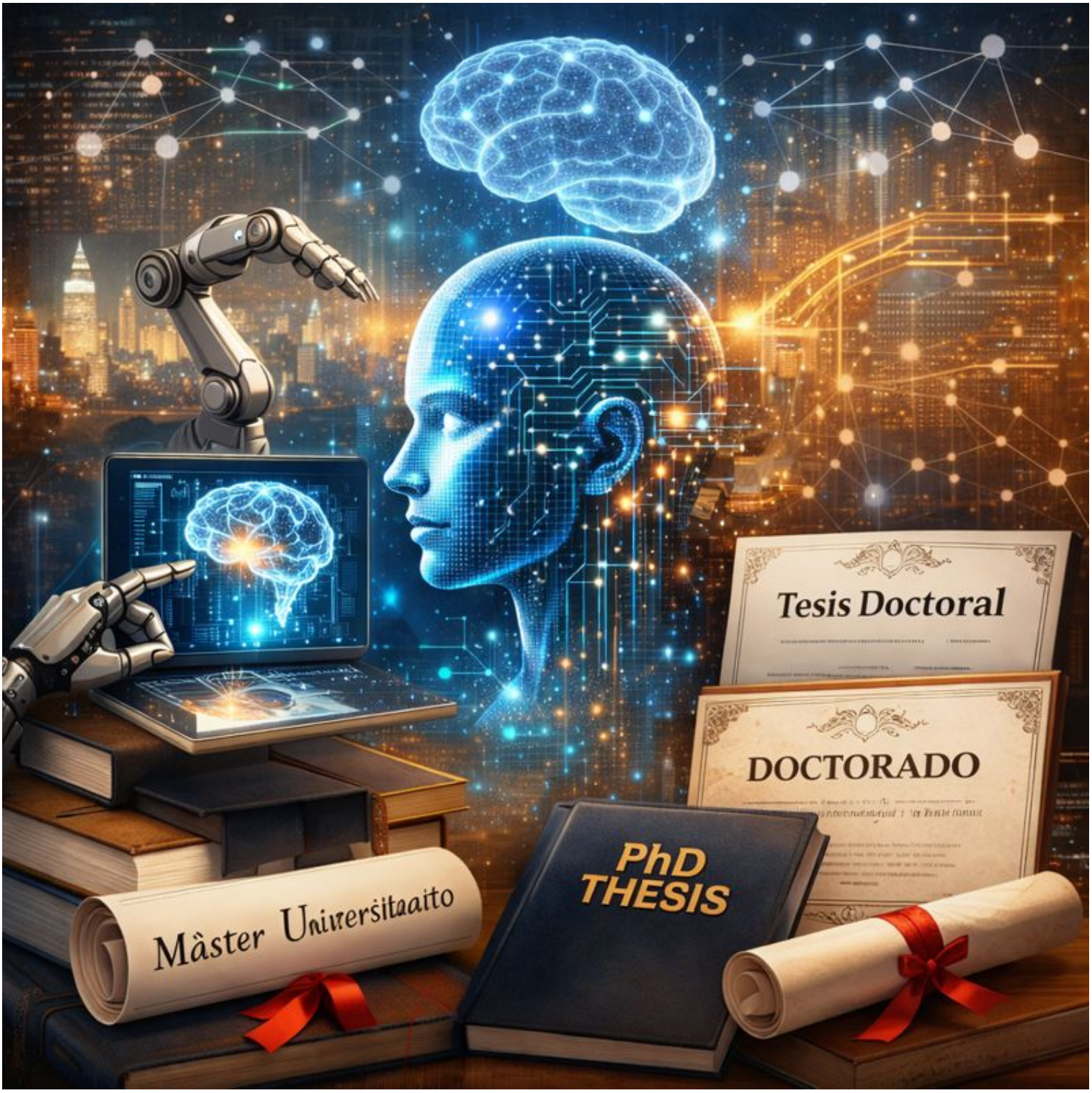


SIMDIA'2025

Proceedings of the Third Ibero-American Symposium of Master and Doctorate in Artificial Intelligence

Edited by: IBERAMIA
(Sociedad Iberoamericana de Inteligencia Artificial)
ISBN:978-84-09-82038-2

iberamia
Sociedad Iberoamericana de Inteligencia Artificial

SIMDIA'2025

13-14, November, 2025

Proceedings of the

Third Ibero-American Symposium of

Master and Doctorate in Artificial Intelligence

Marcela Quiroz Castellanos.
Universidad Veracruzana. México.

Néstor Darío Duque Méndez.
Universidad Nacional de Colombia (Sede Manizales). Colombia.

Edited by: IBERAMIA

ISBN:978-84-09-82038-2

© IBERAMIA 2026 ***https://simdia.iberamia.org/***

ACTAS SIMDIA'2025

Tercer Simposio Iberoamericano de Maestría y Doctorado en Inteligencia Artificial

Terceiro Simpósio Ibero-Americano de Mestrado e Doutorado em Inteligência Artificial

Third Ibero-American Symposium of Master and Doctorate in Artificial Intelligence

9-10, November, 2023 (on line)

Organized by:

Marcela Quiroz Castellanos. Universidad Veracruzana. México.
Néstor Darío Duque Méndez. Universidad Nacional de Colombia (Sede Manizales). Colombia.

Edited by: IBERAMIA (Spain)
ISBN: 978-84-09-82038-2

© IBERAMIA 2026 https://simdia.iberamia.org/
https://www.iberamia.org/
iberamia@iberamia.org

Cover drawing designed by ChatGPT

Preface

In the context of the XIX edition of the Ibero-American Conference on Artificial Intelligence (IBERAMIA'2025),, the *Third Ibero-American Symposium of Master and Doctorate in Artificial Intelligence (SIMDIA'2025)* was held as a preliminary activity, which aim to be a space for socialization, presentation and discussion of academic master's and doctorate works on different topics around Artificial Intelligence (AI). This third edition follows the first edition SIMDIA'2021, and SIMDIA'2023 held within the framework of (IBERAMIA'2022 and IBERAMIA'2024).

The Symposium consisted of the presentation of a set of research papers carried out by postgraduate students (Master's or Doctorate) with a significant contribution to knowledge or presenting innovative experiences in the different areas of AI. The papers submitted went through a peer review and those accepted in the evaluation process were presented. A scientific committee was formed made up of researchers from the different areas of AI and with the support of the different societies that make up IBERAMIA.

The best qualified papers were invited to publish an extended version in the 'Inteligencia Artificial' and 'Inteletica' journals (both edited by IBERAMIA) and granted for the 19th Ibero-American Conference on Artificial Intelligence (Iberamia'2026) to present their works.

These proceedings contain the set of papers accepted and presented at the Third Ibero-American Symposium on Masters and Doctorates in Artificial Intelligence, held online on November 13-14, 2025.

Prefacio

En el marco de la XIX edición de la Conferencia Iberoamericana de Inteligencia Artificial (IBERAMIA'2026), se llevó a cabo como actividad preliminar el *Tercer Simposio Iberoamericano de Maestría y Doctorado en Inteligencia Artificial (SIMDIA'2025)*, que tuvo como objetivo ser un espacio de socialización, presentación y discusión de trabajos académicos de maestría y doctorado sobre diferentes temas en torno a la Inteligencia Artificial (IA). Esta tercera edición sigue a las ediciones SIMDIA'2021 y SIMDIA'2023 celebradas en el marco de IBERAMIA'2022 e IBERAMIA'2024.

El Simposio consistió en la presentación de un conjunto de trabajos de investigación realizados por estudiantes de posgrado (Máster o Doctorado) con una contribución significativa al conocimiento o presentando experiencias innovadoras en las diferentes áreas de la IA. Los trabajos presentados pasaron por una revisión por pares y se presentaron los aceptados en el proceso de evaluación. Se formó un comité científico formado por investigadores de las distintas áreas de la IA y con el apoyo de las distintas sociedades que forman IBERAMIA.

Los trabajos mejor calificados fueron invitados a publicar una versión extendida en las revistas 'Inteligencia Artificial' e 'Intelética' (ambas editadas por IBERAMIA) y becados para la XIX Conferencia Iberoamericana de Inteligencia Artificial (Iberamia'2026) para presentar sus trabajos.

Estas actas contienen el conjunto de trabajos aceptados y presentados en el Tercer Simposio Ibero-Americano de Maestría y Doctorado en Inteligencia Artificial (SIMDIA'2025), realizado en línea el 13- y 14 de noviembre de 2025.

Prefácio

No âmbito da XIX edição da Conferência Ibero-Americana de Inteligência Artificial (IBERAMIA'2024), realizou-se como atividade preliminar o Terceiro Simpósio Ibero-Americano de Mestrado e Doutoramento em Inteligência Artificial (SIMDIA'2025), que pretende ser um espaço de socialização, apresentação e discussão de trabalhos acadêmicos de mestrado e doutorado sobre diferentes temas em torno da Inteligência Artificial (IA). Esta terceira edição segue a edições SIMDIA'2021 e SIMDIA'2023, realizadas no âmbito da (IBERAMIA'2022 e IBERAMIA'2024).

O Simpósio consistiu na apresentação de um conjunto de trabalhos de investigação realizados por alunos de pós-graduação (Mestrado ou Doutoramento) com um contributo significativo para o conhecimento ou apresentando experiências inovadoras nas diferentes áreas da IA. Os trabalhos submetidos passaram por uma revisão por pares e foram apresentados os aceitos no processo de avaliação. Foi formado um comitê científico formado por pesquisadores das diferentes áreas da IA e com o apoio das diferentes sociedades que compõem a IBERAMIA.

Os trabalhos mais qualificados foram convidados a publicar uma versão estendida na revista 'Inteligencia Artificial' e 'Intelética' (editadas pela IBERAMIA) e premiados para a 19ª Conferência Iberoamericana de Inteligência Artificial (Iberamia'2026) para apresentar seus trabalhos.

Estas atas contêm o conjunto de trabalhos aceitos e apresentados no III Simpósio Ibero-Americano de Mestrado e Doutorado em Inteligência Artificial, realizado online em 13 e 14 de novembro de 2025.

Relación de Trabajos Presentados

Speech Restoration from EEG Brain Bio Signals Based on a Novel Deep Learning Approach
Sbaih, Asma, Jorge Garcia-Gutierrez, David

Métrica Pixel Erosion Score y Método CBRISE: Evaluación de la Interpretabilidad en Visión por Computadora
Stanchi, Oscar; Ronchetti, Franco; Quiroga, Facundo

Desarrollo de recursos y modelos para la traducción de Lengua de Señas Argentina
Dal Bianco, Pedro Alejandro; Quiroga, Facundo; Ronchetti, Franco

Fuzzy clustering based on Supervised Classification Pérez Sánchez, Ismay 19 Esquema de Aprendizaje Federado para la Preservación de la Privacidad en el Reconocimiento de Actividades Humanas
Miranda Rucabado, Oscar Alexis

Desarrollo de metaheurísticas aplicadas a genómica estructural y funcional
Rojas, Matías; Carballido, Jessica; Vidal, Pablo 20

Valoración de viviendas urbanas en Argentina mediante técnicas de machine learning e interpretabilidad
Gutierrez, Emiliano ; Lanzarini, Laura; Cecchini, Rocío; Delbianco, Fernando

Forecasting Spatio-Temporal Time Series with Missing Values Using Graph Neural Networks
Martínez Ruiz, Armando; Gomez-Gil, Pilar; FonsecaDelgado, Rigoberto

Hybrid Feature extraction in thermographic images for the detection of abnormal structures in breast tissues
Aburto, Yareli; Gómez Gil, María del Pilar; Altamirano Robles, Leopoldo

Robotic Arm-Based Assisted Feeding System for Patients with Upper Limb Limitations
Leal, Luis; Gutiérrez Giles, Alejandro; López Gutiérrez, Jesús Ricardo

Metodología para la adopción de herramientas de IA en el desarrollo de software
Reyes, Dario; Sanchéz-Torres, Jenny Marcela; Rueda Cáceres, Iván Mauricio

Gestión de la función de pérdida integral y adaptativa en redes neuronales
Avalos Escalante, Karen

Intereses académicos en estudiantes de licenciatura: un análisis con minería de datos
Montejo-Collado, Fatima; Hernández-Torruco, José; Chávez-Bosquez, Oscar

Hand_IA, sistema inteligente para el control de prótesis de mano basado en señales de Electromiografía (EMG)
Padilla, Daniela; Villaseñor, Luis; Gutiérrez , Alejandro

Diagnóstico de Cardiopatía Isquémica Crónica en DMT2 mediante TabTransformer: Contraste de Variables Predictivas con Factores de Riesgo Tradicionales.
Flores-Custodio, Orlando; Pancardo-García, Pablo

Un Modelo Híbrido para la Predicción de Resistencia Antimicrobiana mediante Redes Neuronales Gráficas y PLN
Saldaña Pacheco, Sergio Antonio; Pancardo García, Pablo

Optimización de modelos para identificar arritmia cardiaca con aprendizaje automático
Arias García, SANTIAGO

Neuroevolution of Spiking Neuron Networks
Lopez-Herrera, Carlos-Alberto; Acosta-Mesa, Héctor-Gabriel; Mezura-Montes, Efrén

Design of Classification Models Using Grammatical Evolution and Machine Learning Components Applied to Medical Problem
Zuñiga Núñez, Blanca Verónica; Guerra-Hernández, Erick Israel; Sotelo-Figueroa, Marco Aurelio

An Exploratory Study on the Alignment Between Human Perception and Deep Learning Explainability Methods

Costa, Daniel; de Souza Moura, Pedro Nuno; Cesario de Faria Alvim, Adriana

A Study on Control Schemes in the Evolutionary Process of Grouping Genetic Algorithms
Amador Larrea, Stephanie; Quiroz-Castellanos, Marcela

Professores Na Era Da Inteligência Artificial (IA): Desenvolvimento De Um Repositório Transdisciplinar De Materiais Educacionais Digitais (Med) Na Educação
Oliveira, Arthur

Multi-objective Evolutionary Neural Architecture Search for the design of Convolutional Neural Networks considering explainability mechanisms
Barradas-Palmeros, Jesús Arnulfo; Mezura-Montes, Efrén; Acosta-Mesa, HéctorGabriel

Identificación de biomarcadores en señales de electroencefalogramas para el apoyo al diagnóstico del trastorno depresivo mediante inteligencia artificial
Tiburcio, Juan

Identificación de Factores de Deserción Universitaria Mediante Selección de Características en Modelos de Aprendizaje Automático
Domínguez-Gómez, Daniel; Gonzalez Torres, Juan de Dios; Bolaina Escalante, Ronald Frank

Modelo predictivo para clasificar la infección por hongo fusarium oxysporum en hojas de tomate por medio de redes neuronales profundas y visión transformers
Torres Fernandez, Astrid Ariadna; Hernández Pérez, María Yasmín

SIMDIA ‘2025
Planificación del Simposium

JUEVES 13 DE NOVIEMBRE				VIERNES 14 DE NOVIEMBRE		
Hora Cd. de México (GTM-6)	Id	Titulo	Autores	Id	Titulo	Autores
09:00	1	Speech Restoration from EEG Brain Bio Signals Based on a Novel Deep Learning Approach	sbaih, asma*	16	Metodología para la adopción de herramientas de IA en el desarrollo de software	Reyes, Dario*; Sanchéz-Torres, Jenny Marcela; Rueda Cáceres, Iván Mauricio
09:20	4	Métrica Pixel Erosion Score y Método CB-RISE: Evaluación de la Interpretabilidad en Visión por Computadora	Stanchi, Oscar*; Ronchetti, Franco; Quiroga, Facundo	17	Gestión de la función de pérdida integral y adaptativa en redes neuronales	Avalos Escalante, Karen*
09:40	5	Desarrollo de recursos y modelos para la **traducción de Lengua de Señas Argentina**	Dal Bianco, Pedro Alejandro*; Quiroga, Facundo; Ronchetti, Franco	18	Intereses académicos en estudiantes de licenciatura: un análisis con minería de datos	Montejo-Collado, Fatima*; Hernández-Torruco, José; Chávez-Bosquez, Oscar
10:00	6	Fuzzy clustering based on Supervised Classification	Pérez Sánchez, Ismay*	19	Esquema de Aprendizaje Federado para la Preservación de la Privacidad en el Reconocimiento de Actividades Humanas	miranda rucabado, Oscar Alexis*
10:20	7	Desarrollo de metaheurísticas aplicadas a genómica estructural y funcional	Rojas, Matías*; Carballido, Jessica; Vidal, Pablo	20	Hand_IA, sistema inteligente para el control de prótesis de mano basado en señales de Electromiografía (EMG)	Padilla, Daniela*; Villaseñor, Luis; Gutiérrez , Alejandro
10:40	8	Valoración de viviendas urbanas en Argentina mediante técnicas de machine learning e interpretabilidad	Gutierrez, Emiliano *; Lanzarini, Laura; Cecchini, Rocío; Delbianco, Fernando	22	Diagnóstico de Cardiopatía Isquémica Crónica en DMT2 mediante TabTransformer: Contraste de Variables Predictivas con Factores de Riesgo Tradicionales.	Flores-Custodio, Orlando*; Pancardo-García, Pablo
11:00	9	Forecasting Spatio-Temporal Time Series with Missing Values Using Graph Neural Networks	Martínez Ruiz, Armando*; Gomez-Gil, Pilar; Fonseca-Delgado, Rigoberto	23	Un Modelo Híbrido para la Predicción de Resistencia Antimicrobiana mediante Redes Neuronales Gráficas y PLN	Saldaña Pacheco, Sergio Antonio*; Pancardo García, Pablo
11:20	10	Hybrid Feature extraction in thermographic images for the detection of abnormal structures in breast tissues	Aburto, Yareli*; Gómez Gil, María del Pilar; Altamirano Robles, Leopoldo	25	Optimización de modelos para identificar arritmia cardiaca con aprendizaje automático	Arias García, SANTIAGO*
11:40	11	Robotic Arm-Based Assisted Feeding System for Patients with Upper Limb Limitations	Leal, Luis*; Gutiérrez Giles, Alejandro; López Gutiérrez, Jesús Ricardo	26	Neuroevolution of Spiking Neuron Networks	Lopez-Herrera, Carlos-Alberto*; Acosta-Mesa, Héctor-Gabriel; Mezura-Montes, Efrén
12:00	12	Design of Classification Models Using Grammatical Evolution and Machine Learning Components Applied to Medical Problem	Zuñiga Núñez, Blanca Verónica*; Guerra-Hernández, Erick Israel; Sotelo-Figueroa, Marco Aurelio	27	Multi-objective Evolutionary Neural Architecture Search for the design of Convolutional Neural Networks considering explainability mechanisms	Barradas-Palmeros, Jesús-Arnulfo*; Mezura-Montes, Efrén; Acosta-Mesa, Héctor-Gabriel
12:20	13	An Exploratory Study on the Alignment Between Human Perception and Deep Learning Explainability Methods	Costa, Daniel*; de Souza Moura, Pedro Nuno; Cesario de Faria Alvim, Adriana	29	Identificación de biomarcadores en señales de electroencefalogramas para el apoyo al diagnóstico del trastorno depresivo mediante inteligencia artificial	Tiburcio, Juan*
12:40	14	A Study on Control Schemes in the Evolutionary Process of Grouping Genetic Algorithms	Amador Larrea, Stephanie*; Quiroz-Castellanos, Marcela	31	Identificación de Factores de Deserción Universitaria Mediante Selección de Carasterísticas en Modelos de Aprendizaje Automático	Domínguez-Gómez, Daniel*; Gonzalez Torres, Juan de Dios; Bolaina Escalante, Ronald Frank
13:00	15	Professores Na Era Da Inteligência Artificial (IA): Desenvolvimento De Um Repositório Transdisciplinar De Materiais Educacionais Digitais (Med) Na Educação 5.0	Oliveira, Arthur*	32	Modelo predictivo para clasificar la infección por hongo fusarium oxysporum en hojas de tomate por medio de redes neuronales profundas y visión transformers	TORRES FERNANDEZ, ASTRID ARIADNA*; HERNÁNDEZ PÉREZ, MARÍA YASMÍN

Trabajos Presentados

Speech Restoration from EEG Brain Bio Signals Based on a Novel Deep Learning Approach

Asma Sbaih [1] , Jorge Garcia-Gutierrez [1], David
[1]Engineering department, Faculty of Engineering and Information Technology, Seville University, Spain
✉ asmasbaih2021@gmail.com

Abstract. Restoring intelligible speech from brain signals is a key research goal in neural engineering, with the potential to support individuals affected by severe communication impairments. While significant progress has been made using invasive neural recordings such as electrocorticography (ECoG), fewer studies have successfully achieved high-quality speech reconstruction from stereo-electroencephalography (sEEG) or non-invasive EEG. During this PhD, I aim to develop and evaluate a deep learning framework for reconstructing speech from sEEG signals using the iBIDS dataset. The proposed model will integrate convolutional and bidirectional LSTM layers (CNN–BiLSTM) to capture spatial and temporal dynamics of neural activity. To address data limitations and improve generalization, I will implement a transfer learning strategy by initializing parts of the model with weights pretrained on ECoG data. This research will involve a comparative analysis between the proposed transfer learning model, a baseline CNN–BiLSTM model trained from scratch, and traditional machine learning techniques for speech decoding. A HiFi-GAN neural vocoder will be used to convert predicted Mel spectrograms into intelligible audio, enabling perceptual and quantitative evaluation. The overarching goal is to advance EEG-based brain–computer interfaces by enabling accurate and naturalistic speech reconstruction, ultimately contributing to communication tools for individuals with neurological conditions, particularly Epilepsy.

Keywords: *sEEG, ECoG, CNN-BiLSTM, Transfer Learning, Speech Reconstruction, Brain-Computer Interface. Vocoder, Mel spectrograms.*

a) Problem Statement and Motivation

Restoring speech from brain activity is one of the biggest challenges in neural engineering and brain-computer interface (BCI) research. This is especially true for people who have lost their ability to speak due to neurodegenerative diseases like amyotrophic lateral sclerosis (ALS), stroke, or traumatic brain injuries. Non-invasive electroencephalography (EEG) is a safe and accessible method for brain imaging. However, decoding clear speech from EEG is difficult because of its low spatial resolution, vulnerability to noise, and the complexity of how the brain encodes speech.

Although invasive techniques like electrocorticography (ECoG) and stereo-EEG (sEEG) have shown promising results[1], non-invasive EEG decoding is still far behind in terms of intelligibility and audio quality[2]. Previous studies have mostly concentrated on classifying phonemes or a limited set of words rather than reconstructing continuous speech[3]. There is a pressing need for a scalable, general solution that can reliably restore high-quality speech from EEG signals.

This doctoral research aims to fill that gap by introducing a new deep learning framework for speech restoration from EEG. It uses hybrid neural architectures, transfer learning from invasive methods (ECoG), and neural vocoders to improve the audio quality[4], [5]. To our knowledge, this is the first research to use transfer learning from invasive brain recordings to enhance non-invasive speech restoration.

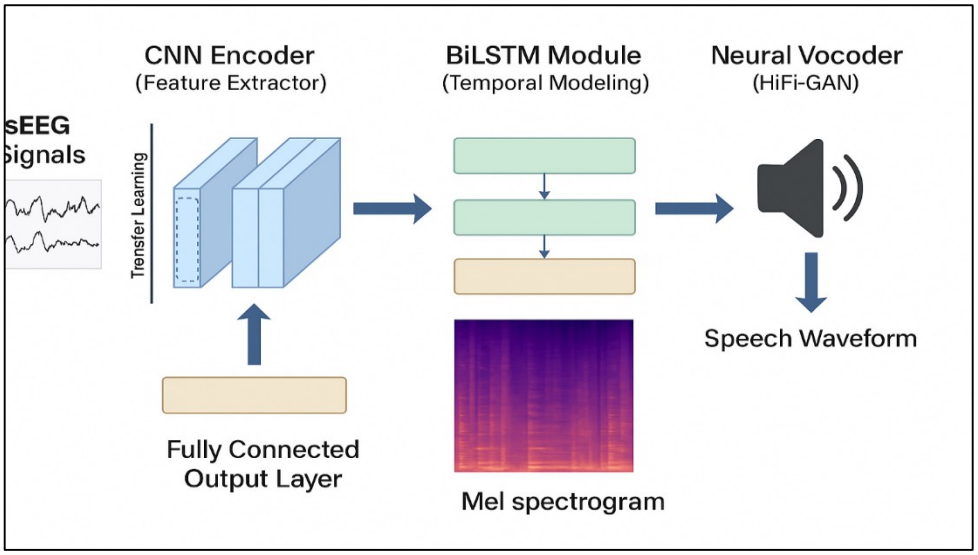


Fig 1.Overview of the proposed speech reconstruction pipeline using EEG, CNN–BiLSTM, transfer learning, and HiFi-GAN vocoder

b) Research Objectives and Hypotheses

The primary objectives of this research are:

1-To develop a robust and scalable deep learning model that reconstructs intelligible speech from EEG and sEEG signals. This objective will be evaluated using standard intelligibility and quality metrics, including the Short-Time Objective Intelligibility (STOI), Perceptual Evaluation of Speech Quality (PESQ), and Mel Cepstral Distortion (MCD) scores. Quantitative improvements in these metrics compared to baseline models will indicate successful speech reconstruction.

2-To investigate the use of transfer learning from ECoG-trained models to improve the performance of EEG-based models[5].

The effectiveness of transfer learning will be assessed by comparing the proposed transfer-learning-based CNN–BiLSTM model with a baseline model trained from scratch on the iBIDS sEEG dataset. Enhanced performance in terms of intelligibility, spectral fidelity, and generalization across subjects will confirm this objective[6].

3-To evaluate the benefit of using advanced neural vocoders (e.g., HiFi-GAN) for enhancing the perceptual quality of speech reconstructed from EEG[7].

This objective will be evaluated through both objective metrics (PESQ, MCD) and subjective perceptual listening tests, focusing on the naturalness, timbre, and audio realism of the reconstructed waveforms generated by the HiFi-GAN vocoder[8].

4-Improve assistive communication technologies for individuals suffering from speech impairments due to neurological disorders [4].

This objective will be fulfilled through benchmarking the proposed CNN–BiLSTM transfer learning framework against SpeechT5, a state-of-the-art Transformer-based speech synthesis model[9], to assess the relative capability of each approach in modeling temporal–spectral dependencies from neural signals. Improved intelligibility and perceived quality will demonstrate the system's potential as an effective assistive communication tool[10].

Central Hypothesis: Incorporating pretrained spatial feature extractors from high-resolution ECoG data into a CNN–BiLSTM framework trained on sEEG signals will significantly improve the intelligibility and naturalness of reconstructed speech, as validated through both objective and perceptual evaluation metrics.

c) Expected Contributions, Datasets, and Related Work

This research is expected to make the following contributions:

-Novel Model Architecture: Introduction of a transfer learning-based CNN–BiLSTM architecture designed for EEG speech restoration[11],[12].

-Cross-Modality Learning: Demonstration of cross-modal knowledge transfer from ECoG to sEEG in neural decoding tasks [5], [1].

-Enhanced Audio Output: Integration of HiFi-GAN vocoders for realistic speech synthesis from Mel spectrograms predicted by the model[13],[14] .

-Benchmark Dataset Usage: Use of the iBIDS sEEG dataset for invasive experiments and public ECoG datasets for pretraining [1], [2].

Dataset Description

-iBIDS sEEG Dataset: The iBIDS dataset consists of intracranial sEEG recordings from 10 native Dutch-speaking adults (5 males, 5 females) undergoing presurgical evaluation for drug-resistant epilepsy [1]. Each participant spoke around 100 unique Dutch words, with neural data recorded at 1024–2048 Hz and corresponding audio at 48 kHz. The dataset includes 1103 depth electrodes, adheres to the iEEG-BIDS standard, and provides rich metadata—making it ideal for speech decoding research.

-ECoG Dataset: The ECoG dataset by Anumanchipalli et al. [2] includes high-density electrode recordings from three English-speaking epilepsy patients. Electrodes were placed over the sensorimotor cortex and Broca's area, with participants reading phonemically balanced sentences. Neural data was sampled at 3052 Hz and downsampled to 200 Hz, with aligned phoneme-level audio and spatial localization information—making it suitable for transfer learning to sEEG.

Previous studies such as Anumanchipalli et al. (2019) [5] and Moses et al. (2021) [4]have demonstrated that speech can be synthesized from high-resolution ECoG signals. These invasive datasets provide a rich representation of neural activity associated with speech articulation. However, few studies have attempted to reconstruct continuous speech from EEG signals. The majority of EEG-based research has instead focused on classification tasks—such as phoneme or word recognition—primarily due to the modality's low spatial resolution, low signal-to-noise ratio, and high inter-subject variability [11],{Citation}. Classification tasks are more computationally manageable and resilient to noise, allowing researchers to sidestep the challenges of waveform synthesis. Nevertheless, these approaches fall short of enabling full communication restoration. This research introduces a novel approach to directly reconstructing speech waveforms from sEEG by leveraging transfer learning from ECoG-trained CNN encoders, setting a new direction for invasive speech prosthetics.

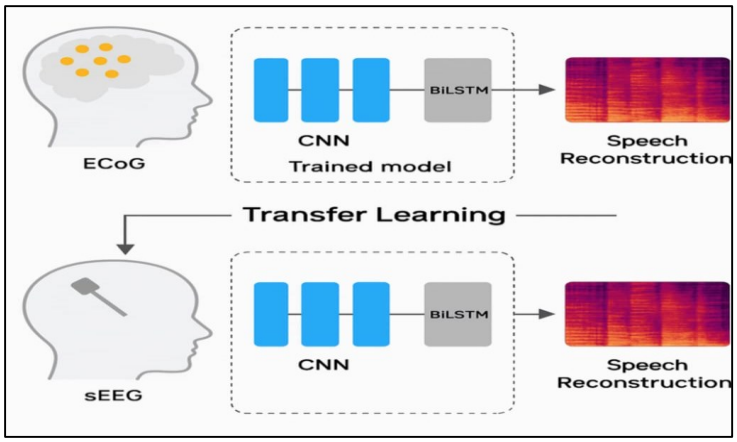


Fig .2. Diagram of CNN–BiLSTM architecture with transfer learning from ECoG pretrained CNN encoder layers

d) Training Strategy and Reproducibility

To ensure methodological transparency and reproducibility, the training strategy will be explicitly defined:

- Optimizer: Adam optimizer with initial learning rate of 1×10^{-4}, decayed using cosine scheduling.
- Loss Function: Combined mean squared error (MSE) and perceptual spectral loss on Mel spectrograms.
- Batch Size: 32 samples per batch; sequence length of 500 ms windows.
- Regularization: Dropout ($p = 0.3$), early stopping, and L2 weight decay.
- Data Augmentation: Gaussian noise injection, random time-shifts, and band-limited filtering to enhance robustness.
- Implementation: Built with PyTorch, trained on NVIDIA A100 GPU, reproducibility ensured through fixed random seeds and open-source GitHub repository release.

e) Expected Limitations

Despite its strengths, several limitations are anticipated:

- Cross-language generalization: Models trained on Dutch/English data may not generalize to other languages.
- Inter-subject variability: Anatomical and physiological differences reduce transferability.
- Signal noise and artifacts: EEG/sEEG signals are highly sensitive to muscle and environmental noise.
- Data imbalance: Limited datasets constrain model diversity.
- Hardware constraints: Training deep neural models on high-dimensional EEG data is computationally intensive.

Tackling these limitations will help guide future improvements, such as multilingual and subject-adaptive learning.

f) Evaluation and Outreach Plan

The proposed models will be evaluated using objective intelligibility and quality metrics including:

STOI (Short-Time Objective Intelligibility)

PESQ (Perceptual Evaluation of Speech Quality)

Mel Cepstral Distortion (MCD)

To validate the effectiveness of the proposed model, we will conduct a comparative analysis between two model variants: (1) a base CNN–BiLSTM model trained from scratch using the iBIDS sEEG dataset, and (2) the same CNN–BiLSTM model enhanced with transfer learning from an ECoG-trained CNN encoder. Both models will be trained on identical data splits and evaluated using the same objective metrics to ensure fair comparison.

Additionally, we plan to benchmark the proposed CNN–BiLSTM transfer learning framework against SpeechT5—a state-of-the-art Transformer-based speech synthesis model—[9]to evaluate the relative effectiveness of our architecture in capturing temporal and spectral dependencies from neural signals. This comparative setup will highlight the impact of transfer learning and architectural design choices on model generalization, phonetic intelligibility, and audio quality [16].

Cross-validation across subjects will be conducted, and ablation studies will evaluate the impact of transfer learning and vocoder integration. For outreach, we plan to submit our results to leading conferences and journals in BCI and AI, such as IEEE EMBC and Journal of Neural Engineering. A GitHub repository with model code and documentation will also be provided to promote reproducibility and collaboration.

Expected Results:

The transfer learning approach is expected to improve STOI by > 10%, reduce MCD by ≈ 15%, and achieve perceptually intelligible speech synthesis via HiFi-GAN.

g) Future Work:

Future directions include:

-Extension to real-time speech BCI systems.

-Expansion to multilingual models.

-Personalization of models using few-shot learning to adapt to individual brain patterns.

-Integration with eye-tracking or EMG for multimodal decoding.

This research lays a foundation for the development of practical EEG-based speech prostheses and advances the state of the art in non-invasive BCI communication.

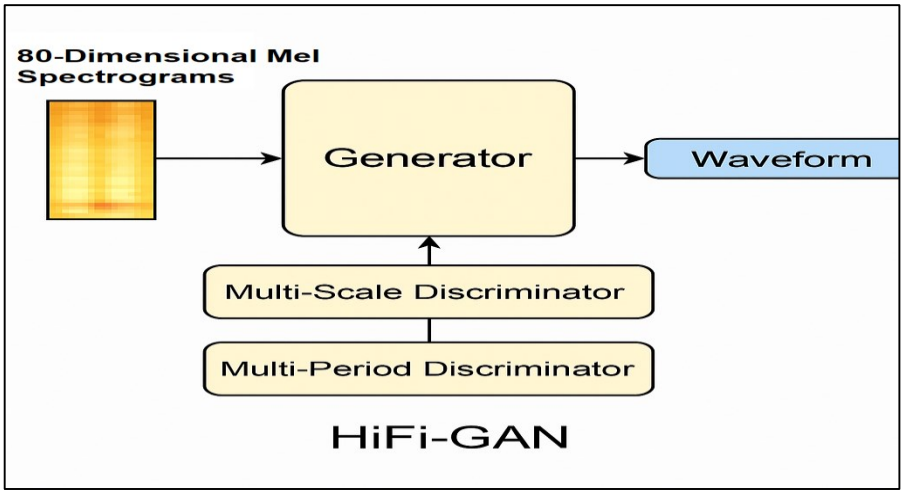


Fig. 3.HiFi-GAN architecture used to convert Mel spectrograms into natural-sounding speech waveforms

References

[1] M. Verwoert *et al.*, "Dataset of Speech Production in intracranial Electroencephalography," *Sci. Data*, vol. 9, no. 1, p. 434, Jul. 2022, doi: 10.1038/s41597-022-01542-9.

