una herramienta adecuada para estudiar tejidos biológicos heterogéneos sometidos a hipertermia magnética.

Los resultados numéricos obtenidos con el método L1 implícito demostraron estabilidad, coherencia física y consistencia con la literatura, permiten analizar la influencia del orden fraccionario α sobre la dinámica térmica. El parámetro α se confirmó como un modulador clave del transporte de calor: valores cercanos a 1 reproducen el modelo clásico con rápida difusión, mientras que α menores generan calentamiento lento y localizado, lo que refleja un comportamiento visco-térmico del tejido.

Las simulaciones espaciales y temporales mostraron que solo el caso $\alpha = 1.0$ permitió alcanzar los 42 °C necesarios para la hipertermia terapéutica, mientras que los modelos fraccionarios ($\alpha < 1$) exhibieron incrementos más lentos y temperaturas inferiores. Esto sugiere que la presencia de memoria térmica puede limitar la propagación del calor, siendo un fenómeno relevante para la planificación de tratamientos térmicos en tumores con características viscoelásticas marcadas.

Las validaciones realizadas con los datos experimentales de Caizer et al. evidenciaron una concordancia directa entre la dinámica térmica simulada y el comportamiento real del calentamiento mediado por nanopartículas, tanto en el ritmo de ascenso térmico como en la estabilización alrededor de los 42–43 °C. La relación entre α y la velocidad de calentamiento reprodujo el efecto de la concentración nanoparticulada observado en laboratorio, confirmando la fidelidad del modelo fraccionario.

La comparación con el modelo tridimensional de Rahpeima & Lin confirmó la capacidad del enfoque fraccionario para reproducir comportamientos térmicos clínicamente relevantes, como ascensos iniciales rápidos, oscilaciones transitorias, estabilización térmica y patrones radiales dependientes de la distribución de nanopartículas. Aunque las temperaturas máximas difieren, la dinámica cualitativa es equivalente y respalda la pertinencia física del modelo usado.

La inestabilidad severa observada con el método L1 explícito mostró que los esquemas numéricos explícitos no son adecuados para resolver ecuaciones fraccionarias de difusión térmica, ya que introducen errores que distorsionan completamente la física del problema. La estabilidad numérica lograda con el método L1 implícito confirma que los esquemas implícitos son indispensables para obtener simulaciones confiables en contextos fraccionarios.

En conjunto, el modelo fraccionario basado en Caputo y resuelto con L1 implícito se validó tanto numéricamente como experimentalmente, lo que demuestra ser capaz de reproducir fenómenos térmicos observados en hipertermia magnética real y predice temperaturas compatibles con la inducción de apoptosis tumoral. Esto respalda la utilidad del enfoque fraccionario como herramienta para la simulación, análisis y

eventual optimización clínica de tratamientos térmicos contra el cáncer de mama ductal.

REFERENCIAS

[1] Huancajulca, R., & del Rocío Mariet, P. (2023). *Risk factors for recurrence of non-metastatic ductal breast carcinoma in patients at the Hospital de Alta Complejidad Virgen de la Puerta (2019–2023)* [Undergraduate thesis, Universidad Privada Antenor Orrego]. https://purl.org/pe-repo/ocde/ford#3.02.27

[2] Urrego-Novoa, J. R., Hincapié-Echeverry, A. L., & Díaz-Rojas, J. A. (2024). *Net costs of breast cancer care in a health promoting entity in Colombia*. *Revista de la Facultad de Medicina, 72*(3), e112282

[3] García-Valdés, N., Casado-Méndez, P. R., Ricardo-Martínez, D., Santos-Fonseca, R. S., Gonsalves-Monteiro, A., & Sambu, Z. (2023). *Prevalence of complications in mastectomized breast cancer patients*. *Revista Médica Electrónica, 45*(2), 250–261.

[4] Tovar, A. D. M., Trujillo, Y. M. C., Huertas, K. T. P., & Sánchez, A. M. C. (2025). *Social determinants influencing healthcare access for women with breast cancer in Huila, Colombia*. *Ciencia Latina Revista Científica Multidisciplinar, 9*(1), 12665–12684.

[5] Fahim, Y. A., Hasani, I. W., & Ragab, W. M. (2025). *Promising biomedical applications with superparamagnetic nanoparticles*. *European Journal of Medical Research, 30*(1), 441.

[6] Périgo, E. A., Hemery, G., Sandre, O., Ortega, D., Garaio, E., Plazaola, F., & Teran, F. J. (2015). *Fundamentals and advances in magnetic hyperthermia* [Preprint]. arXiv. https://arxiv.org/abs/1510.06383

[7] Gilchrist, R. K., Medal, R., Shorey, W. D., Hanselman, R. C., Parrott, J., & Taylor, C. B. (1957). *Selective inductive heating of lymph nodes*. *Annals of Surgery, 146*(4), 596–606. https://doi.org/10.1097/00000658-195710000-00007

[8] Romero Coripuna, R. L., Cordova Fraga, T., Basurto Islas, G., Villaseñor Mora, C., & Guzman Cabrera, R. (2018). Modeling of Temperature Distribution of Fe3O4 Nanoparticles for Oncological Therapy. Computación y Sistemas, 22(4), 1573-1579.

[9] Pennes, H. H. (1948). *Analysis of tissue and arterial blood temperatures in the resting human forearm*. *Journal of Applied Physiology, 1*(2), 93–122.

[10] Jihad-Jebbar, A. (2025). *Application of electromagnetic fields and hyperthermia in cancer treatment* [Doctoral dissertation, University of Valencia, Doctoral Program in Physiology]. https://roderic.uv.es/rest/api/core/bitstreams/3c0c3d74-cfaf-46f9-8761-b8d423036024/content

[11] Caputo, M. (1967). *Linear models of dissipation whose Q is almost frequency independent—II*. *Geophysical Journal International, 13*(5), 529–539.

[12] Goya, G. F., Lima, E., Arelaro, A. D., Torres, T. E., Rechenberg, H. R., Rossi, L., Marquina, C., & Ibarra, M. R. (2013). *Magnetic hyperthermia with Fe_3O_4 nanoparticles: The influence of particle size on energy absorption*. *Journal of Magnetism and Magnetic Materials*. https://arxiv.org/abs/1302.691

[13] Liu, P., et al. (2018). *Size-dependent magnetic and inductive heating properties of Fe_3O_4 nanoparticles*. *Physical Chemistry Chemical Physics, 20*, 801–811.

[14] Hamza, NFA y Aljabair, S. (julio de 2023). Estudio experimental de la mejora de la transferencia de calor mediante un nanofluido híbrido y un inserto de cinta retorcida en intercambiadores de calor. En la IV CONFERENCIA CIENTÍFICA INTERNACIONAL DE CIENCIAS DE LA INGENIERÍA Y TECNOLOGÍAS AVANZADAS (Vol. 2830, N.º 1, p. 070009). AIP Publishing LLC

[15] Zapata Isidro, D. (2022). Síntesis de derivados aminoácidos de naftoquinona y su acoplamiento a nanotubos de carbono funcionalizados. REPOSITORIO NACIONAL CONACYT.

[16] Lemine, O. M., Algessair, S., Madkhali, N., Al Najar, B., & El Boubbou, K. (2023). *Assessing the heat-generation and self-heating mechanism of superparamagnetic Fe_3O_4 nanoparticles for magnetic hyperthermia: Effects of concentration, frequency, and magnetic field*. *Nanomaterials, 13*(3), 453. https://doi.org/10.3390/nano13030453

[17] Pucci, C., Degl'Innocenti, A., Gümüş, M. B., & Ciofani, G. (2022). *Superparamagnetic iron oxide nanoparticles for magnetic hyperthermia: Recent advancements, molecular effects, and future directions*. *Biomaterials Science, 10*, 2103–2121. https://doi.org/10.1039/D1BM01963E

[18] Phong, L. V. H., & Lam, T. D. (2020). *Increase of magnetic hyperthermia efficiency due to optimal size of Fe_3O_4 nanoparticles*. *Journal of Nanoparticle Research, 22*, 20. https://doi.org/10.1007/s11051-020-04986-5

[19] Espinosa Rivas, E. I., & Linares y Miranda, R. (2023). *SAR and temperature increase in a head model composed of several tissues produced by two WiFi devices working on the 2.4 GHz band*. *Científica, 27*(2). Retrieved from https://biblat.unam.mx/es/revista/cientifica-mexico-d-f

[20] Dulf, E. H., Pop, C. I., & Dulf, F. V. (2012). *Fractional calculus in 13C separation column control*. *Signal, Image and Video Processing, 6*(3), 479–485. https://doi.org/10.1007/s11760-012-0335-z

[21] Lathulerie, D. N. N. (2018). *Evaluation of magnetic hyperthermia therapy using perfusion imaging models for glioblastoma multiforme* [Doctoral dissertation, Venezuelan Institute for Scientific Research].

[22] Panda, J. y Das, D. (2025). Nanosistemas basados en nanopartículas superparamagnéticas de óxido de hierro para la teranóstica del cáncer. Medicina Traslacional Global , 4 (2), 31-50.

[23] Quintero, M. (2012). *Mathematical model of hyperthermia procedure for cancer treatment* [Master's thesis, Universidad Nacional de Colombia].

[24] Caizer C, Caizer-Gaitan IS, Watz CG, Dehelean CA, Bratu T, Soica C. High efficacy on the death of breast cancer cells using SPMHT with magnetite cyclodextrins nanobioconjugates. Pharmaceutics. 2023;15:1145. doi:10.3390/pharmaceutics15041145.

[25] Rahpeima, R., & Lin, C. A. (2022). Numerical study of magnetic hyperthermia ablation of breast tumor on an anatomically realistic breast phantom. Plos one, 17(9), e0274801

Agradecimientos a Doney Andrés Peña Julio, estudiante de Ingeniería de Sistemas de la Universitaria Americana. Barranquilla Colombia

AUTHORS

Eder Linares Vargas

Licenciado en Matemática y Física, Magíster en Física Aplicada y Candidato a Doctor en Ciencias Físicas. Cuenta con más de 20 años de experiencia docente en los niveles de Bachillerato, Educación Tecnológica y Educación Universitaria, además de amplia trayectoria en la asesoría metodológica de investigaciones de pregrado y posgrado. Ha participado como ponente en eventos científicos nacionales e internacionales. Actualmente desarrolla investigación en oncología matemática y en el diseño de nanoestructuras con fines teranósticos. Es docente de la Universidad del Atlántico, donde trabaja con la Facultad de Educación en los programas de Licenciatura en Matemáticas y Ciencias Naturales, y de la Corporación Universitaria Americana en Barranquilla, Colombia, adscrito al Departamento de Ciencias Básicas.

ISSN:1390-9266 e-ISSN:1390-9134 LAJC 2026

Green software development using carbon-aware scheduling techniques and energy efficiency metrics throughout the SDLC

ARTICLE HISTORY
Received 31 July 2025
Accepted 19 November 2025
Published 6 January 2026

Mikita Piastou
University of West Georgia
School of Computing, Analytics, and Modeling
Carrollton, United States
mpiasto1@my.westga.edu
ORCID: 0009-0002-7360-2152

ISSN:1390-9266 e-ISSN:1390-9134 LAJC 2026

M. Piastou, "Green software development using carbon-aware scheduling techniques and energy efficiency metrics throughout the SDLC", Latin-American Journal of Computing (LAJC), vol. 13, no. 1, 2026.

Green software development using carbon-aware scheduling techniques and energy efficiency metrics throughout the SDLC

Mikita Piastou
University of West Georgia
School of Computing, Analytics, and Modeling
Carrollton, United States
mpiasto1@my.westga.edu

***Abstract*—**The objective of the present paper is to systematize contemporary approaches of green software development through the prism of carbon-aware scheduling methodologies and energy efficiency metrics at all stages of the software development life cycle (SDLC). The study will analyze English-language, peer-reviewed articles published between 2020 and 2025. The following four carbon-intensive scheduling strategies have been identified: temporal task shifting, geographic load migration, electricity price consideration, and dynamic resource scaling. Experimental data indicates the potential for a 30–70% reduction in the carbon footprint of applications, with only a moderate impact on latency and cost. The metrics employed for evaluating energy efficiency span from low-level measures such as code complexity and measured power consumption to higher-level metrics addressing infrastructure and integration. It has been established that disregarding the initial phases of the SDLC results in an underestimation of the aggregate carbon footprint. The analysis showed that cutting emissions can conflict with maintaining high service quality. It also highlighted problems with standardizing metrics and ensuring accurate carbon-intensity forecasts, especially when significant task shifting is involved. Further unification of metrics, integration of energy monitoring at all stages of the SDLC, and consideration of economic factors are recommended.

Keywords*—green development, software, application development, carbon pollution*

I. Introduction

The development of software is becoming increasingly driven by the integration of environmental sustainability principles, a response to the escalating energy intensity of Information and Communication Technology (ICT) infrastructures. Today, the ICT sector accounts for roughly 2% to 4% of global carbon emissions, thus rendering green software development an urgent challenge in the fight against global warming and environmental concerns [1]. Projections indicate that ICT may account for as much as 8% of the world's energy use by 2030 [2]. The goal of green software development is to create products that minimize energy use and environmental impact at every stage of the SDLC, including design, implementation, maintenance, and eventual decommissioning. Traditionally, green computing has focused primarily on hardware.

However, in the past few years, it has become steadily clear that software architecture, algorithms, and the way systems operate also play a major role in how much energy a system uses. This study focuses on carbon-aware scheduling, which means planning computational and operational tasks with carbon intensity and energy use in mind. It also looks at ways to evaluate energy efficiency throughout the different stages of the SDLC.

Despite growing interest, the field is still emerging. Key challenges remain, such as the lack of standardized metrics, the complexity of isolating software energy consumption from hardware and operating system influences, the integration of sustainable practices within established DevOps pipelines, and clarifying the relationship between energy efficiency and broader sustainability outcomes.

The objective of this review is to analyze carbon-aware scheduling methods in software development and energy efficiency, and to identify key metrics as applied to software development and operation. A detailed analysis of existing studies will provide an up-to-date overview of the field, supporting further research on carbon-aware software development practices.

II. Material and Methods

The literature review followed a structured, multi-stage process: identification, screening, eligibility assessment, and inclusion, following the general principles of PRISMA-style reviews.

The search was conducted across major scientific databases, namely IEEE Xplore, Google Scholar, ScienceDirect, ACM Digital Library, Web of Science, and arXiv. This ensured adequate coverage of both peer-reviewed and preprint research. Searches targeted publications from 2020 to 2025 and were restricted to English-language sources.

The following keyword groups were used:

- *Primary terms:* "carbon-aware scheduling," "green software development," "energy efficiency metrics software lifecycle," "energy efficiency metrics," "sustainable software."
- *Secondary terms:* "energy-aware computing," "low-carbon software design."

Duplicates were removed first, leaving 73 unique articles for screening. Articles were then screened in two stages:

1) Title and abstract screening: 24 papers were excluded due to irrelevance (e.g., not addressing software systems, not

ISSN:1390-9266 e-ISSN:1390-9134 LAJC 2026
DOI: 10.5281/zenodo.17781203

focused on energy or carbon metrics) or ineligible study type (e.g., non-scientific sources such as blogs).

2) Full-text assessment: of the 49 remaining articles, 2 were excluded after detailed evaluation, resulting in 47 articles that met the inclusion criteria. Papers were included if they examined energy-efficiency metrics within the software development lifecycle, addressed carbon-aware or energy-aware scheduling of computational tasks, and provided either quantitative estimates (e.g., energy consumption, CO_2 savings) or qualitative evaluations (method comparisons, limitations).

Particular emphasis was placed on research that explored carbon-aware scheduling and cost-sensitive workload planning, as well as studies presenting or evaluating metrics to assess software sustainability across the SDLC.

III. Result and Discussion

A. Evolution of Key Research Themes

The discourse on sustainable software has evolved considerably over the past two decades. Initially, the concept was tightly coupled with "performance engineering," where reducing resource consumption (CPU cycles, memory) was primarily a means to improve speed and reduce hardware costs, with energy savings being a welcome byproduct.

The first major shift occurred as researchers began to explicitly target energy as a primary optimization goal. This led to the emergence of energy-aware computing. Early work in this phase focused on the operating system and hardware abstraction layers, developing power models for CPUs and other components. The research community then moved up the stack, investigating how programming languages, compilers, and software architectures contribute to energy usage. This phase was characterized by a focus on energy efficiency, minimizing the watts consumed by a software application to perform a given task [3].

More recently, the theme has matured into carbon-aware computing. This represents a more nuanced understanding of environmental impact, recognizing that not all energy is created equal. The carbon intensity of electricity, the amount of greenhouse gas emitted per kilowatt-hour (kWh), varies significantly based on the energy source mix (e.g., renewables versus fossil fuels) of the electrical grid at a given time and location. Carbon-aware software, therefore, does not just aim to use less energy, it aims to consume energy when and where it is "cleanest." This has led to the development of sophisticated scheduling techniques that align computational workloads with periods of low carbon intensity [4], [5].

A holistic perspective is developing -one that considers all stages of the software application lifecycle, including requirements engineering, UI/UX design, deployment, maintenance, and eventual decommissioning. This view argues that sustainability must be a cross-cutting concern, integrated into every stage of software development [6].

In addition to the lifecycle-wide integration of sustainability, recent investigation has also considered the ethical and societal dimensions of green software [7]. As digital services expand globally, disparities in grid cleanliness across regions mean that software systems can inadvertently externalize environmental costs to more carbon-intensive areas. This raises important questions about environmental justice and the responsibilities of cloud providers in minimizing their overall footprint rather than merely shifting it elsewhere.

Interdisciplinary collaborations between software engineers, environmental scientists, and policy experts have begun to shape frameworks for green software governance, suggesting future regulation or certification schemes that could mandate transparency in energy use or emissions reporting. These initiatives, although still emerging, highlight the necessity of embedding sustainability not just as a technical goal but as a societal obligation within software engineering practice.

A foundational challenge in green software is measurement. The adage "you cannot improve what you cannot measure" is particularly salient. Research into energy efficiency metrics has evolved from coarse-grained, hardware-centric measures to fine-grained, software-centric approaches.

Key approaches include the following:

- Early approaches relied on physical power meters or processor-level instrumentation like Intel's Running Average Power Limit (RAPL) to evaluate the energy draw of entire systems. While accurate, these methods often struggle to attribute consumption to specific software processes or lines of code [8].
- To overcome the limitations of physical measurement, researchers developed statistical and machine learning models to estimate software energy consumption based on high-level performance indicators (e.g., I/O operations, CPU utilization, network packets). A study demonstrated a strong correlation between system-level metrics and energy consumption, paving the way for software-based power estimation models [3], [9].
- "Software-energy-label," a multi-dimensional metric that evaluates the energy efficiency of software applications, is similar to the energy labels on appliances. Other studies have focused on defining metrics relevant to specific domains, such as energy per transaction in database systems or energy per user request in web applications [6], [8].

A primary debate revolves around the trade-off between the accuracy and accessibility of metrics. Direct hardware measurement is the gold standard for accuracy but requires specialized equipment and expertise. Model-based approaches are more accessible and scalable but are subject to estimation errors and may require re-calibration for different hardware and software environments. There is currently no universally accepted standard for measuring and reporting the energy consumption of a software application, making it difficult to compare the "greenness" of different products [3].