[2] G. K. Anumanchipalli, J. Chartier, and E. F. Chang, "Speech synthesis from neural decoding of spoken sentences," *Nature*, vol. 568, no. 7753, p. 493, Apr. 2019, doi: 10.1038/s41586-019-1119-1.

[3] M. Bisla and R. S. Anand, "Optimized CNN-Bi-LSTM–Based BCI System for Imagined Speech Recognition Using FOA-DWT," *Adv. Hum.-Comput. Interact.*, vol. 2024, no. 1, p. 8742261, 2024, doi: 10.1155/2024/8742261.

[4] "Neuroprosthesis for Decoding Speech in a Paralyzed Person with Anarthria | New England Journal of Medicine." Accessed: Jun. 30, 2025. [Online]. Available: https://www.nejm.org/doi/full/10.1056/NEJMoa2027540

[5] G. K. Anumanchipalli, J. Chartier, and E. F. Chang, "Speech synthesis from neural decoding of spoken sentences," *Nature*, vol. 568, no. 7753, pp. 493–498, Apr. 2019, doi: 10.1038/s41586-019-1119-1.

[6] J. Li, C. Guo, L. Fu, L. Fan, E. F. Chang, and Y. Li, "Neural2Speech: A Transfer Learning Framework for Neural-Driven Speech Reconstruction," Jan. 31, 2024, *arXiv*: arXiv:2310.04644. doi: 10.48550/arXiv.2310.04644.

[7] A. G. Habashi, A. M. Azab, S. Eldawlatly, and G. M. Aly, "Generative adversarial networks in EEG analysis: an overview," *J. NeuroEngineering Rehabil.*, vol. 20, no. 1, p. 40, Apr. 2023, doi: 10.1186/s12984-023-01169-w.

[8] G. Krishna, C. Tran, M. Carnahan, and A. Tewfik, "Improving EEG based continuous speech recognition using GAN," 2020, *arXiv*. doi: 10.48550/ARXIV.2006.01260.

[9] J. Ao *et al.*, "SpeechT5: Unified-Modal Encoder-Decoder Pre-Training for Spoken Language Processing," May 24, 2022, *arXiv*: arXiv:2110.07205. doi: 10.48550/arXiv.2110.07205.

[10] M. Bisla and R. S. Anand, "Optimized CNN-Bi-LSTM–Based BCI System for Imagined Speech Recognition Using FOA-DWT," *Adv. Hum.-Comput. Interact.*, vol. 2024, no. 1, p. 8742261, Jan. 2024, doi: 10.1155/2024/8742261.

[11] M. Bisla and R. S. Anand, "Optimized CNN-Bi-LSTM–Based BCI System for Imagined Speech Recognition Using FOA-DWT," *Adv. Hum.-Comput. Interact.*, vol. 2024, no. 1, p. 8742261, 2024, doi: 10.1155/2024/8742261.

[12] V. Singh, S. K. Sahana, and V. Bhattacharjee, "A novel CNN-GRU-LSTM based deep learning model for accurate traffic prediction," *Discov. Comput.*, vol. 28, no. 1, p. 38, Apr. 2025, doi: 10.1007/s10791-025-09526-0.

[13] J. Kong, J. Kim, and J. Bae, "HiFi-GAN: Generative Adversarial Networks for Efficient and High Fidelity Speech Synthesis," Oct. 23, 2020, *arXiv*: arXiv:2010.05646. doi: 10.48550/arXiv.2010.05646.

[14] R. G. Kamiloğlu and D. A. Sauter, "Voice Production and Perception," in *Oxford Research Encyclopedia of Psychology*. Accessed: Jun. 30, 2025. [Online]. Available: https://oxfordre.com/psychology/display/10.1093/acrefore/9780190236557.001.0001/acrefore-9780190236557-e-766

[15] "Speech synthesis from intracranial stereotactic Electroencephalographyusinganeuralvocoder | Request PDF," *ResearchGate*, doi: 10.36244/ICJ.2024.1.6.

[16] J. S. Brumberg, K. M. Pitt, and J. D. Burnison, "A Non-Invasive Brain-Computer Interface for Real-Time Speech Synthesis: The Importance of Multimodal Feedback," *IEEE Trans. Neural Syst. Rehabil. Eng. Publ. IEEE Eng. Med. Biol. Soc.*, vol. 26, no. 4, pp. 874–881, Apr. 2018, doi: 10.1109/TNSRE.2018.2808425.

Métrica Pixel Erosion Score y Método CB-RISE: Evaluación de la Interpretabilidad en Visión por Computadora

Oscar Agustín Stanchi[1,3], Franco Ronchetti[1,2], and Facundo Quiroga[1,2]

[1] Instituto de Investigación en Informática LIDI, Facultad de Informática, Universidad Nacional de La Plata (UNLP), La Plata, Argentina
[2] Comisión de Investigaciones Científicas de la Provincia de Buenos Aires (CIC-PBA), La Plata, Argentina
[3] Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), La Plata, Argentina

Resumen La interpretabilidad en modelos de aprendizaje profundo es crucial para generar confianza en aplicaciones críticas. Este trabajo presenta dos contribuciones principales a la interpretabilidad en visión por computadora. Primero, se introduce la métrica Pixel Erosion Score (PeS), que evalúa cuantitativamente la robustez de las explicaciones mediante erosión iterativa de píxeles relevantes, mejorando la evaluación y comparación objetiva entre diferentes métodos de interpretabilidad. Como resultado, se lleva a cabo la primera comparación sistemática y cuantitativa de métodos de interpretabilidad en VGG16 y ResNet18, lo que permite establecer un marco de referencia sólido para futuros estudios. Segundo, se propone CB-RISE, una extensión del algoritmo RISE que incorpora detección automática de convergencia, perturbaciones difuminadas y normalización basada en predicción (PRN). Esta mejora reduce significativamente el costo computacional (speedup de 2.54×) mientras incrementa la calidad de los mapas de importancia mediante el uso de máscaras Gaussianas que evitan el efecto del *distribution shift* causado por ocultar mediante parches negros. Los experimentos en VGG16 y ResNet18 con ImageNet-1K demuestran que CB-RISE supera a métodos tradicionales en interpretabilidad, especialmente en VGG16 donde alcanza un AUC de 0.80, proporcionando herramientas más eficientes y confiables para la interpretabilidad en modelos de caja negra.

Keywords: Interpretabilidad · Aprendizaje Profundo · Visión por Computadora · Mapas de importancia · Métrica de Evaluación

1. Introducción y Trabajo Relacionado

Los modelos de Aprendizaje Profundo son considerados cajas negras debido a su complejidad y falta de interpretabilidad [3]. Sin embargo, la investigación en interpretabilidad está en auge, buscando entender las justificaciones detrás de sus salidas [3]. Esta comprensión es esencial para generar confianza entre usuarios y partes interesadas, especialmente en sistemas de alto riesgo [7].

Los métodos de atribución de características, que asignan una importancia a cada característica según su contribución a la predicción, son ampliamente estudiados [8]. Estos métodos suelen generar mapas de importancia, comúnmente representados como mapas de calor, para visualizar las áreas críticas que afectan las decisiones del modelo [13].

Uno de los métodos más populares de interpretabilidad para datos de imágenes es el algoritmo RISE (Randomized Input Sampling for Explanation) [6]. RISE es un enfoque post-hoc y local diseñado para modelos de caja negra, que genera mapas de importancia suaves que indican la relevancia de cada píxel en una predicción. Estos mapas se calculan mediante la ejecución repetida del modelo con entradas enmascaradas, ponderando las máscaras según la salida.

Aunque RISE es eficaz, su alto costo computacional representa una limitación significativa. El número de iteraciones se determina mediante prueba y error, variando según la imagen y el modelo, lo que dificulta su aplicación en escenarios reales [10]. Además, el ruido causado por parches negros extremos y la aleatoriedad de las máscaras afecta la precisión de los mapas de importancia, comprometiendo la exactitud en la atribución de relevancia de cada píxel [11].

Por otro lado, los mapas de importancia no explican cómo el modelo utiliza la información para hacer sus predicciones [8], y modelos con precisión similar pueden generar atribuciones muy diferentes [15].

En su esencia, la interpretabilidad es un concepto subjetivo, influenciado por el contexto y las percepciones de quienes emiten y reciben la explicación [5]. El problema central radica en que la mayoría de los estudios dependen de evaluaciones cualitativas, lo que no solo dificulta comparaciones sistemáticas, sino que también limita la posibilidad de establecer criterios objetivos y reproducibles para valorar distintos métodos [14]. Ante la falta de métricas estándar, es crucial contar con indicadores cuantitativos para evaluar, comparar y seleccionar métodos de interpretabilidad de manera objetiva [9], asegurando que las explicaciones sean robustas y adecuadas para su uso en entornos críticos [8]. Sin estas métricas, el desarrollo y la adopción de técnicas interpretables se ven limitados [5].

Petsiuk et al. [6] propusieron la métrica de eliminación (*deletion*), que evalúa automáticamente las explicaciones midiendo la disminución en la probabilidad predicha por el modelo a medida que se eliminan secuencialmente los píxeles considerados más importantes. Esta métrica asume que una eliminación eficaz de regiones relevantes debería provocar una reducción progresiva en la confianza del modelo respecto a su predicción original.

No obstante, en [12] se observa que la eliminación secuencial de píxeles en la métrica *deletion* puede provocar un aumento inesperado en la precisión del modelo, lo que pone en evidencia inconsistencias y una fiabilidad limitada en su capacidad de evaluación.

El **objetivo principal** de este trabajo es fortalecer las herramientas de interpretabilidad en modelos de clasificación de visión por computadora, con énfasis en aquellos basados en atribución para modelos de caja negra. Con este fin, se introduce la métrica Pixel Erosion Score (PeS), diseñada para evaluar de forma

cuantitativa y objetiva la calidad y robustez de las explicaciones, facilitando la comparación entre diferentes métodos de interpretabilidad. Como caso de aplicación y validación de esta métrica, se desarrolla CB-RISE, una extensión del algoritmo RISE que incorpora mejoras en eficiencia, calidad y coherencia de los mapas de importancia.

La combinación de PeS y CB-RISE permite facilitar el análisis del comportamiento de los modelos en tareas de clasificación de imágenes y contribuye al desarrollo de sistemas más confiables y explicables, especialmente en contextos críticos. Al mismo tiempo, se presenta la primera comparación sistemática de métodos de interpretabilidad en modelos como VGG16 y ResNet18.

2. Contribuciones

2.1. PeS: Pixel Erosion Score

La métrica Pixel Erosion Score (PeS) evalúa la robustez de las explicaciones del modelo mediante el proceso de erosión iterativa aplicado a los mapas de importancia generados por los métodos de interpretabilidad. Este proceso comienza con una máscara binaria, que resalta los píxeles más relevantes de la imagen, y en cada iteración, se eliminan píxeles relevantes según un elemento estructurante. La salida del modelo se calcula multiplicando la imagen original por la máscara erosionada en cada paso.

La fórmula para la operación de erosión es:

$$h_t(I) = h_{t-1}(I) \ominus B, \quad h_0(I) = \{S_f(I) \geq \tau\} \subseteq [0,1]^{H \times W} \tag{1}$$

donde B es el elemento estructurante (un kernel cuadrado de 3×3 píxeles), $\ominus$ denota la operación de erosión, y $S_f(I)$ es el mapa de importancia generado por el método de interpretabilidad sobre el modelo f para la imagen I.

El AUC $\in [0,1]$ de la curva generada por este proceso mide cómo cambia la predicción del modelo a medida que se eliminan píxeles. Un AUC cercano a 1 indica que el modelo mantiene la precisión por más tiempo, ya que las características críticas permanecen intactas a lo largo de varios ciclos de erosión, lo que demuestra una mayor robustez ante la eliminación de píxeles no relevantes.

2.2. CB-RISE: Convergence Detection and Blurred Perturbations for RISE

2.2.1. Limitaciones de RISE El algoritmo RISE requiere un número fijo de iteraciones definido por prueba y error, lo que implica una alta carga computacional innecesaria, ya que sus mapas de importancia suelen estabilizarse antes de completarlas [11]. Además, su diseño rígido impide el uso de perturbaciones alternativas, como el difuminado, lo que limita su flexibilidad y afecta la calidad de las explicaciones [1].

El uso de máscaras con parches negros puede inducir entradas artificialmente dispersas (*sparse input features*), que degradan la activación de la red, alteran la distribución original de los datos y generan cambios en la predicción no atribuibles a la eliminación de información relevante, sino a un *distribution shift* [2].

Este fenómeno compromete la validez de las explicaciones y sugiere la necesidad de normalizaciones más robustas y perturbaciones más naturales.

2.2.2. Mejoras Propuestas CB-RISE es una extensión de RISE diseñada para mejorar tanto la eficiencia computacional como la calidad de las explicaciones. Incorpora tres innovaciones principales: un mecanismo de detección de convergencia, el uso de perturbaciones difuminadas y una nueva estrategia de normalización basada en la predicción.

Para detectar convergencia de forma automática y evitar un número arbitrario de iteraciones, se aplica el algoritmo en línea de Welford, que permite monitorear la varianza del mapa de importancia generado en cada paso. El usuario puede controlar este proceso mediante tres parámetros: ρ (frecuencia con la que se evalúa la convergencia), ε (cambio mínimo de varianza para considerar un píxel como estable) y τ (porcentaje máximo de píxeles que aún pueden cambiar sin detener la iteración). El algoritmo concluye cuando el porcentaje de píxeles cuyo cambio de varianza es superior a ε es menor o igual a τ, lo que permite detener la generación del mapa sin pérdida de calidad.

En cuanto a las perturbaciones, CB-RISE sustituye los parches negros por máscaras Gaussianas, obtenidas mediante la interpolación de máscaras binarias con una versión suavizada de la imagen. De este modo, cada perturbación se genera combinando de forma gradual la imagen original con su versión difuminada, logrando transiciones suaves en lugar de cortes abruptos. Este enfoque reduce los efectos indeseados del *distribution shift* y evita problemas asociados con entradas altamente dispersas (*sparse inputs*).

Por último, se propone una nueva normalización denominada Prediction-Relative Normalization (PRN), en la que cada máscara se pondera con un factor calculado como el mínimo entre la predicción con la máscara aplicada y la predicción original, normalizado por el puntaje de la clase objetivo. Esto limita la influencia de regiones que, al ocultarse, incrementan artificialmente la confianza del modelo, y permite interpretar cada valor del mapa como una proporción relativa justa respecto a la predicción inicial del modelo.

3. Experimentos y Resultados

Se realizó un estudio de robustez para evaluar cuantitativamente las variaciones de RISE en interpretabilidad y eficiencia, usando 48 imágenes filtradas del conjunto de validación de ImageNet-1K con accuracy mayor a 0.7 [4]. Los experimentos se ejecutaron sobre las arquitecturas ResNet18 y VGG16, comparando cinco métodos: Activations, Grad-CAM, Occlusion, RISE y CB-RISE. Para RISE y sus variantes se usaron 4096 máscaras con tamaños iniciales de 4×4 y 8×8; en CB-RISE se emplearon parámetros $\sigma = 10$, $\rho = 64$, $\varepsilon = 10^{-3}$ y $\tau = 0{,}3$. La métrica PeS se calculó con umbral $\tau = 0{,}5$ y perturbación de desenfoque Gaussiano $\sigma = 10$, deteniendo el proceso al llegar a 100 iteraciones o al reducir a 1 % los píxeles blancos.

Los resultados cuantitativos de erosión, presentados en el Cuadro 1, muestran que un descenso rápido en la métrica PeS indica mapas de importancia deficientes, pues la eliminación de píxeles esenciales afecta la clasificación.

En VGG16, los métodos black-box basados en RISE superan a los white-box (Activations y Grad-CAM) en la calidad del mapa, con CB-RISE destacándose especialmente en tamaño de máscara 8×8 alcanzando un Erosion AUC promedio de 0.80, lo que evidencia una mejor identificación de características relevantes gracias al desenfoque Gaussiano.

Para ResNet18, los métodos white-box presentan un desempeño ligeramente superior a los basados en RISE, aunque CB-RISE se acerca a ellos. En general, ResNet18 mostró menor interpretabilidad en erosión que VGG16 para métodos black-box, reflejado en valores más bajos del área bajo la curva (AUC).

Respecto a la eficiencia, CB-RISE logró una significativa reducción en el número de iteraciones necesarias para converger en comparación con RISE original, que siempre usa las 4096 máscaras completas. Por ejemplo, en VGG16 con máscaras 4×4, CB-RISE promedió 1609.83, frente a las 4096 de RISE. Esto representa un speedup de 2.54× para CB-RISE en esta configuración, y un speedup menor pero aún relevante para máscaras de 8×8 (1.30×). Es por que ello que se puede afirmar que este mecanismo de convergencia reduce notablemente el costo computacional sin sacrificar la calidad explicativa.

Cuadro 1. AUC promedio de la métrica PeS para métodos de interpretabilidad en VGG16 y ResNet18, comparando técnicas white-box y black-box, con el número de iteraciones requerido. Valores mayores de AUC indican mejor interpretabilidad.

Método		**Activations**	**Grad-CAM**	**Occlusion**	**RISE**		**CB-RISE**	
					(4×4)	(8×8)	(4×4)	(8×8)
Black-Box		X	X	✓	✓		✓	
Iteraciones		1	2	170	4096	4096	1614.67	3160.00
AUC ↑	**VGG16**	0.64	0.46	0.52	0.75	0.70	0.79	**0.80**
	ResNet18	0.68	**0.72**	0.51	0.67	0.62	0.70	0.59

4. Conclusiones y Trabajos Futuros

Introducimos la métrica Pixel Erosion Score (PeS) como el método principal para cuantificar la robustez de explicaciones y validamos CB-RISE, que mejora RISE mediante detección de convergencia, perturbaciones difuminadas y una normalización basada en la predicción, incrementando la calidad de los mapas de importancia y la eficiencia computacional. Los resultados en VGG16 y ResNet18 muestran mejoras significativas (p. ej. AUC 0.80 en VGG16 y un speedup de 2.54× en ciertas configuraciones), y, además, este trabajo ofrece la primera comparación sistemática y cuantitativa entre métodos y modelos para esta tarea. En conjunto, PeS y CB-RISE ofrecen herramientas objetivas y eficientes para evaluar y explicar decisiones de modelos de visión por computadora, facilitando su auditoría y confianza en entornos críticos. Como trabajos futuros se plantea la automatización de parámetros, evaluar enfoques con máscaras adaptativas mediante metaheurísticas, y aplicar estrategias de optimización para favorecer su adopción en producción y su aplicabilidad en problemas del mundo real.

Los resultados se divulgarán mediante artículos científicos, presentaciones en congresos, y en revistas especializadas para alcanzar un público aún más amplio.

Referencias

1. Dabkowski, P., Gal, Y.: Real time image saliency for black box classifiers. Advances in neural information processing systems **30** (2017)
2. Hooker, S., Erhan, D., Kindermans, P.J., Kim, B.: A benchmark for interpretability methods in deep neural networks. Advances in neural information processing systems **32** (2019)
3. Lipton, Z.C.: The mythos of model interpretability: In machine learning, the concept of interpretability is both important and slippery. Queue **16**(3), 31–57 (2018)
4. Lombrozo, T.: Explanatory preferences shape learning and inference. Trends in cognitive sciences **20**(10), 748–759 (2016)
5. Nguyen, A.p., Martínez, M.R.: On quantitative aspects of model interpretability. arXiv preprint arXiv:2007.07584 (2020)
6. Petsiuk, V., Das, A., Saenko, K.: Rise: Randomized input sampling for explanation of black-box models. In: Proceedings of the British Machine Vision Conference (BMVC) (2018)
7. Rudin, C.: Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nature machine intelligence **1**(5), 206–215 (2019)
8. Salahuddin, Z., Woodruff, H.C., Chatterjee, A., Lambin, P.: Transparency of deep neural networks for medical image analysis: A review of interpretability methods. Computers in biology and medicine **140**, 105111 (2022)
9. Schmidt, P., Biessmann, F.: Quantifying interpretability and trust in machine learning systems. arXiv preprint arXiv:1901.08558 (2019)
10. Stanchi, O., Ronchetti, F., Dal Bianco, P., Rios, G., Ponte Ahon, S., Hasperué, W., Quiroga, F.: Cb-rise: Improving the rise interpretability method through convergence detection and blurred perturbations. In: Conference on Cloud Computing, Big Data & Emerging Topics. Springer (2024)
11. Stanchi, O., Ronchetti, F., Quiroga, F.: The implementation of the rise algorithm for the captum framework. In: Conference on Cloud Computing, Big Data & Emerging Topics. pp. 91–104. Springer (2023)
12. Stanchi, O.A.: Métrica PeS/PdS y Método CB-RISE: Nuevos enfoques en la evaluación de interpretabilidad en modelos de visión por computadora. Ph.D. thesis, Universidad Nacional de La Plata (2025)
13. Stanchi, O.A., Ronchetti, F., Dal Bianco, P.A., Ríos, G.G., Hasperué, W., Puig Valls, D., Rashwan, H., Quiroga, F.M.: Quantitative evaluation of white & black box interpretability methods for image classification. In: XXX Congreso Argentino de Ciencias de la Computación (CACIC)(La Plata, 7 al 11 de octubre de 2024) (2024)
14. Strumbelj, E., Kononenko, I.: An efficient explanation of individual classifications using game theory. The Journal of Machine Learning Research **11**, 1–18 (2010)
15. Yilmaz, E.B., Mader, A.O., Fricke, T., Peña, J., Glüer, C.C., Meyer, C.: Assessing attribution maps for explaining cnn-based vertebral fracture classifiers. In: Interpretable and Annotation-Efficient Learning for Medical Image Computing: Third International Workshop, iMIMIC 2020, Second International Workshop, MIL3ID 2020, and 5th International Workshop, LABELS 2020, Held in Conjunction with MICCAI 2020, Lima, Peru, October 4–8, 2020, Proceedings 3. pp. 3–12. Springer (2020)

Fuzzy clustering based on Supervised Classification*

Ismay Pérez-Sánchez[1 [0000−0003−0917−8764]], Raúl Monroy[1 [0000−0002−3465−995X]], and Miguel Angel Medina-Pérez[2 [0000−0003−4511−2252]]

[1] School of Engineering and Sciences, Tecnologico de Monterrey, Carretera al Lago de Guadalupe Km. 3.5, Atizapán de Zaragoza, Estado de México 52926, México
{ismayps@gmail.com,raulm@tec.mx}

[2] Kayak Analytics, The Dome Tower, Jumeirah Lake Towers, Dubai, UAE
miguel.medina.perez@gmail.com

Abstract. Fuzzy clustering, in contrast to hard clustering, does not assign the objects to a specific cluster; instead, it keeps a membership degree for all of them. This feature is advantageous because these algorithms can discover those objects with special conditions, like the outliers, and the objects that can belong to more than one cluster. Nevertheless, the clustering algorithms rely directly on dissimilarity measures, which condition the distribution of objects in the clusters and limit their generalization power. In this work, we propose a new fuzzy clustering algorithm assisted by supervised classifiers to improve the performance of other fuzzy clustering algorithms. The results of previous fuzzy clustering algorithms are used as the input data, which we enrich with an artificial class heuristically created to train and validate supervised classifiers. Then, we propose to evaluate the new fuzzy clustering with a fuzzy cluster validity index previously selected from a thoroughly analyzed study. Finally, those results will be evaluated statistically to determine if our algorithm outperforms the initial clustering algorithms on more than 80 data sets.

Keywords: Fuzzy Clustering · Fuzzy Cluster Validity Index · Supervised Classification.

1 Introduction

The amount of data generated by modern information technologies has grown significantly [9]. It has never been more important to extract useful insights from all of that data because it brings unprecedented sources of innovation and performance boost. We can see this phenomenon in any industry area, from marketing [8], engineering [13], or weather forecasting [10]. One of the most

* Supported by the National Council of Humanities, Sciences, Technologies, and Innovation (CONAHCyT) of Mexico and utilizing the computational resources of the GIEE-ML (Machine Learning) research group at Tecnologico de Monterrey.

widely used techniques to perform exploratory analysis and extract information about the organization of the data is clustering [28, 9].

Fuzzy clustering techniques relax the constraints that determine the borders of the clusters [19]. Hence, fuzzy clustering deals more effectively than hard clustering with situations like an overlapping group of objects, uncertain cluster membership of the objects (inliers), and objects that should not belong to any cluster at all (outliers) [19]. It provides a more realistic belonging of objects to each cluster, and higher accuracy of clustering compared to hard approaches [29].

1.1 Problem definition

Fuzzy clustering algorithms have been developed since its introduction by Bezdek [2], more than five decades ago. Nowadays, the research community keeps showing interest in fuzzy clustering [12, 21, 31, 25, 4, 11]. Nonetheless, there are a series of issues in these solutions. The first problem is about how much fuzzy clustering algorithms rely uniquely on the data points' dissimilarity measures [23]. As a consequence, there exist many algorithms developed to fit in specific geometrical structures like curves and surfaces [25, 7]. A further attempt to generalize the formulation includes the use of kernel functions [22, 6] where authors propose to detect patterns in nonlinear data, but they need the correct kernel function beforehand.

Another fundamental problem of fuzzy clustering algorithms, in general, is the parameters needed. Hopefully, in a tiny number of cases, the fuzzy clustering algorithm provides automatic ways to set some parameter values [14, 18], but not all of them. In all cases, there is always at least one parameter that should be set by the user [3, 15] besides the number of clusters. Therefore, those values are selected from initial experiments [14, 18], or inherited from previous experiments in algorithms that share a common theoretical basis [30, 6]. The problem is that in those previous works, the number of experiments, data sets involved, and evaluation methods used are very low, and hence, it can be considered that the results may contain bias if they are incorporated in algorithms applied to a broader context than the one represented by the data used.

Another problem is the selection of the cluster validity index (CVI) to verify the results of the clustering algorithm. This process, in most cases, is based on previous CVI studies that have the problem mentioned above. In the case of the evaluation of the clustering, they use a set of CVIs and analyze in which data sets a given clustering algorithm performs better than others [26, 5, 27]. As the metric is counting the number of those data sets, it is easy to grasp that the analysis will not result in anything conclusive that could be extrapolated to different settings. It is reported in previous studies [1, 16, 17] that all these CVIs are based on similarity or dissimilarity measures, and, for that reason, they cannot distinguish every type of geometry in the data. Which remains true, but their conclusions are reached by the wrong procedure.

2 Objectives and hypothesis

The **general objective** of the present research is to create a fuzzy clustering algorithm that improves previous fuzzy clustering approaches on the most appropriate validation index. This new algorithm is based on supervised classifiers, and the most appropriate fuzzy cluster validation index will be determined through a study in this research.

All this research requires the following **specific objectives**:

1. To measure the most used CVIs to study their behavior and create a method to select the most appropriate one given a group of data sets, minimizing the possibility of biases.
2. To create a robust fuzzy clustering based on supervised classifiers with better performance than the state-of-the-art fuzzy clustering algorithms, measured by the most appropriate cluster validity index.
3. To create a new fuzzy cluster validity index based on supervised classifiers to assess the robustness of the fuzzy clustering algorithms. This new index will not be used to assess the performance of the new algorithm, but to show the suitability of the general framework proposed (the use of supervised learning in the context of fuzzy clustering).

The **hypothesis** of this research is based on the rationale that combining diverse supervised classifiers [24] allows for better inference of the underlying structure of the data than the state-of-the-art fuzzy clustering algorithms. We claim that supervised classifiers learn on multiple relations among objects (e.g., partial comparison of objects, projecting objects in higher dimensions, and selecting features), not only on dissimilarity functions between objects, as the traditional fuzzy clustering algorithms work.

3 Methodology

The proposed new clustering algorithm will be based on supervised classifiers to improve the initial state-of-the-art clustering. The figure 1 shows a diagram with an overview of the algorithm. There, we propose to generate the initial clusterings of a given data set with a selected fuzzy clustering algorithm. The resulting clustering will be used to create an artificial class for each object of the data set using the membership degrees obtained in that clustering. Next, the algorithm augments the data set with the membership degrees of each object and their corresponding artificial class, and the new data set becomes the input of q supervised classifiers. Next, we take from the supervised classifiers' result the probability of the objects belonging to each class instead of the class that the classifier determines. The probabilities are the membership matrix of the fuzzy clustering, and the classification is its hardened version. Finally, the selected fuzzy CVI evaluates the fuzzy clusterings, and the best clustering becomes the output.

The artificial class of the object x_k is defined by the equation 1.

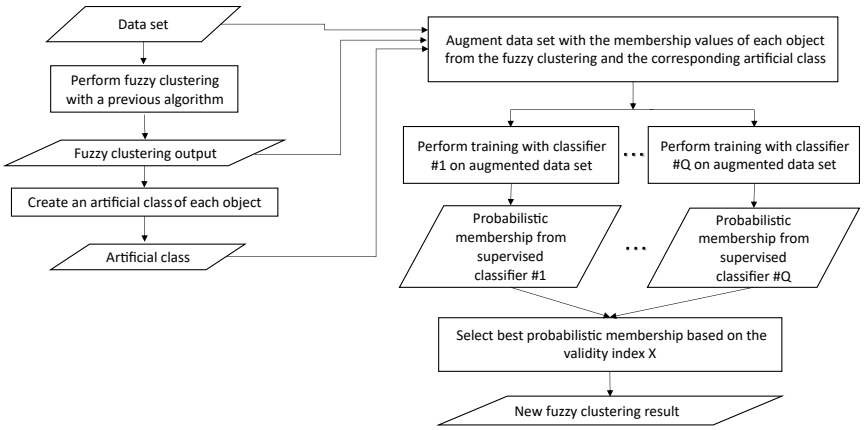


Fig. 1. General scheme of the proposed new fuzzy clustering algorithm.

$$\text{class}(x_k) = \begin{cases} \arg\max_j(u_{jk})\ u_{pk} - u_{qk} > 1/c \\ \arg\max_j |P_{kj}| \text{ otherwise} \end{cases}, \tag{1}$$

where $u_{pk} = \max_i(u_{ik})$ is the best membership value of x_k corresponding to the i-th cluster, $u_{qk} = \max_{i \neq p}(u_{ik})$ is the second best, and the operator $||$ is the module of the set. Also, P_{kj} represents a subset of objects from the $\lceil n/c \rceil$ nearest neighbors of the object x_k that have j as the cluster with higher membership degree, where n and c are the number of objects in the data set and the number of clusters specified, respectively.

The selected fuzzy CVI is obtained from a new evaluation method for indices, designed to answer the 1st objective.

Regarding the comparison of the new fuzzy clustering algorithm with the state-of-the-art, it will be used a statistical test to verify if there is a significant statistical difference in favor of the new algorithm.

It is important to note that the artificial class is only an instrument to make possible the use of the supervised classifiers, and hence, plays a minor role in the fulfillment of the hypothesis. The main role falls on the use of multiples supervised classifiers because with this method we intent to ameliorate the consequences of the traditional iterative methods, where the closely tied relationship between the objective functions and the equations to update the centroids and the membership values determines the geometry recognized on the data.

Finally, for the 3rd objective, a new fuzzy CVI is proposed to be developed following the same strategy as the new fuzzy clustering algorithm (i.e., it includes the use of supervised classifiers too). The difference resides in the evaluation step. Instead of using a CVI, it will use the Area Under the ROC curve (AUC) as the metric.

4 Partial results

The setting of the experiments comprises 84 data sets from the UCI repository, for all the new methods developed in this research.

In the new evaluation method for CVIs [20], developed by the authors as solution to the 1st objective, there were assessed 31 CVIs. The new CVI evaluation method uses the FCM algorithm to generate partitions of the data sets and the Adjusted Rand Index (ARI) as the external CVI to introduce sound comparisons using a ground truth. ARI is needed because the results of the 31 CVIs have different ranges, and a common range is needed to perform a fair comparison using the statistical tests. Therefore, for each fuzzy CVI and data set, it is selected the best partition and the value used for the final comparison is the value returned by ARI in that partition.

The experiments have four settings defined by the value of the fuzzyfier (m). In three of them, the CVI KPBM outperformed the others in the Wilcoxon signed-rank test after the Friedman test failed for all settings. The table 1 shows in detail the performance of each CVI. Let us clarify that the first row indicating the number of CVIs involved means that we included only those CVIs that were on the top of the Friedman tests with no significant difference. Extensive details can be found in Pérez-Sánchez et. al [20].

Table 1. Wilcoxon test summary results with the number of times each CVI outperforms individually the other CVIs by fuzzifier in two different levels of significance. The first row indicates the number of CVIs that were compared in each experiment setup.

CVI	Fuzzyfiers							
	m=1.5		m=2.0		m=2.5		m=3.0	
	$\alpha = 0.05$	$\alpha = 0.025$	$\alpha = 0.05$	$\alpha = 0.025$	$\alpha = 0.05$	$\alpha = 0.025$	$\alpha = 0.05$	$\alpha = 0.025$
CVIs involved	16		19		16		12	
KPBM	**13**	**10**	**10**	8	**12**	**12**	**7**	**5**
OS	6	3	0	0	0	0	1	1
SC2	3	2	**10**	8	0	0		
CS2	4	2	**10**	**10**	0	0		
VF	3	2	0	0			0	0
VR	2	2						
PCAES	3	3	**10**	**10**	3	1	3	3
ICC	3	3	**10**	**10**				
KWON	1	1	**10**	**10**	0	0	0	0
XB	1	1	**10**	**10**			0	0
SI	1	1	**10**	9	9	7	0	0
MPC	1	0	0	0				
CL	1	0						
FS	0	0	9	5	0	0	1	1
PBM	0	0	0	0	0	0		
CWB	0	0	0	0	6	4	4	4
APD			1	0	2	0		
PD			2	1	0	0		
FHV			1	0	0	0		
TSS			0	0	0	0	1	1
WSJ			0	0	2	2	1	1
SC					0	0		
CS							0	0

Also, the new fuzzy clustering is in the experimental phase, and the new fuzzy CVI is in development.

References

1. Arbelaitz, O., Gurrutxaga, I., Muguerza, J., Pérez, J. M., and Perona, I. An extensive comparative study of cluster validity indices. *Pattern Recognition 46*, 1 (2013), 243 – 256.
2. Bezdek, J. C. Numerical taxonomy with fuzzy sets. *Journal of Mathematical Biology 1*, 1 (1974), 57–71.
3. Bezdek, J. C., Ehrlich, R., and Full, W. Fcm: The fuzzy c-means clustering algorithm. *Computers & Geosciences 10*, 2 (1984), 191 – 203.
4. Bi, X., Lin, J., Tang, D., Bi, F., Li, X., Yang, X., Ma, T., and Shen, P. Vmd-kfcm algorithm for the fault diagnosis of diesel engine vibration signals. *Energies 13*, 1 (2020), 228.
5. Chang, S.-T., Lu, K.-P., and Yang, M.-S. Stepwise possibilistic c-regressions. *Information Sciences 334-335* (2016), 307 – 322.
6. Dang, T. H., Mai, D. S., and Ngo, L. T. Multiple kernel collaborative fuzzy clustering algorithm with weighted super-pixels for satellite image land-cover classification. *Engineering Applications of Artificial Intelligence 85* (2019), 85 – 98.
7. Davé, R. N. Fuzzy shell-clustering and applications to circle detection in digital images. *International Journal of General Systems 16*, 4 (1990), 343–355.
8. Desai, P., and Arora, B. Combined k-hierarchy clustering to know the buying pattern of customer and provide them with freebies by online site for future shopping. *TEST Engineering & Management 82* (2020), 7844–7853.
9. D'Urso, P. Informational paradigm, management of uncertainty and theoretical formalisms in the clustering framework: A review. *Information Sciences 400-401* (2017), 30 – 62.
10. Egrioglu, E., Bas, E., Yolcu, U., and Chen, M. Y. Picture fuzzy time series: Defining, modeling and creating a new forecasting method. *Engineering Applications of Artificial Intelligence 88* (2020), 103367.
11. Feng, Q., Chen, L., Chen, C. L. P., and Guo, L. Deep fuzzy clustering—a representation learning approach. *IEEE Transactions on Fuzzy Systems 28*, 7 (2020), 1420–1433.
12. Gharib, A., Sadoghi-Yazdi, H., and hossein Taherinia, A. Robust heterogeneous c-means. *Applied Soft Computing 86* (2020), 105885.
13. Izadi, H., Sadri, J., Hormozzade, F., and Fattahpour, V. Altered mineral segmentation in thin sections using an incremental-dynamic clustering algorithm. *Engineering Applications of Artificial Intelligence 90* (2020), 103466.
14. Krishnapuram, R., and Keller, J. M. A possibilistic approach to clustering. *IEEE Transactions on Fuzzy Systems 1*, 2 (1993), 98–110.
15. Krishnapuram, R., and Keller, J. M. The possibilistic c-means algorithm: insights and recommendations. *IEEE Transactions on Fuzzy Systems 4*, 3 (Aug 1996), 385–393.
16. Milligan, G. W. A monte carlo study of thirty internal criterion measures for cluster analysis. *Psychometrika 46*, 2 (1981), 187–199.
17. Milligan, G. W., and Cooper, M. C. An examination of procedures for determining the number of clusters in a data set. *Psychometrika 50*, 2 (Jun 1985), 159–179.
18. Pal, N. R., Pal, K., Keller, J. M., and Bezdek, J. C. A possibilistic fuzzy c-means clustering algorithm. *IEEE Transactions on Fuzzy Systems 13*, 4 (2005), 517–530.

19. Peters, G., Crespo, F., Lingras, P., and Weber, R. Soft clustering – fuzzy and rough approaches and their extensions and derivatives. *International Journal of Approximate Reasoning 54*, 2 (2013), 307 – 322.
20. Pérez-Sánchez, I., Medina-Pérez, M. A., Monroy, R., Loyola-González, O., and Gutierrez-Rodríguez, A. E. New evaluation method for fuzzy cluster validity indices. *IEEE Access 13* (2025), 22728–22744.
21. Quan, Z., and Chen, S. Robust convex clustering. *Soft Computing 24*, 2 (2020), 731–744.
22. Svetlova, L., Mirkin, B., and Lei, H. Mfwk-means: Minkowski metric fuzzy weighted k-means for high dimensional data clustering. In *IEEE 14th International Conference on Information Reuse Integration (IRI)* (2013), pp. 692–699.
23. Tamir, D. E., Rishe, N. D., and Kandel, A. *Fifty years of fuzzy logic and its applications*, vol. 326. Springer, 2015.
24. Thiyagarajan, S., and Chakravarthy, T. Classification with ensemble methods-a review. *Journal of Scientific Computing 8*, 12 (2019), 119–126.
25. Ullah, K., Garg, H., Mahmood, T., Jan, N., and Ali, Z. Correlation coefficients for t-spherical fuzzy sets and their applications in clustering and multi-attribute decision making. *Soft Computing 24*, 3 (2020), 1647–1659.
26. Wang, T., and Hung, W.-L. A generalized possibilistic approach to shell clustering of template-based shapes. *Journal of Statistical Computation and Simulation 87*, 3 (2017), 423–436.
27. Xie, Z., Wang, S., and Chung, F. L. An enhanced possibilistic c-means clustering algorithm epcm. *Soft Computing 12*, 6 (2008), 593–611.
28. Xu, D., and Tian, Y. A comprehensive survey of clustering algorithms. *Annals of Data Science 2*, 2 (2015), 165–193.
29. Xu, R., and Wunsch, D. C. Survey of clustering algorithms. *IEEE Transactions on Neural Networks* (2005).
30. Zhang, H., and Lu, J. Semi-supervised fuzzy clustering: A kernel-based approach. *Knowledge-Based Systems 22*, 6 (2009), 477 – 481.
31. Zhou, K., and Yang, S. Effect of cluster size distribution on clustering: a comparative study of k-means and fuzzy c-means clustering. *Pattern Analysis and Applications 23*, 1 (2020), 455–466.

Desarrollo de metaheurísticas aplicadas a genómica estructural y funcional

Matías Gabriel Rojas[1], Jessica Andrea Carballido[2], and Pablo Javier Vidal[1,3]

[1] Instituto Interdisciplinario de Ciencias Básicas (ICB), Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Universidad Nacional de Cuyo, Padre Jorge Contreras 1300, Mendoza, M5502JMA, Mendoza, Argentina.
mrojas@mendoza-conicet.gob.ar, pjvidal@conicet.gov.ar
[2] Instituto de Ciencias e Ingeniería de la Computación, Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Universidad Nacional del Sur, San Andres 800, Bahía Blanca, 8000, Buenos Aires, Argentina
jac@cs.uns.edu.ar
[3] Facultad de Ingeniería, Universidad Nacional de Cuyo, Centro Universitario, Mendoza, M5502JMA, Mendoza, Argentina.

Abstract. El volumen y la complejidad de los datos biológicos hacen necesario el desarrollo de enfoques computacionales avanzados que permitan analizarlos, generando nuevas perspectivas y preguntas que conduzcan a experimentos in vivo o in vitro. Entre estas aproximaciones, las metaheurísticas se han consolidado como herramientas eficaces para abordar problemas bioinformáticos complejos, tales como el ensamblado de fragmentos de ADN, el alineamiento múltiple de secuencias, la selección de genes relevantes para clasificar muestras cancerígenas, la optimización de métodos de machine learning aplicados al diagnóstico de enfermedades y el diseño de nuevos fármacos con características específicas. En este trabajo, se propone el desarrollo de métodos metaheurísticos que incluyen el diseño de operadores especializados, la hibridación con otros enfoques y estrategias multiobjetivo, con el fin de aportar avances en los problemas mencionados. Los resultados obtenidos evidencian la efectividad de las técnicas implementadas para las distintas aplicaciones.

Keywords: Bioinformática · Metaheurísticas · Algoritmos de Optimización.

1 Descripción del problema

La bioinformática integra biología, informática, matemática y estadística para procesar datos experimentales, generar nuevas perspectivas y formular preguntas que motiven nuevos ensayos [23]. El volumen masivo y heterogéneo de información experimental exige tecnologías capaces de organizar, analizar y distribuir datos biológicos eficiéntemente [10]. En este contexto, técnicas de inteligencia artificial, como las metaheurísticas, se han consolidado como herramientas clave en diversos problemas bioinformáticos.

El secuenciado de ADN constituye la base para digitalizar información genética y posibilitar estudios bioinformáticos. De esta actividad se desprenden dos grandes

desafíos. El primero es el *ensamblado de fragmentos de ADN* (DNA-FAP, *DNA Fragment Assembly Problem*), una tarea combinatoria NP-difícil que consiste en recomponer fragmentos de ADN, segmentados durante el secuenciado, sin información sobre el orden o la disposición final. El segundo es el *alineamiento múltiple de secuencias*, orientado a comparar tres o más secuencias relacionadas para identificar similitudes, diferencias y regiones conservadas de interés biológico. En ambos casos, se han obtenido avances significativos mediante metaheurísticas, como los propuestos en [1,7] para el DNA-FAP y en [3,8] para el alineamiento múltiple. Sin embargo, aún existe un amplio margen para el desarrollo de hibridaciones y nuevas variantes que mejoren estos procesos.

Las metaheurísticas también han demostrado utilidad en la clasificación de datos ómicos y biomédicos. Un caso central es la *selección de genes*, donde se busca identificar genes relevantes y descartar los redundantes o ruidosos para mejorar el rendimiento del clasificador. Algunos avances significativos reportados en la literatura son [4,5]. Además, debido al ruido, la alta dimensionalidad y el desbalance de los datos biomédicos, los métodos tradicionales de entrenamiento de redes neuronales (ej. back-propagation) han mostrado dificultades, por lo que las metaheurísticas aparecieron como alternativas, logrando propuestas como [2,22]. Sin embargo, aún existen oportunidades para diseñar nuevas estrategias específicas e híbridas para abordar eficientemente estos problemas.

Otra línea de investigación que ha cobrado gran impulso es el diseño de fármacos orientados a blancos específicos. Se trata de una tarea costosa en términos económicos, técnicos y temporales, lo que resalta la necesidad de estrategias más eficientes. Las metaheurísticas han adquirido un rol central en este campo, ya que pudieron adaptarse a las cualidades específicas de este problema [6,9]. Aun así, se trata de una línea reciente, con un amplio margen para explorar nuevas combinaciones y enfoques, tanto en blancos específicos como genéricos.

En este trabajo se abordan los problemas descriptos mediante metaheurísticas, con especial énfasis en el diseño de nuevos operadores y la integración con métodos de inteligencia artificial. En la Sección 2 se presentan los objetivos, en la Sección 3 se describen los avances obtenidos y en la Sección 4, se presentan las conclusiones planteadas y trabajos futuros.

2 Objetivos

El objetivo general de este trabajo es el diseño y desarrollo de metaheurísticas en combinación con otras técnicas de inteligencia computacional y artificial, con el propósito de contribuir al área de Ciencias de la Computación aplicando el conocimiento desarrollado a problemas de bioinformática.

Tomando como base al objetivo general, se definen los objetivos específicos:

1. Proponer alternativas para el ensamblado de fragmentos de ADN utilizando metaheurísticas combinadas con otras técnicas de búsqueda local.
2. Abordar el problema del alineamiento múltiple de secuencias de ADN, ARN o proteínas a través de metaheurísticas y sus hibridaciones.

3. Implementar metaheurísticas y/o hibridaciones para seleccionar y clasificar grandes conjuntos de genes definidos por sus valores de expresión.
4. Diseñar metaheurísticas para el ajuste fino de parámetros e hiperparámetros de algoritmos de *Machine Learning* aplicados al diagnóstico médico.
5. Proponer aproximaciones metaheurísticas para el diseño y descubrimiento de fármacos novedosos que satisfagan múltiples objetivos.

3 Aproximaciones propuestas y resultados obtenidos

Esta sección reporta los resultados logrados a lo largo del desarrollo del trabajo, alcanzando un 95% de cumplimiento de los objetivos específicos.

Para el problema del DNA-FAP, en [11] se presentó una comparación entre la versión canónica del algoritmo de pájaro cuco (CSA, *Cuckoo Search Algorithm*) y sus hibridaciones con métodos de búsqueda local: el recocido simulado (SA) y la búsqueda local consciente del problema (PALS, *Problem Aware Local Search*). Los experimentos, realizados sobre conjuntos de fragmentos de ADN de distintos tamaños, mostraron que la hibridación con PALS obtuvo el mejor desempeño, destacándose también por su robustez. Los resultados con dos conjuntos de fragmentos se muestran en la Fig. 1.

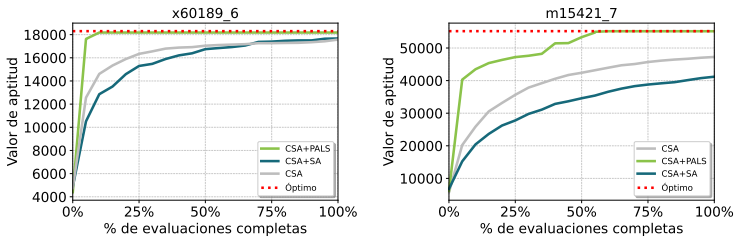


Fig. 1. Evolución del valor de aptitud a medida que se completan las evaluaciones, con los conjuntos de datos x60189_6 y m15421_7.

En [12] se introdujo una variante memética del Algoritmo Genético Celular (AGC) para el alineamiento múltiple de secuencias, aplicado a cadenas de ADN del gen de respuesta de transactivación del virus de HIV y a segmentos del sitio de entrada interno del ribosoma del Picornavirus. Se combina el proceso del AGC con operadores de búsqueda local basados en el movimiento de huecos (discrepancias entre secuencias, producto de la evolución) para mejorar el alineamiento. El enfoque propuesto alcanza soluciones cercanas al óptimo en tiempos de ejecución aceptables, logrando una calidad de solución final igual o incluso mejor que otros algoritmos existentes.

Con respecto a la selección de genes relevantes para la clasificación de muestras cancerígenas se han presentado propuestas con resultados destacados. En [15] se introdujo una variante memética denominada AGCM, que combina el AGC

con un operador de búsqueda local inspirado en la búsqueda por vecindario variable. Esta propuesta fue aplicada a datos de cáncer de colon, linfoma y leucemia, empleando k-vecinos próximos como clasificador. Los resultados, presentados en la Tabla 1, mostraron que el AGCM supera a otras técnicas de la literatura en términos de cantidad de genes mantenidos y precisión de clasificación.

Table 1. Valor promedio del número de genes (entre paréntesis se informa porcentaje del total de genes) y la precisión de clasificación para algoritmo evaluado.

Algoritmo	Colon		Linfoma		Leucemia	
	Precisión	# Genes	Precisión	# Genes	Precisión	# Genes
ACPB	1.00	53.05 (2.65%)	0.90	1100.40 (27.33%)	0.84	2342.80 (32.86%)
AG	1.00	4.60 (0.23%)	0.91	538.75 (13.38%)	0.84	1577.35 (22.13%)
AGC	1.00	11.15 (0.56%)	0.91	643.25 (15.98%)	0.84	1734.70 (24.33%)
AGCM	1.00	12.60 (0.63%)	1.00	51.50 (1.28%)	0.96	175.70 (2.47%)
RS	0.98	5.30 (0.27%)	0.92	278.65 (6.92%)	0.84	1185.90 (16.63%)

En [16] se proponen dos versiones meméticas del Algoritmo Micro-Genético (AMG), cada una con un operador de búsqueda local distinto. La primera variante introduce modificaciones más controladas en la cantidad de genes seleccionados, mientras que la segunda aplica una reducción más agresiva. La variante más agresiva destacó por alcanzar una marcada reducción de genes asociada a valores de precisión competitivos. En [21] se propuso una solución alternativa al problema de la baja cantidad de muestras de expresión génica cancerígenas, utilizando redes generativas adversarias de Wasserstein. El objetivo es aumentar el espacio muestral con muestras artificiales que reproduzcan la distribución de las reales, mejorando la robustez de los métodos estadísticos. Los resultados demostraron que la propuesta logra reproducir adecuadamente el rango de valores y los patrones de expresión de los conjuntos de datos originales.

Asimismo, se han desarrollado propuestas metaheurísticas para optimizar hiperparámetros en métodos de aprendizaje automático orientados al diagnóstico médico. En [13] se optimizaron hiperparámetros de una SVM con kernel Wavelet para retinopatía diabética, comparando al algoritmo genético (AG), el AGC, la evolución diferencial (ED) y el recocido simulado (SA), junto con kernels tradicionales. Los resultados mostraron que la propuesta superó a las configuraciones clásicas de SVM, destacando a AG y AGC como los optimizadores más efectivos. Otros de nuestros trabajos se enfocaron en modificaciones de metaheurísticas para optimizar pesos y sesgos en perceptrones multicapa. En [14] se introdujo una versión modificada del algoritmo de enjambres múltiples de partículas, adaptando la función de actualización de posiciones para evitar la pérdida de información durante el proceso iterativo. En [18] se propuso un nuevo operador de cruza para el AGC, inspirado en el oscilador armónico. En [19] se presentó una variante genética del algoritmo de optimización por hormigas león, incorporando parámetros para regular el balance entre exploración y explotación del espacio de búsqueda. En [17] se diseñó un operador genético que cruza bloques de conocimiento entre perceptrones multicapa, con el fin de identificar representaciones internas complejas que favorezcan la clasificación. En todos los

casos, las variantes propuestas demostraron resultados robustos y que mejoran la calidad de clasificación de las alternativas reportadas en la literatura.

Finalmente, en [20] se propuso al PSOSA, un algoritmo memético que combina el algoritmo de enjambre de partículas (PSO) con el recocido simulado (SA) para el diseño de fármacos con propiedades específicas. La función objetivo consideró tanto la probabilidad de que el fármaco pueda ser suministrado a humanos, medida mediante el QED (*Quantitative Estimation of Druglikeness*), como la accesibilidad sintética (AS) o la facilidad de sintetizar químicamente los compuestos generados. En las comparaciones, mostradas en la Fig. 2, se aprecia que la propuesta superó a otros modelos del estado del arte, incluyendo redes neuronales generativas y metaheurísticas diseñadas para este problema, con grandes probabilidades de ser suministrable a humanos ($0.94 \leq QED \leq 0.95$) y con facilidad de síntesis química ($AS < 2$).

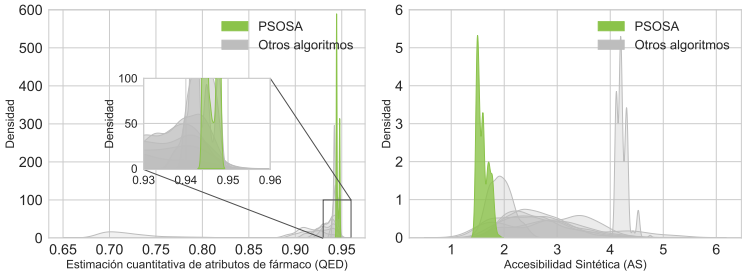


Fig. 2. Valores de atributos de fármaco y accesibilidad sintética logrados por las moléculas generadas por el PSOSA.

4 Conclusiones y trabajos futuros

En este trabajo se desarrollaron distintas estrategias metaheurísticas aplicadas a problemas de bioinformática. Las propuestas incluyeron adaptaciones de algoritmos originales, hibridaciones con técnicas de inteligencia artificial y enfoques multiobjetivo. En conjunto, estas alternativas ofrecieron nuevas perspectivas y soluciones más eficientes para problemas de gran relevancia en la literatura.

Como trabajo futuro, se pretende avanzar en la propuesta de nuevos enfoques multiobjetivos para la selección de genes determinantes en muestras de cáncer y la optimización de hiperparámetros en métodos de aprendizaje automático.

Agradecimientos

El presente trabajo fue realizado mediante el proyecto 80020240100021UN (Res, 1651/2025) de la SIIP-UNCUYO, el PGI 24/N052 de la SGCyT-UNS y los proyectos PIIBA 2022-2023 28720210101070CO y PIP 2025-2027 del CONICET. Los autores agradecen la colaboración de la Dra. Ana Carolina Olivera.

Referencias

1. Abdel-Basset, M., Mohamed, R., Sallam, K.M., Chakrabortty, R.K., Ryan, M.J.: An Efficient-Assembler Whale Optimization Algorithm for DNA Fragment Assembly Problem: Analysis and Validations. IEEE Access **8**, 222144–222167 (2020). https://doi.org/10.1109/ACCESS.2020.3044857
2. Bagchi, J., Si, T.: Artificial Neural Network Training Using Marine Predators Algorithm for Medical Data Classification. In: Tiwari, R., Mishra, A., Yadav, N., Pavone, M. (eds.) Proceedings of International Conference on Computational Intelligence. pp. 137–148. Algorithms for Intelligent Systems, Springer, Singapore (2022). https://doi.org/10.1007/978-981-16-3802-2_11
3. Belattar, K., Zemali, E.A., Baouni, S., Dehni, S.: Parallel multiple DNA sequence alignment using genetic algorithm and asynchronous advantage actor critic model. International Journal of Bioinformatics Research and Applications **18**(5), 460–478 (2022). https://doi.org/10.1504/ijbra.2022.128236
4. Cubillos-Chaparro, J., Dorn, M., Villalobos-Cid, M., Inostroza-Ponta, M.: A Multiobjective Evolutionary Algorithm for Colon Cancer Biomarkers Identification on Gene Expression Data. In: 2024 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB). pp. 1–8. IEEE, Natal, Brazil (Aug 2024). https://doi.org/10.1109/CIBCB58642.2024.10702102
5. Das, A., Neelima, N., Deepa, K., Özer, T.: Gene Selection Based Cancer Classification With Adaptive Optimization Using Deep Learning Architecture. IEEE Access **12**, 62234–62255 (2024). https://doi.org/10.1109/ACCESS.2024.3392633
6. Dong, T., You, L., Chen, C.Y.C.: Multi-objective drug design with a scaffold-aware variational autoencoder. Chemical Science **16**(29), 13352–13367 (2025). https://doi.org/10.1039/D4SC08736D
7. Fang, J.K., Lin, Y.F., Huang, J.H., Chen, Y., Fan, G.M., Sun, Y., Feng, G., Guo, C., Meng, T., Zhang, Y., Xu, X., Xiang, J., Li, Y.: Divide-and-Conquer Quantum Algorithm for Hybrid d e n o v o Genome Assembly of Short and Long Reads. PRX Life **2**(2), 023006 (Apr 2024). https://doi.org/10.1103/PRXLife.2.023006
8. Ibrahim, M.K., Yusof, U.K., Eisa, T.A.E., Nasser, M.: Enhanced Genetic Method for Optimizing Multiple Sequence Alignment. Mathematics **11**(22), 4578 (Jan 2023). https://doi.org/10.3390/math11224578
9. Jain, N., Hornback, A., Wang, M.D.: EMxDesign: A Genetic Algorithm for High Affinity Drug Design. In: Proceedings of the Genetic and Evolutionary Computation Conference Companion. pp. 439–442. GECCO '24 Companion, Association for Computing Machinery, New York, NY, USA (Aug 2024). https://doi.org/10.1145/3638530.3654423
10. Lesk, A.M.: Introduction to bioinformatics, pp. 1–43. Oxford University Press, Oxford, United Kingdom, fifth edition edn. (2019)
11. Rojas, M.G., Carballido, J.A., Olivera, A.C., Vidal, P.J.: Hybrid cuckoo search for solving dna fragment assembly problem. In: IV International Congress of Computer Sciences and Information Systems, IV CICCSI 2020. Universidad Champagnat (11 2020)
12. Rojas, M.G., Carballido, J.A., Olivera, A.C., Vidal, P.J.: A memetic cellular genetic algorithm for multiple sequence alignment. In: 2020 IEEE Biennial Congress of Argentina, ARGENCON 2020. IEEE Inc. (12 2020). https://doi.org/10.1109/ARGENCON49523.2020.9505544
13. Rojas, M.G., Carballido, J.A., Olivera, A.C., Vidal, P.J.: Optimización de support vector machine mediante metaheurísticas para clasificación de retinopatia di-

abética. In: Jornadas Argentinas de Informática e Investigación Operativa, JAIIO 2020. Sociedad Argentina de Informática (10 2020)
14. Rojas, M.G., Carballido, J.A., Olivera, A.C., Vidal, P.J.: A modified multi-swarm optimization for weights and biases of a multi-layer perceptron for medical data classification. In: 2022 IEEE Biennial Congress of Argentina, ARGENCON 2022. pp. 1–8. IEEE Inc. (9 2022). https://doi.org/10.1109/ARGENCON55245.2022.9940066
15. Rojas, M.G., Olivera, A.C., Carballido, J.A., Vidal, P.J.: A memetic cellular genetic algorithm for cancer data microarray feature selection. IEEE Latin America Transactions **18**, 1874–1883 (11 2020). https://doi.org/10.1109/TLA.2020.9398628
16. Rojas, M.G., Olivera, A.C., Carballido, J.A., Vidal, P.J.: Memetic micro-genetic algorithms for cancer data classification. Intelligent Systems with Applications **17**, 200173 (1 2023). https://doi.org/10.1016/j.iswa.2022.200173
17. Rojas, M.G., Olivera, A.C., Carballido, J.A., Vidal, P.J.: A new micro-genetic crossover operator for effective training of medical neural networks. Springer Nature Computer Sciences **6**, 539 (6 2025). https://doi.org/10.1007/s42979-025-04077-z
18. Rojas, M.G., Olivera, A.C., Vidal, P.J.: Optimising multilayer perceptron weights and biases through a cellular genetic algorithm for medical data classification. Array **14**, 100173 (7 2022). https://doi.org/10.1016/j.array.2022.100173
19. Rojas, M.G., Olivera, A.C., Vidal, P.J.: A genetic operators-based ant lion optimiser for training a medical multi-layer perceptron. Applied Soft Computing **151**, 111192 (1 2024). https://doi.org/10.1016/j.asoc.2023.111192
20. Rojas, M.G., Olivera, A.C., Vidal, P.J.: Memetic Genetic Particle Swarm Optimization for Druglike Molecule Discovery. IEEE Latin America Transactions **23**(3), 216–222 (Mar 2025). https://doi.org/10.1109/TLA.2025.10879174
21. Rojas, M.G., Olivera, A.C., Vidal, P.J., Carballido, J.A.: Massive Production of Cancer Synthetic RNA-Seq Gene Expression Samples. SN Computer Science **6**(6), 714 (8 2025). https://doi.org/10.1007/s42979-025-04251-3
22. Si, T., Bagchi, J., Miranda, P.B.C.: Artificial Neural Network training using metaheuristics for medical data classification: An experimental study. Expert Systems with Applications **193**, 116423 (May 2022). https://doi.org/10.1016/j.eswa.2021.116423
23. Zhang, Y., Cheng, L., Chen, G., Alghazzawi, D.: Evolutionary Computation in bioinformatics: A survey. Neurocomputing **591**, 127758 (Jul 2024). https://doi.org/10.1016/j.neucom.2024.127758

Valoración de viviendas urbanas en Argentina mediante técnicas de machine learning e interpretabilidad.

Emiliano Gutiérrez[1,2][0000−0002−6424−996X], Laura Lanzarini[3][0000−0001−7027−7564], Rocío Cecchini[1,4][0000−0003−4214−6852], and Fernando Delbianco[2,5][0000−0002−1560−2587]

[1] Instituto de Ciencias e Ingeniería de la Computación (ICIC-UNS CONICET)
[2] Departamento de Economía. Universidad Nacional del Sur (UNS)
[3] Instituto de Investigación en Informática LIDI. Facultad de Informática (UNLP-CIC)
[4] Departamento de Ciencias e Ingeniería de la Computación. Universidad Nacional del Sur (UNS)
[5] Instituto de Matemática de Bahía Blanca (INMABB-UNS CONICET)

Resumen Este trabajo explora la predicción de precios de apartamentos en la Ciudad de Buenos Aires a partir de más de 60 000 avisos de venta combinados con datos geográficos y de movilidad. Se entrenaron modelos de machine learning (XGBoost, Random Forest, redes neuronales, LASSO, CART) y se compararon con una regresión lineal. XGBoost logró el mejor desempeño, siendo las características del inmueble los variables más relevantes. La aplicación de conformal prediction permitió evaluar la incertidumbre y evidenció mayor variabilidad en propiedades de gran tamaño, aportando información útil para políticas de vivienda y análisis del mercado inmobiliario.

Keywords: Mercado inmobiliario · Conformal prediction· Argentina

1. Introducción

La vivienda suele ser el activo de mayor valor en el patrimonio de un hogar, caracterizándose por dos aspectos particulares: su independencia de la inserción laboral de los propietarios y su capacidad de ser heredada [6].

En el caso Argentino, el mercado inmobiliario presenta profundas imperfecciones, junto con un significativo grado de desregulación [1]. Es característico que los precios de la mayoría de las viviendas en venta estén valuados en dólares estadounidenses [4]. Su dinámica ha mostrado una marcada expansión desde principios del siglo XXI[13].

Tradicionalmente, la valoración de un inmueble se realizaba mediante modelos de precios hedónicos, que asumían una relación lineal entre el precio y las características de la vivienda. Sin embargo, con la disponibilidad actual de algoritmos de machine learning, es posible formular modelos basados en relaciones no lineales mejorando significativamente la capacidad predictiva.

Se postula como objetivo la aplicación de técnicas machine learning para la generación y evaluación de modelos para predicción del precio de propiedades, enfocándose en el precio de apartamentos en la Ciudad de Buenos Aires (Argentina). Concretamente, se busca recuperar los anuncios de venta publicados en línea y combinar esta información con datos de diversas bases de datos públicas, con el fin de formular modelos que permitan estimar el precio final de un apartamento.

Dada la naturaleza de "caja negra" de algunos de estos algoritmos, se aplicarán técnicas de interpretabilidad para identificar el impacto de las variables explicativas utilizadas en las predicciones, permitiendo comprender mejor cómo cada factor influye en el valor estimado [9, 12].

Además, para cuantificar la incertidumbre asociada a las predicciones generadas por los modelos, se empleará un enfoque de conformal prediction, a fin de delimitar intervalos de confianza a partir de las predicciones realizadas [5, 15].

2. Datos y metodología

El modelo de precios hedónicos se basa en la idea de que el valor final de un bien puede explicarse por los precios implícitos de sus atributos. Estos precios se revelan a partir de la formulación de una función de precio p, que depende de z atributos de la forma $(p(z) = p(z_1, z_2, ..., z_i)$. Esta relación entre precio y características fue abordada previamente por Lancaster [7]. Problemas como la incorrecta especificación del modelo, la heterocedasticidad y la presencia de outliers pueden obstaculizar la efectividad de una regresión hedónica para lograr predicciones precisas [16]. Por esta razón, el uso de algoritmos de machine learning se presenta como una alternativa complementaria, capaz de explorar relaciones no lineales entre los atributos de una vivienda y su precio [10, 2].

No obstante, la literatura en Argentina es escasa en cuanto a aportes que vinculen la valorización final de un inmueble mediante machine learning.

En lo que respecta a los inmuebles relevados, estos fueron extraídos del portal *Mercado Libre* mediante la implementación de un algoritmo de web scraping. Se consideraron solamente aquellos departamentos en venta dentro de la ciudad de Buenos Aires. La fecha de extracción de la información corresponde al 11 de agosto de 2023. Tras su preprocesamiento, la cantidad de observaciones válidas fue de 63,017. La totalidad del dataset se encuentra disponible en este un repositorio de Github.Asimismo un análisis geolocalizable puede consultarse en este link, donde se presenta el precio por m^2 mediante un *heatmap*. Los atributos analizados se agruparon en cuatro categorías detalladas en la Tabla 1. Una descripción pormenorizada de cada variable utilizada se encuentra disponible en este link.

Así, se parte de la idea de que el modelo a desarrollar tiene una función dependiente de cuatro grupos de variables, con la siguiente forma funcional:

$$\log(\text{precio}) = f(\text{Características del inmueble, Barrio, Proximidad geográfica, Movilidad}) \quad (1)$$

Tabla 1: Variables utilizadas.

Categoría	Cantidad de atributos	Descripción	Fuente
Características del inmueble	46	Características del inmueble asociadas a su precio, tamaño, habitaciones y amenidades	Avisos de Mercado Libre
Barrio	48	Barrio en donde el departamento se halla localizado	Buenos Aires Data
Proximidad geográfica	7	Distancia en metros a cárceles, hospitales, subte, tren, escuelas y espacios verdes	Instituto Geográfico Nacional
Movilidad	6	Registros de movilidad en sectores de las comunas	Google

El modelo se estimará utilizando distintos enfoques: regresión lineal por Mínimos Cuadrados Ordinarios (MCO), Classification and Regression Trees (CART), Random Forest (RF), XGBoost, Least Absolute Shrinkage and Selection Operator (LASSO) y Red Neuronal Artificial (RNA). Asimismo, dado que algunos algoritmos de machine learning, aunque presentan mayor flexibilidad al formular sus predicciones, enfrentan una crítica significativa: la falta de explicabilidad de sus resultados, debido a su naturaleza de çaja negra"[17]. Para ello, se recurrirá al uso de . Shapley Additive Explanations (SHAP), como técnica de modelización agnóstica que proveerá interpetabilidad en los resultados obtenidos.

3. Principales resultados

El conjunto de entrenamiento para la estimación de los modelos entrenados representó el 80 % de la totalidad de observaciones ascendiendo a un total de 51241 avisos, mientras que el conjunto de testeo compuesto por el 20 % de publicaciones restantes posee 9042 observaciones. Como estrategia de remuestreo sobre los datos de entrenamiento se utilizó validación cruzada, con 10 pliegues.

Todo el flujo de trabajo de los datos utilizados, fue realizado mediante el lenguaje Python, siendo utilizadas las librerías *statsmodels* [14]*scikit-klearn*[11], *shap*[8] y *MAPIE* [3].

Para la calibración de los modelos basados en árboles (CART, RF y XGBoost), la seleccion de hiperparámetros fue evaluada a partir de combinaciones de profundidad máxima, número de árboles, cantidad mínima de observaciones por nodo y parámetros de regularización. Para RNA , se aplicó optimización bayesiana considerando variaciones en el número de neuronas de la capa oculta, función de activación, tasa de aprendizaje, algoritmo de optimización, tamaño de lote y número de épocas. En el modelo LASSO, se analizaron 100 valores de λ en una escala logarítmica desde 10^{-5} hasta 10^{5}. Los hiperparámetros óptimos se determinaron mediante validación cruzada de 5 pliegues, priorizando la minimización del error cuadrático medio (MSE). En la Tabla 2 resume los valores seleccionados para cada modelo.

Tabla 2: Hiperparámetros óptimos obtenidos para cada modelo

Modelo	**Hiperparámetros óptimos**
CART	Profundidad máxima = 5; Mínimo de observaciones por división = 1000; Mínimo de observaciones por hoja = 5.
Random Forest	Número de árboles = 2000; Profundidad máxima = 20; Máximo de atributos por árbol = 10; Mínimo de observaciones por división = 1000.
XGBoost	Número de árboles = 1000; Tasa de aprendizaje = 0.1; Profundidad máxima = 20; Regularización $\gamma = 0.5$.
LASSO	λ óptimo = 2.5×10^{-5}.
Red Neuronal (RNA)	Neuronas en capa oculta = 64; Función de activación = sigmoide; Tasa de aprendizaje = 0.001; Optimizador = Adam; tamaño de batch = 64; Número de épocas = 1000.

Los resultados de la Tabla 3 muestran que los modelos de machine learning superaron en desempeño a las aproximaciones lineales en la predicción del precio de departamentos en la Ciudad de Buenos Aires. El mejor rendimiento correspondió a XGBoost, seguido por LASSO, RF y RNA, mientras que MCO y CART presentaron menor capacidad de generalización.