ISSN:1390-9266 e-ISSN:1390-9134 LAJC 2026
DOI: 10.5281/zenodo.17781203

TABLE I. EVOLUTION OF RESEARCH IN SUSTAINABLE SOFTWARE

Theme	Sustainability Perspective		
	Main Focus	***Representative Techniques***	***Maturity***
Performance Engineering	Optimize resource use for speed and cost	CPU and memory optimization, HW tuning	High
Energy-Aware Computing	Software-level energy optimization	OS power models, RAPL, efficient algorithms	Medium-High
Carbon-Aware Computing	Use energy where carbon is lowest	Time and geo shifting, carbon forecasting	Medium
Lifecycle Sustainability	Sustainability across SDLC	Green requirements, energy-aware design	Low-Medium
Ethical Sustainability	Transparency and environmental justice	Emissions accounting, governance models	Low

B. Carbon-Aware Scheduling

There is a number of methodologies that can be employed to account for carbon intensity within computational processes. These include time-based scheduling, geographic shift, price-aware scheduling, and flexible scaling of resources. Each of these factors deserves individual consideration.

Time-based scheduling is an approach that involves delaying batch and time-insensitive tasks to periods of low carbon intensity on the energy grid. It has been observed that users of cloud services frequently migrate batch tasks to periods of low carbon intensity. A comparable approach is employed within GAIA (Green Aware Instance Allocation), an environmentally oriented scheduler for batch tasks that has been demonstrated to produce substantial emission reductions while exerting a moderate impact on performance and cost. This approach finds application across a wide range of use cases, including data backup processes, machine learning tasks, data distribution, batch processing, and more [7].

The geographic shift approach entails the distribution of computational tasks across data centers or regions that exhibit a reduced carbon footprint. Souza et al. developed CASPER for distributed web services, which dynamically allocates load between geographic regions depending on local carbon intensity and network latency. A series of experiments has been conducted, yielding findings that demonstrate the potential for a carbon reduction of up to 70% while concurrently ensuring the maintenance of Service Level Objectives (SLOs) concerning latency [10]. In a similar vein, Lechowicz et al. proposed PCAPS, a scheduler for computational processes that takes into account both time-dependent carbon intensity and geographical location, as well as task prioritization and ordering. A PCAPS prototype in a cluster of 100 nodes reduced the carbon footprint to 32.9% of the baseline, with no noticeable loss of efficiency [11].

Price-aware scheduling is a price-conscious approach. It is evident from a substantial set of research publications that the importance of the prices of computing resources and services is frequently emphasized [12], [13], [14], [15], [16]. Z. Miao et al. suggest that cloud service and computing providers should take into consideration the carbon intensity of electricity costs and the fluctuations in renewable energy across different locations and times. Indeed, models such as ECMR take into account both carbon intensity and local electricity prices simultaneously, thus minimizing emissions at an acceptable monetary cost. The ECMR algorithm for distributed machine learning tasks has been demonstrated to enhance renewable energy utilization by up to 90.8% while simultaneously reducing carbon emissions by 30% in comparison with the baseline carbon-aware ML methods [17].

In the context of resource allocation, the concept of flexible scaling has been proposed by Hanafy et al. This approach involves the dynamic adjustment of the computing cluster's capacity in response to variations in carbon intensity. In circumstances where the carbon intensity is minimal, the cluster will allocate a greater quantity of resources. Conversely, in instances of elevated emissions, the cluster will allocate a reduced quantity of resources. The prediction of carbon intensity is achieved through the analysis of historical data or the utilization of machine learning techniques. This approach precludes the simultaneous preparation of all tasks, thus circumventing the "buffalo herd" effect, wherein the adoption of a similar low-carbon timeframe can surpass computational capacity, consequently leading to increased carbon emissions. CarboneFlex has demonstrated a 57% reduction in emissions in comparison with conventional task scheduling [18].

Another promising development is the integration of carbon-aware strategies into container orchestration platforms. Kubernetes, for instance, is being extended through plugins and custom schedulers to enable energy-aware and carbon-aware task placements. Research prototypes have demonstrated the feasibility of integrating carbon-intensity forecasts as a scheduling signal, allowing pods to be launched in regions or at times that minimize carbon emissions. These advances open the door to mainstreaming sustainability features in cloud-native systems, though they still face technical barriers in standardization, performance impact, and developer adoption [19].

The study emphasizes that architectural patterns and microservice granularity can substantially impact energy consumption [20]. Fine-grained microservices often result in elevated network traffic and resource duplication, thereby increasing runtime energy use. Xiao et al. show that selecting service co-location or modular reuse patterns can mitigate these inefficiencies and enhance energy performance [21].

Hybrid strategies that combine multiple scheduling techniques (time-based and geographic shifting with price-aware models) have shown superior results in experimental settings, offering flexible trade-offs across cost, latency, and emissions [7], [21], [22]. However, these models demand high-quality, real-time data pipelines for energy pricing and carbon intensity, which remain unreliable or unavailable in many regions [23]. As such, future research must also focus on data infrastructure and interoperability standards to enable wider deployment of carbon-aware systems.

Despite promising results, carbon-aware scheduling has some fundamental problems:

- Geographic shifting itself consumes energy and generates network traffic, the carbon footprint of which must be considered, according to Y. Guo et al. [24]. In

ISSN:1390-9266 e-ISSN:1390-9134 LAJC 2026
DOI: 10.5281/zenodo.17781203

some cases, the carbon cost of data transmission can negate the benefits of cleaner energy.

- The Rebound Effect refers to the phenomenon where efforts to maximize green energy efficiency result in increased overall computation, which can ultimately cause a rise in total energy consumption instead of a reduction [25].
- Designing and implementing complex process schedulers requires significant effort and changes to existing container management platforms such as Kubernetes. As noted by P. Wiesner et al., most current systems do not have built-in mechanisms to account for carbon intensity. It can be concluded that time-based shifting is a relatively simple and efficient method [26].

TABLE II. CARBON-AWARE SCHEDULING TECHNIQUES

Technique	Concept Overview		
	Core Idea	*Examples*	*Key Limitations*
Time-Based Scheduling	Delay tasks to low-carbon periods	GAIA, batch deferral	Latency, unsuitable for real-time tasks
Geographic Shifting	Run in cleaner regions	CASPER, PCAPS	Network latency, data transport cost
Price-Aware Scheduling	Use electricity price signals	ECMR	Requires accurate price and carbon data
Flexible Scaling	Scale resources by carbon intensity	CarboneFlex	Needs forecasting, throughput impact
Carbon-Aware Orchestration	Carbon signals in K8s schedulers	Custom K8s plugins	Lack of standardization
Hybrid Models	Combine time, geo, and price	Multi-factor schedulers	Complexity unreliable data streams

Geographic shifting, however, has the potential to significantly reduce emissions, but reliable data on carbon intensity in different regions and accounting for network delays are prerequisites. It is evident that the aforementioned methods frequently presuppose information regarding task duration, power price dynamics, computing resource prices, local carbon emissions, or the capacity for deferred loading of computing resources. This complicates the practical implementation for a substantial number of tasks, including those of significant importance.

C. Energy efficiency metrics in the SDLC phases

The assessment of software energy efficiency is a fundamental task, without which progress in the field of green engineering is impossible. At present, there is an absence of a standardized metric to assess the energy efficiency and environmental effectiveness of computational tasks, as well as software development and maintenance activities. It is acknowledged that a variety of metrics may be implemented during the different phases of software development. Each of these metrics possesses its own unique characteristics, advantages, and disadvantages. Current research focuses on developing precise metrics for the various phases of the SDLC. A review of the existing literature shows that certain approaches and metrics are far more commonly used than others.

In the initial phases of the SDLC, the direct measurement of energy consumption is often challenging. Consequently, researchers propose the use of indirect metrics. To achieve this objective, static code characteristics are analyzed and correlated with CPU and memory resource consumption. These characteristics include cyclomatic complexity, the use of specific data structures, and code length [27]. Several studies have demonstrated a strong correlation between these metrics and energy efficiency [28], [29]. However, other studies have shown that compilation and processor-level optimizations can make these dependencies non-linear and unpredictable [30], [31].

In the following section, a series of more direct approaches are proposed for the testing phase. For instance, incorporating energy profiling tools, such as Intel Power Gadget, into Continuous Integration/Continuous Delivery (CI/CD) pipelines enables the automated assessment of energy expenditure during the execution of tests. This allows for the identification of "energy regressions," i.e., code changes that inadvertently increase energy consumption [32]. The primary limitation in this context is that results depend heavily on the specific characteristics of the hardware and software environment, which hinders comparison and generalization.

As Kruglov and Succi observe, in the initial phases of development, metrics such as module complexity, coupling, and cohesion can be evaluated. The authors note that metrics of code cohesion show a stronger correlation with energy consumption than metrics of size or inheritance [33]. This helps identify "dark zones" of potential inefficiency during the code design and implementation stages. However, significant energy consumption data only becomes available at later phases, i.e., during software testing and deployment. Therefore, end-to-end tracking of metrics across the entire SDLC is necessary.

Direct metrics of power consumption use either hardware meters or software models, such as RAPL, which provide estimates of power usage by processor components. The issue with models like RAPL is that they do not account for the consumption of RAM, disks, NICs, and other components, which can lead to underestimates of total energy use [34].

In the context of software deployment, it is imperative to meticulously measure two critical metrics: the power consumption of services and the load on servers. In operational mode, the carbon footprint (CO_2-equivalent) and energy consumption in kWh per unit of workflow, such as per request or transaction, are frequently utilized. As widely acknowledged in the academic community, prevailing green infrastructure metrics, such as Power Usage Effectiveness (PUE), Data Center Infrastructure Efficiency (DCiE), and Carbon Usage Effectiveness (CUE), focus on data centers. However, these metrics do not account for software aspects or load fluctuations at the application level [35].

New initiatives propose considering the efficiency "inside" servers (SPUE) or calculating Software Carbon Intensity (SCI), the normalized carbon footprint of software per functional unit [36], [37], [38]. The trend toward

ISSN:1390-9266 e-ISSN:1390-9134 LAJC 2026
DOI: 10.5281/zenodo.17781203

standardization of SCI in ISO 14064/21031 reflects recognition of the need for software-level metrics [39].

The SCI can be calculated using the following formula:

$$SCI = (E \times I + M) / R \quad (1)$$

where E is the energy consumed by the software, I is the carbon intensity, M is the carbon footprint associated with hardware production, and R is the functional unit (e.g., number of users or API requests).

A limitation of the SCI metric is the difficulty of accurately measuring the value of M and defining a relevant functional unit R for complex systems, as previously outlined.

T. Simon et al. emphasize that their evaluation model distributes the overall impact between the "development" and "use" phases and demonstrate that the optimization of just one phase can result in a shift of the burden to the other phase [40]. In their example, the significance of the impact of the development phase was given greater weight, despite the focus traditionally being on operations. It is imperative to recognize that the evaluation of any metric at a specific stage of the SDLC is crucial. As Kruglov and Succi have highlighted, a comprehensive assessment of a project's performance and environmental impact can only be achieved through the integration of measurements across all developmental stages [33].

A considerable number of methodologies have been demonstrated to result in substantial emission reductions; however, it is crucial to acknowledge that these outcomes are frequently accompanied by trade-offs. For instance, Hanafy et al. demonstrate that as carbon savings increase, task delays and costs also rise due to idling reserves. The authors observe that their algorithms achieve a twofold increase in carbon savings for every percentage point increase in cost, concomitantly reducing the additional delay by 26% [7]. In other words, the consistent reduction in energy demands frequently necessitates the allocation of resources and additional time, thereby constraining implementation in critical systems.

The efficacy of such estimation and control methods is constrained by assumptions regarding the availability of data on load dynamics and energy sources. Algorithms frequently presuppose precise prediction of network carbon intensity and task duration. In the event of such data being inaccurate, the decisions made may be suboptimal. The issue of delayed task execution is also pertinent. It is noteworthy that not all studies account for network delay in geographic migration, although the issue is addressed in CASPER [10].

Another critical direction involves the automation of sustainability evaluation within development workflows. Upcoming tools seek to provide real-time energy feedback to developers by integrating estimators and profilers directly into IDEs and version control systems. For example, plug-ins can highlight energy hotspots in code as developers write it, allowing for just-in-time greenness corrections. While still in the early stages, such tooling has the potential to transform sustainability from a late-stage consideration to a core part of the coding process.

Additionally, recent work explores how AI-assisted refactoring tools might recommend low-energy alternatives for common patterns or inefficient loops [41]. These innovations reflect the increasing alignment between green software engineering and developer productivity ecosystems. Cross-field research is beginning to explore the psychological and behavioral factors that influence how developers respond to energy metrics, suggesting that future tools must be not only accurate but also actionable and motivating to drive change in software design practices.

Nevertheless, the identified works form the foundations of green software engineering approaches, indicating directions for further research, integrating metrics throughout the SDLC, automating code greenness control, and considering economic factors in resource scheduling.

Another rising aspect of energy efficiency in the SDLC is the integration of sustainability considerations into software architecture and design patterns. Research has shown that architectural choices, such as the adoption of microservices versus monolithic structures, can have a significant impact on energy consumption [19]. For example, microservice-based systems may increase network traffic and idle time due to container overhead and distributed communication, whereas monolithic designs, while less scalable, can result in lower baseline energy use under certain conditions. This highlights the need for sustainability-aware architecture trade-off analysis, where energy implications are considered alongside maintainability, performance, and scalability.

Similarly, the choice of programming language and runtime environment has been scrutinized in recent studies [42], [43], [44]. For instance, compiled languages such as C++ or Rust typically produce more energy-efficient executables than interpreted languages like Python or JavaScript, though the development speed and ecosystem support may differ. Benchmarks across common workloads (e.g., compression, parsing, web serving) confirm that language-level decisions are not trivial in the context of energy use. The growing interest in domain-specific languages (DSLs) for energy-constrained environments (such as IoT and edge computing) further exemplifies this direction, suggesting the co-evolution of tools, languages, and sustainable practices. The field is also beginning to explore the long-term effects of software bloat and feature creep on sustainability.

As software systems accumulate features, dependencies, and technical debt, they tend to grow in size and complexity, often requiring more resources to run, update, and maintain. This phenomenon, known as "code rot" or "software obesity", introduces persistent overheads, especially when running on cloud infrastructure where idle resources still consume electricity [45]. Lean software engineering principles are being revisited through a sustainability lens, encouraging minimalist, modular, and refactorable designs as mechanisms for long-term energy savings.

TABLE III. ENERGY OPTIMIZATION APPROACHES

SDLC Phase	**Evaluation Framework**		
	Optimization Approach	***Methods***	***Effectiveness***
Planning	High-level energy goals	Green requirements, sustainability guidelines	Medium
Analysis	Evaluate potential energy impact	Software modeling, architectural trade-offs	Medium

ISSN:1390-9266 e-ISSN:1390-9134 LAJC 2026
DOI: 10.5281/zenodo.17781203

SDLC Phase	Evaluation Framework		
	Optimization Approach	***Methods***	***Effectiveness***
Design	Energy-aware architecture	Component selection, modularity, low-power design patterns	Medium-High
Implementation	Efficient code development	Energy-efficient algorithms, linters, static analysis	Medium-High
Testing	Energy regression and monitoring	RAPL, Intel Power Gadget, automated energy tests	Medium
Maintenance	Reduce long-term energy cost	Refactoring, code bloat reduction, continuous monitoring	Low-Medium

Optimizing reuse without bloating systems is thus an active area of exploration. Finally, education and cultural change within the software engineering profession are becoming central to the green software movement. Studies have shown that many developers remain unaware of the energy implications of their design and implementation decisions, or lack the tools and incentives to prioritize sustainability [46], [47]. This has spurred the creation of educational materials, guidelines (e.g., the Green Software Foundation's principles), and even university courses on sustainable computing.

Bridging the knowledge gap between energy modeling experts and everyday developers is crucial if green practices are to become mainstream rather than niche. As sustainability becomes a shared responsibility, fostering a culture that values efficiency, transparency, and accountability will be just as important as advancing technical solutions.

IV. Conclusion

The review demonstrates that carbon-aware scheduling and energy efficiency metrics are active research areas for green software development from 2020 onwards. It is vital to employ critical scheduling strategies, including temporal shifting of tasks, geographic and price-based load balancing, and dynamic resource scaling. Experimental evidence shows that these techniques can reduce the carbon footprint of applications and computational processes by tens of percent. However, many of these methods remain at the prototype stage, and their deployment often involves trade-offs between emissions, performance, and cost.

A key finding is the necessity for end-to-end measurement across all stages of the SDLC. Current metrics inadequately capture software-level energy efficiency or account for application performance. The Software Carbon Intensity (SCI) metric, while promising, remains immature and requires rigorous validation. Future work should focus on developing standardized, cross-platform metrics that integrate energy, carbon, and performance indicators, enabling fair comparison and benchmarking of software systems.

Key strategies to promote sustainable software development include:

1) Integration of energy metrics into development workflows: Incorporate energy profiling, SCI estimators, and carbon-aware alerts directly into IDEs, CI/CD pipelines, and testing frameworks to enable developers to optimize energy use as they code.

2) Adoption of carbon-aware scheduling in cloud environments: Encourage cloud providers to expose real-time carbon intensity data and pricing signals, enabling applications to dynamically shift workloads across time and geography.

3) Standardization and benchmarking: Establish open datasets for carbon intensity, shared benchmarks for energy efficiency, and guidelines for SCI reporting to foster transparency and comparability.

4) Education and cultural change: Train software engineers on sustainable design patterns, energy-efficient programming practices, and the environmental implications of software architecture choices.

5) Policy and regulatory alignment: Encourage policymakers to incentivize sustainable software development through certifications, disclosure requirements, or carbon-aware procurement policies.

6) Interdisciplinary collaboration: Promote partnerships among academia, industry, and environmental science to co-develop frameworks that balance performance, cost, and sustainability in real-world software systems.

Looking ahead, the successful realization of green software systems will require a synergy of technical, organizational, and policy innovations. Scalable, interoperable infrastructures that operationalize sustainability without compromising functionality or accessibility must become the norm. As digital services underpin critical societal functions, software energy efficiency is no longer a niche concern and has become a crucial enabler of climate action. Only through coordinated efforts across stakeholders can the software industry make a meaningful contribution to global emissions reduction goals.

References

[1] J. C. T. Bieser and S. Peters, "A review of assessments of the greenhouse gas footprint of digital devices," *Telecommun. Policy*, vol. 47, no. 5, 2023.

[2] World Bank / ITU, *Measuring the Emissions & Energy Footprint of the ICT Sector*, World Bank, 2024.

[3] W. Wysocki, I. Miciuła, and P. Plecka, "Methods of Improving Software Energy Efficiency: A Systematic Literature Review and the Current State of Applied Methods in Practice," *Electronics*, vol. 14, no. 7, p. 1331, 2025, doi: 10.3390/electronics14071331.

[4] T. Anderson, A. Belay, M. Chowdhury, A. Cidon, and I. Zhang, "Treehouse: A Case For Carbon-Aware Datacenter Software," in *Proc. HotCarbon '22*, 2022.

[5] Y. G. Kim, U. Gupta, A. McCrabb, Y. Son, V. Bertacco, D. Brooks, and C.-J. Wu, "GreenScale: Carbon-Aware Systems for Edge Computing," *arXiv*, Apr. 2023, doi: 10.48550/arXiv.2304.00404.

[6] S. McGuire, E. Shultz, B. Ayoola, and P. Ralph, "Sustainability is Stratified: Toward a Better Theory of Sustainable Software Engineering," in *Proc. ICSE '23*, May 2023, pp. 1996–2008, doi: 10.1109/ICSE48619.2023.00169.