Tabla 3: Desempeño de los modelos entrenados en los conjuntos de entrenamiento y testeo.

Modelo	Entrenamiento					Testeo				
	MSE	RMSE	MAE	MAPE	R^2	MSE	RMSE	MAE	MAPE	R^2
MCO	0,090	0,301	0,224	0,019	0,816	0,403	0,635	0,478	0,041	-0,375
CART	0,101	0,319	0,241	0,020	0,794	0,115	0,339	0,244	0,021	0,608
LASSO	0,090	0,301	0,224	0,019	0,816	0,102	0,319	0,228	0,019	0,653
Random Forest	0,086	0,294	0,218	0,018	0,825	0,103	0,322	0,226	0,019	0,647
XGBoost	0,031	0,176	0,136	0,011	0,937	0,080	0,283	0,203	0,017	0,726
RNA	0,066	0,257	0,196	0,016	0,865	0,084	0,290	0,207	0,018	0,714

El análisis de interpretabilidad permitió identificar los factores que más influyen en el valor de los inmuebles. Tal como se observa en la Figura 1, que presenta la descomposición SHAP de la contribución de cada atributo, la superficie fue las variable con mayor peso en la formación del precio junto con la presencia de amenidades como balcón, cochera, pileta o gimnasio. Otros factores de entorno y la proximidad a transporte público, acceso a espacios verdes, revelarían un impacto menos relevante.

Al momento de cuantificar la incertidumbre de las predicciones mediante conformal prediction, fue utilizado el enfoque de Jackknife siguiendo un remuestreo de tipo Leave-One-Out, sobre el modelo de XGboost. Dada la necesidad

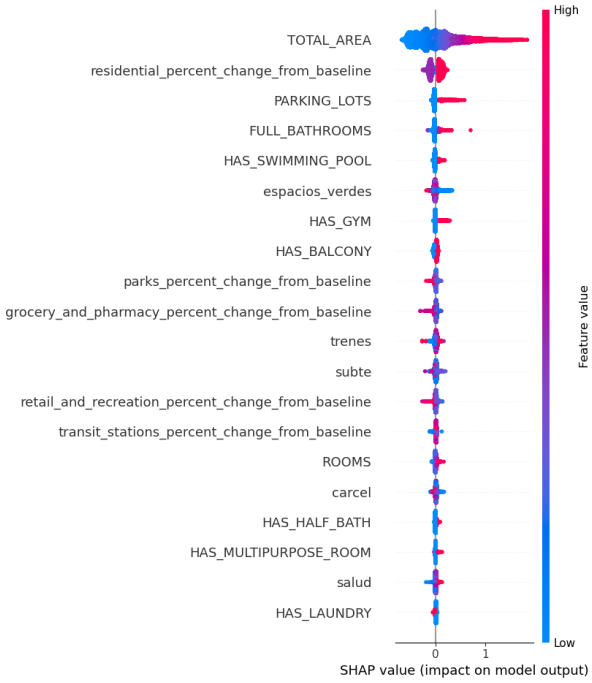


Figura 1: Valores SHAP (XGboost)
Fuente: Elaboración propia.

de hacer uso de un conjunto de calibración y otros de evaluación, los datos de testeo en dos conjuntos, siendo entonces que 1086 observaciones corresponden a calibración y 7956 al conjunto de predicción. A fin de evitar sesgos en los resultados, los datos de entrenamiento del modelo no fueron utilizados. Por otra parte el porcentaje de cobertura esperado debería ser del 90 % dado que el valor de confianza a utilizar (α) correspondió a 0,1. Como non-conformity score S_i, se utilizó el valor de los residuos, es decir la diferencia en valor absoluto entre el valor real y el predecido.

La cobertura global obtenida, fue de 91,38 %. Asimismo En la Tabla 4, se presentan los resultados condicionales. La cobertura en el caso del tamaño del inmueble como de la cantidad de baños, fue por debajo del 90 %, para los segmentos medios y altos. Cuando los valores de estos atributos son menores, el intervalo de confianza se encuentra dentro de la predicción real en una mayor proporción.

4. Conclusiones

Los hallazgo obtenidos aportan evidencia cuantitativa y socialmente relevante para la política habitacional, ya que la vivienda constituye una de las decisiones de inversión más significativas para los hogares.

Tabla 4: Cobertura condicional (Jacknife).

Variable	Cobertura		
	Bajo	Medio	Alto
TOTAL_AREA	0,916	0,760	0,400
FULL_BATHROOM	0,924	0,890	0,763
espacios verdes	0,925	0,893	0,932
residential_percent_change_from_baseline	0,801	0,911	0,943
subte	0,931	0,905	0,862

Variable	Amplitud media del intervalo		
	Bajo	Medio	Alto
TOTAL_AREA	0,776	0,780	0,786
FULL_BATHROOM	0,776	0,778	0,779
espacios verdes	0,776	0,776	0,778
residential_percent_change_from_baseline	0,778	0,776	0,775
subte	0,776	0,776	0,777

Los modelos entrenados, dieron cuenta que la utilización de machine learning, presenta una capacidad predictiva superior a las estimaciones tradicionales. Por limitaciones de tiempo y dado el carácter exploratorio del estudio, no se aplicó validación cruzada espacial. Se reconoce esta limitación y se prevé su incorporación en trabajos futuros.

Asimismo, la adopción de SHAP mostró que las características físicas del inmueble (superficie, cantidad de baños y estacionamientos) son los factores de mayor influencia, con impacto positivo sobre el precio.

Además, la aplicación de conformal prediction permitió evaluar la incertidumbre de las predicciones, revelando que los departamentos de gran superficie presentan coberturas más bajas y, por ende, predicciones menos precisas.

Finalmente cabe destacar que los precios utilizados provienen de publicaciones activas y no de transacciones efectivas, lo cual podría introducir un sesgo de sobreestimación. Asimismo, al tratarse de un corte temporal único, no se aplicó validación temporal; futuras extensiones incluirán modelos con actualización periódica para captar tendencias.

Disclosure of Interests. Los autores declaran no tener ningún conflicto de intereses en relación con este manuscrito. Este estudio reconoce los posibles impactos éticos del uso de modelos predictivos en el mercado inmobiliario, particularmente en relación con la gentrificación y la desigualdad territorial. No se utilizaron variables sensibles ni proxies de variables socioeconómicas individuales. Los datos empleados son públicos y anonimizados.

Bibliografía

[1] Luis Baer and Mark Kauw. Mercado inmobiliario y acceso a la vivienda formal en la ciudad de buenos aires, y su contexto metropolitano, entre 2003 y 2013. *Eure (Santiago)*, 42(126):5–25, 2016.

[2] JM Caridad y Ocerin and N Ceular Villamandos. Un análisis del mercado de la vivienda a través de redes neuronales artificiales. *Estudios de economía aplicada*, 18(2):67, 2001.

[3] Thibault Cordier, Vincent Blot, Louis Lacombe, Thomas Morzadec, Arnaud Capitaine, and Nicolas Brunel. Flexible and Systematic Uncertainty Estimation with Conformal Prediction via the MAPIE library. In *Conformal and Probabilistic Prediction with Applications*, 2023.

[4] Alejandro Gaggero and Pablo Nemiña. El origen de la dolarización inmobiliaria en la argentina. *Sociales en debate*, (5), 2013.

[5] Isaac Gibbs, John J Cherian, and Emmanuel J Candès. Conformal prediction with conditional guarantees. *arXiv preprint arXiv:2305.12616*, 2023.

[6] Karin Kurz. *Home ownership and social inequality in comparative perspective*. Stanford University Press, 2004.

[7] Kelvin J Lancaster. A new approach to consumer theory. *Journal of political economy*, 74(2):132–157, 1966.

[8] Scott M Lundberg and Su-In Lee. A unified approach to interpreting model predictions. *Advances in neural information processing systems*, 30, 2017.

[9] Christoph Molnar. *Introduction to Conformal Prediction with Python: A Short Guide for Quantifying Uncertainty of Machine Learning Models*. Christoph Molnar c/o MUCBOOK, Heidi Seibold, 2023.

[10] Timothy Oladunni and Sharad Sharma. Hedonic housing theory—a machine learning investigation. In *2016 15th IEEE International Conference on Machine Learning and Applications (ICMLA)*, pages 522–527. IEEE, 2016.

[11] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al. Scikit-learn: Machine learning in python. *Journal of machine learning research*, 12(Oct):2825–2830, 2011.

[12] Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. Model-agnostic interpretability of machine learning. *arXiv preprint arXiv:1606.05386*, 2016.

[13] Juan Pablo del Río, Federico Langard, and Diego José Arturi. La impronta del mercado inmobiliario en el período neodesarrollista. *Realidad Económica*, 283, 2014.

[14] Skipper Seabold and Josef Perktold. statsmodels: Econometric and statistical modeling with python. In *9th Python in Science Conference*, 2010.

[15] Glenn Shafer and Vladimir Vovk. A tutorial on conformal prediction. *Journal of Machine Learning Research*, 9(3), 2008.

[16] Mahdieh Yazdani. Machine learning, deep learning, and hedonic methods for real estate price prediction. *arXiv preprint arXiv:2110.07151*, 2021.

[17] Qingyuan Zhao and Trevor Hastie. Causal interpretations of black-box models. *Journal of Business & Economic Statistics*, 39(1):272–281, 2021.

Forecasting Spatio-Temporal Time Series with Missing Values Using Graph Neural Networks

Armando Martinez-Ruiz[1][0009−0005−7343−7339],
Pilar Gomez-Gil[1][0000−0003−1550−6218], and
Rigoberto Fonseca-Delgado[2][0000−0002−8890−3911]

[1] Department of Computer Science, National Institute of Astrophysics, Optics and Electronics, México {armandomtzruiz, pgomez}@inaoep.mx
[2] School of Mathematical and Computational Sciences, Yachay Tech University, Ecuador rfonseca@yachaytech.edu.ec

Abstract. Spatio-temporal data is fundamental in applications as traffic prediction, energy demand or environmental monitoring. Although deep learning models, such as RNNs and LSTMs can capture temporal dependencies, they tend to ignore the graph structure that describe spatial relations. Graph Neural Networks (GNNs) and their Spatio-Temporal variants (ST-GNNs) approach this limitation, but their performance degrades when there are missing values. To address this limitation, our work proposes a unified ST-GNN algorithm which approaches jointly forecasting tasks in time series with missing values. The proposed approach will integrate imputation techniques and spectral based regularization within the model. The evaluation will be done using standard baseline datasets (METR-LA, PEMS-BAY, AQI) under different levels of missing data. The expected contributions include: i) a learning algorithm for ST-GNNs which integrates imputation and prediction tasks, ii) an empirical analysis of the effect of imputation on the accuracy of predictions and iii) the creation of open source ST-GNN library for reproducible results. Preliminary results with synthetic data show that spatial modeling improves forecasting robustness over baseline models such as ARIMA and LSTM, highlighting the potential of the proposed approach for real-world data with missing values.

Keywords: Spatio-temporal forecasting · Graph Neural Networks · Spatio-Temporal Graph Neural Networks · Missing data handling.

1 Introduction

Many critical areas in science and engineering as traffic analysis, urban transport planning or energy demand utilizes spatio-temporal data. In the last decade, deep learning has been a top choice for the analysis of this kind of data, particularly through architectures like Recurrent Neural Networks (RNNs) and Long Short-Term Memory units (LSTMs) [1], which have proven to be effective to capture complex temporal patterns. However, these architectures usually ignore

the inherent graph structure that reflects the spatial relations between the observed entities, which limits their capacity to model phenomena where spatial dependency is as relevant as temporal dependency.

To answer these limitations, Graph Neural Networks (GNNs) [2] and their variant, Spatio-Temporal Graph Neural Networks (ST-GNNs) [3], were designed to handle the spatial processing in forecasting tasks by considering the inherent graph structure associated to the data. This paradigm offers the jointly processing of spatial and temporal dependencies in forecasting tasks based on a message-passing scheme between the nodes of the graph. However, an important issue arises in real world applications: the presence of missing values. These missing values can be caused due to failures in the sensor readings , perturbed transmissions or human mistakes. This incompleteness of the data may be solved with imputation methods, but in spatiotemporal forecasting tasks affects directly the representations of the graph, with weakened connections between nodes that negatively affect the message-passing process between the nodes of the graph [5]. Treating spatio-temporal forecasting and imputation as separate tasks limits real-world application, as forecasting models often assume complete data, which is unrealistic.

We propose a learning algorithm for ST-GNNs that integrates missing value handling directly into forecasting. This approach, using spectral regularization and imputation-based techniques, aims to enhance accuracy and build more robust models, outperforming baselines in critical contexts like traffic, energy, and environmental monitoring.

2 Research Proposal

Problem Statement The analysis of spatiotemporal data is fundamental in critical areas like urban traffic, energy demand and environmental monitoring. Although deep learning models such as RNNs and LSTMs have shown effectiveness to capture temporal dependencies, they ususally ignore the underlying graph structure that describe the spatial relations between entities. GNNs and their variants ST-GNNs have been designed to be more suited alternatives to integrate spatial and temporal dependencies. However, their performance degrades when they face missing data, which is a common scenario in real world applications due to failure in sensors, transmission errors or incomplete registers. The missing data can affect the connectivity of the graph negatively, which limits the message passing between the nodes. So far, forecasting and imputation tasks in time series through ST-GNNs have been treated independently, which leads to solutions that do not represent real conditions.

General Objective Develop a learning algorithm for ST-GNNs that handles partial observations for forecasting tasks, integrating training strategies to improve robustness and accuracy in time series forecasting with missing values while leveraging inherent spatial dependencies.

Expected contributions Our work expects to have the following contributions: Develop a learning algorithm for ST-GNNs which handles partial observations and forecasting tasks in time series, an empirical analysis of the impact of imputation on downstream forecasting tasks and the integration of training strategies based on spectral based regularization and imputation to enhance the robustness of the model.

3 Methodology

The proposed approach is to implement a ST-GNN with the Python library *tsl* [11], which is designed to define graph neural networks for spatio-temporal data processing. Fig. 1 shows the basic architecture of a ST-GNN; the initial approach towards mitigating the missing value issue will be to use a Transformers approach[12] to model the temporal dependencies while leveraging the self-attention modeling of time series imputation. The spatial modelling is planned to follow the graph convolution scheme focused on spectral based methods . The METR-LA, PEMS-BAY and AQI [6] datasets will be used for training, testing and validation of the proposed architecture in missing values scenarios, where metrics like mean absolute error (MAE), mean absolute percentage error (MAPE) and root mean squared error (RMSE) will be used for this task. The missing values will be added to the datasets in various levels of missingness (20%, 30%, 40% and 50% of missing values) to assess the robustness of the proposed approach.

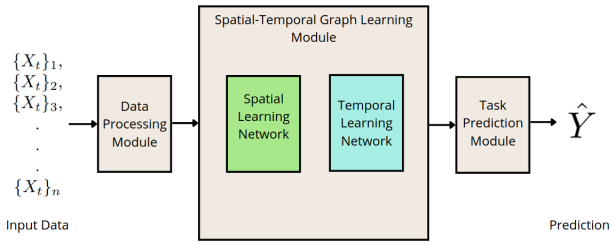


Fig. 1. Basic architecture of a Spatio-Temporal Graph Neural Network (ST-GNN) for time series forecasting.

There will be three kinds of baselines: traditional time series models (ARIMA) [7], deep learning models for imputation and forecasting (RNN, LSTM) [8], GNN models for spatio-temporal data (STGCN [3], Graph Wavenet [4], TMF-GNN [9]).

4 Preliminary Results

The goal of the current experiments is to become familiar with the ST-GNN model in forecasting tasks utilizing synthetic datasets to evaluate and compare its performance against traditional baseline models. It is important to mention that these results have been accepted to be presented at the Mexican International Conference on Artificial Intelligence (MICAI) 2025 [10]

4.1 Datasets

The datasets were designed following a sinusoidal function with additive Gaussian noise, which can be represented with Equation 1. Each time series simulates the behavior of a node in a graph over time, with individual patterns designed to introduce varying levels of correlation and complexity.

$$x_i(t) = \sin(\frac{10t}{T} + i) + \epsilon_i(t), \text{for } i \in \{1, 2, 3, 4\}, \tag{1}$$

where $T = 1000$ are the time steps, $x_i(t)$ is the value of the time series associated to node i at time step t and $\epsilon_i(t) \sim \mathcal{N}(0, \sigma^2)$.

An additional time series, $x_0(t)$, was constructed using three different schema to reflect varying degrees of inter-node dependency: Sum-based dependency ($x_0(t) = \sum_{i=1}^{4} x_i(t)$), Mean-based dependency ($x_0(t) = \frac{1}{4}\sum_{i=1}^{4} x_i(t)$) and Product-based dependency ($x_0(t) = \prod_{i=1}^{4} x_i(t)$).

These settings introduce an explicit way to represent the dependency structure, considering linear and non linear relations, allowing us to test the ability of the model to capture node relationships and exploit spatial correlations in time series. The graph structure was defined based on the Pearson correlation coefficient between each time series with time series $x_0(t)$; the weight of each edge represents statistically significant correlations.

4.2 Discussion

Table 1 summarizes the prediction accuracy in all synthetic settings. For deep learning models (LSTM and ST-GNN), the reported values correspond to the mean and standard deviation over 10 independent runs. For the ARIMA model, forecasts were generated independently for each of the five time series, and the reported metric corresponds to the average performance with the overall standard deviation.

The results indicate that ST-GNN clearly outperforms ARIMA in all synthetic scenarios. Compared with LSTM, both models achieve comparable performance overall. However, ST-GNN tends to achieve lower MAE values, while LSTM achieves lower RMSE values. This tendency may suggest that LSTM is more robust to occasional outliers, whereas ST-GNN provides predictions with smaller and more consistent absolute errors. Fig. 2 shows the results of forecasting in the mean-based dependency setting.

Table 1. Forecasting performance of ARIMA, LSTM, and ST-GNN models (mean ± std) on three synthetic scenarios where Node 0 is the mean, sum or product of other time series.

Setting	Model	RMSE	MAE
Mean	ST-GNN	0.098 ± 0.0003	**0.076 ± 0.0003**
	ARIMA	1.073 ± 0.640	1.038 ± 0.650
	LSTM	**0.091 ± 0.0002**	0.091 ± 0.002
Sum	ST-GNN	0.131 ± 0.0005	**0.098 ± 0.0005**
	ARIMA	1.230 ± 0.450	1.190 ± 0.480
	LSTM	**0.117 ± 0.001**	0.116 ± 0.001
Product	ST-GNN	0.095 ± 0.001	**0.074 ± 0.001**
	ARIMA	1.053 ± 0.667	1.017 ± 0.679
	LSTM	**0.087 ± 0.001**	0.087 ± 0.001

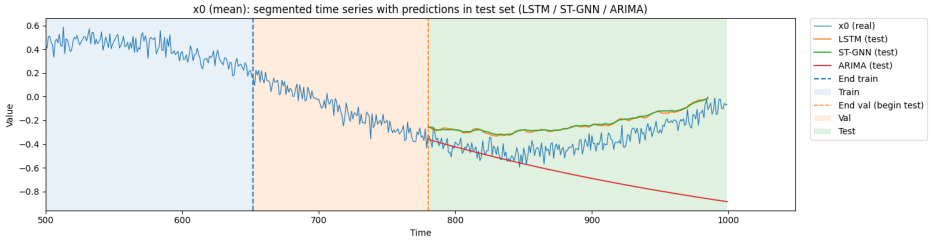


Fig. 2. Forecasting results for the dependent time series under the mean-based dependency setting [10]. The plot illustrate the ground truth versus model predictions.

5 Expected Results and Potential Challenges

This research aims to contribute to the field of spatio-temporal learning through the design of a unified ST-GNN-based algorithm for time series forecasting with missing values. The proposed approach is expected to improve accuracy in scenarios with incomplete data, while providing insights into the impact of imputation on forecasting tasks. In addition, the use of training techniques based on spectral regularization is expected to enhance model robustness. An open-source ST-GNN library for reproducible results is also envisioned, making the contribution directly useful to the research community.

Nevertheless, the project faces several challenges. A major concern is the high computational complexity of training ST-GNNs, which may become problematic when the graph size exceeds 10,000 nodes. Another difficulty lies in ensuring generalization across domains and scaling the model for real-time applications. These challenges, however, also open opportunities to explore innovative optimization, validation, and training methodologies for spatio-temporal learning models.

References

1. I. D. Mienye, T. G. Swart, and G. Obaido: Recurrent neural networks: A comprehensive review of architectures, variants, and applications. Information (Basel) **15**, pp. 517, 2024.
2. F. Scarselli: The graph neural network model. IEEE Transactions on Neural Networks **20**(1), 2021.
3. B. Yu, H. Yin, and Z. Zhu: Spatio-Temporal graph convolutional networks: A deep learning framework for traffic forecasting. Proceedings of the Twenty- Seventh International Joint Conference on Artificial Intelligence, 2018.
4. Z. Wu, S. Pan, G. Long, J. Jiang, and C. Zhang: Graph wavenet for deep spatial-temporal graph modeling. In Proceedings of the 28th International Joint Conference on Artificial Intelligence (IJCAI'19), 2019
5. M. Jin: A survey on graph neural networks for time series: Forecasting, clas- sification, imputation and anomaly detection. IEEE Transactions on Pattern Analysis and Machine Intelligence **46**(12), pp 10466-85, 2024.
6. Y. Li, R. Yu, C. Shahabi, and Y. Liu, "Diffusion convolutional recurrent neu- ral network: Data-driven traffic forecasting," in International Conference on Learning Representations, 2018
7. P. J. Brockwell and R. A. Davis, Time Series: Theory and Methods. Springer, 2013.
8. A. Casolaro, V. Capone, G. Iannuzzo, and F. Camastra, "Deep learning for time series forecasting: Advances and open problems," Information (Basel),**14**, p. 598, Nov. 2023.
9. S. Kim, "Tmf-gnn: Temporal matrix factorization-based graph neural network for multivariate time series forecasting with missing values," Systems with Ap- plications, vol. 275, 2025.
10. Martinez-Ruiz, A., Gomez-Gil, P., Fonseca-Delgado, R. (2025). An Analysis of Spatio-Temporal Graph Neural Networks Based on Synthetic Time Series with Known Structural Dependencies. In: Martínez-Villaseñor, L., Vázquez, R.A., Ochoa-Ruiz, G. (eds) Advances in Soft Computing. MICAI 2025 Posters Track. MICAI 2025. Communications in Computer and Information Science, vol 2712. Springer, Cham. https://doi.org/10.1007/978-3-032-08704-1_24
11. PyG Documentation x2014; pytorch geometric documentation — pytorch- geometric.readthedocs.io." https://pytorch-geometric.readthedocs.io/ en/latest/. [Accessed 20-06-2025]
12. A. Y. Yıldız, E. Koç and A. Koç, "Multivariate Time Series Imputation With Transformers," in IEEE Signal Processing Letters, vol. 29, pp. 2517-2521, 2022, doi: 10.1109/LSP.2022.3224880

Hybrid Feature extraction in thermographic images for the detection of abnormal structures in breast tissues

Yareli Aburto Sánchez[1 [0009−0002−0091−8455]], María del Pilar Gómez Gil,[1 [0000−0003−1550−6218]], and Leopoldo Altamirano Robles[1 [0000−0003−0965−6420]]

Department of Computer Science, National Institute of Astrophysics, Optics and Electronics, México {yarelia,pgomez,robles}@inaoep.mx

Abstract. Breast cancer remains a predominant cause of mortality among women. On the other hand infrared thermography has emerged as a non-invasive alternative to mammography but its application in practical settings is limited by fluctuations in accuracy and other reliability issues. These challenges are attributed, among other factors, to the variability of thermal patterns influenced by physiological fluctuations, individual differences, and external conditions, complicates the identification malignant lesions, resulting in a thermogram interpretation highly dependent on the experience and skills of human experts, and in this way limiting their utility as an autonomous diagnostic aid. This dissertation aims to develop an innovative approach designed to enhance the robustness of automatic abnormal structure detection in breast thermographic images. The proposed methodology will incorporate hybrid feature engineering, combining both manually-designed features and features automatically extracted from deep foundation models, alongside transfer learning and data augmentation strategies to mitigate the challenges posed by data sparsity and inter-patient variability inherent to thermal image features. Specifically, we propose an innovative strategy termed “focused feature learning" which involves deriving thermal features from a RadiImageNet neural network model, fine-tuned through various phases of transfer learning, to improve both sensitivity and specificity performance in thermogram classification.

Keywords: Breast Cancer· Classification · Feature extraction · RadImageNet· Focused feature learning · Thermography.

1 Introduction

Early detection of breast cancer is crucial for public health,in 2022, México documented approximately 23,790 cases in women over 20 years old, with an incidence rate of 51.92 per 100,000 and a mortality rate of 17.48 per 100,000 [5]. Although mammography is the diagnostic gold standard, it has drawbacks, particularly for

younger women and those with dense breast tissue, due to lower sensitivity, potential discomfort, and radiation exposure. Infrared thermography (IRT) offers a non-invasive, radiation-free alternative that identifies tumor-induced physiological changes through analysis of skin thermal profiles, making it applicable to patients of all ages [7, 8, 11, 10]. However, the automatic analysis of thermographic images faces substantial challenges, such as high physiological variability, limited datasets, and data drift/shift that impede model generalization. Therefore, this research seeks a unified strategy to enhance the automatic detection of abnormal structures in breast thermograms.

2 Research Proposal

Problem Statement Detecting breast cancer remains a pressing public health issue. Infrared thermography is a promising non-invasive technique. However, variability in thermal signatures and limited data hinder the development of robust, generalizable models. Data drift and data shift [1] evolving data patterns and distribution changes create further obstacles. These problems highlight the need for adaptive computational methods for a more reliable identification of breast cancer in thermographic images.

General Objective The primary goal of this doctoral research is to create an integrated learning approach that combines hybrid feature engineering with targeted transfer learning, thereby increasing the sensitivity and specificity of breast abnormality detection in thermography compared to current state-of-the-art models.

Expected Contributions Our work is expected to have the following contributions:

- **A novel hybrid feature engineering framework** that integrates manually crafted features with features extracted via transfer learning techniques.
- **The adaptation and assessment of foundational models** specialized in the field of medical imaging to breast thermography images, by transfer learning strategies.
- **A Generative Adversarial Network based on WGAN-GP** able to produce high-fidelity synthetic breast thermography images.
- **A set of human-designed and deep-generated image features** able to represent essential thermal information will improve the sensitivity and specificity of an automatic classifier for breast cancer.

[1] Data drift involves changes in the statistical properties of data over time, often due to protocol variations, population shifts, or equipment updates, impacting model performance in real settings [6]. Data shift refers to differences in data distributions between training and deployment environments, such as the case of using a model trained on limited datasets in diverse clinical contexts [13].

3 Related work

Advances in breast cancer detection using thermography. Several studies have utilized deep neural networks to analyze thermographic images for the early detection of breast cancer. Abdulla et al. [2] employed Inception V3 and V4 models, achieving high accuracy but reporting that their small dataset size limited the generalization of their models. Civilibal et al. [3] developed a segmentation and classification system, but pointed out that high computational demand was a constraint hindering real-time applicability. Aidossov et al.[1] used transfer learning to improve robustness, but they emphasized limited availability of varied thermal image datasets as a limitation.

Integration of manual and automatic features. The use of manually extracted features as thermal asymmetry and textural attributes alongside neural network-learned features enhances discrimination and interpretation[14]. Gupta et al. [4] found that age-specific feature selection improves thermography detection. Veerlapalli et al. [17] combined GANs for synthetic data with CNNs, boosting detection by addressing data scarcity and imbalance.

Transfer learning with pretrained models in medical images. Models pre-trained on medical databases as RadImageNet outperform those trained from scratch or on non-medical datasets. Sheikh et al.[15] compared RadImageNet and ImageNet, emphasizing the importance of domain-specific datasets for anomaly detection in breast thermal images.

Limitations and challenges in the state of the art. These studies face common limitations: small, homogeneous datasets lead to overfitting and limited generalization, due to a lack of rigorous cross-validation, and interpretability issues hinder clinical adoption. Moreover, physiological and technical variations, along with data drift and displacement, pose additional challenges not yet addressed.

4 Methodology

The proposed model will be divided into three phases:

- **Pre-training and model fine-tuning.** The process starts with a foundational ResNet50 model that has been previously trained on RadImageNet[9]. This model will be subsequently fine-tuned utilizing medical thermographic images, including those of the plantar region of the foot and veins, enabling it to extract general thermal features related to abnormal conditions, that are embedded into the resulting deep feature representation.
- **Feature generation, extraction, and fusion.**
  - A Generative Adversarial Network (GAN) will be used to create synthetic thermographic images from real breast thermographic images.
  - The deep model will be customized for thermographic images, enabling autonomous extraction of deep features from the datasets, while manual features will be simultaneously computed using traditional methods such as Histogram of Oriented Gradients (HOG), Local Binary Pattern (LBP), and Gray Level Cooccurrence Matrix (GLCM).

- Finally, these two sources of features (deep and manual) will be fused to form a combined set of features (fused features). We will evaluate different fusion methods to identify the most effective solution.

– **Classification and evaluation.** The fused features will be used to train an SVM classifier. The final model will be evaluated using key performance metrics such as accuracy, sensitivity, and specificity.

5 Preliminary Results

Database There are currently two public databases: Database for Mastology Research with Infrared Image (DMR-IR) [16], containing 287 patients, healthy or sick; and Mammary Thermography Dataset of Cali (Cali) [12], which contains thermographic images of the chest area of 119 women. At this point of the research, a subset from the DMR-IR database composed of 1,000 frontal thermographic images, 500 of each class, was used to train a basic CNN model with 5,611,961 trainable parameters. After training, the Cali dataset was used for the assessment of such model, in order to analyze its generalization capability under cross-dataset conditions.

CNN Model Training and Evaluation A convolutional neural network (CNN) was trained for thermogram classification using DMR-IR database with the following data split: 80% training, 10% validation, 10% testing. Table 1 presents the obtained performance metrics Accuracy, Sensitivity and Specificity.

Table 1. Performance of CNN using cross-dataset evaluation.

Dataset	Data size	Accuracy	Sensitivity	Specificity
DMR-IR	50 of each class	99.6%	99.2%	100%
Cali	28 of each class	64.3%	59.1%	83.3%

Transfer Learning with RadImageNet A RadImageNet pre-trained with clinical data was adjusted using a data split of 80% training, 10% validation and 10% testing. Table 2 presents the obtained performance.

Table 2. Performance of RadImageNet using cross-dataset evaluation.

Dataset	Data size	Accuracy	Sensitivity	Specificity
DMR-IR	50 of each class	99%	98%	100%
Cali	28 of each class	48.0%	68%	29%

Feature Extraction A combination of manual features (SIFT, HOG, LBP, and statistical properties from the grayscale co-occurrence matrix) and deep features obtained from a fine-tuning of a VGG16 model were used to train three SVM

classifiers, in order to evaluate the pertinence on the use of hybrid features. The involved features for each classifier are: HOG, LBP, statistical properties for C1; HOG, LBP, SIFT, statistical properties for C2 and deep features (VGG16) for C3. Classifiers were built using both a linear and a Radial Basis Function (RBF)-based kernel. The experiment was conducted using an expanded, balanced dataset with 500 cancer-free and 500 cancerous images.Table 3 presents the results obtained by the three best classifiers.

Table 3. Best results of the preliminary feature extraction experiment so far using DMR-IR dataset.

Classifier	**Accuracy**
C1	79.10 (linear) / 73.88 (RBF)
C2	70.01 (linear) / 72.36 (RBF)
C3	74.63 (linear) / 72.73 (RBF)

Discussion Preliminary results show that although the CNN model trained for thermogram classification achieved high internal accuracy of up to 99.6% and excellent sensitivity and specificity metrics(see Table 1), it showed a marked drop in generalization when evaluated on external sets, with accuracy dropping to 64.3%, highlighting the need for more robust methodologies to handle intra-class variability and data imbalance. Transfer learning using pre-trained models in RadImageNet improved feature extraction and representation, outperforming models trained from scratch (see Table 2), and the integration of manual and deep features improved model discrimination. However, limitations associated with data variability and small database sizes persist, pointing to the importance of continuing with advanced synthetic data generation techniques and incremental model fine-tuning to strengthen the robustness and generalization ability of the system.

5.1 Expected Results and Potential Challenge

This research will develop a method that integrates hybrid feature engineering with transfer learning to enhance anomaly detection on breast thermographic images. Its objective is to improve robustness and generalization amidst physiological variability and limited data by producing synthetic images through GANs (WGAN-GP). The system will extract thermal features such as asymmetry, vascular tortuosity, and temperature gradients to increase accuracy, sensitivity, and specificity on publicly available datasets, addressing issues of data drift and shift. Challenges encountered include limited and diverse thermographic data, high variability, and external influences. Technical obstacles involve stabilizing the quality of synthetic images, preventing GAN mode collapse, and refining feature extraction and fusion processes. Effectively managing data drift, maintaining clinical accuracy, and minimizing false results are crucial for ensuring reliability and facilitating adoption.