[7] W. A. Hanafy, Q. Liang, N. Bashir, A. Souza, D. Irwin, and P. Shenoy, "Going Green for Less Green: Optimizing the Cost of Reducing Cloud Carbon Emissions," in *Proceedings of the 29th ACM International*

Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS '24), La Jolla, CA, USA, Apr. 2024, pp. 479–496.

[8] P. Thamm, "Strategies to Measure Energy Consumption Using RAPL During Workflow Execution on Commodity Clusters," *arXiv*, May 2025, doi: 10.48550/arXiv.2505.09375.

[9] G. Raffin and D. Trystram, "Dissecting the software-based measurement of CPU energy consumption: a comparative analysis," *arXiv*, Jan. 2024, doi: 10.48550/arXiv.2401.15985.

[10] A. Souza, S. Jasoria, B. Chakrabarty, A. Bridgwater, A. Lundberg, F. Skogh, A. Ali-Eldin, D. Irwin, and P. Shenoy, "CASPER: Carbon-Aware Scheduling and Provisioning for Distributed Web Services," in *Proc. 14th Int'l Green & Sustainable Computing Conf. (IGSC '23)*, Toronto, ON, Canada, Oct. 2023, pp. 67–73, doi: 10.1145/3634769.3634812.

[11] A. Lechowicz, R. Shenoy, N. Bashir, M. Hajiesmaili, A. Wierman, and C. Delimitrou, "Carbon- and Precedence-Aware Scheduling for Data Processing Clusters," *CoRR*, vol. abs/2502.09717, Feb. 2025, doi: 10.48550/arXiv.2502.09717.

[12] J. Bader, J. Irion, J. Kappel, J. Witzke, N. Fomin, D. Sherifi, and O. Kao, "Learning Process Energy Profiles from Node-Level Power Data," *arXiv preprint arXiv:2511.13155*, Nov. 2025.

[13] M. Raeisi-Varzaneh, O. Dakkak, Y. Fazea, and M. G. Kaosar, "Advanced cost-aware Max–Min workflow tasks allocation and scheduling in cloud computing systems," *Cluster Computing*, vol. 27, pp. 13,407–13,419, Dec. 2024.

[14] X. Sun, Z. Wang, Y. Wu, H. Che, and H. Jiang, "A price-aware congestion control protocol for cloud services," *J. Cloud Comput.*, vol. 10, no. 1, Art. 55, Nov. 2021.

[15] S. Tuli, G. Casale, and N. R. Jennings, "MetaNet: Automated dynamic selection of scheduling policies in cloud environments," *arXiv preprint arXiv:2205.10642*, May 2022.

[16] S. G. Ahmad, T. Iqbal, and E. U. Munir, "Cost optimization in cloud environment based on task deadline," *J. Cloud Comput.*, vol. 12, Art. 9, Jan. 2023.

[17] Z. Miao, L. Liu, H. Nan, W. Li, X. Pan, X. Yang, M. Yu, H. Chen, and Y. Zhao, "Energy and carbon-aware distributed machine learning tasks scheduling scheme for the multi-renewable energy-based edge-cloud continuum," *Science and Technology for Energy Transition*, vol. 79, Article 82, 2024, doi: 10.2516/stet/2024076.

[18] W. A. Hanafy, L. Wu, D. Irwin, and P. Shenoy, "CarbonFlex: Enabling Carbon-aware Provisioning and Scheduling for Cloud Clusters," *CoRR*, vol. abs/2505.18357, May 2025.

[19] X. Xiao, C. Gao, and J. Bogner, "On the Effectiveness of Microservices Tactics and Patterns to Reduce Energy Consumption: An Experimental Study on Trade-Offs," in *Proc. 22nd Int'l Conf. on Software Architecture (ICSA '25)*, Odense, Denmark, Apr. 2025, pp. 164–175, doi: 10.1109/ICSA65012.2025.00025.

[20] O. Poy, M. Á. Moraga, F. Garcia, and C. Calero, "Impact on energy consumption of design patterns, code smells and refactoring techniques: A systematic mapping study," *J. Syst. Softw.*, vol. 222, no. 15, p. 112303, Dec. 2024, doi: 10.1016/j.jss.2024.112303.

[21] X. Xiao, "Architectural Tactics to Improve the Environmental Sustainability of Microservices: A Rapid Review," *CoRR*, vol. abs/2407.16706, July 2024.

[22] E. Breukelman, S. Hall, G. Belgioioso, and F. Dörfler, "Carbon-Aware Computing in a Network of Data Centers: A Hierarchical Game-Theoretic Approach," in *Proc. 2024 European Control Conference (ECC)*, Stockholm, Sweden, Jun. 25–28, 2024, pp. 798–803, doi: 10.23919/ECC64448.2024.10591261.

[23] N. Asadov, V. C. Coroamă, M. Franzil, S. Galantino, and M. Finkbeiner, "Carbon-Aware Spatio-Temporal Workload Shifting in Edge–Cloud Environments: A Review and Novel Algorithm," *Sustainability*, vol. 17, no. 14, p. 6433, Jul. 2025, doi: 10.3390/su17146433.

[24] Y. Guo, A. Tomlinson, R. Su, and G. Porter, "The Effect of the Network in Cutting Carbon for Geo-shifted Workloads," CoRR, vol. abs/2504.14022, Apr. 2025.

[25] N. L. Woodruff, D. Schall, M. F. P. O'Boyle, and C. Woodruff, "When Does Saving Power Save the Planet?," in *Proc. HotCarbon '23*, Jul. 2023, pp. 1–6, doi: 10.1145/3604930.3605719.

[26] P. Wiesner, M. Steinke, H. Nickel, Y. Kitana, and O. Kao, "Software-in-the-Loop Simulation for Developing and Testing Carbon-Aware Applications," *Software: Practice and Experience*, vol. 53, no. 12, pp. 2362–2376, 2023, doi: 10.1002/spe.3275.

[27] F. Hussin, S. A. N. Md Rahim, N. S. M. Hatta, M. K. Aroua, and S. A. Mazari, "A systematic review of machine learning approaches in carbon capture applications," *Journal of CO_2 Utilization*, vol. 71, Art. 102474, May 2023, doi: 10.1016/j.jcou.2023.102474.

[28] J. Mancebo, C. Calero, and F. García, "Does maintainability relate to the energy consumption of software? A case study," *Software Qual. J.*, vol. 29, no. 1, pp. 101–127, Jan. 2021, doi: 10.1007/s11219-020-09536-9.

[29] H. M. Alvi, H. Majeed, H. M. Mujtaba, and M. O. Beg, "MLEE: Method level energy estimation — A machine learning approach," *Sustain. Comput. Inform. Syst.*, vol. 31, p. 100594, 2021, doi: 10.1016/j.suscom.2021.100594.

[30] K. Chan-Jong-Chu, T. Islam, M. M. Exposito, S. Sheombar, C. Valladares, O. Philippot, E. M. Grua, and I. Malavolta, "Investigating the Correlation between Performance Scores and Energy Consumption of Mobile Web Apps," in *Proc. 24th Evaluation & Assessment in Software Engineering (EASE '20)*, Trondheim, Norway, Apr. 2020, pp. 190–199, doi: 10.1145/3383219.3383239.

[31] N. Schmitt, J. Bucek, J. Beckett, A. Cragin, K.-D. Lange, and S. Kounev, "Performance, power, and energy-efficiency impact analysis of compiler optimizations on the SPEC CPU 2017 benchmark suite," in *Proc. 2020 IEEE/ACM 13th Int. Conf. on Utility and Cloud Computing (UCC)*, 2020, pp. 292–301, doi: 10.1109/UCC48980.2020.00047.

[32] B. Prieto, J. J. Escobar, J. C. Gómez-López, A. F. Díaz, and T. Lampert, "Energy Efficiency of Personal Computers: A Comparative Analysis," *Sustainability*, vol. 14, no. 19, Art. 12829, Oct. 2022, doi: 10.3390/su141912829.

[33] A. Kruglov, G. Succi, and Z. Kholmatova, "Metrics of Sustainability and Energy Efficiency of Software Products and Process," in Developing Sustainable and Energy-Efficient Software Systems, *SpringerBriefs in Computer Science*, Cham, Switzerland, 2023, pp. 19–26, doi: 10.1007/978-3-031-11658-2_2.

[34] Z. Zhang, S. Liang, F. Yao, and X. Gao, "Red Alert for Power Leakage: Exploiting Intel RAPL-Induced Side Channels," in *Proc. 2021 ACM Asia Conference on Computer and Communications Security (ASIA CCS '21)*, Hong Kong, June 2021, pp. 162–175, doi: 10.1145/3433210.3437517.

[35] A. Safari, H. Sorouri, A. Rahimi, and A. Oshnoei, "A Systematic Review of Energy Efficiency Metrics for Optimizing Cloud Data Center Operations and Management," *Electronics*, vol. 14, no. 11, Art. 2214, May 2025, doi: 10.3390/electronics14112214.

[36] Green Software Foundation. "Software Carbon Intensity (SCI) Specification v1.0," *Green Software Foundation*, 2021. [Online]. Available: https://sci.greensoftware.foundation/

[37] UBS, "Baselining Software Carbon Emissions: UBS Use Case," *Green Software Foundation*, 2023. [Online]. Available: https://greensoftware.foundation/articles/baselining-software-carbon-emissions-ubs-use-case/

[38] A. Schmidt, G. Stock, R. Ohs, L. Gerhorst, B. Herzog, and T. Hönig, "carbond: An Operating-System Daemon for Carbon Awareness," in *Proc. 2nd Workshop on Sustainable Computer Systems, HotCarbon '23*, Boston, MA, USA, 2023, pp. 1-10, doi: 10.1145/3604930.3605707.

[39] ISO/IEC 21031:2024. *Information technology - Software carbon intensity - Measurement and reporting framework*, International Organization for Standardization, Geneva, Switzerland, 2024. [Online]. Available: https://www.iso.org/standard/86612.html/

[40] T. Simon, P. Rust, R. Rouvoy, and J. Penhoat, "Uncovering the environmental impact of software life cycle," in *Proc. 2023 IEEE Int. Conf. on ICT for Sustainability (ICT4S '23)*, 2023, pp. 176–187, doi: 10.1109/ICT4S58814.2023.00026.

[41] A. Imran, T. Kosar, J. Zola, and M. F. Bulut, "Towards Sustainable Cloud Software Systems through Energy-Aware Code Smell Refactoring," in *Proc. 2024 IEEE 17th International Conference on Cloud Computing (CLOUD '24)*, Shenzhen, China, 2024, pp. 223–234, doi: 10.1109/CLOUD62652.2024.00034.

[42] N. Marini, L. Pampaloni, F. Di Martino, R. Verdecchia, and E. Vicario, "Green AI: Which Programming Language Consumes the Most?," *arXiv*, 2024.

[43] N. van Kempen, H.-J. Kwon, D. T. Nguyen, and E. D. Berger, "It's Not Easy Being Green: On the Energy Efficiency of Programming Languages," *arXiv*, 2024.

[44] N. Fatema Reya, A. Ahmed, T. Zaman, and Md. M. Islam, "GreenPy: Evaluating Application-Level Energy Efficiency in Python for Green Computing," *Annals of Emerging Technologies in Computing*, vol. 3, 2023, pp. 92–110, doi:10.33166/aetic.2023.03.005.

[45] M. Couto, D. Maia, J. Saraiva, and R. Pereira, "On Energy Debt: Managing Consumption on Evolving Software," in *Proc. 3rd IEEE/ACM Int. Conf. on Technical Debt (TechDebt '20)*, Seoul, Republic of Korea, June 2020, pp. 62–66, doi: 10.1145/3387906.3388628.

[46] S. U. Lee, N. Fernando, K. Lee & J.-G. Schneider, "A Survey of Energy Concerns for Software Engineering," *Journal of Systems and Software*, vol. 210, article no. 111944, 2024, doi: 10.1016/j.jss.2023.111944.

[47] H. Noman, M. Freed, and S. Jain, "An Exploratory Study of Software Sustainability at Early Stages," *Sustainability*, vol. 14, no. 14, p. 8596, Jul. 2022, doi: 10.3390/su14148596

AUTHORS

Mikita Piastou

Mikita Piastou is a senior full-stack software engineer with over eight years of experience in the technology industry, specializing in both data and software development. He holds a Master's degree in Computer Science from the University of West Georgia and has contributed to several publications on artificial intelligence, software development, and broader computer science topics. Throughout his career, Mikita has designed, implemented, and maintained highly scalable and reliable software solutions, integrated cutting-edge AI technologies, and optimized complex data systems for various organizations across multiple sectors. He is also a member of IEEE, actively engaging with the professional community. Beyond his technical expertise, he serves as a judge in hackathons and provides guidance to fellow developers. Outside of work, he enjoys keeping up with the latest technology trends, experimenting with new programming frameworks, working on creative side projects that contribute to the advancement of society, and exploring entrepreneurship through startup competitions and collaborative innovation ventures.

ISSN:1390-9266 e-ISSN:1390-9134 LAJC 2026

Agile Development and Usability Evaluation of an Educational Application Prototype to Foster Traditional and Digital Literacy

ARTICLE HISTORY
Received 11 June 2025
Accepted 19 August 2025
Published 6 January 2026

Lucrecia Llerena
Quevedo State University
Software Engineering Program
Quevedo, Ecuador
lllerena@uteq.edu.ec
ORCID: 0000-0002-4562-6723

Steffany Loor
Quevedo State University
Software Engineering Program
Quevedo, Ecuador
sloors@uteq.edu.ec
ORCID: 0009-0009-2263-2010

Nancy Rodríguez
Quevedo State University
Software Engineering Program
Quevedo, Ecuador
nrodriguez@uteq.edu.ec
ORCID: 0000-0002-0861-4352

ISSN:1390-9266 e-ISSN:1390-9134 LAJC 2026

Desarrollo ágil y evaluación de usabilidad de un prototipo de aplicación educativa para fomentar el alfabetismo tradicional y digital

Agile Development and Usability Evaluation of an Educational Application Prototype to Foster Traditional and Digital Literacy

Lucrecia Llerena
Quevedo State University
Software Engineering Program
Quevedo, Ecuador
lllerena@uteq.edu.ec

Steffany Loor
Quevedo State University
Software Engineering Program
Quevedo, Ecuador
sloors@uteq.edu.ec

Nancy Rodríguez
Quevedo State University
Software Engineering Program
Quevedo, Ecuador
nrodriguez@uteq.edu.ec

Resumen —El analfabetismo tanto tradicional y digital continúa limitando la integración social y el acceso equitativo a oportunidades educativas y laborales. En respuesta, se desarrolló PixelABC, un prototipo de aplicación interactiva basado en Windows Forms, orientado a fortalecer habilidades básicas de lectoescritura y competencias digitales. El desarrollo se estructuró utilizando la metodología ágil Scrum, y su diseño se fundamentó en un mapeo sistemático de literatura (SMS) que identificó estrategias tecnológicas en contextos vulnerables. PixelABC integra recursos como juegos educativos, módulos temáticos, videos y cuestionarios, facilitando el aprendizaje interactivo. La evaluación de usabilidad se realizó mediante entrevistas estructuradas con usuarios, lo que permitió identificar aspectos clave para mejorar la interfaz, la claridad de instrucciones y el rendimiento del sistema. Los resultados destacan el potencial del prototipo para promover la inclusión educativa, aunque se identificaron mejoras necesarias en diseño visual, velocidad de carga y adaptabilidad. Este trabajo concluye que PixelABC es una herramienta viable para fomentar el alfabetismo tradicional y digital. Se proyecta su evolución mediante optimizaciones de interfaz, inclusión de recursos multimedia y adaptación a plataformas móviles, lo cual incrementará su alcance y efectividad en la reducción de brechas digitales.

Palabras clave— *Alfabetismo tradicional, Alfabetismo digital, Windows Forms, Desarrollo de software educativo, Pruebas de usabilidad*

Abstract —Traditional and digital illiteracy continue to hinder social integration and equitable access to educational and employment opportunities. In response, PixelABC was developed as an interactive application prototype based on Windows Forms, designed to strengthen basic literacy and digital skills. The development process followed the agile Scrum methodology and was guided by a Systematic Mapping Study (SMS) to identify technological strategies in vulnerable contexts. PixelABC integrates educational games, thematic modules, videos, and quizzes to facilitate interactive learning. Usability was evaluated through structured interviews with users, which helped identify key areas for improvement in interface design, instructional clarity, and system performance. Results highlight the potential of the prototype to promote educational inclusion, although enhancements are needed in visual design, loading speed, and adaptability. This study concludes that PixelABC is a viable tool for fostering traditional and digital literacy. Future improvements will focus on interface optimization, integration of multimedia resources, and adaptation to mobile platforms, thereby increasing its reach and effectiveness in reducing digital divides.

Keywords— *Traditional Literacy, Digital Literacy, Windows Forms, Educational Software Development, Usability Testing*

I. Introduction

En la actualidad, el alfabetismo tradicional y digital representa un factor decisivo para la integración plena de las personas en la sociedad. La capacidad de leer, escribir y utilizar tecnologías digitales básicas no solo permite el acceso a la educación y al empleo, sino que también fortalece la participación ciudadana, el desarrollo individual y la cohesión social. En un entorno progresivamente digitalizado, estas competencias se han vuelto esenciales para reducir la exclusión y promover oportunidades equitativas en todos los ámbitos de la vida moderna [1], [2].

Sin embargo, aún persisten brechas significativas que impiden que estas habilidades sean accesibles para toda la población. El analfabetismo tradicional, reflejado en la incapacidad de leer y escribir, continúa afectando a millones de personas, y limita sus posibilidades educativas y laborales. Por su parte, el analfabetismo digital (entendido como la falta de habilidades para interactuar con tecnologías básicas) restringe el acceso a la información, los servicios en línea y los entornos virtuales de aprendizaje [3], [4]. Estas formas de analfabetismo suelen coexistir en comunidades vulnerables, lo que genera un ciclo de exclusión social y económica difícil de romper [5].

Ante esta realidad, diversas investigaciones han propuesto la implementación de tecnologías educativas como una estrategia efectiva para enfrentar ambos tipos de analfabetismo. Por ejemplo, Schmidt et al. [6] desarrollaron TalkingBook, un dispositivo de bajo costo que permitió a comunidades rurales de Ghana acceder a contenidos educativos y mejorar su alfabetización funcional y sus prácticas agrícolas. Asimismo, Khan et al. [4] diseñaron una

ISSN:1390-9266 e-ISSN:1390-9134 LAJC 2026
DOI: 10.5281/zenodo.17289287

aplicación educativa móvil durante la pandemia de COVID-19 para mitigar la deserción escolar, y demostraron cómo las herramientas digitales pueden fortalecer la continuidad educativa incluso en contextos de emergencia. Estas propuestas, junto con el uso de metodologías de desarrollo ágil como Scrum, han facilitado la creación de soluciones centradas en el usuario, adaptadas a sus necesidades reales y con potencial para reducir la brecha educativa y tecnológica.