References

1. Aidossov, N., Zarikas, V., Mashekova, A., Zhao, Y., Ng, E.Y.K., Midlenko, A., Mukhmetov, O.: Evaluation of integrated cnn, transfer learning, and bn with thermography for breast cancer detection. Applied Sciences **13**(1), 600 (2023)
2. Al Husaini, M.A.S., Habaebi, M.H., Gunawan, T.S., Islam, M.R., Elsheikh, E.A., Suliman, F.: Thermal-based early breast cancer detection using inception v3, inception v4 and modified inception mv4. Neural Computing and Applications **34**(1), 333–348 (2022)
3. Civilibal, S., Cevik, K.K., Bozkurt, A.: A deep learning approach for automatic detection, segmentation and classification of breast lesions from thermal images. Expert Systems with Applications **212**, 118774 (2023)
4. Gupta, K.K., Vijay, R., Pahadiya, P., Saxena, S., Gupta, M.: Novel feature selection using machine learning algorithm for breast cancer screening of thermography images. Wireless Personal Communications **131**(3), 1929–1956 (2023)
5. INEGI: Estadísticas a propósito del día internacional de la lucha contra el cáncer de mama (19 de octubre), `https://www.inegi.org.mx/contenidos/saladeprensa/aproposito/2023/EAP_CMAMA23.pdf`
6. Iwashita, A.S., Papa, J.P.: An overview on concept drift learning. IEEE access **7**, 1532–1547 (2018)
7. Jacob, G., Jose, I., Sujatha, S.: Breast cancer detection: A comparative review on passive and active thermography. Infrared Physics & Technology p. 104932 (2023)
8. Marzo-Castillejo, M., Vela-Vallespín, C., Bellas-Beceiro, B., Bartolomé-Moreno, C., Melús-Palazón, E., Vilarrubí-Estrella, M., Nuin-Villanueva, M.: Recomendaciones de prevención del cáncer. actualización papps 2018. Atención primaria **50**, 41–65 (2018)
9. Mei, X., Liu, Z., Robson, P.M., Marinelli, B., Huang, M., Doshi, A., Jacobi, A., Cao, C., Link, K.E., Yang, T., et al.: Radimagenet: an open radiologic deep learning research dataset for effective transfer learning. Radiology: Artificial Intelligence **4**(5), e210315 (2022)
10. Moayedi, S.M.Z., Rezai, A., Hamidpour, S.S.F.: Toward effective breast cancer detection in thermal images using efficient feature selection algorithm and feature extraction methods. Biomedical Engineering: Applications, Basis and Communications **36**(02), 2450007 (2024)
11. Rautela, K., Kumar, D., Kumar, V.: A comprehensive review on computational techniques for breast cancer: past, present, and future. Multimedia Tools and Applications **83**(31), 76267–76300 (2024)
12. Rodriguez-Guerrero, S., Loaiza-Correa, H., Restrepo-Girón, A.D., Reyes, L.A., Olave, L.A., Diaz, S., Pacheco, R.: Dataset of breast thermography images for the detection of benign and malignant masses. Data in Brief **54**, 110503 (2024)
13. Sahiner, B., Chen, W., Samala, R.K., Petrick, N.: Data drift in medical machine learning: implications and potential remedies. The British Journal of Radiology **96**(1150), 20220878 (2023)
14. Sayed, B.A., Eldin, A.S., Elzanfaly, D.S., Ghoneim, A.S.: A comprehensive review of breast cancer early detection using thermography and convolutional neural networks. In: 2023 International Conference on Computer and Applications (ICCA). pp. 1–6. IEEE (2023)
15. Sheikh, H., Dey, A., Rajan, S.: Thermography-based breast abnormality detection using pre-trained radimagenet models. In: 2024 International Symposium on Sensing and Instrumentation in 5G and IoT Era (ISSI). vol. 1, pp. 1–6. IEEE (2024)

16. Silva, L., Saade, D., Sequeiros, G., Silva, A., Paiva, A., Bravo, R.d.S., Conci, A.: A new database for breast research with infrared image. Journal of Medical Imaging and Health Informatics **4**(1), 92–100 (2014)
17. Veerlapalli, P., Dutta, S.R.: A hybrid gan-based deep learning framework for thermogram-based breast cancer detection. Scientific Reports **15**(1), 1–33 (2025)

Robotic Arm-Based Assisted Feeding System for Patients with Upper Limb Limitations

L. Y. Leal Ramos[0009−0008−7009−6008], A. Gutiérrez Giles[0000−0002−2166−2463], and J. R. López Gutiérrez[0000−0001−5326−5342]

National Astrophysics, Optics and Electronics, Tonantzintla, Puebla, Mexico

Abstract. The project presented here involves the design and development of an assistive feeding system intended for people with limited upper limb mobility. The system is based on a 6 DOF robotic arm that incorporates an advanced visual perception module to detect the patient's mouth position in real-time using computer vision algorithms.
To achieve adaptive and safe behavior, deep reinforcement learning techniques are employed, allowing the robot to learn personalized feeding behaviors by adjusting its movements according to the posture and actions of the user. This is accomplished through a simulation environment that supports sim-to-real transfer for deployment on a physical robot.
The proposed approach addresses the challenge of reducing patient dependence during feeding tasks. The physical version of this system will be validated through controlled trials with real users in collaboration with the Instituto Nacional de Rehabilitación (INR).

Keywords: Sim-to-Real · Reinforcement Learning · Visual Detection · Safe Assistance · Autonomous System

1 Introduction

The Clasificación Internacional del Funcionamiento, de la Discapacidad y de la Salud (CIF) [1] recognizes that autonomy in basic activities of daily living, such as feeding, is essential for the well-being and social inclusion of people with disabilities. The category E115 groups products and technologies designed to support everyday personal functions, including assistive feeding.

Individuals with upper-limb disabilities caused by events such as cerebrovascular accidents [2], spinal cord injuries [3], or neurodegenerative diseases [4] often rely on caregivers for feeding, which affects their autonomy and quality of life. Despite advances in robotic assistive systems, most current solutions present significant limitations because they rely on predefined trajectories, manual controls, and cannot adapt to variations in the user's posture.

This project proposes an intelligent, safe, and user-centered assistive feeding system developed under the framework of deep reinforcement learning. The system is built around a 6 DOF robotic arm (Kinova MICO) equipped with a real-time computer vision module to detect the patient's mouth. See fig 1.

The integration of deep reinforcement learning enables the robot to learn and refine its behavior through interaction with the environment, adapting to user's posture and providing a high degree of flexibility and personalization.

The training is carried out in a simulation environment with controlled clinical conditions called *Assistive Gym* [5], to develop and test assistive robotics using reinforcement learning, where the RL model of the feeding task is designed allowing the robot to learn to adapt to various scenarios.

Subsequently, a sim-to-real transfer will be performed to deploy the optimal policy on the physical robotic arm without hardware modifications. For this, the *Mediapipe* framework is used [6], a pre-built recognition system that does not require prior training. The entire architecture is implemented on a modular framework based on ROS (Robot Operating System). The controlled physical trials with real users will be carried out by the INR of Mexico.

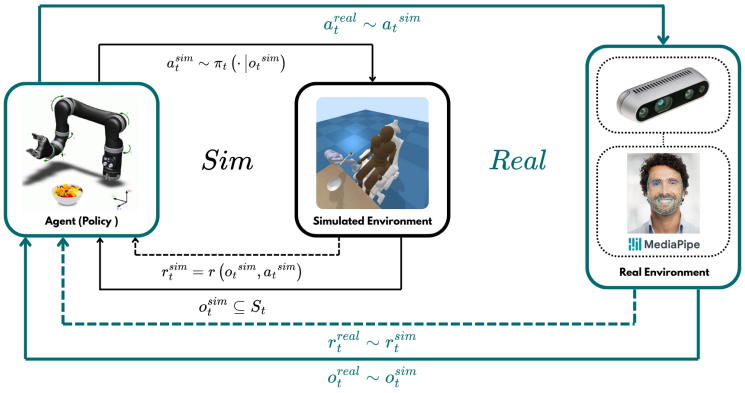


Fig. 1. Sim-to-Real framework proposed for the assisted feeding task.

2 Research Objectives

2.1 General Objective

Develop an autonomous and safe assistive feeding system using computer vision and reinforcement learning.

2.2 Specific Objectives

- Integrate a facial detection network using MediaPipe and a depth camera to locate the patient's mouth position in real time.
- Implement a deep reinforcement learning algorithm and evaluate the obtained results .
- Perform sim-to-real transfer to execute the learned policy on the robot.
- Design a safe feeding routine so that the robotic arm operates according to the learned policies and requested actions.

3 Related Work

The assistive feeding problem involves numerous considerations due to the complexity of the task. Factors such as the user's condition, age, mobility, posture, food type and the utensils used must all be taken into account.

For this reason, current assistive feeding systems propose different approaches to the same problem, each incorporating specific design considerations. Commercial systems such as *OBI Robot* [7] and *Neater Eater* [8] have demonstrated practical usability due to their ease of operation; however, they rely on preprogrammed trajectories and manual controls, which limits their adaptability. I-Feed [9] offers a more advanced solution by avoiding manual control, making it particularly useful for individuals with reduced mobility. Similarly, ICRAFT [10] adopts a comparable strategy but integrates a visual interface that enables the user to select among various robot actions, such as choosing between three different menus, changing dishes, drinking and activation by eye-tracking input.

On the other hand, efforts have been made to implement systems based on active perception, in which the agent observes the environment and generates actions in response. Feel-the-bite [11] is an assistive feeding robot for individuals with severe mobility impairments who cannot lean forward to eat.

Likewise, the AROHIN robot [12] features 6 DOF to assist people in self-feeding. It integrates various controllers (including PID, fuzzy logic, and advanced versions such as FOPID) and an a depth camera to plan trajectories with six intermediate waypoints. The ADA system [13] focuses on facilitating the transfer of food from the spoon or fork to the user's mouth without hand use. It emphasizes the robot's ability to acquire and deliver food naturally.

4 Results Evaluation

Based on the established planning, the following concrete results are expected:

- The generation of smooth, adaptive, and safe feeding trajectories by the agent, which can be executed autonomously. Trajectories will be evaluated using quantitative metrics, adaptability to changes in user posture or head movement, collision avoidance, and task completion efficiency.
- The training of a reinforcement learning agent under the sim-to-real approach, and its validation with respect to quantitative performance metrics, task completion success rates, safety measures, and adaptability to individual user characteristics in real-world scenarios.
- A functional prototype that can be tested with volunteer patients in clinical settings, under the supervision of health professionals from the INR. These trials will ensure the safe and controlled evaluation of the system's performance, usability, and adaptability.

These results will open new avenues for user-centered robotic technologies in personal assistance, leveraging deep reinforcement learning techniques. The goal is to enhance patient independence, addressing basic needs with a proposal aligned with the CFI E115 objectives for personal daily-use technology.

References

1. Centro Mexicano para la Clasificación de Enfermedades (CEMECE) and Dirección General de Información en Salud (DGIS), "Clasificación internacional del funcionamiento, de la discapacidad y de la salud (cif), versión oficial vigente," https://www.gob.mx/salud/documentos/catalogo-cif-clasificacion-internacional-del-funcionamiento-de-la-discapacidad-y-de-la-salud, 2017, cAT_CIF catalog. NOM-024-SSA3-2012 code. Mexico: Ministry of Health. Last update in 2017. Annual review, decennial update.
2. V. L. Feigin, M. Brainin, B. Norrving, S. O. Martins, J. Pandian, P. Lindsay, M. F. Grupper, and I. Rautalin, "World stroke organization: Global stroke fact sheet 2025," *International Journal of Stroke*, vol. 20, no. 2, pp. 132–144, Feb. 2025, epub 2025 Jan 3.
3. W. H. Organization, "Spinal cord injury," 2024, accessed: 2025-05-29. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/spinal-cord-injury
4. ——, "Multiple sclerosis," 2023, accessed: 2025-05-29. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/multiple-sclerosis
5. Z. Erickson, V. Gangaram, A. Kapusta, C. K. Liu, and C. C. Kemp, "Assistive gym: A physics simulation framework for assistive robotics," *CoRR*, vol. abs/1910.04700, 2019. [Online]. Available: http://arxiv.org/abs/1910.04700
6. C. Lugaresi, J. Tang, H. Nash, C. McClanahan, E. Uboweja, M. Hays, F. Zhang, C. Chang, M. G. Yong, J. Lee, W. Chang, W. Hua, M. Georg, and M. Grundmann, "Mediapipe: A framework for building perception pipelines," *CoRR*, vol. abs/1906.08172, 2019. [Online]. Available: http://arxiv.org/abs/1906.08172
7. E. Automation, "Obi," https://meetobi.com/, 2017, accedido en junio de 2025.
8. Neater Solutions Ltd, "Neater eater robotic," https://www.neater.co.uk/neater-eater-robotic, n.d., Accessed in June 20th, 2025.
9. F. Liu, H. Yu, W. Wei, and C. Qin, "I-feed: A robotic platform of an assistive feeding robot for the disabled elderly population," *Technology and Health Care*, vol. 28, no. 4, pp. 425–429, 2020.
10. P. Lopes, R. Lavoie, R. Faldu, N. Aquino, J. Barron, M. Kante, and B. Magfory, "Icraft-eye-controlled robotic feeding arm technology," *Tech. Rep*, 2012.
11. R. K. Jenamani, D. Stabile, Z. Liu, A. Anwar, K. Dimitropoulou, and T. Bhattacharjee, "Feel the bite: Robot-assisted inside-mouth bite transfer using robust mouth perception and physical interaction-aware control," 2024. [Online]. Available: https://arxiv.org/abs/2403.04067
12. P. Parikh, R. Trivedi, J. Dave, K. Joshi, and D. Adhyaru, "Design and development of a low-cost vision-based 6 dof assistive feeding robot for the aged and specially-abled people," *IETE Journal of Research*, vol. 70, no. 2, pp. 1716–1744, 2024. [Online]. Available: https://doi.org/10.1080/03772063.2023.2173665
13. D. Gallenberger, T. Bhattacharjee, Y. Kim, and S. S. Srinivasa, "Transfer depends on acquisition: Analyzing manipulation strategies for robotic feeding," in *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 2019, pp. 267–276.

Design of Classification Models Using Grammatical Evolution and Machine Learning Components Applied to Medical Problems

Blanca Verónica Zuñiga-Núñez[1][0009−0006−3148−9402], Erick Israel Guerra-Hernández[1][0000−0003−1554−0002], and Marco Aurelio Sotelo-Figueroa[2][0000−0002−9795−0138]

[1] Universidad Autónoma Benito Juárez de Oaxaca, Oaxaca, Oaxaca, MX
{zunb941006.fmc,ghernandez.cat}@uabjo.mx
[2] Universidad de Guanajuato, Guanajuato, Guanajuato, MX
masotelo@ugto.mx

Abstract. Artificial intelligence encompasses multiple areas, including machine learning and evolutionary computation. Among these, algorithms such as neural networks, genetic algorithms, and grammatical evolution have been studied to improve their efficiency using different configurations (parameters or structure). In addition, they have been applied in different domains, like healthcare, where applications such as predicting and diagnosing diseases stand out. Although notable progress has been made in the field of automatic design of machine learning models, this continues to present challenges, especially regarding the optimization of model architectures. Existing research has provided valuable advances in hyperparameter optimization and model selection, but the structural design of models remains comparatively less explored.
This study contributes to these ongoing efforts by proposing the use of grammatical evolution for the automated design of machine learning models. The proposal aims to identify and generalize modular components of existing machine learning algorithms and then use them through grammatical evolution, an algorithm that has proved its flexibility to build valid architectures using a grammar.
Expected contributions include more accurate and efficient models for healthcare classification tasks, advancing research on automated machine learning, and supporting the development of practical decision-support systems in medicine.

Keywords: machine learning · grammatical evolution · healthcare · automatic algorithm design

1 Introduction

1.1 Background

Artificial Intelligence Artificial Intelligence (AI) is a broad field of computer science focused on building systems capable of simulating intelligent behaviors

and solving problems autonomously [20]. It encompasses diverse sub-disciplines, among them Machine Learning (ML) and Evolutionary Computation (EC) stand out. ML develops algorithms that learn patterns from data and make predictions without explicit programming for every task [5]. EC, inspired by the theory of species evolution, utilizes operators such as mutation, crossover, and selection to optimize candidate solutions [15].

The intersection of ML and EC has led to advances in many domains, with a strong impact in healthcare [10]. For example, ML methods have been applied to medical image analysis with expert-level accuracy [8], to the study of degenerative diseases such as diabetes [14], and to simulate complex biological processes [9].

A notable algorithm in EC is Grammatical Evolution (GE) [19, 21], a grammar-based variant of Genetic Programming (GP). In GE, candidate solutions are represented as genotypes that are mapped to phenotypes through syntactic rules defined by a grammar, ensuring that generated solutions are both valid and structurally constrained. GE has been successfully applied to problems such as symbolic regression [7, 29], neural network design [2], and biological process modeling [17].

Different studies in GE have proposed variants in the mapping processes, which map phenotypes to syntactically valid genotypes, adding permutations to the process and improving the quality of the solution [24, 27].

Automatic Design of Machine Learning Models The performance of an ML model depends not only on hyperparameters but also on its structural design, that is, how algorithms and their components are organized and combined. Traditionally, model design has been manual, requiring expert knowledge and extensive experimentation. This becomes increasingly complex as the number of parameters grows and the size of the search space increases.

Automatic design approaches aim to overcome this by exploring structural design spaces through optimization strategies. GE in particular is well-suited for this kind of tasks because grammars ensure syntactic validity while allowing flexible search. Existing research has applied grammar-based methods to neural network topologies [13, 16], hyperparameter optimization [26], and to develop hyper-heuristics [22].

1.2 Problem Statement and Motivation

The automatic design of ML algorithms remains an open challenge despite advances in computation. Although significant progress has been made, the structural design of these models, that is, the selection of model architecture and component interactions, is still relatively less explored compared to other aspects in automatic design, such as hyperparameter optimization. ML has been widely used in domains such as health, specifically in tasks such as predicting and diagnosing different types of diseases [10, 1].

On the other hand, several studies have demonstrated the potential of grammar-based approaches to automate various aspects of the general design process, reporting results that are competitive or better than other proposals. For example, in [3], a grammar-based framework was proposed to optimize scikit-learn classification pipelines. In [4], a grammar-based approach was proposed to automate the most effective sequence of data pre-processing and ML algorithms, including the best hyperparameter configuration. Other studies, such as [26], have also studied hyperparameter optimization for neural networks using a grammar approach. The flexibility of grammar-based designs, particularly highlighted in all these examples, demonstrates that this kind of approach is a versatile tool.

Despite advances, the systematic design of ML models through the generalization of modular algorithmic components remains an open challenge. This research seeks to address this gap by formalizing these components, using them to design a grammar, and employing GE to automatically generate models for healthcare classification. For the experimentation of this study, it will use datasets related to heart disease, cancer, and diabetes. These conditions are frequently used as benchmarks in ML research. The following databases have been identified so far. Heart disease [6, 11, 23], cancer [12, 18], diabetes [25], and a benchmark proposed to test classification in ML algorithms [28].

1.3 Objectives

Main Objective: Design ML models applicable to healthcare through the use of GE to combine and generalize modular components of existing algorithms, to obtain more efficient models than those currently reported in the state-of-the-art for medical classification tasks.

Specific Objectives

1. Review the state-of-the-art in automatic ML design and GE to identify limitations and opportunities, especially in the field of healthcare.
2. Identify and categorize modular components of ML algorithms that can be generalized.
3. Design a grammar that integrates these modular components.
4. Develop ML models for healthcare classification using GE and modular components, with the aim of improving efficiency and accuracy compared to state-of-the-art methods.
5. Implement the proposed approach.
6. Evaluate and compare the performance of the proposal with state-of-the-art algorithms using statistical methods.

1.4 Hypothesis

The use of GE to combine the modular components of ML algorithms will design models that are more efficient and accurate than current state-of-the-art approaches in medical classification tasks.

2 Research Proposal

2.1 Methodology

The proposed research will follow an incremental and iterative methodology, allowing continuous refinement of techniques and validation of the results.

1. Literature review and problem definition
   - Conduct a literature review on automatic design of ML algorithms, GE, and grammar design.
   - Identify current challenges, gaps, and limitations in existing approaches.
   - Choose the main ML algorithms to study.
   - Analyze the applications of ML in healthcare.
   - Select benchmark healthcare datasets. The plan is to use data from 3 specific diseases: heart disease, cancer, and diabetes.
2. Identification and generalization of modular components ML.
   - Analyze the major ML algorithms to identify modular components.
   - Generalize these components suitable for integration in a grammar.
3. Grammar design and validation
   - Design a formal grammar using the identified modular ML components.
   - Validate grammar by generating syntactically correct example models.
4. Automatically design ML models using the proposal.
   - Implement the proposal in Python.
   - Explore different configurations for the algorithm.
5. Experimental validation in healthcare data
   - Apply the GE-based approach to automatically design classification models and evaluate them on the selected health benchmarks.
   - Compare the results with baseline ML methods and state-of-the-art automatically generated ML algorithms.
   - Use statistical methods, such as the Wilcoxon test, to make comparisons.
6. Evaluation, improvement, and divulgation
   - Improve the proposal and grammar based on the results.
   - Publish the proposal and results in congresses and scientific articles.
   - Document the proposal, methodology, results, and contributions in a doctoral thesis.

2.2 Expected Contributions and Results

Scientific and Technological Contributions

1. Scientific
   - Formalization of modular components in ML algorithms.
   - Design of a grammar for the ML design.
   - GE applied to the design of ML algorithms.
2. Technological
   - GE framework for the design of ML algorithms applicable to healthcare.

Expected Results

- The models automatically generated through the GE proposal are expected to achieve equal or better performance compared to the state-of-the-art algorithms.
- Develop models with better performance for health classification that can be used in decision support systems to improve diagnostic accuracy.
- At least one article in a high-impact journal and participation in two congresses.

References

1. Abdel-Jaber, H., Devassy, D., Al Salam, A., Hidaytallah, L., EL-Amir, M.: A Review of Deep Learning Algorithms and Their Applications in Healthcare. Algorithms **15**(2), 71 (Feb 2022). https://doi.org/10.3390/a15020071
2. Ahmadizar, F., Soltanian, K., AkhlaghianTab, F., Tsoulos, I.: Artificial neural network development by means of a novel combination of grammatical evolution and genetic algorithm. Engineering Applications of Artificial Intelligence **39**, 1–13 (2015). https://doi.org/10.1016/j.engappai.2014.11.003
3. Assunção, F., Lourenço, N., Ribeiro, B., Machado, P.: Evolution of Scikit-Learn Pipelines with Dynamic Structured Grammatical Evolution. In: Castillo, P.A., Jiménez Laredo, J.L., Fernández De Vega, F. (eds.) Applications of Evolutionary Computation, vol. 12104, pp. 530–545. Springer International Publishing, Cham (2020). https://doi.org/10.1007/978-3-030-43722-0_34
4. Barbudo, R., Ramírez, A., Romero, J.R.: Grammar-based evolutionary approach for automated workflow composition with domain-specific operators and ensemble diversity. Applied Soft Computing **153**, 111292 (2024). https://doi.org/10.1016/j.asoc.2024.111292
5. Bishop, C.M.: Pattern Recognition and Machine Learning. Information Science and Statistics, Springer, New York (2006)
6. Doppala, B.P.: Cardiovascular_disease_dataset (Apr 2021). https://doi.org/10.17632/DZZ48MVJHT.1
7. Espinal, A., Sotelo-Figueroa, M.A., Soria-Alcaraz, J.A.: Grammatical Evolution with Codons Selection Order as Intensification Process. Computación y Sistemas **28**(2) (2024). https://doi.org/10.13053/cys-28-2-5026
8. Esteva, A., Kuprel, B., Novoa, R.A., Ko, J., Swetter, S.M., Blau, H.M., Thrun, S.: Dermatologist-level classification of skin cancer with deep neural networks. Nature **542**(7639), 115–118 (2017). https://doi.org/10.1038/nature21056
9. Ezziane, Z.: Applications of artificial intelligence in bioinformatics: A review. Expert Systems with Applications **30**(1), 2–10 (2006). https://doi.org/10.1016/j.eswa.2005.09.042
10. Ferdous, M., Debnath, J., Chakraborty, N.R.: Machine Learning Algorithms in Healthcare: A Literature Survey. In: 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT). pp. 1–6. IEEE, Kharagpur, India (2020). https://doi.org/10.1109/ICCCNT49239.2020.9225642
11. Halder, R.K.: Cardiovascular disease dataset (2020). https://doi.org/10.21227/7QM5-DZ13
12. Hamdalla F. Al-Yasriy: The IQ-OTH/NCCD lung cancer dataset. https://doi.org/10.34740/KAGGLE/DS/672399

13. Hamdia, K.M., Zhuang, X., Rabczuk, T.: An efficient optimization approach for designing machine learning models based on genetic algorithm. Neural Computing and Applications **33**(6), 1923–1933 (2021). https://doi.org/10.1007/s00521-020-05035-x
14. Kavakiotis, I., Tsave, O., Salifoglou, A., Maglaveras, N., Vlahavas, I., Chouvarda, I.: Machine Learning and Data Mining Methods in Diabetes Research. Computational and Structural Biotechnology Journal **15**, 104–116 (2017). https://doi.org/10.1016/j.csbj.2016.12.005
15. Koza, J., Poli, R.: Genetic Programming, pp. 127–164. The MIT Press (01 2005). https://doi.org/10.1007/0-387-28356-0_5
16. López-Vázquez, G., Ornelas-Rodriguez, M., Espinal, A., Soria-Alcaraz, J.A., Rojas-Domínguez, A., Puga-Soberanes, H.J., Carpio, J.M., Rostro-Gonzalez, H.: Evolutionary Spiking Neural Networks for Solving Supervised Classification Problems. Computational Intelligence and Neuroscience **2019**, 1–13 (2019). https://doi.org/10.1155/2019/4182639
17. Moore, J.H., Sipper, M.: Grammatical Evolution Strategies for Bioinformatics and Systems Genomics. In: Handbook of Grammatical Evolution, pp. 395–405. Springer International Publishing, Cham (2018). https://doi.org/10.1007/978-3-319-78717-6_16
18. Namdari, R.: Breast cancer, `https://www.kaggle.com/datasets/reihanenamdari/breast-cancer/data`
19. O'Neill, M., Ryan, C.: Grammatical Evolution. IEEE Transactions on Evolutionary Computation **5**(4), 349–358 (2001). https://doi.org/10.1109/4235.942529
20. Russell, S.J., Norvig, P.: Inteligencia artificial. Un enfoque moderno. Pearson, 2 edn. (2004)
21. Ryan, C., Collins, J., Neill, M.O.: Grammatical evolution: Evolving programs for an arbitrary language. In: Lecture Notes in Computer Science, pp. 83–96. Springer Berlin Heidelberg, Berlin, Heidelberg (1998). https://doi.org/10.1007/bfb0055930
22. Sabar, N.R., Ayob, M., Kendall, G., Qu, R.: Grammatical Evolution Hyper-Heuristic for Combinatorial Optimization Problems. IEEE Transactions on Evolutionary Computation **17**(6), 840–861 (2013). https://doi.org/10.1109/TEVC.2013.2281527
23. Siddhartha, M.: Heart Disease Dataset (Comprehensive) (Nov 2020). https://doi.org/10.21227/DZ4T-CM36, `https://ieee-dataport.org/open-access/heart-disease-dataset-comprehensive`
24. Sotelo-Figueroa, M.A., Puga Soberanes, H.J., Carpio, J.M., Fraire Huacuja, H.J., Cruz Reyes, L., Soria-Alcaraz, J.A.: Improving the Bin Packing Heuristic through Grammatical Evolution Based on Swarm Intelligence. Mathematical Problems in Engineering **2014**(1) (2014). https://doi.org/10.1155/2014/545191
25. Teboul, A.: Diabetes Health Indicators Dataset, `https://www.kaggle.com/datasets/alexteboul/diabetes-health-indicators-dataset`
26. Vaidya, G., Kshirsagar, M., Ryan, C.: Grammatical Evolution-Driven Algorithm for Efficient and Automatic Hyperparameter Optimisation of Neural Networks. Algorithms **16**(7), 319 (2023). https://doi.org/10.3390/a16070319
27. Verónica Zuñiga, B., Martín Carpio, J., Aurelio Sotelo-Figueroa, M., Espinal, A., Jair Purata-Sifuentes, O., Ornelas, M., Alberto Soria-Alcaraz, J., Rojas, A.: Exploring Random Permutations Effects on the Mapping Process for Grammatical Evolution. Journal of Automation, Mobile Robotics and Intelligent Systems pp. 65–72 (2019). https://doi.org/10.14313/jamris/1-2020/8

28. Yang, J., Shi, R., Ni, B.: MedMNIST Classification Decathlon: A Lightweight AutoML Benchmark for Medical Image Analysis. In: 2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI). pp. 191–195. IEEE, Nice, France (2021). https://doi.org/10.1109/ISBI48211.2021.9434062
29. Zuñiga-Nuñez, B.V., Carpio, J.M., Sotelo-Figueroa, M.A., Soria-Alcaraz, J.A., Purata-Sifuentes, O.J., Ornelas, M., Rojas-Domínguez, A.: Studying Grammatical Evolution's Mapping Processes for Symbolic Regression Problems. In: Studies in Computational Intelligence, pp. 445–459. Springer International Publishing, Cham (2020). https://doi.org/10.1007/978-3-030-35445-9_32

An Exploratory Study on the Alignment Between Human Perception and Deep Learning Explainability Methods

Daniel da Silva Costa[1 [0009−0002−2952−5129]], Pedro Nuno de Souza Moura[2,3 [0000−0001−7799−4718]], and Adriana C. F. Alvim[3 [0000−0002−9486−607X]]

Graduate Program in Informatics (PPGI)
Federal University of the State of Rio de Janeiro, Rio de Janeiro, Brazil
daniel.scosta@edu.unirio.br
{pedro.moura,adriana}@uniriotec.br

Abstract. Deep Learning models have been increasingly used in numerous applications, including those sensitive to human life, which require clear explanations and justifications. This has encouraged several investigations into the explainability of neural networks. Various explainability methods have been proposed, but not many metrics to evaluate them. The Intersection over Union is the most common metric. It is applied between two bounding boxes to measure they similarity. The saliency maps generated by the explainability methods have no known shape. We assume that there must be a better metric that allows us to find an explainability method that best resembles human perception. To this end, we propose applying distance metrics to assess the similarity between human perception and explanation saliency maps. We conducted an investigation using chihuahuas images from the ImageNet dataset. Several CAM-based explainability methods were used to generate saliency maps, which were compared with human perception of the most important parts of each image. With the results of the distance metrics rankings of the best saliency maps were created and compared to the ranking obtained using people's choice for each selected image. This comparison was performed using the Rank-Biased Overlap (RBO) metric. The results indicate the feasibility of our method to find the explainability method that best resembles human perception. In our experiments, the two best metrics were Manhattan and Correlation. Besides, the best explainability methods regarding human perception were LayerCAM, Score-CAM, and IS-CAM.

Keywords: Explainability · Computer Vision · Deep Learning

1 The Problem

In recent years, Deep Learning (DL) models which are based on artificial neural networks (ANNs) have been used in a wide variety of contexts. ANNs can internally and automatically build a complex model of the input data and without the need for human operators to describe the characteristics of the data [15].

Despite significant advances in DL, some application areas require greater care and understanding of the information on which ANNs are based when learning internal representations of input data, since they deal with situations potentially harmful to human life, such as medicine and law, that usually require clear and unequivocal explanations of the systems' decisions. The confidence in the results and the usefulness of ANNs will be compromised if the user does not understand the rules behind the decision process of these models [10].

Generally speaking, Explainability aims to create an explanation that can help the user understand why a given model has a specific decision pattern, about the entire dataset or an instance of this set [10] and can help to increase the reliability of the use of ANNs models [10].

For the task of image classification, in Computer Vision (CV), typically the explainability methods try to show the user which information of the input the model considers most relevant for a specific classification generating a saliency map (or heatmap) that can be superimposed on the original image. Researchers often assess the quality of their explainability methods by visually comparing their heatmaps with the heatmaps of other methods, which can bias the evaluation. Although objective metrics have been proposed, they are more common than human perception-based metrics and, to the best of our knowledge, evaluating the quality and reliability of these methods remains an open problem [11].

2 The Hypothesis

Our hypothesis is based on the idea that the best and most helpful explainability methods will be the ones that most closely resemble human perception [21, 13, 11–13]. We seek to understand and evaluate the alignment [13] between the explainability methods, such as Class Activation Maps (CAM) [1] with the perception of people of the most important area of each image for a specific classification.

It is common to apply the Intersection over Union (IoU) metric (also called Jaccard similarity) [1] to calculate the similarity between two boxes. The problem is determining the limit of the heatmap which is not necessarily a box. Therefore, the present research ought to use different metrics that allow an investigation of the similarity between human perception and the heatmaps generated by the explainability methods, using all the information of the heatmaps.

3 Related Works and Contributions

CAM-based methods are among the most traditional explainability methods [10, 16] used with convolutional neural networks (CNN) [17] and can be divided into two groups: activation-based and gradient-based methods. Gradient-based (or backpropagation-based [10]) methods calculate the importance of each input feature based on some evaluation of the gradients of the trained neural network. The activation-based (or perturbation-based [10]) methods calculate the importance of each input feature based on the difference of the outputs for an

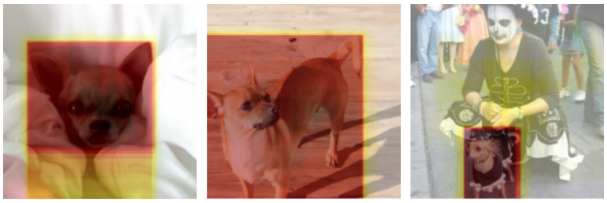


Fig. 1. Examples of the final annotation heatmap for three images: n02085620_1312, n02085620_5542 and n02085620_8174. The darker the shade of red, the more important the area. The lighter the yellow, the less important the area.

image and modified copies of it. Despite many proposals for CAM-based methods in recent years [1–9], to date, there have been no many articles dedicated to exploring and conducting an exhaustive comparison of these methods. Our novelty corresponds to the evaluation of the results of CAM-based explainability methods using human perception as the ground truth, while considering both gradient-based and activation-based methods. The results indicate the feasibility of our method to find the explainability method that best resembles human perception.

4 Methodology

We randomly selected 20 Chihuahua images from the ImageNet [14] dataset and classified each using the pre-trained ResNet50 model [26]. Three images were incorrectly classified. We removed them from the experiments. The remaining 17 images were used in an "annotation experiment" on the Appen crowdsourcing platform[1] where 50 people drew a bounding box around the chihuahua in each image. All the 50 annotations were combined in one "annotation heatmap" to allow us to better capture the importance of each pixel w.r.t. the perception of people. In this way, if a pixel is present, for example, in the 5 bounding boxes, it has a weight of 5 in the annotation heatmap. The values were normalised between 0 and 1. Figure 1 shows three examples of the annotation heatmap.

CAM-based methods, available in the TorchCAM[2] library, were applied to the model to generate "explanation heatmaps". The chosen methods were: (i) Activation-based: CAM (CAM) [1], SS-CAM (SSCAM) [3], IS-CAM (ISCAM) [4], Score-CAM (ScCAM) [2], Grad-CAM (GCAM) [5]; (ii) Gradient-based: Grad-CAM++ (GCAM++) [6], Smooth Grad-CAM++ (SGCAM++) [7], X-Grad-CAM (XGCAM) [8] and LayerCAM (LCAM) [9].

Figure 2 shows the preprocessed image n02085620_5542 and the explanation heatmaps generated by the methods.

We calculate the distance between the annotation heatmaps and the explanation heatmaps using these metrics: Weighted Jaccard (WJ); Wasserstein (WA);

[1] https://www.appen.com/
[2] https://github.com/frgfm/torch-cam

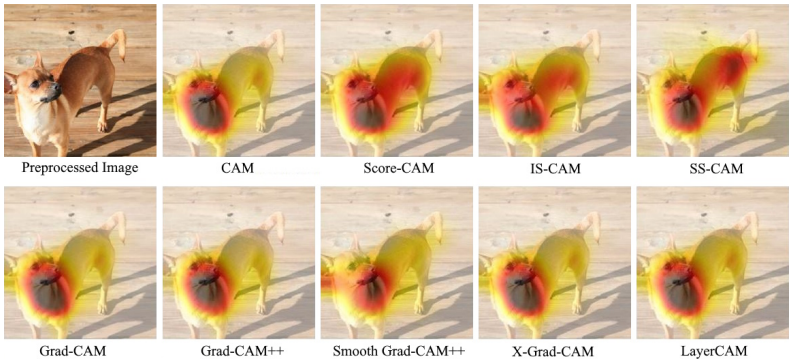


Fig. 2. The preprocessed image n02085620_5542 on the top-left and the overlapped explanation heatmap of each explainability method.