En este marco, el presente trabajo de investigación propone el desarrollo de un prototipo de aplicación educativa interactiva en entorno Windows Forms, denominada PixelABC. Esta herramienta está diseñada para fortalecer simultáneamente las habilidades de lectoescritura y las competencias digitales básicas en personas con baja alfabetización, a través de módulos interactivos que incluyen juegos educativos, cuestionarios y contenidos audiovisuales. Su enfoque didáctico, intuitivo y atractivo busca generar una experiencia de aprendizaje accesible y significativa para usuarios con limitaciones educativas o tecnológicas. El uso de Windows Forms fue clave para estructurar el prototipo, mediante el uso de formularios interactivos y controles gráficos, lo cual facilitó la integración de módulos educativos. Aunque se trata de una tecnología con menor vigencia frente a alternativas como WPF o .NET MAUI, se seleccionó por su facilidad de implementación, bajo requerimiento de hardware y rápida curva de aprendizaje, factores adecuados para un prototipo educativo en contexto académico.

El desarrollo del prototipo se fundamenta en un mapeo sistemático de literatura (SMS), que permitió identificar tanto las principales causas del analfabetismo como las estrategias tecnológicas más efectivas para enfrentarlo. Posteriormente, se aplicó la metodología ágil Scrum para organizar el proceso de diseño, implementación y evaluación del prototipo. Esta estructura metodológica aseguró una planificación iterativa que permitió incorporar retroalimentación constante y mejorar progresivamente las funcionalidades del sistema.

Para evaluar la usabilidad de PixelABC, se utilizó la técnica entrevista estructurada, adaptada al contexto de usuarios con baja alfabetización digital. Esta técnica permitió identificar barreras de interacción, necesidades específicas y oportunidades de mejora en el diseño del sistema. Entre los hallazgos más relevantes se identificaron aspectos como la necesidad de simplificar la interfaz, mejorar la organización de los módulos, optimizar el rendimiento del sistema y ofrecer instrucciones claras durante el uso.

Este estudio plantea a PixelABC como una solución inicial con proyección inclusiva para abordar el analfabetismo desde una perspectiva integral, que combina la enseñanza de habilidades tradicionales y digitales. Su carácter interactivo y su diseño centrado en el usuario la convierten en una herramienta prometedora para reducir la brecha educativa y tecnológica en sectores marginados. La contribución de este trabajo radica en la integración de metodología ágil Scrum y la técnica de usabilidad entrevista estructurada adaptada en el desarrollo de un prototipo educativo orientado al alfabetismo tradicional y digital. Esta combinación metodológica constituye un aporte novedoso en el ámbito de la ingeniería de software educativo y la interacción humano–computador, ya que ofrece evidencia práctica sobre cómo adaptar procesos de desarrollo y evaluación a usuarios con baja alfabetización digital. A través de este prototipo, se busca no solo mejorar la alfabetización funcional de los usuarios, sino también empoderarlos para participar activamente en una sociedad cada vez más conectada y digitalizada.

II. Revisión de Literatura

La presente sección expone la revisión de literatura realizada mediante un Mapeo Sistemático de Literatura (SMS), cuyo propósito fue identificar los principales enfoques tecnológicos aplicados al alfabetismo tradicional y digital, especialmente en contextos vulnerables. El proceso siguió las fases típicas de un SMS: identificación de fuentes, cribado inicial, evaluación de elegibilidad e inclusión final. Para asegurar la claridad y transparencia del proceso de mapeo sistemático, la Figura 1 presenta un diagrama de flujo adaptado al estándar PRISMA 2020 [7], donde se resumen las etapas desarrolladas y el número de registros en cada fase. Esta revisión sustenta la pertinencia del desarrollo del prototipo PixelABC y permite establecer los requisitos funcionales del sistema a partir de soluciones previas documentadas y vacíos persistentes en la literatura.

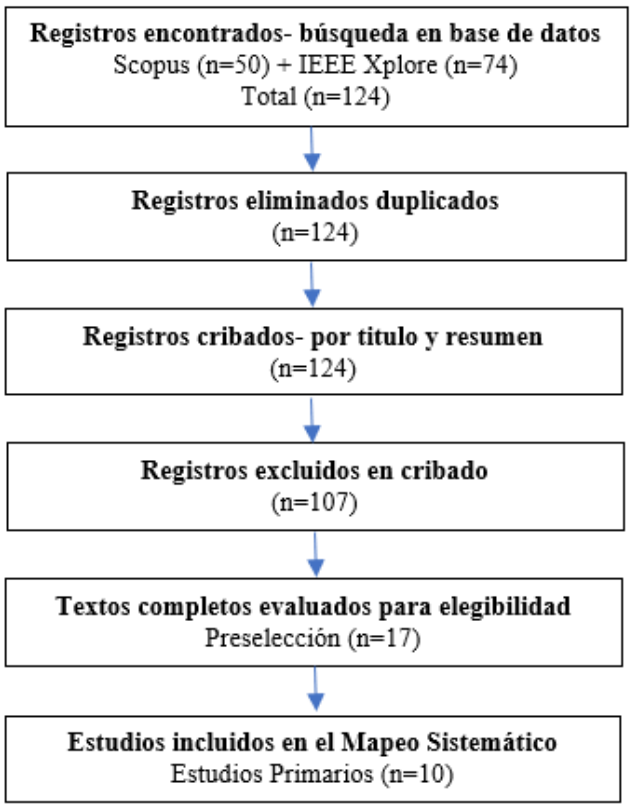

Fig. 1. Diagrama del proceso SMS. Adaptado de [7]

De acuerdo con Kitchenham et al. [8], el SMS es un enfoque útil para organizar y categorizar la evidencia científica disponible, la cual facilita el análisis de tendencias, soluciones aplicadas y líneas emergentes en un área específica del conocimiento. En este estudio, se formuló la siguiente pregunta de investigación: ¿Cómo se puede desarrollar de manera efectiva un prototipo de aplicación educativa para fomentar el alfabetismo tradicional y digital para favorecer la integración social de los usuarios?

ISSN:1390-9266 e-ISSN:1390-9134 LAJC 2026
DOI: 10.5281/zenodo.17289287

Para responder esta interrogante, se realizó una búsqueda sistemática en las bases de datos Scopus e IEEE Xplore, utilizando la cadena: (illiteracy OR digital illiteracy) AND (illiteracy rate OR causes of illiteracy). Se aplicaron los siguientes criterios de inclusión y exclusión los cuales se detallan en la Tabla I.

TABLE I. CRITERIOS DE INCLUSIÓN Y EXCLUSIÓN

Criterios de inclusión:
Estudios publicados entre 2018 y 2024. Estudios que incluyan tecnologías aplicadas al alfabetismo digital o tradicional. Estudios que reporten aplicaciones educativas (móviles o de escritorio) con enfoque en lectoescritura o habilidades tecnológicas.
Criterios de exclusión:
Estudios de menos de dos páginas. Estudios que no aborden herramientas tecnológicas orientadas a la alfabetización. Estudios que no contemplen contextos de uso educativo o social.

Tras aplicar estos criterios, se identificaron inicialmente 124 estudios, de los cuales se preseleccionaron 17 y finalmente, se consideraron 10 estudios primarios (ver Tabla II).

TABLE II. NÚMERO TOTAL DE ESTUDIOS OBTENIDOS DE LA BBDD

BBDD	**Encontrados**	**Preseleccionados**	**Estudios Considerados**
Scopus	50	10	5
IEEE Xplore	74	7	5
Total	124	17	10

A continuación, se sintetizan los hallazgos clave que sustentan esta investigación:

Uno de los trabajos más destacados es el de Khan et al. [4], quienes presentan una aplicación móvil desarrollada en Flutter para disminuir la deserción escolar durante la pandemia de COVID-19. La investigación demuestra cómo las herramientas digitales pueden garantizar la continuidad educativa en situaciones adversas, y reducen la brecha digital y promueven el acceso al conocimiento en entornos virtuales, aún en comunidades con conectividad limitada.

En una línea complementaria, Widiyaningtyas et al. [5] proponen el uso de minería de datos para identificar zonas con alta prevalencia de analfabetismo. Su algoritmo permite mapear grupos sociales vulnerables según niveles de alfabetización, lo que facilita la focalización de esfuerzos educativos y la recolección de datos estratégicos para intervenciones tecnológicas.

Un aporte significativo es el de Schmidt et al. [6], quienes desarrollaron TalkingBook, una tableta de audio de bajo costo implementada en aldeas rurales de Ghana. Este dispositivo ayudó a personas sin acceso a educación formal a mejorar su calidad de vida mediante la adquisición de conocimientos aplicables, especialmente en prácticas agrícolas, evidenciando el impacto positivo de la tecnología contextualizada.

Asimismo, Juditha et al. [1] realizaron un estudio en zonas rurales de Papúa que analizó las causas del analfabetismo digital. El trabajo destaca el papel clave de las TIC como medio para fortalecer habilidades digitales básicas y promover la inclusión tecnológica en contextos aislados y con baja inversión en educación tecnológica.

En relación con los factores estructurales del analfabetismo, Suhasini et al. [3] resaltan las consecuencias económicas y sociales de la falta de acceso a la educación. La carencia de recursos obliga a muchas familias a relegar la educación, lo que genera ciclos de pobreza intergeneracional y, en los casos más graves, fomenta el trabajo infantil. Estos hallazgos subrayan la urgencia de intervenciones accesibles y adaptadas a estos entornos.

Por otro lado, Schwartz et al. [9] enfatizan que el avance acelerado de la tecnología ha dejado a muchas personas atrás en términos de habilidades digitales. La falta de competencias tecnológicas constituye una barrera seria para la inclusión social y laboral. Su investigación propone programas de alfabetización digital que fomenten la autonomía y la participación activa en entornos digitales, y contribuyen a la reducción de desigualdades estructurales.

Desde una perspectiva macroeconómica, Viviana et al. [2] evidencian que reducir la brecha digital puede impactar positivamente en el PIB de un país. En el caso ecuatoriano, la mejora de habilidades digitales en la población podría aumentar la innovación, productividad y generación de empleo, convirtiendo la alfabetización digital en una herramienta de desarrollo nacional.

A nivel educativo, Bataller Català et al. [10] señalan que las habilidades de lectoescritura son esenciales para la participación social, el acceso a información útil y la mejora de la calidad de vida. Mejorar estas competencias facilita la participación en su entorno, abre oportunidades laborales y refuerza su integración en la comunidad.

En una línea similar, Kalman et al. [11] proponen un enfoque centrado en la interacción y la práctica para el desarrollo de la lectoescritura. Argumentan que el aprendizaje efectivo se basa en la apropiación activa de conocimientos, lo cual exige diseñar software educativo que permita a los usuarios interactuar de forma significativa con los contenidos.

Finalmente, el trabajo de Traversini C. [12] aporta una dimensión sociopsicológica al problema del analfabetismo. Su estudio sobre el Programa Alfabetización Solidaria en Brasil identifica la autoestima como un factor determinante en la permanencia de los estudiantes en procesos formativos. Las estrategias discursivas y metodológicas orientadas a fortalecer la autoconfianza contribuyen a una mayor retención, compromiso y éxito en los programas de alfabetización.

En conjunto, esta revisión muestra que existen múltiples enfoques tecnológicos, educativos y sociales que abordan los desafíos del analfabetismo. Sin embargo, también revela que persisten vacíos en cuanto a herramientas accesibles, validadas y orientadas a usuarios con baja alfabetización digital. Con base en los estudios identificados en el SMS, se presenta la Tabla III, la cual sintetiza los aportes más relevantes y permite una comparación sistemática entre autores, año de publicación, propuesta planteada, tecnología empleada y contexto de aplicación.

ISSN:1390-9266 e-ISSN:1390-9134 LAJC 2026
DOI: 10.5281/zenodo.17289287

TABLE III. ESTUDIOS SELECCIONADOS EN EL SMS Y SU CARACTERIZACIÓN COMPARATIVA

Autor	Año	Propuesta	Tecnología empleada	Contexto de aplicación
Khan et al.	2023	Aplicación móvil educativa para reducir la deserción escolar	Flutter	Estudiantes en comunidades con baja conectividad
Widiyaningtyas et al.	2023	Identificación de zonas con analfabetismo	Minería de datos (K-Means)	Políticas educativas y sociales
Schmidt et al.	2011	TalkingBook, dispositivo de audio de bajo costo	Tableta de audio	Aldeas rurales de Ghana
Juditha et al.	2018	TIC para combatir analfabetismo digital	Aplicaciones TIC	Comunidades rurales de Papúa
Suhashi et al.	2013	Microdonaciones para prevenir la explotación infantil	Plataforma digital de donaciones	Comunidades pobres sin acceso educativo
Viviana & Wilman-Santiago	2022	Relación entre la brecha digital y PIB	Análisis económico con TIC	Ecuado
Bataller Català et al.	2019	Propuestas metodológicas de alfabetización	Métodos pedagógicos	Personas adultas
Kalman	2008	Enfoque de interacción y práctica en lecto escritura	Software educativo	Contextos de aprendizaje activo
Traversini	2009	Autoestima como factor clave en la alfabetización	Programas educativos	Brasil
Schwartz et al.	2024	Programas de alfabetización digital para fomentar autonomía y participación en entornos digitales	Plataformas y herramientas digitales	Inclusión social y laboral, reducción de desigualdades estructurales

A partir de estos hallazgos sintetizados se definieron los requisitos funcionales y no funcionales, los cuales orientaron el diseño del prototipo PixelABC, los mismos se presentan en la siguiente sección.

III. METODOLOGÍA DE DESARROLLO DE SOFTWARE

Para el desarrollo del prototipo educativo PixelABC, orientado a la enseñanza de lectoescritura y alfabetización digital básica, se adoptó la metodología ágil Scrum, ampliamente reconocida por su adaptabilidad en proyectos educativos y tecnológicos. Este enfoque permite gestionar el desarrollo de manera iterativa e incremental, con entregas continuas de valor y con la incorporación de retroalimentación de usuarios en cada etapa del proyecto. La estructura del proceso Scrum se ilustra en la Figura 2.

Fig. 2. Proceso de la Metodología Scrum. Adaptada de [13]

A. Definición de Requisitos

Los resultados obtenidos en el Mapeo Sistemático de la literatura (SMS) orientaron la definición de los requisitos funcionales, como juegos educativos interactivos, así como también cuestionarios, y los requisitos no funcionales, como la necesidad de una interfaz intuitiva y optimizar el rendimiento del hardware, esto teniendo en cuenta las prácticas y funcionalidades clave identificadas en soluciones similares. Además, se aplicaron entrevistas semiestructuradas a usuarios potenciales, con el objetivo de adaptar el desarrollo a las necesidades reales del público objetivo, compuesto por personas en contextos de vulnerabilidad educativa y tecnológica.

1) *Recolección de Requisitos:* Los requisitos identificados fueron clasificados en dos categorías: requisitos funcionales relacionados con las acciones que el sistema debe ejecutar y requisitos no funcionales relacionados con el rendimiento del hardware.
2) *Documentación de Requisitos:* Las Tablas IV y V presentan los requisitos definidos para el sistema. En la Tabla IV, se presentan los requisitos funcionales definidos para el sistema.

TABLE IV. REQUISITOS FUNCIONALES DEL PROTOTIPO PIXELABC

Requisito	Descripción
Registro de Usuario	Permitir el ingreso de nuevos usuarios mediante datos básicos.
Acceso con Usuario y Clave	Validar credenciales para ingresar a la plataforma.
Juegos de Lectoescritura	Incluir ejercicios interactivos que refuercen habilidades básicas.
Cuestionarios de Práctica	Evaluar el progreso mediante preguntas de selección y respuesta corta.
Reproductor de Contenido	Facilitar el acceso a contenidos visuales y auditivos para reforzar temas.
Registro de Resultados	Guardar automáticamente los avances del usuario en una base de datos local.

La Tabla V, por su parte, recoge los requisitos no funcionales, que fueron planteados con el fin de que el sistema cumpla con los objetivos establecidos de eficiencia y calidad del servicio.

ISSN:1390-9266 e-ISSN:1390-9134 LAJC 2026
DOI: 10.5281/zenodo.17289287

TABLE V. REQUISITOS NO FUNCIONALES DEL PROTOTIPO PIXELABC

Requisito	Descripción
Interfaz Intuitiva	Diseñar una interfaz visual amigable para personas con baja alfabetización.
Tiempo de Respuesta	Garantizar fluidez al cambiar de actividades o pantallas.
Bajo Requerimiento	Optimizar el rendimiento para equipos con especificaciones limitadas.
Instalación Sencilla	Permitir una instalación rápida, sin conexión a Internet.

B. Planificación del Proyecto

1) *Elaboración del Product Backlog:* Se elaboró un Product Backlog que incluye las funcionalidades a desarrollar, priorizadas según el impacto educativo y la viabilidad técnica. Estos fueron gestionados como tareas de sprint, de tal modo que en cada una de las iteraciones se incluyeron actividades concretas para avanzar en las funcionalidades priorizadas.
2) *Planificación de Sprints:* Se definieron sprints de 2 semanas, cada uno con metas específicas.
3) *Desarrollo Iterativo:* La Tabla VI presenta la planificación de los sprints que detalla las tareas asignadas, los objetivos definidos y los tiempos estimados de ejecución. Este enfoque iterativo permite una implementación progresiva del sistema, y facilita revisiones continuas, retroalimentación frecuente y validaciones constantes del producto en desarrollo.

Además, los requisitos se documentaron como historias de usuario, y aseguraron la trazabilidad entre los hallazgos del SMS, los requisitos y las tareas de los sprints. Ejemplo: "Como usuario con baja alfabetización digital, quiero instrucciones claras y visuales en cada módulo para comprender mejor las actividades". Esta historia se tradujo en el requisito no funcional de interfaz intuitiva (Tabla V) y se gestionó en el Sprint 1 (Tabla VI).

TABLE VI. PLANIFICACIÓN DE SPRINTS

Sprint	Fase	Tiempo (Semana)	Tareas	Objetivos
0	Preparación	0	Instalación de herramientas y planificación inicial	Establecer el entorno y objetivos del proyecto
1	Diseño de Interfaz Gráfica	1–2	Bocetos, maquetas y selección de colores/tipografía	Crear una UI intuitiva para usuarios con bajo nivel educativo
2	Desarrollo del Módulo de Registro	3–4	Implementar formularios de registro e inicio de sesión	Controlar acceso seguro al sistema
3	Juegos y Actividades	5–6	Programar juegos de lectura y escritura, validar lógica	Favorecer el aprendizaje lúdico e interactivo
4	Cuestionarios y Evaluación	7–8	Desarrollar pruebas de práctica con retroalimentación	Medir progreso y reforzar el aprendizaje
5	Integración y Ajustes Finales	9–10	Unificar módulos, corrección de errores, validación de flujo	Garantizar funcionamiento integral y experiencia positiva del usuario

C. Mecanismos de automatización de procesos

El sistema automatiza diversos procesos pedagógicos, entre ellos:

- **Almacenamiento de progreso:** Guarda automáticamente los resultados obtenidos por el usuario al finalizar cada actividad.
- **Evaluación inmediata:** Muestra retroalimentación instantánea en cuestionarios, lo que motiva el aprendizaje continuo.
- **Adaptabilidad de actividades:** Según el rendimiento registrado, el sistema puede ofrecer juegos de refuerzo u opciones avanzadas.

IV. EVALUACIÓN DE USABILIDAD DEL PROTOTIPO PIXELABC

En esta sección, se detalla la estrategia de evaluación de la usabilidad del prototipo PixelABC, una aplicación interactiva desarrollada en Windows Forms, orientada a la enseñanza de habilidades básicas de lectoescritura y alfabetización digital. Para evaluar su efectividad y facilidad de uso, se implementó la técnica de entrevista estructurada, la cual permitió obtener retroalimentación directa de los usuarios sobre su experiencia de interacción con el sistema.