Bray-Curtis (BC); Canberra (CA); Chebyshev (CY); Manhattan (MA); Correlation (CR); Cosine (CS); Euclidean (EU); Jensen-Shannon (JS); Minkowski (MI); and squared Euclidean (SE).

We used the crowdsourcing platform called Prolific[3] to create an "validation experiment" where we asked people to select the heatmap that best explained the most important parts of the chihuahua. The images were shown with each overlapped heatmap, similar to Fig. 2 but without the methods names. The results were used to create a human ranking of the best explainability methods and in the same way the rankings with the results of each distance metric ("metric ranking"). We applied the Rank-Biased Overlap (RBO) [25] metric to compare the rankings. RBO is a similarity measure for indefinite rankings.

5 Results and Discussion

Table 1 shows the scores for each distance metric calculated between the annotation heatmap and each explanation heatmap for the image n02085620_5542. The lower the score, the better. The best results were highlighted in bold. In this case, the activation-based methods were better. We did the same for all the 17 images.

SSCAM obtained the best result in 8 metrics, followed by ISCAM (3) and ScCAM (2). The Chebyshev metric did not help select the best method, as the best value, 0.0000, appeared for several methods. It is also possible to see that the XGCAM scores were the same as the GCAM scores for all metrics. This was observed for all 17 chihuahua images.

Table 2 shows the number of times each metric ranking was the most similar to the human ranking. Each value shows the total of images where that metric obtained the best RBO. The best values are highlighted in bold. For some images, there were two or more rankings with the same score result.

[3] https://www.prolific.com/

Table 1. Calculated distance normalized values between the annotation heatmaps and the explanation heatmaps, for the image n02085620_5542.

	Explainability Methods								
Metrics	CAM	SSCAM	ISCAM	ScCAM	GCAM	GCAM++	SGCAM++	XGCAM	LCAM
WJ	0.6331	**0.0000**	0.1291	0.1186	0.7394	1.0000	0.5915	0.7394	0.6478
WA	0.6197	**0.0000**	0.2219	0.1947	0.7477	1.0000	0.5777	0.7477	0.6494
BC	0.6129	**0.0000**	0.1198	0.1099	0.7225	1.0000	0.5707	0.7225	0.6280
CA	1.0000	0.4674	**0.0000**	0.0053	0.1518	0.1911	0.7140	0.1518	0.7374
CY	**0.0000**	0.3333	**0.0000**	**0.0000**	**0.0000**	1.0000	**0.0000**	**0.0000**	**0.0000**
MA	0.6513	0.1041	0.0024	**0.0000**	0.7259	1.0000	0.5922	0.7259	0.6330
CR	0.8450	0.7031	**0.0000**	0.1530	0.7889	1.0000	0.5291	0.7889	0.7351
CS	0.6289	**0.0000**	0.2410	0.3332	0.7582	1.0000	0.3428	0.7582	0.6684
EU	0.6648	**0.0000**	0.2623	0.2966	0.7767	1.0000	0.5565	0.7767	0.6952
JS	0.5859	**0.0000**	0.1576	0.2927	0.7162	1.0000	0.2342	0.7162	0.6455
MI	0.5370	**0.0000**	0.8306	0.6830	0.7437	1.0000	0.4393	0.7437	0.6185
SE	0.6438	**0.0000**	0.2441	0.2771	0.7604	1.0000	0.5334	0.7604	0.6753

Table 2. The total of images where each metric obtained the best RBO score

Metric Ranking	**p=0.0**	**p=0.5**	**p=0.8**	**p=0.9**	**p=1.0**
WJ	7	4	4	2	3
WA	5	5	5	3	4
BC	7	4	4	2	3
CA	4	3	2	1	2
CY	1	0	0	1	1
MA	6	**6**	**6**	4	5
CR	**8**	**6**	**6**	**6**	5
CS	6	4	4	4	**6**
EU	6	4	4	2	3
JS	4	2	2	2	4
MI	6	4	5	4	4
SE	6	4	4	2	3

6 Future Work

We proposed a method in which we applied different distance metrics between the annotation heatmap created on several bounding boxes made by humans and the explanation heatmaps created using the CAM-based methods. We conducted an experiment to ask people to choose the best explanations and built a ranking with the results to compare with the results of the distance metrics, using the RBO metric. Our results suggest that: (i) the best metrics are two: Manhattan and Correlation; and (ii) the best explainability methods were LayerCAM, Score-CAM, and IS-CAM, for 4 images each, and SS-CAM, for 3, out of the total of 17 chihuahuas images randomly selected from the ImageNet dataset.

For future work, we intend to do different experiments, such as: (i) using different CNN architectures; (ii) collecting annotations in polygon shapes; (iii) collecting annotations created using Vision Language Models; and (iv) using images from different classes or other datasets.

References

1. Zhou, Bolei and Khosla, Aditya and Lapedriza, Agata and Oliva, Aude and Torralba, Antonio, Learning Deep Features for Discriminative Localization, 2016, 2921-2929, IEEE Computer Society
2. Wang, Haofan and Wang, Zifan and Du, Mengnan and Yang, Fan and Zhang, Zijian and Ding, Sirui and Mardziel, Piotr and Hu, Xia, Score-CAM: Score-Weighted Visual Explanations for Convolutional Neural Networks, 2020, 111-119, IEEE Computer Society
3. Haofan Wang and Rakshit Naidu and Joy Michael and Soumya Snigdha Kundu, SS-CAM: Smoothed Score-CAM for Sharper Visual Feature Localization, 2020, arXiv, https://arxiv.org/abs/2006.14255, last accessed 2025/09/14
4. Rakshit Naidu and Ankita Ghosh and Yash Maurya and Shamanth R Nayak K and Soumya Snigdha Kundu, IS-CAM: Integrated Score-CAM for axiomatic-based explanations, 2020, arXiv, https://arxiv.org/abs/2010.03023, last accessed 2025/09/14
5. Selvaraju, Ramprasaath R. and Cogswell, Michael and Das, Abhishek and Vedantam, Ramakrishna and Parikh, Devi and Batra, Dhruv, Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization, 2017, 618-626
6. Chattopadhay, Aditya and Sarkar, Anirban and Howlader, Prantik and Balasubramanian, Vineeth N, Grad-CAM++: Generalized Gradient-Based Visual Explanations for Deep Convolutional Networks, 2018, 839-847
7. Daniel Omeiza and Skyler Speakman and Celia Cintas and Komminist Weldermariam, Smooth Grad-CAM++: An Enhanced Inference Level Visualization Technique for Deep Convolutional Neural Network Models, 2019, arXiv, https://arxiv.org/abs/1908.01224, last accessed 2025/09/14
8. Ruigang Fu and Qingyong Hu and Xiaohu Dong and Yulan Guo and Yinghui Gao and Biao Li, Axiom-based Grad-CAM: Towards Accurate Visualization and Explanation of CNNs, 2020, arXiv, https://arxiv.org/abs/2008.02312, last accessed 2025/09/14
9. Jiang, Peng-Tao and Zhang, Chang-Bin and Hou, Qibin and Cheng, Ming-Ming and Wei, Yunchao, LayerCAM: Exploring Hierarchical Class Activation Maps for Localization, 2021, 30, 5875-5888
10. Ras, Gabrielle and Xie, Ning and van Gerven, Marcel and Doran, Derek, Explainable Deep Learning: A Field Guide for the Uninitiated, 2022, May 2022, J. Artif. Int. Res., 68
11. Colin, Julien and FEL, Thomas and Cadene, Remi and Serre, Thomas, What I Cannot Predict, I Do Not Understand: A Human-Centered Evaluation Framework for Explainability Methods, 2832–2845, Curran Associates, 35, 2022
12. Kim, Sunnie S. Y. and Meister, Nicole and Ramaswamy, Vikram V. and Fong, Ruth and Russakovsky, Olga, HIVE: Evaluating the Human Interpretability of Visual Explanations, 2022, Computer Vision – ECCV 2022: 17th European Conference, 280–298
13. Prasad, Grusha and Nie, Yixin and Bansal, Mohit and Jia, Robin and Kiela, Douwe and Williams, Adina", To what extent do human explanations of model behavior align with actual model behavior?, Proceedings of the Fourth BlackboxNLP Workshop on Analyzing and Interpreting Neural Networks for NLP, 2021, 1–14
14. Russakovsky, Olga and Deng, Jia and Su, Hao and Krause, Jonathan and Satheesh, Sanjeev and Ma, Sean and Huang, Zhiheng and Karpathy, Andrej and Khosla, Aditya and Bernstein, Michael and Berg, Alexander C. and Fei-Fei, Li, ImageNet

Large Scale Visual Recognition Challenge, International Journal of Computer Vision, 115, 211–252, 2015
15. Goodfellow, Ian and Bengio, Yoshua and Courville, Aaron, Deep Learning, 2016, The MIT Press,
16. Ibrahim, Rami and Shafiq, M. Omair, Explainable Convolutional Neural Networks: A Taxonomy, Review, and Future Directions, 2023, ACM Comput. Surv., 55
17. LeCun, Y. and Boser, B. and Denker, J. S. and Henderson, D. and Howard, R. E. and Hubbard, W. and Jackel, L. D., Backpropagation Applied to Handwritten Zip Code Recognition, Neural Computation, 1989, 1, 541-551
18. Morrison, Katelyn and Mehra, Ankita and Perer, Adam, Shared Interest... Sometimes: Understanding the Alignment between Human Perception, Vision Architectures, and Saliency Map Techniques, 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2023, 3776-3781
19. Kapishnikov, Andrei and Bolukbasi, Tolga and Viegas, Fernanda and Terry, Michael, XRAI: Better Attributions Through Regions, 2019 IEEE/CVF International Conference on Computer Vision (ICCV), 2019, 4947-4956
20. Feldman, Vitaly, Does learning require memorization? a short tale about a long tail, 2020, Proceedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing, 954–959
21. Zhang, Zijian and Singh, Jaspreet and Gadiraju, Ujwal and Anand, Avishek, Dissonance Between Human and Machine Understanding, 2019, Proc. ACM Hum.-Comput. Interact., 3
22. Sina Mohseni and Jeremy E. Block and Eric D. Ragan, A Human-Grounded Evaluation Benchmark for Local Explanations of Machine Learning, 2020, arXiv, https://arxiv.org/abs/1801.05075, last accessed 2025/09/14
23. Angie Boggust and Benjamin Hoover and Arvind Satyanarayan and Hendrik Strobelt, Shared Interest: Measuring Human-AI Alignment to Identify Recurring Patterns in Model Behavior, 2022, arXiv, https://arxiv.org/abs/2107.09234, last accessed 2025/09/14
24. Ioffe, Sergey, Improved Consistent Sampling, Weighted Minhash and L1 Sketching, 2010 IEEE International Conference on Data Mining, 2010, 246-255
25. Webber, William and Moffat, Alistair and Zobel, Justin, A similarity measure for indefinite rankings, 2010, ACM Trans. Inf. Syst., 28
26. He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian, Deep Residual Learning for Image Recognition, Cornell University Library, 2015

A Study on Control Schemes in the Evolutionary Process of Grouping Genetic Algorithms

Stephanie Amador-Larrea[1][https://orcid.org/0009−0008−8167−9102], Marcela Quiroz-Castellanos[1][https://orcid.org/0000−0001−8078−9491]

Universidad Veracruzana, Instituto de Investigaciones en Inteligencia Artificial
stephanieamadorlarrea@gmail.com, maquiroz@uv.mx

Abstract. Grouping Genetic Algorithms (GGAs) are among the most effective approaches for combinatorial grouping problems, but their performance is susceptible to parameter settings and operator behavior. This dependence often causes premature convergence and limits generalization. This work applies the GGA-CGT to the One-Dimensional Bin Packing Problem (1D-BPP), a classical NP-hard problem with broad practical relevance, and proposes an adaptive control mechanism to mitigate such sensitivity. A key contribution is an adaptive mutation strategy based on a dynamic threshold that uses population feedback to balance random and structured ordering, thereby enhancing search effectiveness. Experiments on classical and feature indexed benchmarks show that the proposed scheme increases the number of optimal solutions, reduces repeated fitness values, and improves robustness and diversity across instance classes. These findings lay the groundwork for extending adaptive control to other operators, such as crossover, and for building causal models that connect instance features, algorithmic choices, and performance.

Keywords: Bin Packing Problem · Grouping Genetic Algorithm · Controle esquemes.

1 Introduction

When tackling optimization problems, one often faces high computational complexity, which makes exact methods impractical within reasonable time limits. Many of these problems are NP-Hard, with grouping problems standing out for their practical relevance and intrinsic difficulty. A notable example is 1D-BPP (Martello, 1990), whose combinatorial complexity and practical relevance have motivated extensive research. In particular, metaheuristic approaches such as Evolutionary Computation, and especially Genetic Algorithms (GAs), have been widely applied to address this problem.

A specialized family, Grouping Genetic Algorithms (GGAs), was introduced by Falkenauer and Delchambre (1992), using group-based representations and variation operators specifically designed for this representation. Among them, the Grouping Genetic Algorithm with Controlled Gene Transmission (GGA-CGT) (Quiroz-Castellanos et al., 2015) is considered state-of-the-art, as it promotes

the propagation of high-quality groups. However, its performance is highly sensitive to parameter settings and operator design. Configurations that work well for one instance set may fail on others, limiting portability and robustness.

This research addresses this limitation by studying control schemes in the evolutionary process of the GGA-CGT, aiming to enhance adaptability, robustness, and performance on the 1D-BPP. Previous studies have shown that changes in instance features, such as item distribution, bin capacity, or problem size, significantly affect outcomes, often requiring manual adjustment or redesign of components (Amador-Larrea, 2022; Carmona-Arroyo G., 2021; Vázquez-Aguirre et al., 2024).

The guiding research question is: Is it possible to build models that describe the relationships between instance features, heuristic strategies, evolutionary dynamics, and the final performance of the GGA-CGT? Based on this, two hypotheses are posed: **H1:** Deterministic and adaptive control schemes can be implemented to select and configure heuristic strategies and parameters online, relying on models that capture relationships between instance features and algorithm behavior. **H2:** Incorporating such control schemes will improve the GGA-CGT by increasing solution quality and the proportion of optimal solutions, reflected in greater robustness across classes and higher efficiency in terms of computational cost.

Both hypotheses will be tested through systematic experiments on representative benchmarks, comparing the GGA-CGT with and without control schemes using metrics of quality, diversity, robustness, and efficiency. The general objective is to design deterministic and adaptive control schemes that integrate knowledge of both the problem domain and algorithmic behavior, thereby strengthening the robustness and effectiveness of the GGA-CGT on the 1D-BPP. Figure 1 summarizes the conceptual framework uniting instance features, control schemes, and evaluation metrics.

The study evaluates the GGA-CGT with different types of published 1D-BPP problems, which are considered very challenging (Carmona-Arroyo G., 2021; Falkenauer, 1996; Schoenfield, 2002; Scholl et al., 1997; Schwerin & Wäscher, 1997; Wiischer & Gau, 1996). These benchmarks allow testing under varying item distributions, bin capacities, and scales. Experiments compare baseline and controlled GGA-CGT versions, using quality, robustness, diversity, and efficiency as metrics, with statistical validation of results.

Methodologically, this research adopts an experimental and analytical design, combining benchmarking on classical and feature-indexed datasets, structural solution analysis, and statistical comparison of configurations. Results show that the proposed adaptive scheme increases the number of optimal solutions, reduces the frequency of repeated fitness values, and enhances robustness and diversity across instance classes. These findings support the working hypothesis that feedback-based control schemes can systematically reduce parameter sensitivity in GGAs, laying the groundwork for extending adaptive control to other operators, such as crossover, and for constructing causal models that connect instance features, algorithmic decisions, and performance.

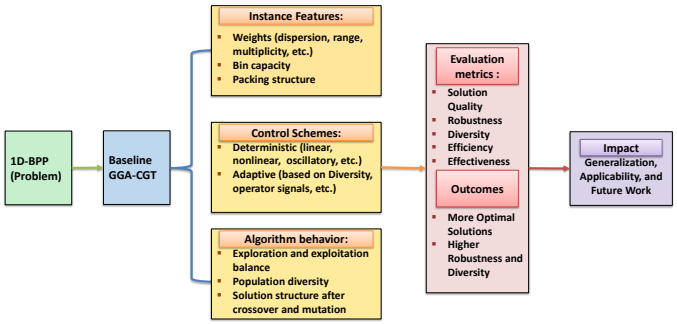


Fig. 1. Conceptual framework illustrating the relationship between instance features, algorithmic behavior, and control schemes in the GGA-CGT.

2 Related Work

This section is divided into two parts: control schemes in evolutionary algorithms, and techniques for solving the 1D-BPP.

Regarding control schemes, most approaches focus on parameter adaptation. Deterministic schedules vary crossover and mutation rates over time (L. Agustín-Blas et al., 2012; L. E. Agustín-Blas et al., 2011), while adaptive rules adjust them according to instance features or search progress (Chaurasia & Singh, 2014; Srinivas & Patnaik, 1994). In grouping contexts, the GGA-CGT integrates adaptive mutation via controlled gene transmission (Quiroz-Castellanos et al., 2015), and deterministic rules based on the generational gap have also been explored (Jawahar & Subhaa, 2017). Recent surveys highlight the increasing role of learning-based control, including reinforcement learning, to automatically tune parameters (de Lacerda et al., 2021). In addition, several studies have proposed adaptive operator variants: dynamic mutation in Fixed Job Scheduling (Rossi et al., 2010), alternating crossover and mutation in clustering (Peddi & Singh, 2011), multiple mutation types for microaggregation (Balasch-Masoliver et al., 2014), and adaptive crossover control in grouping contexts (Mutingi & Mbohwa, 2017).

The 1D-BPP has been extensively studied using both exact and approximate methods. Among heuristic approaches, the GGA-CGT (Quiroz-Castellanos et al., 2015) stands out for its specialized representation and operators, while the GGA-CGT/D variant (J. E. González-San-Martín, 2021) introduces instance-aware mechanisms that further enhance performance. Comparative studies confirm that Arc-Flow, CNS_BP, and GGA-CGT achieve leading results on classical datasets, although their relative ranking depends on instance features (J. González-San-Martín et al., 2023).

3 Objectives and Achieved Results

To accomplish the general objective, this research was structured into specific objectives. The first one consisted in analyzing the baseline performance of the

GGA-CGT on 1D-BPP benchmark instances, using different parameter configurations while keeping the standard heuristic settings. This study revealed an interesting behavior: the algorithm's performance varied significantly with respect to the chosen parameter configuration. Although it is well known in the area that the effectiveness of an algorithm strongly depends on the tuning of its numerical parameters (Eiben & Smith, 2015), it was remarkable to observe that, for some test cases, a crossover rate close to 100% increased the number of optimal solutions, while in others it caused a substantial decrease. In particular, some classes were almost completely solved optimally without applying crossover, but their performance deteriorated when crossover was increased. Regarding mutation, the analysis showed that higher mutation rates were directly associated with an increased number of optimal solutions, which is an interesting finding for a genetic algorithm, since mutation is usually expected to be kept relatively low.

The second specific objective was to evaluate the performance of the GGA-CGT on different benchmark sets. As mentioned earlier, two benchmarks with distinct characteristics were employed. A controlled experimentation was conducted, focusing on cases in which the algorithm struggled to reach the optimal solution. Characterization indexes from the literature (Carmona-Arroyo G., 2021) were used to measure the properties of the instances. Four main factors were identified as sources of difficulty: (i) the percentage of unused space in the optimal solution, where lower free space implies higher difficulty; (ii) the bin capacity, whose increase makes the problem harder; (iii) the distribution of weights and the range between minimum and maximum values, where larger ranges increase complexity; and (iv) weight multiplicity, where lower multiplicity (fewer items with the same weight) also increases difficulty (Carmona-Arroyo G., 2021; Vázquez-Aguirre et al., 2024). An additional index was also introduced, related to algorithmic performance: the number of unique-fitness solutions, where smaller values indicated more challenging instances.

The third specific objective consisted in measuring the impact of different heuristic strategies on the performance of the GGA-CGT. GGAs are generally composed of several components (initial population strategies, crossover and mutation operators, reproduction techniques), each with multiple possible alternatives. In this work, different techniques were tested for the crossover operator, considering ways of ordering parental genes, inheriting genes, and repairing solutions. In some cases, randomization produced better results, while structured strategies were more effective in others. For the mutation operator, different methods were analyzed for repairing solutions and ordering genes, while for reproduction strategies, several options were explored to integrate offspring into the population. The best performance was obtained when random solutions were added to the population.

Finally, in relation to control schemes, two perspectives were studied: the control of numerical parameters and the control of strategies. Regarding parameter control, both deterministic techniques (guided by fixed rules independent of population dynamics) and adaptive techniques (guided by feedback from the population) were considered (E et al., 1999). Linear, trigonometric, and nonlinear

functions were tested, applying oscillatory variations. Deterministic functions showed no significant differences compared to fixed parameters. In contrast, adaptive functions produced improvements when applied to mutation. In this case, the function depended on the number of unique-fitness individuals (some results are reported in Flores-Torres et al., 2025). However, not all cases followed the expected pattern of decreasing unique-fitness individuals during execution, meaning that this proposal may not be universally applicable. Based on this observation, current work focuses on metrics that depend exclusively on the algorithmic behavior of operators, such as the number of bins filled to maximum capacity, the number of released items, and their proportions. These indexes, which consistently appear across all cases, are used to design and test an adaptive control scheme for the crossover rate in the GGA-CGT, currently under systematic experimental evaluation.

On the other hand, regarding control of strategies, structural aspects of both crossover and mutation operators were studied. This analysis considered evolutionary process indicators such as the number of unique-fitness individuals and the structure of solutions after crossover. In the first control scheme, two phases of mutation were targeted: how to order the genes in solutions before applying mutation, and how to order the free items left after removing bins during mutation. Results showed significant performance improvements, increasing both the number of unique-fitness solutions and the number of optimal ones, as reported in Amador-Larrea et al., 2025. The results for the crossover operator are currently under systematic analysis as part of the proposed adaptive control mechanism.

4 Conclusions and Future Work

In conclusion, this work has consolidated the experimental, methodological, and analytical foundations of the research. Performance metrics were systematically examined, control schemes were implemented and analyzed, and distinct behaviors of the GGA-CGT were identified across different 1D-BPP instance classes. Results confirm that adaptive strategies balance exploration and exploitation, improving robustness, diversity, and quality. The research contributes empirical evidence and methodological advances that can guide the design of evolutionary approaches for other NP-hard grouping problems.

Future work will extend the study to the crossover operator within the same adaptive control approach, including a formal validation of its structural components and their interaction with feedback-based adaptive rules. In parallel, a causal analysis based on Bayesian networks will be developed to explicitly model the relationships between instance features, control variables, and algorithmic performance. This causal perspective aims not only to explain the dynamics observed in the experiments but also to support predictive and generalizable control mechanisms for GGAs and related evolutionary algorithms applied to NP-hard grouping problems.

References

Agustín-Blas, L., Salcedo-Sanz, S., Jiménez-Fernández, S., Carro-Calvo, L., Del Ser, J., & Portilla-Figueras, J. (2012). A new grouping genetic algorithm for clustering problems. *Expert Systems with Applications*, *39*(10), 9695–9703. https://doi.org/https://doi.org/10.1016/j.eswa.2012.02.149

Agustín-Blas, L. E., Salcedo-Sanz, S., Ortiz-García, E. G., Portilla-Figueras, A., Pérez-Bellido, Á. M., & Jiménez-Fernández, S. (2011). Team formation based on group technology: A hybrid grouping genetic algorithm approach. *Computers Operations Research*, *38*(2), 484–495. https://doi.org/https://doi.org/10.1016/j.cor.2010.07.006

Amador-Larrea, S. (2022). *Un estudio de operadores genéticos de cruza para el problema de empacado de objetos en contenedores*.

Amador-Larrea, S., Quiroz-Castellanos, M., & Ramos-Figueroa, O. (2025). An experimental study of strategies to control diversity in grouping mutation operators: An improvement to the adaptive mutation operator for the gga-cgt for the bin packing problem. *Mathematical and Computational Applications*, *30*(2). https://doi.org/10.3390/mca30020031

Balasch-Masoliver, J., Muntés-Mulero, V., & Nin, J. (2014). Using genetic algorithms for attribute grouping in multivariate microaggregation. *Intelligent data analysis*, *18*(5), 819–836.

Carmona-Arroyo G., Q.-C. M., Vázquez-Aguirre J.B. (2021). *One-dimensional bin packing problem: An experimental study of instances difficulty and algorithms performance* (Vol. 940). Springer, Cham.

Chaurasia, S. N., & Singh, A. (2014). A hybrid evolutionary approach to the registration area planning problem. *Applied intelligence*, *41*(4), 1127–1149.

de Lacerda, M. G. P., de Araujo Pessoa, L. F., de Lima Neto, F. B., Ludermir, T. B., & Kuchen, H. (2021). A systematic literature review on general parameter control for evolutionary and swarm-based algorithms. *Swarm and Evolutionary Computation*, *60*, 100777.

E, E. Á., Robert, H., & Zbigniew, M. (1999). Parameter control in evolutionary algorithms. *IEEE Transactions on evolutionary computation*, *3*(2), 124–141.

Eiben, A. E., & Smith, J. E. (2015). *Introduction to evolutionary computing* (2nd). Springer Publishing Company, Incorporated.

Falkenauer, E. (1996). A hybrid grouping genetic algorithm for bin packing. *Journal of heuristics*, *2(1)*, 5–30. https://doi.org/https://doi.org/10.1007/BF00226291

Falkenauer, E., & Delchambre, A. (1992). A genetic algorithm for bin packing and line balancing. *Proceedings 1992 IEEE International Conference on Robotics and Automation*, 1186–1192. https://doi.org/10.1109/ROBOT.1992.220088

Flores-Torres, L., Quiroz-Castellanos, M., Ramos-Figueroa, O., & Amador-Larrea, S. (2025). An experimental approach to design deterministic and adaptive control schemes for grouping genetic algorithms. *Neural Computing and Applications*. https://doi.org/10.1007/s00521-025-11270-x

González-San-Martín, J., Cruz-Reyes, L., Gómez-Santillán, C., Fraire, H., Rangel-Valdez, N., Dorronsoro, B., & Quiroz-Castellanos, M. (2023). Comparative study of heuristics for the one-dimensional bin packing problem. Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-28999-6_19

González-San-Martín, J. E. (2021). *Un estudio formal de heurísticas para el problema de empacado de objetos de una dimensión* [Tesis de Maestría]. Instituto Tecnológico de Ciudad Madero [Repositorio del Instituto Tecnológico de Ciudad Madero].

Jawahar, N., & Subhaa, R. (2017). An adjustable grouping genetic algorithm for the design of cellular manufacturing system integrating structural and operational parameters. *Journal of Manufacturing Systems*, *44*, 115–142.

Martello, S. (1990). *Knapsack problems: Algorithms and computer implementations*. J. Wiley & Sons.

Mutingi, M., & Mbohwa, C. (2017). Grouping genetic algorithms. *Advances and Applications. Switzerland: Springer International Publishing*, *243*.

Peddi, S., & Singh, A. (2011). Grouping genetic algorithm for data clustering. *International Conference on Swarm, Evolutionary, and Memetic Computing*, 225–232.

Quiroz-Castellanos, M., Cruz-Reyes, L., Torres-Jimenez, J., Gómez S., C., Huacuja, H. J. F., & Alvim, A. C. (2015). A grouping genetic algorithm with controlled gene transmission for the bin packing problem. *Comput. Oper. Res.*, *55*(100), 52–64. https://doi.org/10.1016/j.cor.2014.10.010

Rossi, A., Singh, A., & Sevaux, M. (2010). A metaheuristic for the fixed job scheduling problem under spread time constraints. *Computers & operations research*, *37*(6), 1045–1054.

Schoenfield, J. E. (2002). Fast, exact solution of open bin packing problems without linear programming. *Draft, US Army Space and Missile Defense Command, Huntsville, Alabama, USA*.

Scholl, A., Klein, R., & Jürgens, C. (1997). Bison: A fast hybrid procedure for exactly solving the one-dimensional bin packing problem. *Computers Operations Research*, *24*(7), 627–645. https://doi.org/https://doi.org/10.1016/S0305-0548(96)00082-2

Schwerin, P., & Wäscher, G. (1997). The bin-packing problem: A problem generator and some numerical experiments with ffd packing and mtp. *International Transactions in Operational Research*, *4*(5), 377–389. https://doi.org/https://doi.org/10.1016/S0969-6016(97)00025-7

Srinivas, M., & Patnaik, L. M. (1994). Adaptive probabilities of crossover and mutation in genetic algorithms. *IEEE Transactions on Systems, Man, and Cybernetics*, *24*(4), 656–667.

Vázquez-Aguirre, J. B., Carmona-Arroyo, G., Quiroz-Castellanos, M., & Cruz-Ramírez, N. (2024). Causal analysis to explain the performance of algorithms: A case study for the bin packing problem. *Mathematical and Computational Applications*, *29*(5). https://doi.org/10.3390/mca29050073

Wiischer, G., & Gau, T. (1996). *Orspe m heuristics for the integer one-dimensional cutting stock problem: A computational study*.

Professores Na Era Da Inteligência Artificial (IA): Desenvolvimento De Um Repositório Transdisciplinar De Materiais Educacionais Digitais (Med) Na Educação 5.0

Arthur Marques de Oliveira [0000-0001-9755-4277] Patrícia da Silva Campelo Costa Barcellos [0000-0002-5142-4730] and Rosa Maria Vicari [0000-0002-6909-6405]

[1] Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre/RS, BR
athur_marques@outlook.com

Abstract. This doctoral research addresses the challenge of preparing teachers to critically and effectively integrate Artificial Intelligence (AI) into their pedagogical practices in line with the principles of Education 5.0. Despite the potential of AI to support personalization, inclusion, and innovation in teaching, teacher education in Brazil still lacks consistent policies and systemic practices that incorporate AI from both critical and practical perspectives. To address this gap, the study proposes the development and validation of a transdisciplinary Pedagogical Model (PM) supported by action research, combined with the creation of an open and collaborative repository of Digital Educational Materials (MEDIAR) produced with AI. The novelty of the proposal lies in the articulation of three complementary dimensions: (i) a theoretical-methodological model that integrates AI and teacher education across disciplines, (ii) the implementation of MEDIAR as the first national repository dedicated to AI-generated educational resources, and (iii) the iterative validation of the model through cycles of action research involving pre-service and in-service teachers as active participants. Expected contributions include the advancement of theoretical and methodological approaches, the promotion of critical and ethical teacher training, and the technological development of scalable resources. The project also aims to generate evidence of how AI can foster personalization and innovation in education while ensuring transparency, ethics, and accessibility.

Keywords: Artificial Intelligence, Teacher Education, Transdisciplinarity.

1 Sobre a pesquisa

1.1 Contextualização

Nas últimas décadas, a educação tem sido profundamente impactada pelas transformações tecnológicas. A presença crescente da Inteligência Artificial (IA) nos ambientes escolares e acadêmicos tem remodelado práticas pedagógicas, abrindo caminhos para personalização do ensino, automação de tarefas administrativas e construção de experiências de aprendizagem mais inclusivas e centradas no estudante. Ao mesmo tempo, tais mudanças suscitam debates éticos e metodológicos: até que ponto a IA pode ser

utilizada sem comprometer a mediação humana? Como garantir que os docentes estejam preparados para empregar criticamente as novas ferramentas?

No contexto da Educação 5.0, paradigma que articula inovação tecnológica e humanização do processo formativo [1][2], torna-se urgente repensar a formação docente. Professores em formação inicial e continuada precisam não apenas dominar ferramentas digitais, mas também compreender seus impactos sociais, éticos e pedagógicos [3]. Esse cenário evidencia a necessidade de propostas que integrem a IA como tecnologia-ferramenta pedagógica, favorecendo a criação de recursos digitais inovadores.

Cabe destacar que o uso da Inteligência Artificial na educação não é um fenômeno recente. Pesquisas em Sistemas Tutores Inteligentes (STI), Ambientes Inteligentes de Aprendizagem e Inteligência Artificial na Educação (IAED) vêm sendo desenvolvidas há mais de cinco décadas [5][7][9], contribuindo para a personalização e o acompanhamento adaptativo de estudantes. O presente estudo reconhece esse legado histórico e busca ampliá-lo, incorporando as possibilidades abertas pela IA generativa e pela Educação 5.0, com ênfase na co-inteligência humano-máquina e na curadoria crítica de materiais educacionais digitais.

1.2 Problema de pesquisa, relevância e fundamentação

O problema que orienta esta pesquisa é compreender de que modo o desenvolvimento e a implementação de um Modelo Pedagógico (MP) transdisciplinar, articulado a um repositório digital de Materiais Educacionais Digitais (MEDs) produzidos com Inteligência Artificial (IA), podem mediar a capacitação docente para práticas pedagógicas personalizadas, éticas e críticas, em consonância com os princípios da Educação 5.0. Esse problema revela-se particularmente relevante porque a formação de professores ainda carece de políticas e práticas que contemplem o uso pedagógico da IA de forma crítica, estruturada e sistemática.

Como apontam estudos [4], grande parte dos docentes conclui sua formação inicial sem contato com metodologias digitais emergentes, enquanto a formação continuada, quando existente, mostra-se fragmentada, tecnicista e pouco conectada com as demandas reais do ensino. Essa lacuna compromete a qualidade e a inovação das práticas pedagógicas, tornando os professores pouco preparados para integrar tecnologias inteligentes em seus contextos de atuação. Nesse cenário, reforça-se que, embora a IA possua enorme potencial para personalizar o ensino e enriquecer a aprendizagem, sua adoção ainda é incipiente e desigual nos sistemas educacionais brasileiros [5].

A inovação da presente proposta está no fato de que ela não se limita a oferecer cursos ou treinamentos pontuais sobre o uso de tecnologias digitais, mas articula di- mensões que ainda não foram exploradas em conjunto pela literatura e pela prática. A primeira delas consiste no desenvolvimento de um Modelo Pedagógico transdiscipli- nar, fundamentado em Behar [8] e em abordagens críticas que ultrapassam as fronteiras disciplinares, buscando redefinir metodologias de ensino em diálogo com os princípios da Educação 5.0. Essa perspectiva rompe com iniciativas fragmentadas que tratam a IA como recurso isolado e propõe sua integração transversal aos processos formativos. A segunda dimensão refere-se à criação do repositório MEDIAR, que será inédito no Brasil, concebido para reunir, organizar e disponibilizar MEDs produzidos com apoio da IA em diferentes áreas do conhecimento. Ainda que existam repositórios educacionais nacionais e internacionais, como os de Recursos Educacionais Abertos (REA), não há registros de um ambiente sistemático voltado especificamente à produção de MEDs com IA, validados por professores em formação inicial e continuada.