A. Descripción de la técnica de usabilidad Entrevista Estructurada

La técnica denominada entrevista estructurada en el ámbito de la usabilidad de software consiste en aplicar un conjunto de preguntas previamente definidas a los usuarios, con el objetivo de recopilar información sobre su experiencia de interacción con el sistema, así como identificar sus necesidades y posibles dificultades [14]. Esta técnica resulta especialmente útil porque permite al investigador mantener el control del desarrollo de la entrevista y obtener datos sistemáticos y consistentes. El término "estructurada" implica que las preguntas han sido diseñadas de antemano, en función de los aspectos específicos que se desean evaluar [15]. La Tabla VII presenta los pasos y actividades que conforman la aplicación de esta técnica, la cual detalla su estructura metodológica para garantizar la validez de los resultados obtenidos.

TABLE VII. PASOS Y TAREAS DE LA TÉCNICA ENTREVISTA ESTRUCTURADA SEGÚN HIX [16]

N°	Nombre del paso	Tareas
1	Definición de los objetivos	Se debe identificar el objetivo de aplicar la técnica para mejorar la usabilidad de un proyecto open-source.
2	Identificar el público objetivo	Se deben seleccionar a los participantes para aplicar la técnica.

ISSN:1390-9266 e-ISSN:1390-9134 LAJC 2026
DOI: 10.5281/zenodo.17289287

3	Diseñar las preguntas	Las preguntas deben ser claras, concisas y relevantes para los objetivos de la entrevista
4	Validar las preguntas	Solicita a otros usuarios que revisen las preguntas para asegurarse de que sean claras y comprensivas
5	Analizar los datos	Utiliza técnicas estadísticas apropiadas para el análisis de datos y de esta forma extraer conclusiones

B. Adaptaciones de la técnica de usabilidad Entrevista Estructurada

La técnica entrevista estructurada no representa una dificultad significativa, ya que se utiliza con frecuencia en evaluaciones de usabilidad [14]. No obstante, su correcta implementación requiere la participación de un evaluador con experiencia en usabilidad, así como habilidades interpersonales que le permitan conducir la entrevista de manera eficaz y objetiva [17].

En la Tabla VIII, se presentan las principales condiciones adversas encontradas durante la aplicación de esta técnica, junto con las adaptaciones realizadas para garantizar la calidad y confiabilidad de los datos recopilados.

TABLE VIII. CONDICIONES ADVERSAS Y ADAPTACIONES DE LA TÉCNICA ENTREVISTA ESTRUCTURADA SEGÚN HIX [16]

N°	Nombre del paso	Condiciones adversas	Adaptaciones Propuestas
1	Definición de los objetivos	Es indispensable contar con un experto en usabilidad para aplicar la técnica.	El experto en usabilidad es sustituido por: Un estudiante o grupo de estudiantes de la UTEQ (bajo la supervisión de un mentor).
2	Identificar el público objetivo	Es necesaria la participación de los usuarios para aplicar la técnica.	Los usuarios participan remotamente a través de: foros o correos electrónicos, o, realizando comentarios en el blog.
3	Diseñar las preguntas	Es indispensable contar con un experto en usabilidad para aplicar la técnica.	El experto en usabilidad es sustituido por: Un estudiante o grupo de estudiantes de la UTEQ (bajo la supervisión de un mentor).
4	Validar las preguntas		
5	Analizar los datos		

La técnica entrevista estructurada [16] fue adaptada debido a las dificultades encontradas durante el desarrollo del estudio. Para superar estas limitaciones contextuales, se estableció una versión modificada de la técnica, compuesta por seis pasos específicos que integran recursos y estrategias concretas para asegurar su adecuada implementación. A continuación, se detallan los pasos que conforman esta adaptación metodológica.

- Paso 1: Prueba piloto. Se realizó una prueba piloto con el objetivo de verificar el funcionamiento general del proceso y validar los insumos desarrollados. Esta fase permitió detectar posibles ajustes en los instrumentos antes de aplicarlos en el estudio de usabilidad definitivo.
- Paso 2: Selección de herramientas de comunicación y colaboración. Debido a que el proyecto se gestionó con un equipo distribuido, se emplearon herramientas virtuales (Gmail para el intercambio de documentos y WhatsApp como canal de soporte durante la aplicación de la técnica) que facilitaron la interacción con los participantes.
- Paso 3: Adaptación del formato de entrevista. Se elaboró una primera versión del cuestionario con preguntas centradas en factores de usabilidad específicos. Estas preguntas fueron revisadas y validadas con el mentor del proyecto antes de su implementación definitiva.
- Paso 4: Diseño de plantillas y documentos de apoyo. Las preguntas validadas fueron integradas en un documento estructurado en Word, el cual fue enviado a los participantes tras la ejecución de las tareas asignadas, garantizando una recolección organizada de los datos.
- Paso 5: Ejecución de la evaluación de usabilidad. Se estableció una fecha para la evaluación y se envió a los usuarios un paquete de insumos que incluía: enlace a una videollamada en Google Meet, grupo de WhatsApp, consentimiento informado, guía de instalación, tareas a realizar, cuestionario de evaluación y un video explicativo. Esta organización permitió una participación remota efectiva y estructurada.
- Paso 6: Análisis de datos y sistematización de resultados. Una vez finalizada la evaluación, se procedió al análisis de las respuestas. Los datos fueron depurados, eliminando duplicados, y agrupados en categorías comunes que se integraron en un documento titulado "Entrevista Estructurada", el cual sirvió como insumo clave para retroalimentar el diseño del sistema.

La Tabla IX sintetiza los pasos y tareas implementadas en la adaptación de la técnica entrevista estructurada, aplicada específicamente para la evaluación de usabilidad en un entorno OSS.

TABLE IX. PASOS Y TAREAS DE LA TÉCNICA ADAPTADA ENTREVISTA ESTRUCTURADA

N	Nombre del paso	Tarea
1	Realizar una prueba piloto	• Realizar una prueba piloto para probar los insumos a utilizar en la aplicación de la técnica.
2	Utilizar herramientas de comunicación y colaboración	• Definir el perfil de usuario que será requerido para la aplicación de la técnica. • Enviar un correo solicitando la participación de los usuarios. • Reclutar usuarios a través de redes sociales.
3	Adaptar el formato de las entrevistas	• Diseñar una serie de preguntas para aplicar la entrevista según los elementos que se desean evaluar.
4	Diseñar plantillas para aplicar las entrevistas	• Usar herramientas como Google forms para diseñar la lista de preguntas.
5	Realizar la evaluación de usabilidad	• Realizar la evaluación mediante una reunión remota, en donde se envían los insumos necesarios para los usuarios.
6	Realizar un análisis de los datos y agrupar los comunes	• Agrupar los resultados en el documento Entrevista Estructurada

ISSN:1390-9266 e-ISSN:1390-9134 LAJC 2026
DOI: 10.5281/zenodo.17289287

V. RESULTADOS

En esta sección se presentan los resultados obtenidos a partir de la aplicación de la técnica de evaluación de usabilidad mediante entrevistas estructuradas, desarrolladas con usuarios reales en un entorno virtual. El objetivo fue obtener retroalimentación cualitativa sobre el uso de la herramienta "PixelABC" en relación con su diseño, funcionalidades, y experiencia de usuario.

A. Resultados de las pruebas de usabilidad

Las pruebas se realizaron con 4 estudiantes universitarios de primer nivel (18–20 años) de la carrera de Software en la Universidad Técnica Estatal de Quevedo. La muestra estuvo conformada por participantes seleccionados por conveniencia, quienes contaban con conocimientos básicos de lectoescritura y alfabetización digital, simulando el perfil objetivo del sistema. Las sesiones se llevaron a cabo de manera remota, con el soporte técnico del equipo desarrollador y el acompañamiento de un mentor académico. Se reconoce como limitación metodológica la ausencia de usuarios en procesos reales de alfabetización, lo que restringe la validez externa; esta limitación será abordada en futuras iteraciones con una muestra más representativa.

A los participantes se les proporcionó previamente el instalador del sistema, una guía de instalación, un video tutorial, las tareas a ejecutar y el formulario de evaluación. Tras completar las tareas de interacción con la aplicación, los usuarios respondieron un cuestionario estructurado que permitió recopilar observaciones relevantes. La Figura 3 muestra la interacción remota con la herramienta.

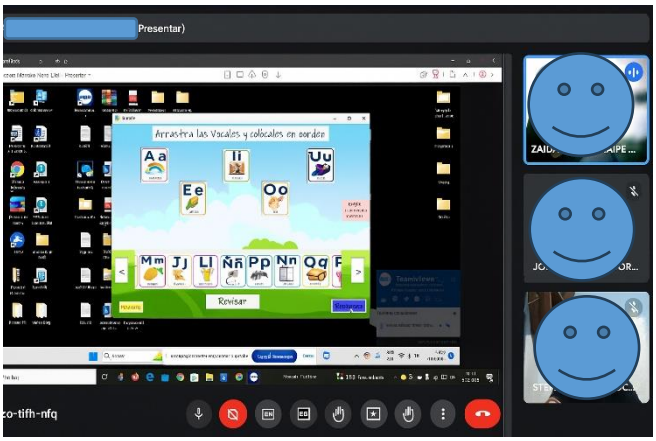

Fig. 3. Reunión remota con los usuarios para la entrevista

La Tabla X presenta el conjunto de preguntas formuladas a los usuarios en el proceso de recolección de datos mediante la técnica de entrevista estructurada.

TABLE X. PREGUNTAS FORMULADAS PARA LA TÉCNICA ADAPTADA ENTREVISTA ESTRUCTURADA

N.	Pregunta
1	¿Cuáles son los principales problemas en las funcionalidades que encontraste?
2	¿Tienes alguna propuesta de mejora para la interacción con la herramienta?
3	¿Tienes alguna crítica en torno a la interfaz de usuario?
4	¿Cómo piensas que la interfaz de usuario (o una parte de ella) podría ser rediseñada?
5	¿Hubo alguna característica o proceso difícil de usar o de entender de la herramienta?

B. Análisis de respuestas de la Entrevista Estructurada

Se analizaron las respuestas de los participantes en la evaluación de usabilidad, buscando identificar patrones comunes, dificultades percibidas y oportunidades de mejora. A continuación, se presentan los hallazgos agrupados según las preguntas formuladas durante la entrevista estructurada.

1. ¿Cuáles son los principales problemas en las funcionalidades que encontraste? Los participantes identificaron varios aspectos a mejorar en el prototipo. Un usuario sugirió la elaboración de un video instructivo para facilitar la comprensión inicial. Otro participante mencionó que el prototipo no resultó funcional ni atractivo al ser probado en un dispositivo móvil. Asimismo, se reportó que la interfaz visual no era coherente con el propósito educativo de la aplicación, y se recomendó incorporar fondos temáticos y permitir trazos manuales en lugar de presionar botones. También se mencionó que algunos módulos resultan tediosos y poco intuitivos, lo que dificultó la navegación.

2. ¿Tienes algunas propuestas de mejora para la interacción con la herramienta? Los usuarios propusieron diversas mejoras enfocadas en la experiencia interactiva. Se destacó la necesidad de evitar confusiones con nombres similares a plataformas existentes, como arbolabc.com. Además, se sugirió rediseñar la interfaz para hacerla más atractiva, incorporar instrucciones claras en cada módulo, y agregar mensajes que indiquen la finalización de cada actividad. En el módulo de "Ordenar vocales y consonantes", se recomendó que las letras aparezcan de forma automática, y en el módulo de dibujo, se planteó reemplazar los botones por una función que permita trazar letras manualmente.

3. ¿Tienes alguna crítica o queja de la interfaz de usuario? Todos los participantes coincidieron en que el diseño de la interfaz presenta debilidades significativas. En particular, se mencionó que los fondos utilizados tienden a opacar los elementos funcionales, y reduce la claridad del contenido educativo. Asimismo, el menú fue percibido como sobrecargado y desorganizado, lo cual dificulta la navegación dentro del sistema.

4. ¿Cómo piensas que la interfaz de usuario (o una parte de ella) podría ser rediseñada? En cuanto a las sugerencias de rediseño, los entrevistados indicaron que es necesario optimizar el diseño visual general de la herramienta. Se propuso mejorar la estética del sistema, cambiar la tipografía utilizada para mejorar la legibilidad y rediseñar módulos específicos como el de "Trivia", que fue considerado visualmente limitado por un usuario.

5. ¿Hubo alguna característica o proceso difícil de usar dentro de la herramienta? Respecto a los desafíos de uso, algunos participantes manifestaron dificultades al arrastrar letras en un orden específico, lo cual generó frustración. Otro usuario señaló que el uso del prototipo desde un teléfono móvil limitó significativamente la experiencia. Se reportaron también demoras en la carga del módulo de juegos de palabras, y aunque un participante no encontró mayores dificultades, indicó que la herramienta en general era algo lenta.

C. Análisis de observaciones dadas por el usuario

Las observaciones proporcionadas permitieron identificar áreas críticas de mejora, especialmente en la interfaz gráfica y en la personalización de la experiencia para distintos perfiles de usuario. La Tabla XI resume los hallazgos organizados por

ISSN:1390-9266 e-ISSN:1390-9134 LAJC 2026
DOI: 10.5281/zenodo.17289287

categoría y frecuencia de mención por parte de los participantes. En conjunto, todos estos hallazgos revelan algunos patrones comunes, como la necesidad de simplificar las interfaces de usuario, optimizar la navegabilidad y mejorar el rendimiento, factores clave para su adopción.

TABLE XI. INFORMACIÓN RECOPILADA MEDIANTE LA ENTREVISTA A USUARIOS, CLASIFICADA POR CATEGORÍAS, HALLAZGOS Y FRECUENCIA

Categoría	Hallazgo	Frecuencia
Problemas	Retroalimentación ambigua en módulo de escritura	2
	Fallo ocasional del sonido de ayuda	1
Propuestas de Mejora	Agregar tutorial interactivo al inicio	2
	Incluir elementos visuales motivacionales	2
Interfaz	Modernizar colores y fuentes	3
	Mejorar visibilidad de botones	1
Usabilidad	Dificultades en instalación inicial del sistema	2

D. Diseño de Interfaces resultantes

A partir de las observaciones recopiladas en la fase de evaluación de usabilidad, se realizaron mejoras visuales y funcionales al prototipo interactivo "PixelABC". Las interfaces resultantes buscan proporcionar una experiencia educativa accesible, lúdica y atractiva para el usuario, e integrar principios de usabilidad y elementos motivacionales en cada módulo.

La Figura 4 presenta la pantalla principal del prototipo, compuesta por una interfaz colorida e intuitiva que invita al usuario a explorar los cuatro módulos disponibles. Esta pantalla fue diseñada con colores vibrantes, formas dinámicas y efectos sonoros envolventes para fomentar el aprendizaje a través del juego. Los módulos accesibles desde esta vista son: (1) Dibuja las letras y números, (2) Ordena las letras del abecedario, (3) Ordena las vocales y consonantes, y (4) Trivia digital.

Fig. 4. Interfaz principal del prototipo PixelABC

La Figura 5 muestra el módulo Ordena las vocales y consonantes, un espacio que promueve el reconocimiento y la clasificación de letras mediante una actividad interactiva. Los usuarios deben identificar, arrastrar y ordenar las vocales y consonantes, lo cual contribuye al fortalecimiento de habilidades lingüísticas básicas.

Fig. 5. Interfaz "ordena las vocales y consonantes"

La Figura 6 corresponde al módulo Dibuja las letras y los números. En esta sección, los usuarios pueden practicar la escritura mediante una pizarra digital que estimula el desarrollo de la motricidad fina, la familiarización con la forma de los caracteres y el reconocimiento visual de letras y cifras.

Fig. 6. Interfaz "dibujar las letras y los números"

En la Figura 7, se ilustra el módulo Ordena las letras del abecedario, donde se reta a los usuarios a organizar correctamente las letras del alfabeto. Esta actividad está orientada al refuerzo del orden alfabético y al desarrollo de capacidades cognitivas relacionadas con la memoria y la secuenciación.

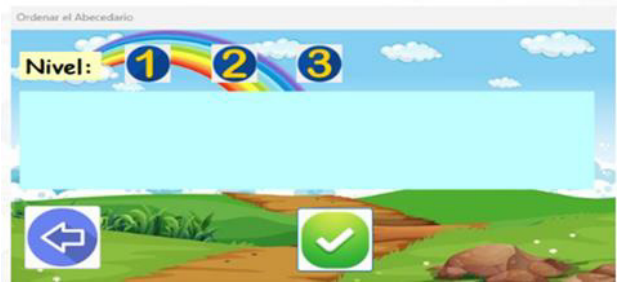

Fig. 7. Interfaz "ordenar las letras del abecedario"

ISSN:1390-9266 e-ISSN:1390-9134 LAJC 2026
DOI: 10.5281/zenodo.17289287

Finalmente, la Figura 8 representa el módulo Trivia digital, un juego educativo de preguntas y respuestas basado en contenidos audiovisuales sobre alfabetización digital. El usuario debe visualizar breves videos explicativos y responder preguntas relacionadas. Esto promueve el aprendizaje significativo sobre temas tecnológicos actuales.

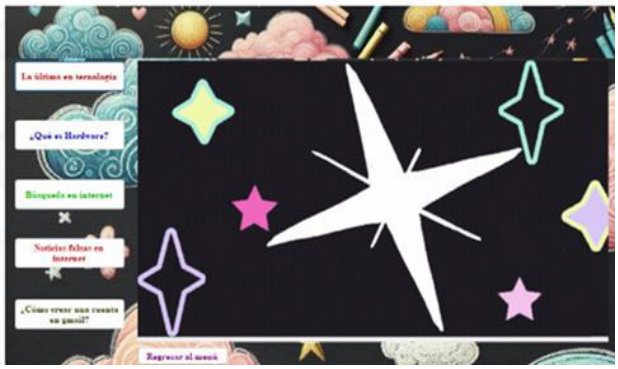

Fig. 8. Interfaz “realizar una trivia digital”

VI. Discusión de Resultados

Los resultados obtenidos evidencian que el prototipo “PixelABC” cumple con los objetivos propuestos, y se destaca como una herramienta pedagógica funcional y accesible que puede contribuir significativamente a la alfabetización tradicional y digital. Su diseño centrado en módulos interactivos permite una experiencia de aprendizaje lúdica e inclusiva, especialmente útil para usuarios con conocimientos tecnológicos limitados.

Una de las fortalezas principales identificadas fue la integración efectiva de contenidos educativos mediante actividades gamificadas, trivias, ejercicios de ordenamiento y escritura, lo que refuerza su aplicabilidad en contextos educativos que enfrentan desafíos relacionados con la brecha digital. La combinación de componentes visuales, auditivos e interactivos potencia la motivación del usuario, y se alinean con enfoques pedagógicos modernos que promueven el aprendizaje activo.

No obstante, las pruebas de usabilidad revelaron áreas de mejora relevantes. Se detectaron limitaciones en la interfaz gráfica de usuario, particularmente relacionadas con la sobrecarga visual y la ausencia de instrucciones explícitas, aspectos que dificultaron la navegación fluida e intuitiva, especialmente para usuarios novatos. Asimismo, se identificaron problemas de rendimiento en algunos módulos, asociados a la velocidad de carga, lo que sugiere la necesidad de optimizar tanto el código como la estructura del sistema para entornos de bajos recursos.

Si bien la muestra se limitó a estudiantes universitarios, los resultados obtenidos ofrecen evidencia práctica sobre los ajustes de interfaz y usabilidad necesarios, constituyendo una base sólida para validar el prototipo en futuras pruebas con usuarios en procesos reales de alfabetización.

Asimismo, en comparación con soluciones similares, como “TalkingBook” desarrollado por Schmidt et al. [6], que utiliza dispositivos de bajo costo para mejorar la alfabetización en comunidades rurales, “PixelABC” se distingue por su enfoque interactivo y su diseño multiplataforma basado en entornos de escritorio. Sin embargo, a diferencia de “TalkingBook”, que presenta una alta adaptabilidad a contextos sin conectividad, “PixelABC” aún requiere mejoras en portabilidad y rendimiento para extender su alcance a dispositivos móviles y zonas con recursos limitados.

A pesar de estas limitaciones, el desarrollo de “PixelABC” representa un avance significativo hacia la inclusión digital. Su enfoque integral puede fortalecer los procesos de enseñanza-aprendizaje en sectores marginados y ofrecer una alternativa tecnológica viable para apoyar la alfabetización. Las futuras líneas de trabajo incluyen la mejora del diseño visual, la integración de recursos multimedia didácticos, la reducción del peso del ejecutable y la migración a plataformas móviles, lo cual potenciará el impacto social y educativo del sistema. Así, “PixelABC” se perfila como una solución prometedora que puede contribuir activamente al cierre de brechas educativas en entornos vulnerables.

VII. Conclusiones

Los resultados preliminares sugieren que el prototipo “PixelABC” evidencia un potencial prometedor como herramienta educativa orientada a la alfabetización tanto tradicional como digital, especialmente en contextos de exclusión tecnológica. La aplicación integra componentes pedagógicos centrados en el juego, la escritura y la resolución de trivias, lo que podría facilitar la adquisición de habilidades fundamentales mediante una experiencia dinámica y accesible.

El diseño de módulos funcionales como “Dibuja letras y números”, “Ordena vocales y consonantes”, y “Trivia digital”, permitió evaluar de manera práctica los niveles de interacción, comprensión y retención de contenidos por parte de los usuarios. Los resultados obtenidos a través de pruebas de usabilidad respaldan la pertinencia del enfoque didáctico del prototipo, así como su capacidad para estimular la participación de los usuarios, incluso aquellos con escaso dominio tecnológico.

No obstante, la evaluación de usabilidad también puso de manifiesto áreas de mejora relevantes, como la necesidad de optimizar el rendimiento en ciertos módulos, simplificar la navegación y rediseñar elementos de la interfaz para reducir la sobrecarga visual. Estos hallazgos ofrecen una base sólida para futuras iteraciones del sistema.

La adaptación de la técnica entrevista estructurada a un entorno OSS representa un aporte metodológico significativo, al permitir evaluar de manera sistemática la experiencia del usuario en condiciones no presenciales y con recursos limitados. Esta experiencia aporta evidencia práctica sobre cómo implementar técnicas de evaluación de usabilidad en proyectos orientados a la inclusión digital.

A futuro, se proyecta una ampliación del prototipo mediante la incorporación de recursos multimedia adicionales, tutoriales interactivos y una versión ejecutable para dispositivos móviles, con el fin de ampliar su cobertura y mejorar su adaptabilidad en contextos educativos diversos. Con estas mejoras, “PixelABC” podría consolidarse como una herramienta clave en la reducción de la brecha digital y en la promoción de una educación más inclusiva y accesible.

Como aporte específico, este estudio constituye un avance preliminar en la integración de metodologías ágiles y técnicas de usabilidad en el desarrollo de prototipos de software educativo orientados a la alfabetización tradicional y digital. Aunque los resultados actuales son exploratorios, aportan

ISSN:1390-9266 e-ISSN:1390-9134 LAJC 2026
DOI: 10.5281/zenodo.17289287

evidencia práctica de su pertinencia y abren el camino para una validación futura con usuarios en procesos reales de alfabetización, lo que permitirá ampliar el alcance y la solidez de los hallazgos y fortalecer su impacto en el campo de la tecnología educativa.

REFERENCES

[1] C. Juditha and M. J. Islami, "ICT Development Strategies for Farmer Communities in Rural Papua," in *2018 International Conference on ICT for Rural Development (IC-ICTRuDev)*, IEEE, Oct. 2018, pp. 105–111. doi: 10.1109/ICICTR.2018.8706869.

[2] T.-D. Viviana and O.-M. Wilman-Santiago, "Digital illiteracy, internet use at work and economic growth at the provincial level in Ecuador," in *2022 17th Iberian Conference on Information Systems and Technologies (CISTI)*, IEEE, Jun. 2022, pp. 1–6. doi: 10.23919/CISTI54924.2022.9820485.

[3] Suhasini, B. Patil, P. Pavan Kumar, and A. Usman, "Micro donation preventing child exploitation and envisaging the plan to save humanity," in *2013 IEEE Global Humanitarian Technology Conference (GHTC)*, IEEE, Oct. 2013, pp. 279–284. doi: 10.1109/GHTC.2013.6713696.

[4] S. Khan, R. Usman, W. Haider, S. Murtaza Haider, A. Lal, and A. Q. Kohari, "E-Education Application Using Flutter: Concepts and Methods," in *2023 Global Conference on Wireless and Optical Technologies (GCWOT)*, IEEE, Jan. 2023, pp. 1–10. doi: 10.1109/GCWOT57803.2023.10064660.

[5] T. Widiyaningtyas, A. Ludfianto, D. Widhiyanuriyawan, M. A. Fritama, and W. S. Pratama, "K-Means Algorithm in Illiteracy Clustering," in *2023 8th International Conference on Electrical, Electronics and Information Engineering (ICEEIE)*, IEEE, Sep. 2023, pp. 1–5. doi: 10.1109/ICEEIE59078.2023.10334692.

[6] C. Schmidt, T. Gorman, A. Bayor, and M. S. Gary, "Impact of Low-Cost, On-demand, Information Access in a Remote Ghanaian Village," in *2011 IEEE Global Humanitarian Technology Conference*, IEEE, Oct. 2011, pp. 419–425. doi: 10.1109/GHTC.2011.88.

[7] M. J. Page *et al.*, "The PRISMA 2020 statement: An updated guideline for reporting systematic reviews," *Bmj*, vol. 372, 2021, doi: 10.1136/bmj.n71.

[8] B. A. . Kitchenham, D. . Budgen, and Pearl. Brereton, *Evidence-based software engineering and systematic reviews*. CRC Press, Taylor & Francis Group, 2020. Accessed: Aug. 15, 2024. [Online]. Available: https://www.routledge.com/Evidence-Based-Software-Engineering-and-Systematic-Reviews/Kitchenham-Budgen-Brereton/p/book/9780367575335

[9] S. A. Schwartz, "Adult illiteracy and the distortion of American culture," *EXPLORE*, vol. 20, no. 2, pp. 155–157, Mar. 2024, doi: 10.1016/j.explore.2023.12.010.

[10] A. Bataller Català and J. Ballester Roca, "Los retos de la alfabetización de las personas adultas. Creencias de docentes peruanos y propuestas metodológicas," *REDU. Revista de Docencia Universitaria*, vol. 17, no. 1, p. 153, Jun. 2019, doi: 10.4995/redu.2019.9758.

[11] J. Kalman, "BEYOND DEFINITION: CENTRAL CONCEPTS FOR UNDERSTANDING LITERACY," *International Review of Education*, vol. 54, no. 5–6, pp. 523–538, Nov. 2008, doi: 10.1007/s11159-008-9104-1.

[12] C. S. Traversini, "Autoestima e alfabetização: o que há nessa relação?," *Cadernos de Pesquisa*, vol. 39, no. 137, pp. 577–595, Aug. 2009, Accessed: Jun. 08, 2025. [Online]. Available: https://publicacoes.fcc.org.br/cp/article/view/237

[13] Tatevikarm, "Instalaré y configuraré el proyecto Scrum en Jira," Fiverr. Accessed: Jun. 08, 2025. [Online]. Available: https://www.fiverr.com/tatevikarm/install-and-configure-scrum-project-in-jira

[14] C. Lopezosa, "Entrevistas semiestructuradas con NVivo: pasos para un análisis cualitativo eficaz," in *Methodos Anuario de Métodos de Investigación en Comunicación Social, 1*, Universitat Pompeu Fabra, 2020, pp. 88–97. doi: 10.31009/methodos.2020.i01.08.

[15] M. Mijancos, "Evaluación de la usabilidad de una aplicación para alfabetización digital," Universidad Politécnica de Madrid (UPM), Madrid, 2022. Accessed: Feb. 10, 2025. [Online]. Available: https://oa.upm.es/69836/

[16] D. Hix and R. Hartson, *Developing User Interfaces: Ensuring Usability Through Product and Process*, vol. 4, no. 1. Wiley, 1994. doi: 10.1002/STVR.4370040109.

[17] J. W. Castro, "Incorporación de la Usabilidad en el Proceso de Desarrollo Opencas Source Software," Tesis Doctoral, Escuela Politécnica Superior, Universidad Autónoma de Madrid, Madrid, España, 2014.

AUTHORS

Lucrecia Llerena

Lucrecia Llerena finalizó su Doctorado en Informática y Telecomunicaciones con mención CUM LAUDE, y obtuvo también el Máster Universitario en Investigación e Innovación en Tecnologías de la Información y las Comunicaciones (I2TIC), ambos en la Escuela Politécnica Superior de la Universidad Autónoma de Madrid (UAM). Además, cursó una Maestría en Educación a Distancia y Abierta, así como su título de Ingeniera en Sistemas, en la Universidad Autónoma de Los Andes (Ecuador). Actualmente se desempeña como profesora titular en la Facultad de Ciencias de la Computación y Diseños Digitales de la Universidad Técnica Estatal de Quevedo (UTEQ), donde labora desde el año 2001. Ha dirigido varios proyectos FOCICYT y tesis de pregrado y posgrado en las universidades UTEQ y UPSE. Sus líneas de investigación se centran en la ingeniería de software, los procesos de desarrollo, la integración de la usabilidad, los sistemas inteligentes y la educación en entornos e-learning.

Steffany Loor

Steffanny del Rocío Loor Suárez es estudiante de Ingeniería de Software en la Universidad Técnica Estatal de Quevedo (UTEQ) en Quevedo, Ecuador.
Se especializa en el desarrollo de software educativo orientado a la inclusión y la mejora de procesos de aprendizaje. A lo largo de su experiencia ha aplicado metodologías ágiles, particularmente Scrum, para gestionar proyectos de manera iterativa y flexible, lo que le ha permitido garantizar entregas funcionales ajustadas a los requerimientos de los usuarios. También posee dominio en el modelado de software mediante diagramas UML, recurso que emplea para estructurar soluciones claras y eficientes.

Un eje fundamental de su trabajo es el diseño de interfaces interactivas, priorizando siempre la usabilidad y la experiencia del usuario. En este ámbito ha explorado diferentes entornos para la construcción de prototipos educativos que integran recursos lúdicos y didácticos. Su interés académico y profesional se orienta hacia la creación de soluciones tecnológicas innovadoras que respondan a diversos contextos sociales, contribuyendo a reducir brechas de alfabetismo tradicional y digital mediante propuestas que combinan accesibilidad, funcionalidad y creatividad.

ISSN:1390-9266 e-ISSN:1390-9134 LAJC 2026

AUTHORS

Nancy Rodriguez

Nancy Rodríguez obtuvo su título de Máster en Investigación e Innovación en Tecnologías de la Información y las Comunicaciones en la Universidad Autónoma de Madrid (España), donde actualmente cursa un Doctorado en Ingeniería Informática y de Telecomunicaciones. Cuenta con más de diez años de experiencia profesional en desarrollo de software y actualmente se desempeña como profesora en la Facultad de Ciencias de la Computación y Diseño Digital de la Universidad Técnica Estatal de Quevedo (UTEQ) en Ecuador. Ha impartido una variedad de asignaturas a nivel de pregrado y posgrado, particularmente en las áreas de programación, ingeniería de software, bases de datos y tecnologías web. Su trabajo académico incluye la participación en proyectos de investigación FOCICYT-UTEQ, enfocados en sistemas inteligentes, educación digital y tecnologías para el envejecimiento activo, orientadas a mejorar el bienestar de los adultos mayores. También ha sido ponente en conferencias nacionales e internacionales en el campo de la informática educativa y el aprendizaje mediado por tecnologías. Sus principales áreas de investigación incluyen los procesos de desarrollo de software, la usabilidad en sistemas de código abierto, los entornos de aprendizaje en línea, y los cursos en línea masivos y abiertos (MOOC).

L. Llerena, S. Loor, and N. Rodríguez, "Agile Development and Usability Evaluation of an Educational Application Prototype to Foster Traditional and Digital Literacy", Latin-American Journal of Computing (LAJC), vol. 13, no. 1, 2026.

ISSN:1390-9266 e-ISSN:1390-9134 LAJC 2026

Application of Convolutional Neural Networks in the Automatic Detection of Cutaneous Melanoma

ARTICLE HISTORY
Received 15 May 2025
Accepted 19 August 2025
Published 6 January 2026

José Alberto León Alarcón
Universidad Técnica de Manabí
Instituto de Lenguas Modernas
Portoviejo, Ecuador
jose.leon@utm.edu.ec
ORCID: 0009-0004-6190-0990

Roly Steeven Cedeño Menéndez
Universidad Técnica de Manabí
Instituto de Lenguas Modernas
Portoviejo, Ecuador
roly.cedeno@utm.edu.ec
ORCID: 0009-0004-1571-9410

ISSN:1390-9266 e-ISSN:1390-9134 LAJC 2026

J. León and R. Cedeño,
"Application of Convolutional Neural Networks in the Automatic Detection of Cutaneous Melanoma",
Latin-American Journal of Computing (LAJC), vol. 13, no. 1, 2026.

Aplicación de Redes Neuronales Convolucionales en la Detección Automática de Melanoma Cutáneo

Application of Convolutional Neural Networks in the Automatic Detection of Cutaneous Melanoma

José Alberto León Alarcón
Universidad Técnica de Manabí
Instituto de Lenguas Modernas
Portoviejo, Ecuador
jose.leon@utm.edu.ec

Roly Steeven Cedeño Menéndez
Universidad Técnica de Manabí
Instituto de Lenguas Modernas
Portoviejo, Ecuador
roly.cedeno@utm.edu.ec

***Resumen*—** El diagnóstico temprano del melanoma es crucial para mejorar la tasa de supervivencia, lo que ha impulsado el desarrollo de modelos de aprendizaje profundo para su detección automatizada. Esta investigación tiene como objetivo evaluar el rendimiento de una red neuronal convolucional (CNN) en la clasificación de imágenes dermoscópicas de lesiones en la piel, comparando su precisión con la de expertos en dermatología. Para lograr esto, se entrenó una CNN utilizando un conjunto de imágenes que fueron preprocesadas para mejorar la capacidad de generalización del modelo. La evaluación se llevó a cabo mediante el uso de métricas de calidad como exactitud, precisión, sensibilidad y F1-score. Además, se utilizó la curva ROC y la matriz de confusión para analizar el equilibrio entre los falsos positivos y falsos negativos en la clasificación. Los resultados mostraron que la CNN superó el rendimiento de los dermatólogos en términos de especificidad y sensibilidad, con un área bajo la curva (AUC) cercana a 1, lo que indica una gran capacidad discriminativa. La matriz de confusión reveló que la clasificación fue correcta en la mayoría de los casos, minimizando los errores de tipo I y II. En conclusión, la implementación de redes neuronales en el diagnóstico de melanoma representa una herramienta prometedora para la asistencia médica. No obstante, se identificaron oportunidades de mejora, como el ajuste de umbrales de decisión y la optimización del preprocesamiento de imágenes, lo que permitirá incrementar la precisión del modelo en aplicaciones clínicas futuras.

***Palabras Clave*—** *Redes Neuronales Convolucionales (CNN), Melanoma, Aprendizaje profundo, Preprocesamiento de Imágenes, Diagnóstico Automatizad.*

***Abstract*—** Early diagnosis of melanoma is crucial for improving survival rates, which has driven the development of deep learning models for its automated detection. This research aims to evaluate the performance of a convolutional neural network (CNN) in classifying dermoscopic images of skin lesions, comparing its accuracy with that of dermatology experts. To achieve this, a CNN was trained using a set of images that were preprocessed to improve the generalization ability of the model. The evaluation was carried out by means of quality metrics such as accuracy, precision, sensitivity, and F1-score. In addition, the ROC curve and confusion matrix were used to analyze the balance between false positives and false negatives in the classification. The results showed that the CNN outperformed dermatologists in terms of specificity and sensitivity, with an area under the curve (AUC) close to 1, indicating high discriminatory power. The confusion matrix revealed that the classification was correct in most cases, minimizing type I and type II errors. In conclusion, the implementation of neural networks in melanoma diagnosis represents a promising tool for medical care. However, opportunities for improvement were identified, such as adjusting decision thresholds and optimizing image preprocessing, which will increase the accuracy of the model in future clinical applications.

Keywords*—Convolutional Neural Networks; Melanoma; Deep Learning; Preprocessing images; Automatic diagnostic*

I. Introducción

El melanoma representa una de las variantes más peligrosas del cáncer cutáneo, identificarlo en sus etapas iniciales resulta fundamental para mejorar las probabilidades de supervivencia de quienes lo padecen. Según la Sociedad Americana Contra el Cáncer (ACS), el cáncer de piel destaca como la forma de cáncer más común entre todas. Aunque el melanoma constituye apenas el 1% de los diagnósticos de cáncer cutáneo, es el principal causante de fallecimientos relacionados con esta afección. Hasta el 2025, se estima que habrá alrededor de 104,960 nuevos casos de melanoma en los Estados Unidos (aproximadamente 60,550 en hombres y 44,410 en mujeres). Se prevé que cerca de 8,430 personas (5,470 hombres y 2,960 mujeres) fallecerán debido a este tipo de cáncer [1]

Actualmente, los dermatólogos utilizan métodos tradicionales como la inspección visual y la dermatoscopia para la evaluación de lesiones cutáneas. Sin embargo, estos enfoques están fuertemente condicionados por el nivel de experiencia del profesional y pueden verse afectados por factores subjetivos. Además, en regiones con acceso limitado a dermatólogos capacitados, la detección temprana se ve comprometida, lo que aumenta el riesgo de diagnósticos tardíos.

Por ello, en los últimos años, ha crecido el uso de diversas técnicas de análisis automatizado de imágenes por ordenador para mejorar la precisión y la reproducibilidad del diagnóstico del melanoma en comparación con los resultados clínicos obtenidos de imágenes dermatoscópicas. La inteligencia artificial (IA) se está posicionando como una tecnología con un enorme potencial en el campo de la medicina, especialmente cuando se trata de interpretar imágenes clínicas. Las CNN han revolucionado la detección de enfermedades gracias a su habilidad para extraer características clave y realizar clasificaciones con una

ISSN:1390-9266 e-ISSN:1390-9134 LAJC 2026
DOI: 10.5281/zenodo.17288648

precisión impresionante. Estas arquitecturas han demostrado un rendimiento que, en muchas ocasiones, es comparable e incluso superior al de los especialistas humanos en diversas tareas de diagnóstico por imágenes.