Para fundamentar criticamente a concepção do repositório MEDIAR, será conduzida uma revisão sistemática da literatura sobre repositórios educacionais abertos e iniciativas de formação docente com IA, seguindo o protocolo PRISMA, de modo a identificar lacunas, boas práticas e limitações das experiências já documentadas.

Esse repositório amplia a democratização do acesso a práticas inovadoras, fomenta a colaboração entre docentes e possibilita a constituição de uma comunidade de prática em nível nacional. A terceira dimensão é de ordem metodológica e diz respeito à utilização da pesquisa-ação iterativa como eixo estruturante do estudo. Diferentemente de

investigações restritas a análises bibliográficas ou experiências pontuais, a pesquisa-ação garante que professores em formação inicial e em exercício atuem como protagonistas do processo, participando ativamente de ciclos de diagnóstico, implementação, avaliação e reflexão crítica. Esse movimento assegura validade prática e empírica ao modelo proposto e legitima o conhecimento produzido pelos próprios educadores.

A originalidade e o ineditismo do estudo se manifestam, portanto, em dois planos complementares. No plano teórico-metodológico, o trabalho propõe um Modelo Pedagógico transdisciplinar que integra IA, formação docente e Educação 5.0 em uma estrutura orgânica e validada empiricamente. No plano prático-tecnológico, apresenta a criação do repositório MEDIAR, um ambiente aberto e colaborativo capaz de reunir práticas reais de uso da IA em sala de aula, suprindo uma lacuna ainda não atendida por universidades, políticas públicas ou iniciativas privadas.

Dessa forma, demonstra-se que o problema de pesquisa ainda não foi resolvido por três razões centrais: inexistem modelos pedagógicos empiricamente validados que orientem a formação docente no uso crítico da IA em diferentes áreas do conhecimento; não há, até o momento, um repositório transdisciplinar de MEDs gerados com IA que organize, socialize e democratize práticas pedagógicas inovadoras no Brasil; e a literatura nacional e internacional, embora registre avanços significativos [5][3][7], carece de propostas sistêmicas que unam formação docente, prática pedagógica e disseminação em escala ampla. Em síntese, esta pesquisa não apenas responde a uma necessidade premente, mas apresenta uma solução inovadora e estruturante, com potencial para gerar impactos acadêmicos, pedagógicos e sociais duradouros.

2 Objetivos

No que tangem aos objetivos desta pesquisa é possível dividi-los em gerias e específicos. Desenvolver, implementar e avaliar um Modelo Pedagógico Transdisciplinar apoiado por Inteligência Artificial (IA) e articulado a um repositório digital de Materiais Educacionais Digitais (MEDs), visando à formação crítica e personalizada de professores na Educação 5.0.

Nesse sentido, surgem os objetivos específicos: a) Propor e estruturar o Modelo Pedagógico Transdisciplinar fundamentado em abordagens críticas da IA e da Educação 5.0; b) Implementar e avaliar empiricamente o modelo por meio de um curso teórico-prático com professores em formação inicial e continuada; c) Criar e disponibilizar o repositório MEDIAR como espaço aberto de socialização e curadoria dos MEDs desenvolvidos, promovendo a co-inteligência humano-máquina.s.

3 Contribuições esperadas e trabalhos relacionados

A presente pesquisa busca oferecer contribuições em diferentes níveis que dialogam entre si e respondem tanto às demandas teóricas quanto às práticas da Educação 5.0. Em nível acadêmico e científico, pretende-se propor um Modelo Pedagógico transdisciplinar que articula a Inteligência Artificial à formação docente, validado por meio da

pesquisa-ação e fundamentado em bases críticas e inovadoras. Do ponto de vista pedagógico e social, a contribuição esperada é a promoção de uma formação crítica e ética de professores para o uso da IA fomentando a autonomia docente, a capacidade de reflexão sobre as práticas e a personalização da aprendizagem em contextos diversos. No campo tecnológico, destaca-se a disponibilização de um repositório aberto, colaborativo e transdisciplinar, intitulado MEDIAR, que poderá ser utilizado em diferentes realidades educacionais, reunindo Materiais Educacionais Digitais (MEDs) criados por docentes em formação inicial e continuada.

Diferentemente dos repositórios de Recursos Educacionais Abertos (REA), o MEDIAR é concebido como um espaço de curadoria e validação pedagógica, no qual docentes atuam como coautores e avaliadores dos materiais produzidos com apoio da IA. Assim, o repositório não apenas armazena conteúdos, mas organiza um ecossistema formativo e colaborativo que integra pesquisa, prática e inovação pedagógica, fortalecendo a autoria docente e a circulação de práticas educacionais éticas e contextualizadas.

Em outro trabalho [10], os autores analisaram como soluções baseadas em IA vêm sendo utilizadas para adaptar métodos de instrução, conteúdos e ritmo de aprendizado às necessidades individuais de estudantes do ensino superior. Os resultados apontaram benefícios como maior engajamento, adaptação de conteúdos e eficiência administrativa, mas também revelaram lacunas quanto à padronização das práticas, à ética do uso e à aplicabilidade em contextos distintos, reforçando a necessidade de propostas que articulem formação docente, produção de materiais e criação de repositórios digitais.

investigação [11], foi possível examinar a percepção e o uso de ferramentas de IA por estudantes de medicina em diferentes etapas de sua formação. Foram consi- derados fatores como preocupações éticas, confiabilidade do conteúdo gerado por IA e variações dessas percepções ao longo do progresso acadêmico. Os autores concluíram que, mesmo em áreas com forte tradição científica, existe a necessidade de formações mais estruturadas, de acompanhamento crítico e da definição de diretrizes claras que orientem o uso seguro e pedagógico da IA.

No campo da formação docente, trabalhos [3] [7] [10] já evidenciaram o potencial da IA em promover práticas pedagógicas centradas no estudante, sobretudo na personalização da aprendizagem. Contudo, esses estudos não apresentam propostas concretas que articulem de modo sistêmico a formação docente, a produção de MEDs e a criação de repositórios digitais. É nesse ponto que a presente pesquisa se mostra inovadora, ao propor um ecossistema educacional integrado que combina a elaboração de um modelo pedagógico transdisciplinar, a capacitação docente e a constituição de um repositório digital de acesso aberto. Essa tríade permite não apenas a formação qualificada de professores, mas também a disseminação colaborativa de práticas pedagógicas mediadas por IA ampliando o alcance e a sustentabilidade das inovações educacionais.

4 Avaliação e divulgação científica

A avaliação dos resultados será conduzida em ciclos contínuos de pesquisa-ação, garantindo acompanhamento próximo e ajustes iterativos ao longo do processo. Inicialmente, será realizado um diagnóstico das competências digitais dos participantes, com o intuito de mapear conhecimentos prévios e necessidades formativas. Na sequência, será implementado o curso teórico-prático, que servirá como campo de experimentação do Modelo Pedagógico, possibilitando a análise das práticas desenvolvidas. Os MEDs produzidos serão avaliados quanto à sua relevância pedagógica, aplicabilidade em di-

ferentes áreas do conhecimento e potencial de personalização da aprendizagem. Ao final de cada ciclo, ocorrerá uma etapa de reflexão crítica, permitindo ajustes e aperfeiçoamentos no MP, de modo a assegurar sua validade empírica e pertinência prática.
A divulgação dos resultados se dará por múltiplos caminhos, visando alcançar tanto a comunidade científica quanto os sistemas educacionais. No âmbito acadêmico, serão realizadas publicações em periódicos de alto impacto, além de apresentações em congressos nacionais e internacionais voltados à Informática na Educação e à Educação 5.0. Em termos de impacto social e institucional, está previsto o lançamento público do repositório MEDIAR como espaço de acesso aberto, fortalecendo a democratização de práticas inovadoras. Complementarmente, serão produzidos relatórios de impacto e recomendações voltadas a gestores educacionais e à formulação de políticas públicas, de forma a ampliar a disseminação e a aplicabilidade dos resultados.

5 Resultados esperados e prospecções

Entre os resultados esperados, destaca-se a formação qualificada de licenciandos e docentes em exercício no uso pedagógico da IA promovendo maior autonomia e criticidade na utilização de tecnologias emergentes. Espera-se, ainda, a validação empírica de um Modelo Pedagógico transdisciplinar aplicável a diferentes disciplinas escolares, bem como a disponibilização do repositório MEDIAR com MEDs acessíveis, inovadores e passíveis de reutilização em distintos contextos. Um efeito adicional previsto é a geração de evidências de que a integração da IA pode potencializar processos de personalização e inclusão educacional, tornando o ensino mais responsivo às necessidades dos estudantes.
Para além desses resultados, o estudo projeta linhas de continuidade que poderão ampliar o alcance da pesquisa. Entre elas, destaca-se a expansão do MEDIAR para incluir recursos multimodais, como materiais em áudio, vídeo, realidade aumentada e simulações em realidade virtual, ampliando o espectro de experiências pedagógicas disponíveis. Também se prevê a integração do repositório a ambientes virtuais de aprendizagem institucionais, como Moodle, Canvas ou Google Classroom, fortalecendo sua aplicabilidade em redes escolares. Espera-se que o modelo e o repositório MEDIAR contribuam para consolidar uma cultura de co-inteligência entre humanos e máquinas, inspirando novas políticas públicas de formação docente pautadas em princípios éticos, inclusivos e sustentáveis. A proposta busca demonstrar que a IA, quando compreendida como tecnologia mediadora e não substitutiva, pode potencializar processos de personalização, reflexão e inovação no ensino, fortalecendo a autonomia e o protagonismo dos educadores.
Outra frente será o desenvolvimento de parcerias internacionais para ampliação do acervo e validação do modelo em diferentes contextos culturais e educacionais. Finalmente, o projeto abrirá espaço para explorar a co-inteli- gência artificial e a curadoria informacional como eixos estruturantes para a formação docente crítica e contínua, permitindo que professores não apenas utilizem a IA, mas também compreendam seu funcionamento e façam escolhas pedagógicas conscientes e fundamentadas.

References

1. Moravec, J. W. (Ed.). (2013). Knowmad society. Minneapolis, MN: Education Futures.
2. Luckin, R., Holmes, W., Griffiths, M., Forcier, L.B.: Intelligence Unleashed: An Argument for AI in Education. Pearson, London (2016)
3. Secchin, R. D., Silva, M. C., Costa, A., & Souza, F. (2024). A integração de tecnologias de inteligência artificial no currículo: implicações para a formação de professores. Revista Ibero-Americana de Humanidades, Ciências e Educação, 10(10), 4082–4103. https://doi.org/10.51891/rease.v10i10.16398).
4. Vicari, R. M., Brackmann, C., Mizusaki, L., & Galafassi, C. (2023). Inteligência artificial na educação básica: prática na escola. São Paulo, SP: Novatec.
5. Vicari, R. M. (2004). Inteligência artificial e educação: teoria e prática. [S.l.]: [s.n.].
6. Vicari, R. M. (2025, abril 15). IAgora Entrevista: o poder da IA na sala de aula [Vídeo]. YouTube. https://www.youtube.com/watch?v=WYDyr_a3Cec
7. Vicari, R. M. (Ed.). (2025). Nota Técnica 01 – Sistemas Tutores Inteligentes (Pinto, A. et al.). Porto Alegre, RS: Universidade Federal do Rio Grande do Sul, CINTED/PPGIE. https://www.ppgie.ufrgs.br/notas-tecnicas.
8. Behar, P.A.: Modelos Pedagógicos em Educação a Distância. Artmed, Porto Alegre (2019)
9. Holmes, W., Bialik, M., Fadel, C.: Artificial Intelligence in Education: Promise and Implications for Teaching and Learning. UNESCO, Paris (2019)
10. Merino-Campos, C., López, J., Torres, A.: The impact of artificial intelligence on personalized learning in higher education: A systematic review. Digital Learning Journal 4(2), 1–23 (2025). https://doi.org/10.3390/dlj4020017
11. Sunmboye, O., Chan, L., Adewale, F.: Exploring the influence of artificial intelligence integration on medical students' learning. BMC Medical Education 25(70), 1–13 (2025). https://doi.org/10.1186/s12909-025-07084-z.
12. Yao, Y. (2022). Human–machine co-intelligence through symbiosis in the SMV space. Applied Intelligence, 53(3), 2777–2797. https://doi.org/10.1007/s10489-022-03574-5.

Metodología para la adopción de herramientas de IA en el desarrollo de software

Darío Reyes Reina[1][0000-0002-9431-3344], Jenny Marcela Sanchéz-Torres[2][0000-0001-5284-836X], Iván Mauricio Rueda Cáceres[2][0000-0003-3139-9539]

[1] Estudiante del Doctorado en Ingeniería de Sistemas y Computación, Universidad Nacional de Colombia. Bogotá, Colombia
[2] Departamento de Ingeniería de Sistemas e Industrial, Facultad de Ingeniería de la Universidad Nacional de Colombia. Bogotá, Colombia
{dreyesr, jmsanchezt, imruedac}@unal.edu.co

Resumen. La propuesta doctoral aborda un vacío metodológico sobre la adopción estratégica y responsable de herramientas de inteligencia artificial (IA) en empresas desarrolladoras de software en Colombia. Aunque la adopción de estas tecnologías ofrece grandes oportunidades para optimizar procesos y aumentar la productividad, su incorporación plantea desafíos complejos, especialmente en lo que respecta a la colaboración entre humano-IA. En ese sentido, el objetivo general del estudio es diseñar una metodología para la adopción de herramientas de IA que oriente a los programadores en dicha colaboración. Se espera que los resultados del estudio contribuyan a mejorar los procesos de adopción de estas nuevas tecnologías. Particularmente, orientando la toma de decisiones sobre niveles apropiados de automatización, control y confianza. En otras palabras, delimitar qué tareas pueden delegadas a la IA, cuáles requieren mayor interacción humano–IA, cómo estructurar esta interacción, y cuáles requieren mayor control humano.

Palabras Claves: inteligencia artificial, adopción, desarrollo de software

1 Introducción

El crecimiento en el uso de la inteligencia artificial (IA) representa grandes oportunidades y desafíos. Se ha estimado que la IA tiene el potencial de impactar positivamente el 79% de los Objetivos de Desarrollo Sostenible (ODS), pero al mismo tiempo, que puede causar impactos negativos en un 35% de ellos (1).

Esta doble cara de la IA hace que sea esencial generar una comprensión detallada de la adopción de estas herramientas en contextos particulares y las mejores estrategias para hacerlo de manera adecuada. Específicamente, esta investigación se focaliza en la adopción de herramientas de IA en empresas desarrolladoras de software en Colombia.

En el país, aún hay un conocimiento limitado sobre cómo la IA está siendo utilizada. Cenisoft (Centro de Investigaciones de la Federación Colombia de Software), a partir de una muestra reducida de 69 empresas desarrolladoras, encontró que para 2024 el 68 % de estas estaba priorizando la adopción de IA (2). No obstante, aún persisten grandes vacíos sobre el entendimiento de ese proceso de adopción de la IA: sus desafíos de implementación, estrategias usadas para impulsar la adopción, casos de uso, e incluso, los posibles efectos y cambios que ha generado en los ciclos de desarrollo de software.

2 Planteamiento del problema

Se ha avanzado en una revisión sistemática de literatura (RSL) siguiendo los lineamientos propuestos por Kitchenham (3) para la práctica basada en evidencia en la ingeniería de. De este ejercicio preliminar, ha sido identificado que las empresas de desarrollo de software enfrentan cinco principales dificultades en la adopción de herramientas de IA: (I) falta de lineamientos y soporte institucional para impulsar la adopción de IA, (II) dificultades en la colaboración humano–IA, (III) complicaciones para su articulación con las rutinas de trabajo actuales, (IV) la percepción de riesgos por la adopción de IA, y (V) limitaciones presupuestales que limitan el uso de las herramientas de IA.

En la propuesta doctoral, se decide priorizar el segundo problema: dificultades en la interacción humano–IA. El problema tiene dos caras. Por una parte, se evidencia que hay una gran presión a que haya una adopción más intensiva de la IA con el objetivo de mejorar la productividad (4,5). Pero al mismo tiempo, no están dadas las garantías para hacerlo por la falta de lineamientos organizacionales que clarifiquen la posición de la empresa sobre el uso de IA y orienten su utilización, así como por falta de programas de entrenamiento (4,6–10). A esto se le suma, la misma velocidad con la que evolucionan las herramientas de IA (7,9–11), lo que complejiza incluso la identificación y la creación de consensos sobre que nuevas competencias que deberían adquirir los programadores en un entorno donde la IA tiene un rol central (5,7,9).

Las dificultades en la colaboración adecuada entre Humano-IA tiene algunas consecuencias. La calidad del software puede verse comprometida, al generar inconsistencias o alucinaciones que implique una intervención humana para subsanarlo (4,8,10,12–14). También comprometer la seguridad del software, ya que el código generado puede introducir vulnerabilidades que expongan información sensible (4,7–10,12,13). Adicionalmente, el uso de IA genera desafíos en la reproductibilidad (4,15) y explicabilidad (7,10,16,17) del código generado con ella, hay menor certidumbre sobre la su funcionamiento lo que entorpece trabajos de mantenimiento y evolución.

En este sentido, se plantea la siguiente pregunta de investigación que orientará el desarrollo del estudio doctoral: ¿Cómo diseñar una metodología para la adopción de herramientas de IA que guíe a los programadores en una adecuada colaboración Humano-IA?

3 Objetivo General

Diseñar una metodología para la adopción de herramientas de IA que oriente a los programadores en la colaboración humano–IA en el proceso de desarrollo de software.

3.1 Objetivos específicos

- Identificar los factores determinantes en la adopción de herramientas de IA en el desarrollo de software.
- Caracterizar el estado actual y los problemas en la adopción de herramientas de IA en un grupo de programadores de empresas colombianas de desarrollo de software.
- Definir las fases, actividades e instrumentos de la metodología para la adopción de herramientas de IA orientado a resolver los principales retos de la colaboración humano-IA: niveles adecuados de automatización, control humano y confianza.
- Probar la metodología de adopción con programadores de una empresa colombiana de desarrollo de software con el fin de retroalimentarlo.

4 Propuesta metodológica

Se propone que este estudio tenga un orientación pragmatista que es sugerida para problemas de investigación aplicados y situados en complejos contextos contemporáneos (18). Se tendrá un abordaje metodológico mixto que por medio de la triangulación y complementación de técnicas cualitativas y cuantitativas nos permita tener una visión integral del fenómeno. El diseño será un estudio de caso único con múltiples unidades de análisis. Específicamente, el escenario de estudio será la empresa en la cual trabaja el estudiante doctoral (caso único) que cuenta con áreas que abarcan todo el ciclo de desarrollo de software: Backend, Frontend, QA, Mobile, Arquitectura, CloudOps (múltiples unidades de análisis).

El diseño se justifica ya que se abordará un caso revelador, que consiste en una oportunidad poco común de acceso privilegiado a un fenómeno de interés complejo, que generalmente no está disponible para realizar estudios académicos (19). El investigador principal al estar vinculado profesionalmente con la empresa donde se desarrollará el estudio podrá observar y comprender en profundidad los procesos reales de adopción de herramientas de IA.

Ahora, con este diseño no se busca la generalización de los hallazgos, sino la posible transferibilidad de sus inferencias (20), garantizada mediante una descripción detallada del contexto, la documentación rigurosa del proceso investigativo y la triangulación de fuentes y métodos. Por cada uno de los objetivos definidos se ha elaborado una propuesta inicial de abordaje metodológico (Tabla 1).

Tabla 1. Relación objetivos - Metodología

Objetivo específico	Actividades y técnicas de investigación
Caracterizar el estado actual y los problemas en la adopción de herramientas de IA	-Encuesta aplicando el modelo *UTAUT Extended:* una prueba estandarizada para analizar los procesos de adopción de nuevas tecnologías -Entrevistas semiestructuradas a programadores
Identificar los factores determinantes en la adopción de herramientas de IA	Consolidación de evidencia: uso de hallazgos de revisión, encuestas y entrevistas.
Definir las fases, actividades e instrumentos de la metodología para la adopción de IA	Diseño de la metodología de adopción con los insumos generados
Probar la metodología de adopción con programadores de una empresa colombiana con el fin de retroalimentarlo.	Aplicación de la metodología de adopción en una empresa y retroalimentación de esta: realización entrevistas y notas de campo

Se reconoce la importancia de establecer métricas de evaluación que permitan valorar la efectividad de la metodología de adopción propuesta. No obstante, considerando los criterios de viabilidad y el alcance del trabajo doctoral, se proyecta como una etapa posterior a la tesis. La propuesta doctoral se concentrará en el diseño y retroalimentación de la metodología de adopción.

5 Contribuciones esperadas

Los resultados de este estudio generarán evidencia para implementar de manera estratégica y responsable la colaboración humano-IA en el proceso de desarrollo de software. La IA representa grandes oportunidades y desafíos que deben ser abordadas directamente para, por una parte, sacar provecho de los beneficios que genera, y por otra, minimizar los riesgos y posibles problemas que su uso conlleva.

En este contexto, los hallazgos del estudio orientarán sobre temas críticos en los cuáles hay brechas de conocimiento: la determinación de los niveles de automatización, control y confianza adecuados en la IA. En otras palabras, establecer qué puede y es correcto delegarse a la IA, qué actividades del desarrollo de software deben realizarse por medio de la colaboración humano-IA, cómo hacer esta interacción, y cuáles tareas, para evitar riesgos deben realizarse por los humanos.

Adicionalmente, este estudio hace una contribución al caracterizar la adopción de herramientas de IA y la colaboración humano - IA en un país latinoamericano. Como resultado de la RSL solo se identificó un estudio que tuvo participantes de la región, particularmente, de Brasil (8). Un vacío de conocimiento teniendo en cuenta que las particularidades latinoamericanas y colombianas, en materia de talento humano, cultura organizacional y disponibilidad de recursos, probablemente hacen que el proceso de adopción de estas tecnologías en el trabajo tenga diferencias significativas con otros países.

6 Resultados esperados y divulgación

Primero, una caracterización del estado actual de la adopción de herramientas de IA en programadores de una empresa de desarrollo de software en Colombia. En segundo lugar, generaremos una metodología que oriente a los programadores en la adopción de herramientas de IA. En tercer lugar, la formación doctoral del investigador principal, contribuyendo con personal altamente cualificado que actúa en la industria nacional. Respecto divulgación, esperamos generar dos productos: un artículo de revisión sistemática de literatura y un artículo de resultados. Adicionalmente, se planea la presentación del estudio en un congreso académico y otro en un espacio gremial de la industria de desarrollo de software en Colombia.

Referencias

1. Vinuesa R, Azizpour H, Leite I, Balaam M, Dignum V, Domisch S, et al. The role of artificial intelligence in achieving the Sustainable Development Goals. Nat Commun. 13 de enero de 2020;11(1):233.
2. CENISOFT. Estudio de apropiación tecnológica de la Inteligencia Artificial en las empresas empresas de Software y TI. Bogotá; 2024.
3. Kitchenham B, Charters S, Budgen D, Budgen P, Turner M, Linkman S, et al. Guidelines for performing systematic literature reviews in software engineering. Technical report, ver. 2.3 ebse technical report. ebse; 2007.
4. Banh L, Holldack F, Strobel G. Copiloting the future: How generative AI transforms Software Engineering [Internet]. Vol. 183, Information and Software Technology. 2025. Disponible en: https://www.scopus.com/inward/record.uri?eid=2-s2.0-105002314681&doi=10.1016%2fj.infsof.2025.107751&partnerID=40&md5=70f11299b67ce4a98a1d7fdd1b04ff49
5. Woodruff A, Shelby R, Kelley PG, Rousso-Schindler S, Smith-Loud J, Wilcox L. How Knowledge Workers Think Generative AI Will (Not) Transform Their Industries. En: Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems [Internet]. New York, NY, USA: Association for Computing Machinery; 2024. (CHI '24). Disponible en: https://doi.org/10.1145/3613904.3642700
6. Alenezi M, Akour M. AI-Driven Innovations in Software Engineering: A Review of Current Practices and Future Directions. Appl Sci. 28 de enero de 2025;15(3):1344.
7. Clear T, Cajander Å, Clear A, McDermott R, Daniels M, Divitini M, et al. AI Integration in the IT Professional Workplace: A Scoping Review and Interview Study with Implications for Education and Professional Competencies. En: 2024 Working Group Reports on Innovation and Technology in Computer Science Education [Internet]. New York, NY, USA: Association for Computing Machinery; 2025. p. 34-67. (ITiCSE 2024). Disponible en: https://doi.org/10.1145/3689187.3709607
8. Klemmer JH, Horstmann SA, Patnaik N, Ludden C, Burton C, Powers C, et al. Using AI Assistants in Software Development: A Qualitative Study on Security Practices and Concerns. En: Proceedings of the 2024 on ACM SIGSAC Conference on Computer and Communications Security [Internet]. New York, NY, USA: Association for Computing Machinery; 2024. p. 2726-40. (CCS '24). Disponible en: https://doi.org/10.1145/3658644.3690283
9. Li ZS, Arony NN, Awon AM, Damian D, Xu B. AI Tool Use and Adoption in Software Development by Individuals and Organizations: A Grounded Theory Study [Internet]. arXiv; 2024 [citado 10 de mayo de 2025]. Disponible en: http://arxiv.org/abs/2406.17325

10. Russo D. Navigating the Complexity of Generative AI Adoption in Software Engineering. ACM Trans Softw Eng Methodol [Internet]. junio de 2024;33(5). Disponible en: https://doi.org/10.1145/3652154
11. Suryavanshi P, Kapse M, Sharma V. Integrating ChatGPT into Software Development: Valuating Acceptance and Utilisation Among Developers. Vol. 19, Australasian Accounting, Business and Finance Journal. 2025. p. 96-117.
12. Mbizo T, Oosterwyk G, Tsibolane P, Kautondokwa P. Cautious Optimism: The Influence of Generative AI Tools in Software Development Projects. Vol. 2159 CCIS, Communications in Computer and Information Science. 2024. p. 361-73.
13. Serafini R, Yardim A, Naiakshina A. Exploring the Impact of Intervention Methods on Developers' Security Behavior in a Manipulated ChatGPT Study. En: Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems [Internet]. New York, NY, USA: Association for Computing Machinery; 2025. (CHI '25). Disponible en: https://doi.org/10.1145/3706598.3713989
14. Simaremare M, Edison H. The State of Generative AI Adoption from Software Practitioners' Perspective: An Empirical Study. Proceedings of the Euromicro Conference on Software Engineering and Advanced Applications, EUROMICRO-SEAA. 2024. p. 106-13.
15. Wuisang MC, Kurniawan M, Wira Santosa KA, Agung Santoso Gunawan A, Saputra KE. An Evaluation of the Effectiveness of OpenAI's ChatGPT for Automated Python Program Bug Fixing using QuixBugs. En: 2023 International Seminar on Application for Technology of Information and Communication (iSemantic) [Internet]. Semarang, Indonesia: IEEE; 2023 [citado 11 de agosto de 2025]. p. 295-300. Disponible en: https://ieeexplore.ieee.org/document/10295323/
16. Asal B, Demir MO. Enhancing Software Defect Prediction through Explainable AI: Integrating SHAP and LIME in a Voting Classifier Framework [Internet]. 8th International Artificial Intelligence and Data Processing Symposium, IDAP 2024. 2024. Disponible en: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85207868445&doi=10.1109%2fIDAP64064.2024.10710700&partnerID=40&md5=4b2d37ef83ffeda3db45f694447cdb1c
17. Samhan A, AlHajHassan S, Dabaa't SA, Elrashidi A. A Review of AI-Assisted Impact Analysis for Software Requirements Change: Challenges and Future Directions. En: 2024 25th International Arab Conference on Information Technology (ACIT) [Internet]. Zarqa, Jordan: IEEE; 2024 [citado 11 de agosto de 2025]. p. 1-13. Disponible en: https://ieeexplore.ieee.org/document/10877072/
18. Kaushik V, Walsh CA. Pragmatism as a Research Paradigm and Its Implications for Social Work Research. Soc Sci. 6 de septiembre de 2019;8(9):255.
19. Yin, R. Estudo de caso: planejamento e métodos. 5a ed. Porto Alegre: Bookman; 2015.
20. Hong QN, Fàbregues S. A Critical Reflection of Generalization in Mixed Methods Research. Eval Rev. 4 de abril de 2025;0193841X251331723.

Gestión de la función de pérdida integral y adaptativa en redes neuronales

Karen Kristel Avalos-Escalante[1 [0009−0002−1939−0312]], José Adán Hernández-Nolasco[1 [0000−0003−4671−0350]], Pablo Pancardo[1 [0000−0002−5482−6372]], and Noel Zacarías-Morales[1 [0000−0001−7850−6587]]

División Academica de Ciencias y Tecnologías de la Información, Universidad Juárez Autonóma de Tabasco, Tabasco 86690, Mexico; 241H18007@alumno.ujat.mx (K.K.A.-E.); adan.hernandez@ujat.mx (J.A.H.-N.); pablo.pancardo@ujat.mx (P. P)

Abstract. Las operaciones de perforación petrolera enfrentan eventos imprevistos, como el atrapamiento de la sarta, que generan retrasos y pérdidas económicas significativas. Este estudio propone una función de pérdida integral y adaptativa para redes neuronales secuenciales, orientada a mejorar la predicción temprana de dichos eventos. La propuesta surge ante las limitaciones de pérdidas estándar como MSE y MAE, que resultan insuficientes en contextos con datos secuenciales, no estacionarios y desbalanceados, donde los costos de error son asimétricos. El marco en proceso de desarrollo combina criterios de sensibilidad a clases minoritarias, estabilidad del gradiente, penalización de activaciones extremas y calibración probabilística, con ponderaciones dinámicas aprendibles. El modelo se validara mediante particionado temporal, estudios de ablación y comparaciones frente a pérdidas convencionales, utilizando datos reales de pozos del sur de México. Los resultados esperados incluyen una mejora del 15% en el tiempo de respuesta y un aumento del 5% en la eficiencia respecto a MSE, manteniendo estabilidad y generalización. Este trabajo contribuye con un marco teórico–práctico reproducible, evidencia empírica de su efectividad y una propuesta transferible a otros dominios de datos secuenciales, como mantenimiento predictivo, diagnóstico industrial o monitoreo energético, favoreciendo operaciones más seguras, eficientes y sostenibles.

Keywords: Adaptive loss function · Deep learning · Stuck pipe prediction · Oil well drilling · Time-series modeling · Early warning systems

1 Introducción

Las operaciones de perforación de pozos petroleros enfrentan riesgos operativos críticos, entre ellos el atrapamiento de la sarta de perforación, que ocurre en entornos con datos secuenciales, ruidosos y de alta variabilidad. Para mitigar estos riesgos se requieren sistemas predictivos confiables y adaptativos, capaces de aprender patrones temporales complejos a partir de datos históricos y en tiempo

real. En este contexto, las redes neuronales secuenciales, particularmente las arquitecturas LSTM (Long Short-Term Memory) [10] y GRU (Gated Recurrent Unit) [11], han mostrado una gran capacidad para modelar la dinámica temporal inherente a los procesos de perforación.

Dentro del aprendizaje profundo, la función de pérdida juega un papel esencial al cuantificar el error entre las predicciones del modelo y los valores reales, guiando así el proceso de optimización. Sin embargo, las funciones de pérdida convencionales, como el Mean Squared Error (MSE) y el Mean Absolute Error (MAE), presentan limitaciones en entornos industriales complejos caracterizados por datos no estacionarios, desbalanceados y con costos de error asimétricos. Estas condiciones afectan la estabilidad y la capacidad de generalización de las redes neuronales secuenciales. Autores como *Wang et al.* [1] y *Terven et al.* [2] destacan la tendencia actual de diseñar funciones de pérdida robustas y adaptativas, capaces de ajustarse dinámicamente a la distribución de los datos. En esta línea, *Barron* [3] propuso la *General and Adaptive Robust Loss*, que permite transitar entre MSE, MAE y Cauchy mediante un único parámetro continuo; mientras que *Heydari et al.* [4] desarrollaron *SoftAdapt*, un método que asigna automáticamente pesos a múltiples pérdidas, mejorando la estabilidad durante el entrenamiento.

En aplicaciones industriales, la predicción temprana de eventos adversos, como el atrapamiento de la sarta de perforación que requiere no solo arquitecturas de red adecuadas, sino también funciones de pérdida que integren sensibilidad ante clases minoritarias, estabilidad del gradiente y penalización de activaciones extremas. Este enfoque se relaciona con los principios del aprendizaje costo-sensible [13], especialmente útil cuando los falsos negativos conllevan consecuencias económicas graves.

En este trabajo se propone una función de pérdida integral y adaptativa orientada a mejorar la estabilidad, la calibración [12] y la sensibilidad de las redes neuronales secuenciales bajo condiciones operativas reales. El objetivo es anticipar el atrapamiento de la sarta de perforación, reduciendo los tiempos de respuesta y las pérdidas económicas asociadas. Además, el marco propuesto busca ser extensible a otros dominios con secuencias temporales y altos costos de error, como el mantenimiento predictivo, la detección de fallas industriales, la predicción del consumo energético o el diagnóstico médico basado en series fisiológicas.

2 Planteamiento del problema

En las redes neuronales, la función de pérdida constituye el componente central del riesgo estructural, pues cuantifica la discrepancia entre las predicciones y los valores reales, guiando el ajuste de los parámetros del modelo. Su elección incide directamente en la eficacia del aprendizaje. En contextos con grandes volúmenes de datos o clases desbalanceadas, suele ser necesario combinar distintas pérdidas para equilibrar objetivos; si bien esto mejora la cobertura, también puede reducir la interpretabilidad si no se diseña cuidadosamente [1]. Las pérdidas estándar

para regresión (MSE y MAE), aunque convexas, presentan limitaciones en redes profundas altamente no lineales [2] y no penalizan adecuadamente activaciones excesivas [4].

La investigación en pérdidas adaptativas ha avanzado considerablemente; sin embargo, persisten retos como la convergencia lenta o la ponderación subóptima entre términos, lo que deja un vacío metodológico [3]. Este vacío resulta crítico en entornos industriales con datos secuenciales, como la perforación de pozos, donde anticipar eventos no programados, por ejemplo, el atrapamiento de la sarta, que requiere funciones de pérdida capaces de capturar múltiples criterios bajo distribuciones desbalanceadas.