Un buen ejemplo de esto es la investigación de Kothapalli et al. [2] que describe las CNN como estructuras propias del aprendizaje profundo capaz de reconocer y extraer características automáticamente a partir de grandes volúmenes de datos de imágenes complejas. Esta investigación ha demostrado que las CNN son muy eficaces para identificar y clasificar lesiones cutáneas, incluido el melanoma. Al entrenar las CNN con extensos conjuntos de datos de imágenes de lesiones cutáneas, se pueden enseñar a distinguir las características que separan las lesiones benignas de las malignas.

Del mismo modo, Yalcinkaya & Erbas [3] optan por una arquitectura diferente. En este artículo, se utiliza una arquitectura de detección automática de melanomas que combina un modelo de aprendizaje profundo CNN con un enfoque basado en la lógica difusa. Este enfoque genera un mapa de correlación difusa de los píxeles de la imagen que se introduce en la red CNN. Al probarse en un extenso conjunto de datos ISIC, el modelo demostró una alta precisión, sensibilidad y especificidad en la clasificación en comparación con clasificadores que utilizaban mapas de correlación no difusos.

Este estudio presenta un enfoque fundamentado en CNN para la detección de melanoma, utilizando imágenes dermatológicas preprocesadas y optimización de hiperparámetros para mejorar la precisión del modelo. Se presenta un análisis detallado del conjunto de datos utilizado, el procesamiento de los datos, la arquitectura de la red implementada y la evaluación del desempeño del modelo frente a metodologías previamente desarrolladas.

II. METODOLOGÍA

Para el desarrollo del presente estudio, la información empleada fue extraída de la plataforma Kaggle, en concreto del conjunto de datos denominado "Melanoma", el cual está disponible de forma pública en esta comunidad en línea dirigida a científicos de datos [4]. Cabe destacar que, para este estudio, se utilizó una muestra representativa del repositorio e imágenes, seleccionada de manera aleatoria con el fin de garantizar la diversidad y representatividad de los datos en relación con el tema de estudio. La selección de esta porción del conjunto de datos se fundamentó en criterios de relevancia y en la disponibilidad de información adecuada para el entrenamiento y validación del modelo propuesto. En la Fig. 1, se puede observar una muestra aleatoria de las imágenes pertenecientes a las clases dentro del conjunto de datos.

Fig. 1. Scarlat A. (2020) Melanoma Dataset [Conjunto de datos]. Kaggle. https://www.kaggle.com/datasets/drscarlat/melanoma

Este conjunto de datos está estructurado en tres carpetas principales: **entrenamiento (train_sep)**, **validación (valid)** y **prueba (test)**. Cada una de estas carpetas se subdivide a su vez en dos subcarpetas, las cuales corresponden a las dos clases de diagnóstico: **Melanoma** (Maligno) y **NotMelanoma** (Benigno). Esta organización se encarga de clasificar las imágenes de manera clara según su función en el ciclo de desarrollo del modelo. Así, se asegura de que los datos de entrenamiento, validación y prueba estén bien categorizados y listos para ser utilizados.

Además, esta estructura jerárquica facilita una gestión eficiente de los datos. Cada subcarpeta alberga únicamente imágenes de una clase específica, lo que hace que la carga y el preprocesamiento de las imágenes sean mucho más sencillos durante la implementación del modelo. La división en tres conjuntos independientes (entrenamiento, validación y prueba) es fundamental para garantizar que el modelo pueda ser entrenado, ajustado y evaluado de manera rigurosa, minimizando el riesgo de sobreajuste y asegurando una generalización adecuada a nuevos datos. A continuación, en la TABLA I se resume la distribución de imágenes en cada conjunto:

TABLA I. CONJUNTO DE DATOS MELANOMA

Directorio	Melanoma	NotMelanoma	Total
Train_sep	1008	1008	2016
Test	336	336	672
Valid	336	336	672
Total	1680	1680	3360

En resumen, el conjunto de datos completo consta de **3360 imágenes**, con una distribución equilibrada entre las clases **Melanoma** y **NotMelanoma** en cada uno de los conjuntos (entrenamiento, validación y prueba), lo que garantiza un desarrollo y evaluación robustos del modelo. La Fig. 2 permite una apreciación más detallada de esta distribución.

ISSN:1390-9266 e-ISSN:1390-9134 LAJC 2026
DOI: 10.5281/zenodo.17288648

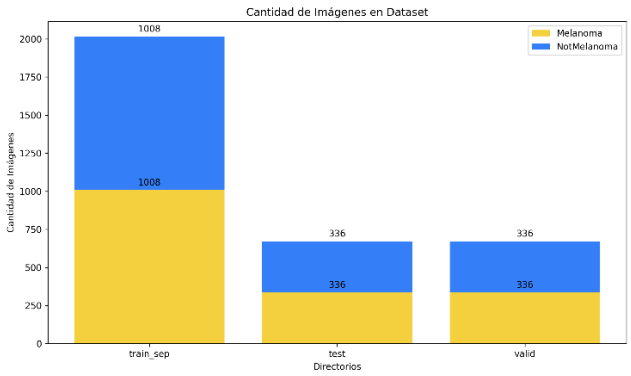

Fig. 2. Distribución del conjunto de datos Melanoma

Tras examinar el dataset, se procedió a realizar el procesamiento de datos. El procesamiento de datos representa un paso clave en la adecuación de las imágenes para el entrenamiento de algoritmos clasificatorios. Se aplicaron diversas técnicas de preprocesamiento con la finalidad de optimizar la nitidez de las imágenes y optimizar la extracción de características relevantes.

Uno de los principales desafíos en el análisis de imágenes dermatológicas es la presencia de vellos, los cuales pueden ocultar detalles clave de la piel. Para abordar este problema, se utilizó el algoritmo DullRazor, una herramienta creada específicamente para eliminar el vello en imágenes médicas. Este algoritmo se encarga de identificar y eliminar las áreas con vello, reemplazándolas con información interpolada de los píxeles vecinos. Este proceso no solo mejora la claridad de las imágenes, sino que también facilita la identificación y clasificación de características relacionadas con el melanoma, lo que contribuye a una mayor precisión en el diagnóstico asistido por computadora. [5]

Una vez implementado el algoritmo DullRazor, se puede observar una diferencia notable entre la imagen original y la imagen procesada por el algoritmo. En la sección izquierda de la Fig. 3, se observa una lesión de la piel recubierta con vellosidad, lo cual puede afectar el desempeño del algoritmo al momento del entrenamiento. Sin embargo, en la sección derecha de la Fig. 3, se puede notar que la misma lesión ya no contiene vellosidad y se puede apreciar de una mejor manera la lesión en la piel.

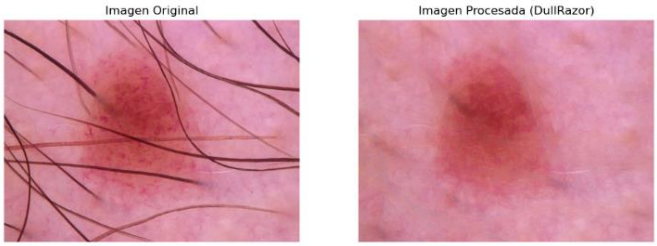

Fig. 3. Imagen procesada a partir de Scarlat A. (2020) Melanoma Dataset [Conjunto de datos]. Kaggle. https://www.kaggle.com/datasets/drscarlat/melanoma

Después de aplicar el algoritmo DullRazor, las imágenes procesadas pasaron por una serie de transformaciones para mejorar su calidad y hacer más fácil su visualización y análisis. Primero, se redimensionaron a un tamaño uniforme de 224x224 píxeles, que es un formato ideal para las arquitecturas de redes neuronales convolucionales preentrenadas, ya que requieren dimensiones específicas de entrada. Este redimensionamiento garantiza que las imágenes mantengan una relación de aspecto adecuada y que no se pierdan las características espaciales importantes durante el proceso.

Luego, se utilizó la técnica de mejora de nitidez llamada "Unsharp Masking", que consiste en crear una versión ligeramente desenfocada de la imagen original usando un filtro Gaussiano [6]. Esta versión desenfocada se mezcla con la imagen original para resaltar los bordes y detalles, lo que mejora la claridad y el enfoque de las imágenes. Este paso es crucial para destacar las características importantes que el modelo necesita identificar durante el entrenamiento y la clasificación.

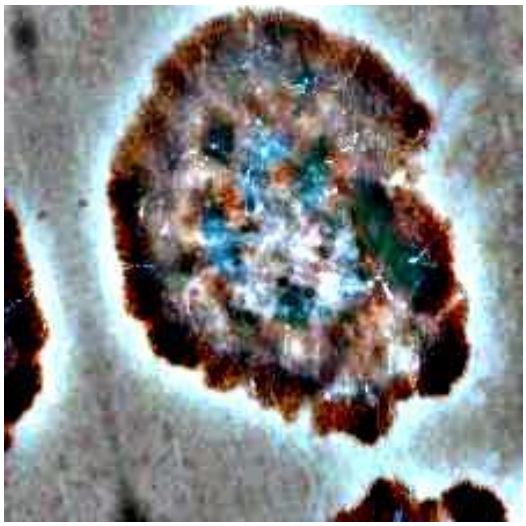

Fig. 4. Lesión con aplicación de técnica Unsharp Masking a partir de: Scarlat A. (2020) Melanoma Dataset [Conjunto de datos]. Kaggle. https://www.kaggle.com/datasets/drscarlat/melanoma

Por último, se realizó una normalización de las imágenes, transformando los valores de píxeles del rango [0, 255] al rango [0, 1]. Esta normalización es una práctica común en redes neuronales, ya que mejora la consistencia en el procesamiento de datos, mejora la eficiencia del modelo y refuerza su estabilidad tanto en la fase de entrenamiento como en la de evaluación.

A continuación, se detalla el proceso de construcción del algoritmo de red neuronal convolucional (CNN) utilizado en este estudio. Para ello, el modelo fue construido mediante la clase Sequential de Keras, proporcionando funciones de entrenamiento para el modelo [7]. A diferencia de redes neuronales preentrenadas que están optimizadas para conjuntos de datos específicos y suelen ser grandes (lo que aumenta el número de parámetros y el consumo de memoria), se optó por desarrollar una CNN personalizada. Esta decisión facilitó la adaptación de la arquitectura a las necesidades del estudio, como el tamaño de las imágenes (224x224x3, donde 3 corresponde a los canales de color) y el tipo de características que se buscaba capturar.

La CNN implementada consta de varias capas convolucionales que cuentan con 8, 16 y 32 filtros, los cuales son clave para la extracción de características. [8]. Se incluyó una capa de normalización por lotes (*Batch Normalization*) con el objetivo de regular las activaciones internas del modelo y agilizar el proceso de entrenamiento [9]. Además, se

ISSN:1390-9266 e-ISSN:1390-9134 LAJC 2026
DOI: 10.5281/zenodo.17288648

complementó la función de activación ReLU (*Rectified Linear Unit*) con el propósito de incorporar no linealidad al modelo, facilitando así el aprendizaje de patrones complejos en los datos [10]. Para reducir la dimensionalidad espacial de las salidas, se aplicó Max Pooling con un tamaño de 2x2 y *strides* de 2, una técnica de muestreo descendente que simplifica los cálculos y reduce el tamaño de los parámetros [11].

Luego de esto, la salida es enviada a una capa Flatten, utilizada para aplanar la salida de la capa anterior sin afectar al lote [12]. La red también incluye una capa Dense completamente conectada con 15 neuronas en una capa oculta, que combina las características aprendidas en las capas anteriores [13]. En la capa de salida, se implementaron 2 neuronas (una por cada clase en la clasificación binaria) junto con la función de activación Softmax, encargada de generar una distribución probabilística que facilita la asignación de clases, facilitando la clasificación binaria [14]. Toda esta arquitectura se presenta en la Fig. 5. Por último, se añadió el optimizador Adam para minimizar el error durante el entrenamiento [15]. Este optimizador según Kingma et al. Se trata de un algoritmo de optimización que emplea gradientes de primer orden para funciones objetivo de naturaleza estocástica, utilizando estimaciones adaptativas de momentos de bajo orden para ajustar dinámicamente el proceso de aprendizaje [16].

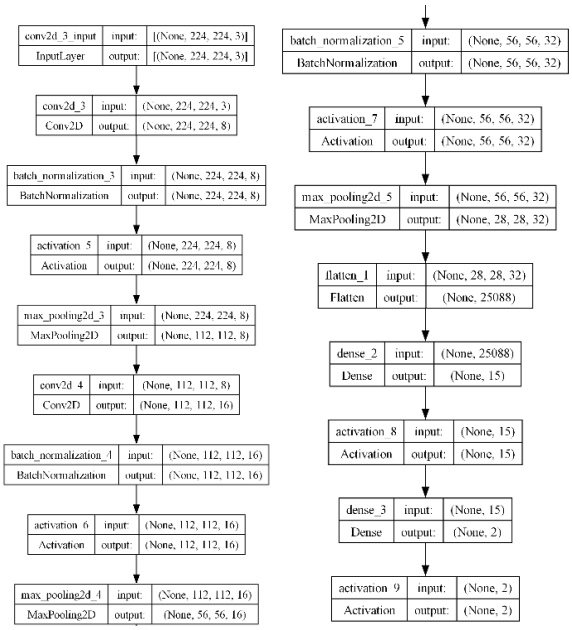

Fig. 5. Arquitectura de Red Neuronal

Para asegurar un óptimo entrenamiento, se configuraron dos callbacks: EarlyStopping, el cual detiene el entrenamiento cuando una métrica supervisada haya dejado de mejorar [17] y ReduceLROnPlateau, que disminuye la tasa de aprendizaje cuando una métrica haya dejado de mejorar [18]. Estos tienen como propósito mejorar el rendimiento de la red neuronal.

En la TABLA II, se presentan los parámetros utilizados en el callback EarlyStopping. Este ayuda a prevenir el sobreajuste y ahorra tiempos y recursos al detener el entrenamiento cuando el rendimiento del algoritmo deja de mejorar.

TABLA II. ARGUMENTOS DE EARLYSTOPPING

Argumento	**Valor**	**Función**
Monitor	val_loss	Métrica para monitorear
Patience	5	Épocas sin progreso.
Restore_best_weights	True	Restablece los pesos del modelo

En la TABLA III, se muestra los parámetros utilizados por ReduceLROnPlateau. Este ayuda a ajustar de manera dinámica la tasa de aprendizaje permitiendo que el modelo converja de mejor manera cuando el progreso sea detenido.

TABLA III. ARGUMENTOS DE REDUCELRONPLATEAU

Argumento	**Valor**	**Función**
Monitor	val_loss	Métrica para monitorear
Factor	0.1	Disminuye la velocidad de aprendizaje (nueva_lr = lr * factor)
Patience	2	Número de épocas sin mejoras
Min_lr	0.0001	Tasa de aprendizaje mínima

Asimismo, en la Tabla IV se exhiben los hiperparámetros manejados para entrenar la red neuronal, incluyendo detalles sobre las configuraciones específicas empleadas en el proceso de aprendizaje, tales como la tasa de aprendizaje, el número de épocas, el tamaño del lote y otros parámetros relevantes que influenciaron el desempeño del modelo.

TABLA IV. HIPERPARÁMETROS DE LA RED NEURONAL

Parámetros	**Valores**
Tamaño de la muestra	2016 (1008 malignas y 1008 benignas)
Épocas	40
Épocas recorridas	21
Tasa de aprendizaje	0.01
Callbacks	EarlyStopping, ReduceLROnPlateau
Tamaño de lote	24

El proceso de entrenamiento se realizó en la plataforma Google Colab Pro, aprovechando sus capacidades avanzadas en términos de recursos computacionales. Esta plataforma proporciona acceso a aceleradores de hardware, como las unidades de procesamiento tensorial (TPU). Se trata de circuitos integrados personalizados, diseñados especialmente para optimizar y acelerar el proceso de entrenamiento de modelos de aprendizaje profundo. [19].

Para el desarrollo del proyecto, se optó por utilizar Python como lenguaje de programación, ya que cuenta con una gran cantidad de bibliotecas y herramientas enfocadas en Deep Learning. Python destaca por su sintaxis intuitiva y su capacidad para implementar algoritmos de aprendizaje profundo de manera eficiente. En particular, se utilizaron las

bibliotecas TensorFlow y Keras para la construcción, entrenamiento y evaluación del modelo.

III. Resultados

Esta sección expone los resultados derivados del proceso de entrenamiento y validación de la CNN. Para medir el desempeño del modelo, se utilizaron métricas de calidad ampliamente empleadas en el área del aprendizaje profundo y la detección de enfermedades, como la exactitud, precisión, sensibilidad y el puntaje F1.

Se llevó a cabo un análisis de los resultados mediante la curva ROC (Característica Operativa del Receptor) y su área bajo la curva (AUC-ROC). Esto facilita la evaluación del rendimiento del modelo en la diferenciación entre imágenes de melanomas malignos y benignos. Finalmente, se muestra la matriz de confusión, la cual nos da una perspectiva detallada sobre los aciertos y errores del modelo, lo que facilita la interpretación de su rendimiento en cuanto a clasificación.

Como se mencionó previamente, se utilizó la curva ROC para evaluar el desempeño del modelo de red neuronal. El resultado fue un impresionante 0.9955 en el valor AUC-ROC, lo que indica una predicción muy precisa. Esta colección de datos de prueba sugiere que el modelo tiene un rendimiento superior en la tarea de clasificación binaria. Un valor cercano a 1 significa que el modelo distingue eficazmente entre las clases positivas y negativas, lo cual muestra una alta sensibilidad y una baja tasa de falsos positivos. En la práctica, un AUC-ROC de 0.99 significa que el modelo tiene un 99% de probabilidad de clasificar correctamente un ejemplo positivo frente a uno negativo al azar; es decir, identifica con claridad de lesiones que pertenecen al melanoma y de lesiones que no pertenecen al melanoma. Por lo tanto, la Fig. 6 muestra el área bajo la curva ROC donde ilustra de forma gráfica el rendimiento del modelo en la predicción.

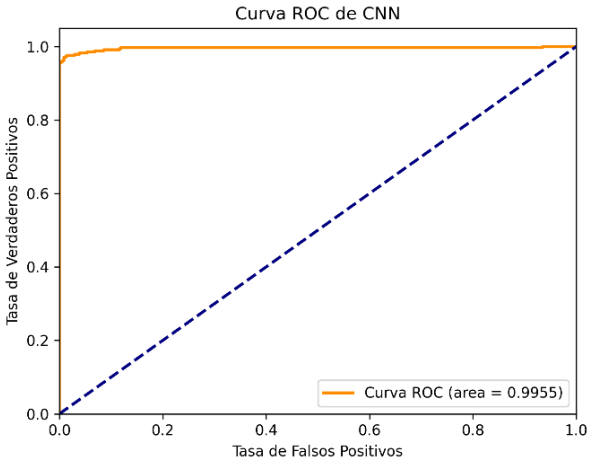

Fig. 6. Curva ROC

En la Fig. 7, se muestra la matriz de confusión, un instrumento primordial para valorar la calidad de la clasificación. Esta permite observar la cantidad de aciertos y errores cometidos por el modelo, distinguiendo entre los aciertos y los errores cometidos al momento de la clasificación.

En particular, la matriz de confusión presenta la distribución de los verdaderos positivos (VP) y verdaderos negativos (VN); es decir, los casos donde el modelo identificó correctamente las lesiones malignas y benignas, respectivamente. Asimismo, permite identificar los errores de tipo I (falsos positivos, FP), que ocurren cuando una lesión benigna es clasificada erróneamente como maligna, y los errores de tipo II (falsos negativos, FN), en los que una lesión maligna es incorrectamente clasificada como benigna.

El análisis de esta matriz resulta clave para evaluar el impacto de los errores en un contexto clínico, dado que un falso negativo podría retrasar el diagnóstico de un melanoma, mientras que un falso positivo podría generar alarmas innecesarias y procedimientos médicos adicionales.

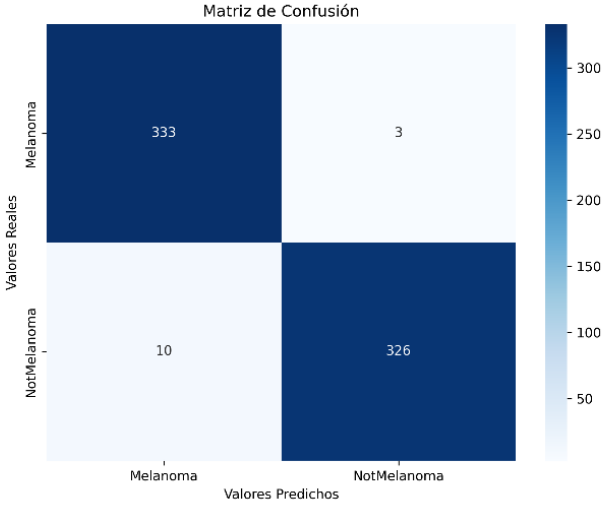

Fig. 7. Matriz de confusión

En la matriz de confusión (Fig. 7), el cuadrante superior izquierdo representa la cantidad de verdaderos positivos (VP); en otras palabras, aquellos casos en los que el modelo clasificó correctamente una lesión como melanoma, sumando un total de 333 muestras. De manera similar, el cuadrante inferior derecho refleja los verdaderos negativos (VN), indicando que 326 casos fueron identificados correctamente como lesiones benignas. Esta distribución evidencia que los valores más altos se concentran en la diagonal principal, lo que sugiere un alto nivel de precisión en la clasificación tanto de melanomas malignos como benignos.

Por otro lado, los errores de tipo I, representados por los falsos positivos (FP), corresponden a aquellas muestras en las que el algoritmo clasificó incorrectamente una lesión benigna como melanoma. En este caso, se identificaron 3 instancias en las que el modelo generó una alerta errónea de la enfermedad. Asimismo, los errores de tipo II, que corresponden a los falsos negativos (FN), ocurrieron en 10 muestras en las que el modelo no logró detectar correctamente el melanoma, y clasificó erróneamente una lesión maligna como benigna. Estos errores tienen un impacto clínico significativo, ya que un falso negativo podría retrasar el diagnóstico y tratamiento oportuno del paciente.

Por último, en base a las predicciones realizadas por el algoritmo y los valores resultantes de la matriz de confusión se obtuvieron diferentes métricas que evalúan el rendimiento de la arquitectura. En la Tabla V, se presentan los valores obtenidos al realizar las predicciones en función del conjunto de datos de testeo.

ISSN:1390-9266 e-ISSN:1390-9134 LAJC 2026
DOI: 10.5281/zenodo.17288648

TABLA V. MÉTRICAS DE EVALUACIÓN

Métrica	Valor
Exactitud	98.07%
Precisión	98.09%
Sensibilidad	98.07%
Puntaje F1	98.07%
AUC	99.55%

En general, los resultados obtenidos reflejan un desempeño sobresaliente del modelo. La elevada precisión, sensibilidad, puntuación F1 y una curva AUC cercana a 1 indican que el modelo no solo es eficaz en la clasificación global, sino que también demuestra solidez en la identificación de casos positivos y en la reducción de errores. Este rendimiento sugiere que el modelo está bien ajustado y es altamente eficiente para la tarea específica de clasificación para la que fue entrenado. Sin embargo, a pesar de los resultados alentadores, todavía hay oportunidades de mejora en la capacidad predictiva y en la evaluación del modelo. Como parte de las limitaciones de este estudio, se aconseja implementar estrategias que optimicen su rendimiento, como ajustar los umbrales de decisión para lograr un mejor equilibrio entre las métricas, aplicar técnicas de data augmentation que enriquezcan la diversidad del conjunto de entrenamiento, y explorar otras metodologías que podrían ayudar a mejorar la capacidad del algoritmo, permitiendo así una clasificación más precisa y confiable.

IV. Discusión

El presente estudio ha demostrado que el modelo apoyado en redes neuronales profundas es altamente eficiente para la clasificación de imágenes dermatológicas, lo que permite la detección precisa del melanoma. La evaluación del modelo a través de métricas de desempeño como exactitud, precisión, sensibilidad (recall), especificidad y F1-score, junto con un área bajo la curva ROC (AUC) indica que la arquitectura implementada logra una capacidad discriminativa robusta entre lesiones malignas y benignas.

Los hallazgos obtenidos guardan coherencia con investigaciones anteriores que han utilizado enfoques de aprendizaje profundo para la identificación de melanoma, donde modelos basados en redes convolucionales han mostrado desempeños comparables a los de especialistas en dermatología. Por ejemplo, en la investigación de Maron et al [20], se desarrolló y entrenó una CNN empleando un total de 11,444 imágenes dermoscópicas, y se evaluó posteriormente su desempeño en un grupo de prueba conformado por 6,390 imágenes con verificación histopatológica. En este estudio, se evidenció que la CNN alcanzó una especificidad del 91.3% en la clasificación binaria de lesiones benignas y malignas, lo cual superó significativamente la especificidad alcanzada por los dermatólogos, correspondiente al 59.8%. Asimismo, la sensibilidad obtenida por la CNN fue similar a la de los especialistas, mientras que sus valores de especificidad y sensibilidad fueron superiores en las pruebas comparativas.

En términos comparables, en la investigación realizada por Haenssle et al. [21], se aborda una comparación similar. Se utilizó una red neuronal convolucional (CNN) aprobada para el mercado europeo como dispositivo médico (Moleanalyzer Pro, FotoFinder Systems) para clasificar imágenes dermoscópicas de lesiones cutáneas. La precisión de la CNN se comparó con la de 96 dermatólogos, quienes evaluaron las mismas imágenes bajo condiciones menos artificiales, incluyendo imágenes clínicas y dermoscópicas, junto con información textual del caso. Los resultados mostraron que la CNN logró una sensibilidad del 95.0% y una especificidad del 76.7%, con un área bajo la curva (AUC) de 0.918. En comparación, los 96 dermatólogos alcanzaron una sensibilidad del 89.0% y una especificidad del 80.7 en su primera evaluación (nivel I), que mejoró significativamente al 94.1% con información adicional (nivel II), mientras que la especificidad se mantuvo prácticamente igual en 80.4%.

Por lo tanto, en la presente investigación el análisis detallado de la matriz de confusión revela la presencia de errores de tipo I (falsos positivos), en los que el modelo clasifica erróneamente lesiones benignas como malignas, y errores de tipo II (falsos negativos), donde casos de melanoma no son correctamente identificados. La existencia de estos errores puede traer consigo implicaciones clínicas significativas. Un falso positivo podría llevar a realizar procedimientos invasivos innecesarios, mientras que un falso negativo podría retrasar un diagnóstico oportuno y, por lo tanto, el tratamiento del paciente.

Uno de los principales desafíos que se ha identificado en este estudio es la posible existencia de sesgos en los datos de entrenamiento. La variabilidad en la calidad de las imágenes, el desequilibrio en la distribución de clases y la variedad en los tipos de lesiones pueden afectar la capacidad del modelo para generalizar en entornos clínicos reales. Es crucial que en futuras investigaciones se aborden estos aspectos, integrando conjuntos de datos más amplios y diversos que incluyan imágenes de diferentes fuentes y poblaciones.

Con el fin de mejorar la robustez del modelo y reducir los errores de clasificación, se recomienda considerar estrategias como la optimización de los umbrales de decisión, la aplicación de técnicas de aumento de datos para diversificar el conjunto de entrenamiento, y la implementación de enfoques de aprendizaje por transferencia, mediante el uso de arquitecturas preentrenadas en grandes conjuntos de datos médicos. Además, combinar técnicas de interpretación de modelos, como Grad-CAM, podría ofrecer una mejor comprensión de las áreas de interés que la red neuronal ha señalado, lo que facilitaría la validación por parte de expertos clínicos.

Esto se ve reflejado en la investigación dada por Salma y Eltrass [22] donde se propone un método basado en filtrado morfológico para la eliminación de vello en imágenes dermatológicas, compuesto por dos fases principales. Primero, se convierte la imagen a escala de grises mediante una transformación ponderada del espacio de color RGB. Luego, el contorno del vello se detecta mediante la transformación morfológica de sombrero negro. Posteriormente, se emplea el Método de Marcha Rápida (FMM) para aplicar una función de inpainting y generar una máscara, donde los píxeles por debajo de un umbral se asignan a 0 y el resto a 1. Tras el preprocesamiento, se implementa una estrategia de aumento de datos mediante rotaciones de 0°, 90°, 180° y 270°, generando cuatro nuevas imágenes por cada original. Este procedimiento permite expandir el conjunto de datos y mitigar la escasez de imágenes etiquetadas.

Otra línea de mejora es la integración de metodologías híbridos que integren algoritmos de aprendizaje profundo con técnicas tradicionales de procesamiento de imágenes y métodos basados en reglas clínicas, lo que podría mejorar la interpretabilidad del sistema y aumentar su confiabilidad en la práctica médica. Además, la implementación de técnicas de calibración probabilística puede contribuir a disminuir la inseguridad en la toma de decisiones, y proporcionar predicciones más confiables para los especialistas en dermatología.

Siendo así, los hallazgos de este estudio refuerzan el potencial del aprendizaje profundo en la detección temprana del melanoma, mostrando métricas de rendimiento comparables con métodos convencionales de diagnóstico. No obstante, la presencia de errores en la clasificación y las limitaciones asociadas a la calidad y diversidad de los datos sugieren la necesidad de futuras optimizaciones para garantizar su aplicabilidad en contextos clínicos reales. La integración de técnicas avanzadas de preprocesamiento, ajuste de hiperparámetros y validación en cohortes independientes permitirá fortalecer la eficacia del modelo, y facilitar su adopción como herramienta de apoyo en el diagnóstico dermatológico.

V. CONCLUSIONES

El presente estudio ha evaluado la capacidad de un algoritmo de redes neuronales del tipo convolucional para la detección y clasificación de lesiones cutáneas, y se obtuvieron resultados altamente satisfactorios en términos de exactitud, precisión, sensibilidad y puntaje F1.

El análisis de la matriz de confusión y las métricas de calidad revelan que el modelo logra un equilibrio adecuado entre la identificación de casos positivos y negativos, con una tasa de falsos positivos y falsos negativos relativamente baja. De esta manera, la adopción del valor del área bajo la curva ROC (AUC) como indicador de rendimiento respalda la capacidad del modelo para distinguir eficazmente entre lesiones benignas y malignas.

A pesar de los resultados obtenidos, se identifican oportunidades de optimización. La implementación de técnicas de preprocesamiento, como la eliminación de artefactos en las imágenes y el uso de aumento de datos, ha contribuido significativamente a la mejora del modelo. No obstante, futuros trabajos podrían explorar la integración de técnicas más avanzadas de segmentación de imágenes, ajustes en los umbrales de decisión y enfoques de aprendizaje activo para mejorar la robustez del modelo en distintos conjuntos de datos.

Aunque los hallazgos de este estudio son prometedores, la adopción del modelo en entornos clínicos requiere validaciones adicionales. Es primordial evaluar su desempeño en escenarios del mundo real, considerando la variabilidad en la calidad de las imágenes y la diversidad de poblaciones de pacientes. La colaboración con especialistas en dermatología será clave para garantizar que el modelo no solo sea preciso, sino también interpretable y útil en la práctica médica.

REFERENCIAS

[1] Sociedad Americana Contra el Cáncer, "Estadísticas importantes sobre el cáncer de piel tipo melanoma," Estadísticas importantes sobre el cáncer de piel tipo melanoma. Accessed: Feb. 07, 2025. [Online]. Available: https://www.cancer.org/es/cancer/tipos/cancer-de-piel-tipo-melanoma/acerca/estadisticas-clave.html

[2] S. P. Kothapalli, P. S. H. Priya, V. S. Reddy, B. Lahya, and P. Ragam, "Melanoma Skin Cancer Detection using SVM and CNN," *EAI Endorsed Trans Pervasive Health Technol*, vol. 9, no. 1, May 2023, doi: 10.4108/eetpht.9.4340.

[3] F. Yalcinkaya and A. Erbas, "Convolutional neural network and fuzzy logic-based hybrid melanoma diagnosis system," *Elektronika ir Elektrotechnika*, vol. 27, no. 2, pp. 69–77, 2021, doi: 10.5755/j02.eie.28843.

[4] Alexander Scarlat MD, "Dataset Melanoma," Kaggle. Accessed: Feb. 20, 2025. [Online]. Available: https://www.kaggle.com/datasets/drscarlat/melanoma

[5] T. Lee, V. Ng, R. Gallagher, A. Coldman, and D. McLean, "Dullrazor®: A software approach to hair removal from images," *Comput Biol Med*, vol. 27, no. 6, pp. 533–543, 1997, doi: https://doi.org/10.1016/S0010-4825(97)00020-6.

[6] C. H. Chen, "Adaptive Image Filtering," *Proceedings - IEEE Computer Society Conference on Pattern Recognition and Image Processing*, pp. 19–31, Jan. 2000, doi: 10.1016/B978-012077790-7/50005-9.

[7] Keras, "The Sequential class." Accessed: Feb. 20, 2025. [Online]. Available: https://keras.io/api/models/sequential/

[8] Keras, "Conv2D layer." Accessed: Feb. 20, 2025. [Online]. Available: https://keras.io/api/layers/convolution_layers/convolution2d/

[9] Keras, "BatchNormalization layer." Accessed: Feb. 20, 2025. [Online]. Available: https://keras.io/api/layers/normalization_layers/batch_normalization/

[10] Keras, "Layer activation functions - ReLU ." Accessed: Feb. 20, 2025. [Online]. Available: https://keras.io/api/layers/activations/#relu-function

[11] Keras, "MaxPooling2D layer." Accessed: Feb. 20, 2025. [Online]. Available: https://keras.io/api/layers/pooling_layers/max_pooling2d/

[12] Keras, "Flatten layer." Accessed: Feb. 20, 2025. [Online]. Available: https://keras.io/api/layers/reshaping_layers/flatten/

[13] Keras, "Dense layer." Accessed: Feb. 20, 2025. [Online]. Available: https://keras.io/api/layers/core_layers/dense/

[14] Keras, "Layer activation functions | Softmax." Accessed: Feb. 20, 2025. [Online]. Available: https://keras.io/api/layers/activations/#softmax-function

[15] Keras, "Adam." Accessed: May 13, 2025. [Online]. Available: https://keras.io/api/optimizers/adam/

[16] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," Dec. 2014, [Online]. Available: http://arxiv.org/abs/1412.6980

[17] Keras, "EarlyStopping." Accessed: Feb. 20, 2025. [Online]. Available: https://keras.io/api/callbacks/early_stopping/

[18] Keras, "ReduceLROnPlateau." Accessed: Feb. 20, 2025. [Online]. Available: https://keras.io/api/callbacks/reduce_lr_on_plateau/

[19] Google Cloud, "Introducción a Cloud TPU." Accessed: Feb. 20, 2025. [Online]. Available: https://cloud.google.com/tpu/docs/intro-to-tpu?hl=es-419

[20] R. C. Maron *et al.*, "Systematic outperformance of 112 dermatologists in multiclass skin cancer image classification by convolutional neural networks," *Eur J Cancer*, vol. 119, pp. 57–65, Sep. 2019, doi: 10.1016/j.ejca.2019.06.013.

[21] H. A. Haenssle *et al.*, "Man against machine reloaded: performance of a market-approved convolutional neural network in classifying a broad spectrum of skin lesions in comparison with 96 dermatologists working under less artificial conditions," *Annals of Oncology*, vol. 31, no. 1, pp. 137–143, Jan. 2020, doi: 10.1016/j.annonc.2019.10.013.

[22] W. Salma and A. S. Eltrass, "Automated deep learning approach for classification of malignant melanoma and benign skin lesions," *Multimed Tools Appl*, vol. 81, no. 22, pp. 32643–32660, Sep. 2022, doi: 10.1007/s11042-022-13081-x.

ISSN:1390-9266 e-ISSN:1390-9134 LAJC 2026
DOI: 10.5281/zenodo.17288648

AUTHORS

José Alberto León Alarcón

José León Alarcón es un profesional especializado en Ciencia de Datos, posee un máster en Sistemas de Información con mención en Data Science por la Pontificia Universidad Católica del Ecuador (PUCE Quito). Su formación académica se complementa con una sólida experiencia en el ámbito de la inteligencia artificial, especialmente en el aprendizaje automático (machine learning) y el aprendizaje profundo (deep learning). A lo largo de su trayectoria profesional, se ha enfocado en el análisis de imágenes médicas, contribuyendo al desarrollo de modelos capaces de apoyar el diagnóstico clínico mediante técnicas avanzadas de procesamiento de imágenes.

Además, ha trabajado en la extracción y análisis de información a partir de datos complejos, aplicando metodologías estadísticas y herramientas computacionales modernas. Sus áreas de interés incluyen la inteligencia artificial, el análisis predictivo y el desarrollo de soluciones innovadoras que permitan transformar grandes volúmenes de datos en conocimiento útil para la toma de decisiones. Se caracteriza por su compromiso con la investigación aplicada y el desarrollo tecnológico orientado a resolver problemas reales.

Roly Steeven Cedeño Menéndez

Ingeniero en Sistemas de Información por la Universidad Técnica de Manabí y Magíster en Sistemas de Información con mención en Data Science por la Pontificia Universidad Católica del Ecuador. Su formación académica y experiencia profesional se enfocan en el análisis de datos, el aprendizaje automático y la aplicación de técnicas avanzadas para la extracción de conocimiento a partir de grandes volúmenes de información. Actualmente se desempeña como técnico docente en la Universidad Técnica de Manabí y cuenta con un año de experiencia adicional como docente en modalidad online.

Ha participado en proyectos de investigación vinculados a la ciencia de datos, destacando su trabajo de tesis de posgrado titulado "Análisis de sentimientos utilizando la red social X (Twitter) para medir el nivel de aceptación del nuevo presidente del Ecuador, Daniel Noboa (noviembre 2023 - abril 2024)". También cuenta con dos artículos académicos publicados. Sus áreas de interés incluyen la inteligencia artificial, la minería de datos y el desarrollo de soluciones basadas en ciencia de datos. Sus objetivos profesionales actuales se centran en mejorar continuamente como docente y consolidarse como investigador en el área, contribuyendo con nuevas publicaciones científicas.

ISSN:1390-9266 e-ISSN:1390-9134 LAJC 2026

LATIN-AMERICAN JOURNAL OF COMPUTING (LAJC), Vol XIII, Issue 1, January 2026

LATIN-AMERICAN JOURNAL OF COMPUTING

Published by

Escuela Politécnica Nacional
Facultad de Ingeniería de Sistemas

Quito-Ecuador

https://lajc.epn.edu.ec/
lajc@epn.edu.ec

January 2026

Vol XIII, Issue 1, January 2026

LATIN-AMERICAN
JOURNAL OF
COMPUTING