3 Justificación

La presente investigación aborda un problema de alta relevancia para la industria petrolera: el atrapamiento de la sarta de perforación, un evento no deseado que genera pérdidas económicas significativas y riesgos operativos elevados [14]. A pesar de los avances en monitoreo y automatización, las técnicas tradicionales aún presentan limitaciones para anticipar este tipo de eventos en entornos con datos secuenciales, ruidosos y de alta variabilidad. La sarta de perforación se define como el conjunto de tuberías metálicas conectadas entre sí que transmiten el movimiento de rotación y el peso desde la superficie hasta la barrena, herramienta encargada de cortar la formación geológica. Su función es, por tanto, permitir la transmisión del torque y del peso sobre la barrena, así como facilitar la circulación del fluido de perforación. Debido a las condiciones extremas de presión, temperatura y fricción, la sarta puede sufrir atrapamientos mecánicos o diferenciales, ocasionando interrupciones costosas en la operación.

En este contexto, las redes neuronales secuenciales han demostrado un notable potencial para modelar dinámicas temporales complejas; sin embargo, su desempeño depende en gran medida de la función de pérdida empleada durante el entrenamiento. Las pérdidas estándar, como el Mean Squared Error (MSE) y el Mean Absolute Error (MAE), suelen ser insuficientes ante datos no estacionarios, desbalanceados o con costos de error asimétricos, lo que limita la capacidad predictiva y la estabilidad del modelo.

El presente trabajo propone una función de pérdida integral y adaptativa para redes neuronales secuenciales aplicadas al proceso de perforación petrolera. La motivación central surge de la necesidad de contar con un mecanismo de optimización capaz de integrar criterios múltiples y ponderaciones dinámicas, mejorando la sensibilidad a clases minoritarias, la estabilidad del gradiente y la calibración probabilística de las predicciones.

Desde una perspectiva científica, la investigación busca sistematizar y ampliar el conocimiento sobre funciones de pérdida robustas, formalizando sus propiedades teóricas y proponiendo esquemas de ponderación que incrementen la eficacia y la resiliencia del aprendizaje profundo. En el plano industrial, se pretende validar un marco genérico que habilite alertas tempranas y estrategias de mitigación de riesgos aplicables en tiempo real, directamente vinculadas con

la operación de perforación. Este enfoque no solo contribuye a la reducción de tiempos no productivos, sino que también favorece la seguridad operativa y la sostenibilidad económica del proceso.

4 Investigaciones previas

El desarrollo de funciones de pérdida adaptativas ha cobrado relevancia como estrategia para mejorar la estabilidad y adaptabilidad de los modelos de aprendizaje profundo en entornos industriales. Por ejemplo, [5] propusieron un esquema de ponderación automática en redes informadas por la física, donde un parámetro λ aprende a equilibrar distintas pérdidas, logrando reducciones significativas del error relativo. Asimismo, [6] desarrollaron variantes de pérdida en el algoritmo XGBoost para problemas industriales con clases críticas, como la "Adaptive Weighted Softmax Loss", que mejora la precisión en escenarios con altos costos de error. Aunque estos estudios pertenecen a dominios distintos, evidencian la efectividad de las funciones de pérdida adaptativas en entornos complejos y desbalanceados, lo que refuerza la pertinencia de la propuesta de este trabajo.

En la industria petrolera ya se han empleado modelos de aprendizaje automático, desde clasificadores clásicos hasta redes neuronales, para anticipar eventos no programados. Estudios recientes han logrado detecciones en tiempo real de síntomas asociados a presión diferencial [7]. En el ámbito comercial, la herramienta Exebenus Spotter analiza datos superficiales en tiempo real y emite alertas tempranas, integrándose con plataformas como SiteCom de Kongsberg Digital [8,9]. Sin embargo, aún no existen marcos generales, explicables y transferibles que aborden distintos mecanismos de atrapamiento bajo diversos contextos geológicos, lo cual constituye la motivación de esta investigación.

5 Objetivos

5.1 Objetivo general

Desarrollar y validar un modelo de red neuronal secuencial optimizado mediante una función de pérdida integral y adaptativa, capaz de mejorar la predicción temprana del atrapamiento de la sarta de perforación en pozos petroleros de la región sur, incrementando la precisión, estabilidad y eficiencia operativa frente a enfoques convencionales.

5.2 Objetivos específicos

1. Diseñar un marco teórico–matemático para la formulación de una función de pérdida integral y adaptativa que combine criterios de sensibilidad, estabilidad del gradiente y calibración probabilística mediante ponderaciones dinámicas aprendibles.

2. Implementar experimentalmente la función propuesta en un modelo de red neuronal secuencial con datos históricos reales de perforación, asegurando su reproducibilidad y escalabilidad.
3. Evaluar el desempeño del modelo mediante métricas de error, estabilidad y generalización, comparando los resultados con pérdidas estándar (MSE y MAE) y analizando los beneficios en la reducción de errores críticos y tiempo de respuesta.
4. Analizar la aplicabilidad del marco propuesto en otros dominios con datos secuenciales y distribuciones desbalanceadas (por ejemplo, mantenimiento predictivo o diagnóstico industrial).

6 Hipótesis

La hipótesis central plantea que incorporar una función de pérdida integral y adaptativa en el entrenamiento de una red neuronal secuencial mejora significativamente la capacidad predictiva, la estabilidad del aprendizaje y la sensibilidad ante eventos infrecuentes, en comparación con funciones convencionales como el Mean Squared Error (MSE). Se espera que el modelo propuesto optimice la detección del atrapamiento de la sarta, reduciendo los tiempos de respuesta y aumentando la eficiencia operativa. Esta mejora será validada mediante métricas de precisión, estabilidad temporal y generalización, aplicadas a datos históricos de pozos petroleros de la región sur de México.

7 Contribuciones esperadas

Las contribuciones de esta investigación se organizan en dos ejes: científico e industrial. Desde la perspectiva científica, se espera demostrar que la función de pérdida integral y adaptativa mejora la estabilidad del entrenamiento, la calibración probabilística y la sensibilidad ante clases minoritarias, superando a pérdidas tradicionales como MSE y MAE. Además, se desarrollará un modelo de red neuronal secuencial optimizado que contribuirá al conocimiento teórico sobre pérdidas robustas y adaptativas y establecerá guías para su aplicación en otros contextos complejos y desbalanceados.

En el ámbito industrial, la propuesta busca evidenciar beneficios tangibles en la reducción de costos operativos, la disminución del tiempo no productivo y la mitigación de incidentes por atrapamiento. El modelo se concibe como una herramienta de apoyo para la toma de decisiones y la emisión de alertas tempranas, contribuyendo a la seguridad, la eficiencia y la sostenibilidad de las operaciones. Se prevé también la generación de propiedad intelectual y la publicación de artículos científicos derivados del proyecto, fortaleciendo el liderazgo académico e institucional de la Universidad Juárez Autónoma de Tabasco.

8 Discusión y trabajo futuro

Los resultados obtenidos hasta el momento permiten identificar aportaciones relevantes tanto en el ámbito científico como en el industrial. Desde la perspectiva

científica, la principal contribución consiste en la formulación y validación inicial de una función de pérdida integral y adaptativa, capaz de combinar criterios de sensibilidad, estabilidad y calibración probabilística mediante ponderaciones dinámicas aprendibles. Este enfoque representa un avance conceptual frente a las pérdidas tradicionales al integrar de forma explícita factores asociados al desbalance de clases y la dinámica temporal de los datos.

Desde la perspectiva industrial, la propuesta demuestra su potencial como herramienta predictiva para la operación segura y eficiente de pozos petroleros, al mejorar la detección temprana del atrapamiento de la sarta de perforación y reducir los tiempos de respuesta en comparación con técnicas de monitoreo tradicionales. Estos resultados aportan evidencia del valor práctico del enfoque en escenarios reales, contribuyendo a la disminución de pérdidas económicas, la optimización de recursos y la mitigación de riesgos operativos. Asimismo, la integración del modelo en un flujo de monitoreo continuo abre la posibilidad de su adopción en entornos de control industrial y toma de decisiones en tiempo real.

Para fortalecer estos hallazgos, se desarrollará una fase de validación ampliada mediante particionado temporal, entrenando el modelo con datos históricos y evaluando su capacidad predictiva en periodos futuros, garantizando la ausencia de fuga de información. De manera complementaria, se aplicará un análisis de ablación para cuantificar el aporte individual de cada componente de la pérdida y se realizarán comparaciones sistemáticas frente a pérdidas estándar.

Como líneas de trabajo futuro, se prevé ampliar la aplicación del modelo a otros mecanismos de atrapamiento y condiciones de perforación, así como explorar su adaptación a diferentes campos geológicos mediante técnicas de transfer learning. De igual forma, se contempla la incorporación de mecanismos de explicabilidad e interpretación que permitan comprender las decisiones del modelo en tiempo real, fortaleciendo su adopción en entornos industriales críticos. Finalmente, se proyecta la extensión del marco propuesto a otros dominios de datos secuenciales, como mantenimiento predictivo, diagnóstico industrial o análisis de comportamiento dinámico en sistemas energéticos, consolidando así su aplicabilidad transversal.

References

1. Wang, Q., Ma, Y., Zhao, K., Tian, Y.: A comprehensive survey of loss functions in machine learning. *Annals of Data Science* **9**(2), 187–212 (2022). https://doi.org/10.1007/s40745-020-00253-5
2. Terven, J., Córdova-Esparza, D.M., Romero-González, J.A., Ramírez-Pedraza, A., Chávez-Urbiola, E.A.: A comprehensive survey of loss functions and metrics in deep learning. *Artificial Intelligence Review* **58**, 195 (2025). https://doi.org/10.1007/s10462-025-11198-7
3. Barron, J.T.: A general and adaptive robust loss function. In: *Proc. IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR)*, pp. 4331–4339. IEEE, Long Beach (2019). https://doi.org/10.1109/CVPR.2019.00446

4. Heydari, A.A., Thompson, C.A., Mehmood, A.: SoftAdapt: techniques for adaptive loss weighting of neural networks with multi-part loss functions. *arXiv preprint arXiv:1912.12355* (2019). https://arxiv.org/abs/1912.12355
5. Son, H., Cho, S.W., Hwang, H.J.: Enhanced physics-informed neural networks with augmented Lagrangian relaxation method (AL-PINNs). *arXiv preprint arXiv:2205.01059* (2022). https://arxiv.org/abs/2205.01059
6. Bukowski, M., Kurek, J., Świderski, B., Jegorowa, A.: Custom loss functions in XGBoost algorithm for enhanced critical error mitigation in drill-wear analysis of melamine-faced chipboard. *Sensors* **24**(4), 1092 (2024). https://doi.org/10.3390/s24041092
7. Meor Hashim, M.M., Yusoff, M.H., Arriffin, M.F., *et al.*: Utilizing artificial neural network for real-time prediction of differential sticking symptoms. In: *International Petroleum Technology Conference (IPTC)* (2021). https://doi.org/10.2523/IPTC-21221-MS
8. Payrazyan, V.K., Robinson, T.S.: Leveraging targeted machine learning for early warning and prevention of stuck pipe, tight holes, pack offs, hole cleaning issues and other potential drilling hazards. In: *Offshore Technology Conference (OTC)*, Houston (2023). https://doi.org/10.4043/32169-MS
9. Exebenus AS, Kongsberg Digital: Strategic partnership to enhance oil and gas operations efficiency. Press release (2023). https://kongsbergdigital.com/news/exebenus-as-and-kongsberg-digital-form-strategic-partnership-to-enhance-oil-and-gas-operations-efficiency
10. Hochreiter, S., Schmidhuber, J.: Long short-term memory. *Neural Computation* **9**(8), 1735–1780 (1997). https://doi.org/10.1162/neco.1997.9.8.1735
11. Chung, J., Gulcehre, C., Cho, K., Bengio, Y.: Empirical evaluation of gated recurrent neural networks on sequence modeling. *arXiv preprint arXiv:1412.3555* (2014). https://arxiv.org/abs/1412.3555
12. Guo, C., Pleiss, G., Sun, Y., Weinberger, K.Q.: On calibration of modern neural networks. In: *Proc. 34th Int. Conf. on Machine Learning (ICML)*, pp. 1321–1330 (2017).
13. Elkan, C.: The foundations of cost-sensitive learning. In: *Proc. 17th Int. Joint Conf. on Artificial Intelligence (IJCAI)*, pp. 973–978 (2001).
14. Lake, L.W., *et al.*: *Petroleum Engineering Handbook*. Society of Petroleum Engineers, Richardson, TX (2018).

Minería de datos para identificar intereses académicos en estudiantes universitarios: un estudio preliminar

Fátima Guadalupe Montejo-Collado[0009−0007−2694−0307], Oscar Chávez-Bosquez[0000−0002−0324−9886] y José Hernández-Torruco[0000−0003−3146−9349]

División Académica de Ciencias y Tecnologías de la Información, Universidad Juárez Autónoma de Tabasco, 86690. Cunduacán, Tabasco, México
252H23002@alumno.ujat.mx, {oscar.chavez,jose.hernandezt}@ujat.mx
https://www.ujat.mx

Resumen Conocer los intereses académicos de los estudiantes es fundamental para fortalecer la pertinencia educativa, la motivación y la permanencia escolar. Este estudio identificó dichos intereses en alumnos de licenciatura de la DACyTI (UJAT) mediante cuestionarios diseñados para cada carrera. Los datos obtenidos se analizaron con técnicas de minería de datos empleando clasificadores como Random Forest, J48, JRip, OneR, KNN y SVM, así como filtros de selección de atributos (CFS, Chi-Squared, *Information Gain* y *Random Forest Importance*). Las métricas consideradas fueron Accuracy, Kappa, sensibilidad, especificidad y *Balanced accuracy*. Los resultados mostraron desempeños elevados en la mayoría de los modelos, sin embargo, estos valores se explican por el desbalance del conjunto de datos, que genera patrones altamente memorables y conduce a sobreajuste. Los métodos filtro coincidieron en que las calificaciones y la experiencia previa en el área seleccionada son los atributos más influyentes para predecir el interés académico. Este hallazgo sugiere que la afinidad de los estudiantes hacia un área está fuertemente relacionada con su desempeño previo.

Palabras clave: Inteligencia Artificial, Análisis exploratorio de datos, Instrumentos educativos.

1 Introducción

Comprender los intereses académicos de los estudiantes es fundamental para diseñar programas pertinentes, orientar vocacionalmente y reducir la deserción; sin embargo, estos intereses suelen ser diversos y difíciles de identificar mediante métodos tradicionales. La minería de datos permite analizar grandes volúmenes de información y descubrir patrones útiles para la toma de decisiones.

En este estudio se aplicó un cuestionario a estudiantes de la División Académica de Ciencias y Tecnologías de la Información (DACyTI) de la Universidad Juárez Autónoma de Tabasco (UJAT), con al menos 40% de avance curricular. En total

se recolectaron 81 variables demográficas, académicas, motivacionales y técnicas. Este conjunto de datos permitió identificar tendencias sobre las preferencias formativas de los estudiantes.

Los resultados del análisis ofrecen a la institución una herramienta analítica para optimizar planes de estudio, fortalecer líneas de especialización y alinear la formación académica con las expectativas profesionales, contribuyendo a una educación más pertinente y motivante.

1.1 Trabajos relacionados

Diversos estudios han aplicado técnicas de minería de datos en el ámbito educativo. En [2], se analizaron 54 atributos académicos, demográficos y psicológicos para predecir deserción, evaluando algoritmos como Naïve Bayes, SVM, Árboles de Decisión y Métodos de ensamble. En [1], se emplearon modelos supervisados para recomendar especialidades universitarias, destacando Random Forest y Gradient Boosting. Por otro lado, [7] desarrollaron un modelo basado en Árboles de Decisión para predecir deserción en universidades mexicanas, alcanzando 86% de precisión. En conjunto, estos trabajos muestran la utilidad del aprendizaje automático para analizar perfiles estudiantiles y apoyar la toma de decisiones educativas.

2 Planteamiento y justificación del problema

2.1 Definición del problema

Conocer los intereses académicos de los estudiantes es esencial para diseñar programas pertinentes y reducir la deserción; sin embargo, estos intereses son variados y difícilmente detectables mediante métodos tradicionales. Para atender esta limitación, se construyó un conjunto de datos a partir de encuestas aplicadas en la DACyTI con el fin de utilizar minería de datos que permita identificar patrones no evidentes y apoyar la toma de decisiones institucionales relacionadas con la oferta académica.

2.2 Objetivos

Objetivo general: Realizar un análisis preliminar de los intereses académicos de estudiantes de licenciatura aplicando minería de datos.

Objetivos específicos:
- Diseñar el instrumento y aplicar la encuesta.
- Crear un dataset de los intereses académicos de estudiantes de licenciatura.
- Descubrir patrones preliminares en el dataset creado.

2.3 Metodología

En la Figura 1 se muestra la metodología utilizada.

Fig. 1: Flujo metodológico del estudio.

2.4 Justificación

La identificación de intereses académicos suele basarse en observación o métodos tradicionales que no permiten detectar patrones complejos. Por ello, se empleó la minería de datos, técnica ampliamente utilizada en educación para predecir rendimiento y deserción, con el fin de descubrir tendencias ocultas en las preferencias estudiantiles. Esta aproximación permite generar modelos analíticos que apoyan la actualización curricular, el diseño de programas más atractivos y la orientación personalizada.

2.5 Instrumento

Ante la falta de un instrumento específico, se diseñó una encuesta diferenciada para las cuatro carreras de la DACyTI: Licenciatura en Tecnologías de la Información (LTI), Licenciatura en Sistemas Computacionales (LSC), Ingeniería en Sistemas Computacionales (ISC) e Ingeniería en Informática Administrativa (IIA)]. El cuestionario estaba dividido en una sección general y otra específica por área formativa. La mayoría de las preguntas fueron de opción múltiple, su contenido fue validado por expertos y aprobado por el Comité de Ética. La encuesta se aplicó en línea a 152 estudiantes con al menos 40% de avance curricular.

2.6 Conjunto de datos y preprocesamiento

El dataset final contiene 152 instancias y 81 variables, se integraron los archivos de las cuatro carreras, se simplificaron los nombres de los atributos y las celdas vacías se codificaron con un valor de 99. Las variables `AreaInteres` y `Genero` se transformaron a factor y se agregó la variable `Lic` para identificar la carrera, el conjunto de datos quedó conformado por 75 variables para su análisis en RStudio.

3 Experimentos y resultados

Con el fin de evaluar el desempeño de los modelos bajo diferentes configuraciones del dataset, se realizaron 10 experimentos entre los que destacan: el uso de todas las variables, reducción mediante métodos filtro y análisis por licenciatura (ver Tabla 1). En cada caso, los modelos se entrenaron con $\frac{2}{3}$ de los datos y se evaluaron con el $\frac{1}{3}$ restante, repitiendo cada ejecución 30 veces para obtener estimaciones estables. Los clasificadores utilizados fueron Random Forest, J48, JRip, OneR, KNN y SVM (lineal, polinomial, radial y sigmoide), evaluados mediante *accuracy*, especificidad, sensibilidad, Kappa y *accuracy balanceado*.

3.1 Resultados generales de los modelos

En varios experimentos se obtuvieron métricas muy altas, pero estos valores no reflejan modelos robustos. La variable `AreaInteres` está fuertemente desbalanceada y algunos atributos presentan alta correlación con la clase, lo que reduce la complejidad del problema y favorece predicciones que parecen perfectas pero no lo son. Por ello, las métricas elevadas responden más a las características del dataset que al desempeño real de los clasificadores. Aun así, Random Forest, J48 y JRip mostraron los resultados más consistentes dentro de este análisis exploratorio.

3.2 Selección de atributos

Los métodos CFS, Chi-Squared, Information Gain y Random Forest Importance coincidieron en señalar las calificaciones y la experiencia previa en el área seleccionada como los atributos más relevantes. CFS incluso redujo el conjunto a un solo atributo, lo que evidencia una fuerte correlación entre desempeño previo e interés académico. Esto también explica las métricas perfectas observadas.

3.3 Resultados por carrera

El análisis por carrera mostró que en programas con mayor número de estudiantes (como ISC) se obtuvieron métricas perfectas debido al desbalance y la homogeneidad. En carreras con pocas instancias (LTI, LSC, IIA) algunas métricas por clase no pudieron calcularse, evidenciando insuficiencia de datos.

3.4 Hallazgos principales

Un resultado consistente en todos los experimentos fue que los estudiantes eligen como área de interés aquella en la que obtienen mejores calificaciones y poseen experiencia previa. Este patrón fue validado por los métodos filtro y por el comportamiento de los modelos, constituyendo el principal aporte del análisis.

La Tabla 1 resume los 10 experimentos realizados y muestra algunas de las variables del conjunto de datos.

Table 1: Experimentos realizados

Exp	Objetivo	Núm. de variables usadas	Métricas con mejor resultado	Hallazgo
1	Evaluar si las variables generales del cuestionario permiten predecir el área de interés académico.	Primeras 20 variables (sección general): Genero, Edad, Semestre, Lic, RazonEleccion, FormacionPrevia, CapacitacionPrevia, BloquePrevio, AreaLabora, PorGusto, MetodoAprende, ActExtraCur, ModTit, ActAcadInteres, NumConvenios, MaestriaPrefiere, AreaInteres.	Accuracy: 0.88, Especificidad: 0.92, Sensibilidad: 0.50, Kappa: 0.39, Acc. balanceado: 0.71.	J48 obtuvo el mejor desempeño; sensibilidad baja por desbalance de clases.
2	Evaluar el desempeño con todas las variables del dataset preprocesado.	75 variables.	Todas las métricas = 1.	Sobreajuste por alto número de predictores respecto al tamaño de la muestra.
3	Evaluar selección de atributos con Random Forest Importance.	51 variables relevantes según RF (X9CalificacionEntornoSocial, X2ExperienciaPreviaEntornoSocial, X7QueTeMotivaProgramacion, X8CalificacionProgramacion, X6ContinuarEstProgramacion, BloquePrevio, NumConvenios, AreaLabora, Genero).	Todas las métricas = 1.	Variables críticas generan separación casi perfecta entre clases.
4	Evaluar impacto de selección de atributos con método CFS.	1 variable relevante: X9CalificacionEntornoSocial.	Todas las métricas = 1.	Modelos memorizaron patrones simples.
5 y 6	Evaluar selección de atributos con Chi-Squared e information gain.	18 variables relevantes (Genero, X7QueTeMotivaProgramacion, X8CalificacionProgramacion, X6ContinuarEstProgramacion, X9CalificacionEntornoSocial, X9VecesReprobadasProgramacion, X10VecesReprobadasEntornoSocial).	Todas las métricas = 1.	Variables informativas producen modelos altamente deterministas.
7	Evaluar desempeño en subset de datos de Ingeniería en Informática Administrativa (IIA).	74 variables, 32 instancias.	Accuracy, Kappa, Especificidad = 1; Sensibilidad y Acc. balanceado = 0.83.	Los modelos con mejor desempeño fueron RF, J48 y JRip, con las métricas mencionadas anteriormente. esto quiere decir que hicieron una clasificación correcta, aunque algunas clases con pocos estudiantes impiden calcular métricas por clase.
8	Evaluar desempeño en subset de datos de Ingeniería en Sistemas Computacionales (ISC).	74 variables, 103 instancias.	Todas las métricas = 1 (RF y SVM).	Homogeneidad favorece clasificaciones perfectas.
9	Evaluar desempeño en subset de datos de Licenciatura en Sistemas Computacionales (LSC).	47 variables, 11 instancias.	Accuracy = 1; métricas por clase = NaN.	Desbalance extremo.
10	Evaluar desempeño en subset de datos de Licenciatura en Tecnologías de la Información (LTI).	53 variables, 6 instancias.	Accuracy = 1; SVM sigmoide = 0; métricas por clase = NaN.	Clases muy pequeñas e irregulares.

Las métricas igual a 1.0 se explican por el fuerte desbalance del dataset, especialmente en las áreas de interés con muy pocos estudiantes, lo que provoca el sobreajuste y genera precisiones erróneamente altas. Por otro lado, se aplicaron métodos filtro para identificar variables relevantes. La Figura 2 muestra la cantidad de atributos seleccionados por cada uno.

De los métodos filtro, CFS fue el método que seleccionó menos variables, priorizando las calificaciones del área de interés. Por su parte, Chi-Squared e Information Gain coincidieron en la mayoría de los experimentos al identificar prácticamente el mismo conjunto de variables relevantes.

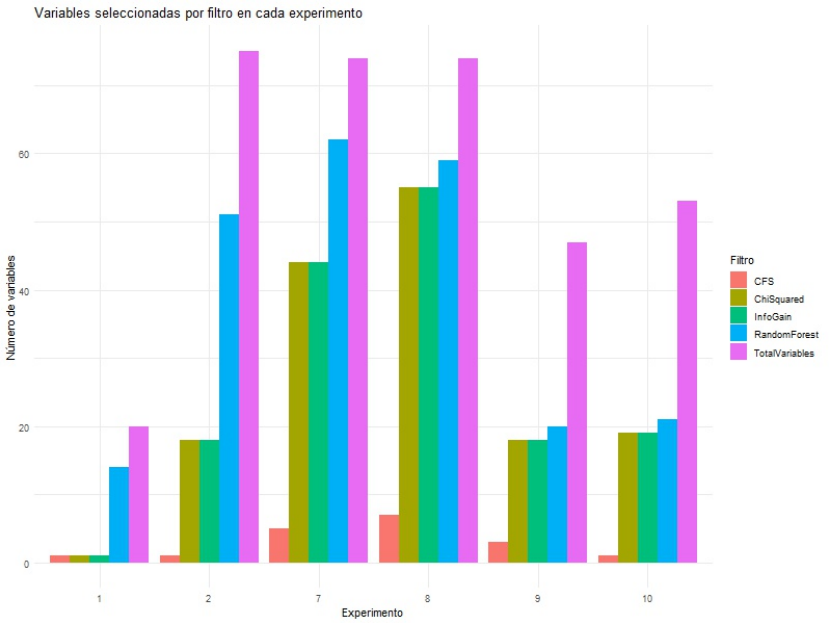


Fig. 2: Variables relevantes encontradas por los métodos filtro.

4 Conclusiones y trabajos futuros

Este estudio analizó los intereses académicos de los estudiantes de la DACyTI mediante técnicas de minería de datos aplicadas a un conjunto propio de encuestas. A través de 10 experimentos con distintos conjuntos de variables y análisis por carrera, se observó que la calificación y la experiencia previa en un área constituyen los factores más influyentes para determinar el interés académico de los estudiantes. Los métodos filtro coincidieron en este hallazgo, y los modelos de clasificación replicaron dicho comportamiento.

Aunque varios modelos mostraron métricas perfectas, estos resultados se deben al desbalance del dataset y a la existencia de variables fuertemente correlacionadas con la clase objetivo, lo que produjo el sobreajuste. Por lo tanto, los resultados podemos interpretarlos como un análisis exploratorio preliminar con capacidad de mejora.

Como trabajos futuros se propone: aplicar técnicas de balanceo de clases para generar modelos más robustos, integrar metaheurísticas y métodos de ensamble para mejorar la selección de atributos, validar y ampliar el instrumento aplicado, y explorar estudios adicionales sobre deserción, motivación y trayectorias académicas. Este estudio está limitado por el tamaño reducido y el desbalance del dataset, lo que restringe la generalización de los resultados.

References

1. Alsayed, A.O., Rahim, M.S.M., AlBidewi, I., Hussain, M., Jabeen, S.H., Alromema, N., Hussain, S., Jibril, M.L.: Selection of the right undergraduate major by students using supervised learning techniques. applied sciences **11**(22), 10639 (2021), https://doi.org/10.3390/app112210639
2. Cruz, E., González, M., Rangel, J.C.: Técnicas de machine learning aplicadas a la evaluación del rendimiento ya la predicción de la deserción de estudiantes universitarios, una revisión. Prisma Tecnológico **13**(1), 77–87 (2022), https://doi.org/10.33412/pri.v13.1.3039
3. César Pérez López, D.S.G.: Minería de datos, técnicas y herramientas. (2007), https://books.google.com.mx/books?id=wz-D_8uPFCEC&pg=PA699&dq=definicion+de+redes+neuronales&hl=es-419&sa=X&ved=2ahUKEwifkI7Q8rr8AhUCIOQIHYm9AfwQuwV6BAgIEAg#v=onepage&q=definicion%20de%20redes%20neuronales&f=false
4. LLevot, J.R.: Medicina predictiva, aprendizaje automático y anestesia (2020), https://pesquisa.bvsalud.org/portal/resource/pt/ibc-200721
5. Martínez, G.: Minería de datos. Cómo hallar una aguja en un pajar. Ingenierías **14**(53), 53–66 (2001), https://www.cs.buap.mx/~bbeltran/NotasMD.pdf
6. Moine, J.M., Haedo, A.S., Gordillo, S.E.: Estudio comparativo de metodologías para minería de datos. In: XIII Workshop de Investigadores en Ciencias de la Computación (2011), https://sedici.unlp.edu.ar/handle/10915/20034
7. Rodrıguez-Maya, N.E., Lara-Álvarez, C., May-Tzuc, O., Suárez-Carranza, B.A.: Modeling students' dropout in mexican universities. Research in Computing Science **139**, 163–175 (2017)
8. Rojas, E.M.: Machine learning: análisis de lenguajes de programación y herramientas para desarrollo. Iberian Journal of Information Systems and Technologies (2020), https://www.proquest.com/docview/2388304894?pq-origsite=gscholar&fromopenview=true
9. Tan, P.N., Steinbach, M., Kumar, V.: Data mining introduction. People's Posts and Telecommunications Publishing House, Beijing (2006), https://books.google.com.mx/books/about/Introduction_to_Data_Mining.html?id=KZQ0jgEACAAJ&redir_esc=y
10. Zumarraga Espinosa, M.R., Castro Quimbiulco, M.I., Romero Cruz, J.C., Escobar Paredes, P.B., Boada Suraty, M.J., Armas Caicedo, R.H., Luzuriaga Mera, J.C., Gonzalez Rivadeneira, Y.E.: Medición de intereses profesionales en estudiantes universitarios y un abordaje exploratorio de su relación con el desempeño académico. In: 7ma Conferencia Latinoamericana sobre el Abandono de la Educación Superior (VII-CLABES 2017) (2017)

Esquema de Aprendizaje Federado para la Preservación de la Privacidad en el Reconocimiento de Actividades Humanas

Oscar Alexis Miranda Rucabado[1], Jessica Beltrán Márquez[1], and Ciro A. Martínez García-Moreno[2]

[1] Centro de Investigación en Matemáticas Aplicadas (CIMA), Universidad Autónoma de Coahuila, Saltillo, México
[oscar.miranda,jessicabeltran]@uadec.edu.mx
[2] Instituto Politécnico Nacional
ciro@citedi.mx

Abstract. Este trabajo aborda la viabilidad del Aprendizaje Federado aplicado al Reconocimiento de Actividades Humanas, considerando la preservación de la privacidad y la ejecución local en dispositivos. Se plantea un plan de evaluación con los conjuntos de datos UCI-HAR, PAMAP2 y KU-HAR, empleando un esquema de preprocesamiento basado en ventanas temporales, particionado por sujeto y un modelo bajo un esquema federado. El análisis contempla la comparación con un enfoque centralizado, discutiendo el compromiso entre privacidad y desempeño, así como los desafíos que plantea la heterogeneidad de los clientes. El objetivo es establecer condiciones prácticas que favorezcan la adopción del Aprendizaje Federado en escenarios cotidianos.

Keywords: Aprendizaje Federado · Reconocimiento de Actividades Humanas · Privacidad · Datos heterogéneos

1 Introducción

El Reconocimiento de Actividades Humanas (RAH) con sensores de *smartphones* y *wearables* constituye un área ampliamente documentada, con aplicaciones en salud y entornos inteligentes [1, 9, 10]. Entre los conjuntos de datos más utilizados destacan UCI-HAR [1], PAMAP2 [9] y, KU-HAR [13]. En enfoques de aprendizaje automático tradicional, los datos crudos suelen transferirse a servidores para su entrenamiento, lo que plantea retos de privacidad, seguridad y costos de procesamiento. A este tipo de entrenamiento se le conoce como centralizado, ya que todo se procesa en una unidad central.

El Aprendizaje Federado (AF) permite entrenar modelos sin compartir los datos crudos a una unidad central, compartiendo únicamente actualizaciones locales del modelo [6, 4, 16]. En entornos de *Internet of Things* (IoT) y comunicación inalámbrica, la variación entre clientes y las limitaciones de recursos plantean desafíos adicionales, especialmente cuando los datos no siguen la misma distribución entre usuarios.

Este trabajo busca evaluar la viabilidad práctica de un despliegue de AF para RAH, con el objetivo de reducir la transferencia de datos sensibles y mantener un desempeño competitivo frente a esquemas centralizados. Cabe mencionar que el AF enfrenta limitaciones frente al enfoque centralizado en RAH, ya que en este escenario, existe alta heterogeneidad de como los usuarios realizan actividades, además de una comunicación distribuida debido al uso de dispositivos IoT, que complican la convergencia y reducen el control de calidad de los datos [5, 7, 6].

2 Pregunta de investigación e hipótesis

La pregunta de investigación de este trabajo es ¿Cómo puede implementarse el aprendizaje federado en el reconocimiento de actividades humanas para integrar datos heterogéneos de múltiples fuentes y distribuciones, logrando un equilibrio entre privacidad, heterogeneidad y rendimiento frente a los métodos centralizados?

A continuación se muestran el objetivo general y los objetivos específicos de este trabajo de investigación.

Objetivo general. Diseñar y evaluar un esquema de AF aplicado al RAH, capaz de preservar la privacidad de los datos locales y mantener una exactitud con una diferencia no mayor al 15 % respecto a un modelo centralizado, en un contexto con datos heterogéneos.

Objetivos específicos.

- Analizar técnicas actuales de AF aplicadas a RAH y sus retos de privacidad.
- Prototipar un esquema de AF para clasificar actividades humanas
- Evaluar estrategias de mitigación de heterogeneidad y desbalanceo midiendo su impacto en la convergencia global y desempeño local en AF.
- Evaluar el desempeño con métricas estándar y compararlo contra un *baseline* centralizado.

3 Trabajo relacionado

Los desarrollos iniciales en AF se enfocaron en aplicaciones como la predicción de texto en teclados móviles, según se menciona en [4]. En el ámbito de la salud, el uso del AF sigue creciendo significativamente, especialmente en la interpretación de imágenes médicas y dermatología en donde el AF muestra un desempeño cercano a la forma tradicional centralizada, con una diferencia menor del 10 %, mostrando el potencial del uso del AF. [15, 3]. Para RAH y dispositivos con limitaciones de recursos, se están discutiendo alternativas como el entrenamiento en el borde (*edge omputing*) y los sistemas con clientes heterogéneos, tal como se presenta en los estudios [14, 5, 7]. En particular, en RAH, se han utilizado ampliamente datasets como el como UCI-HAR [1] y PAMAP2 [9], más recientemente, KU-HAR [13] ha aportado actividades y condiciones más diversas para evaluar

generalización de los modelos. En presencia de heterogeneidad en distribuciones de datos, en la literatura reportan técnicas que buscan mejorar la alineación entre clientes, incluyendo enfoques basados en transferencia de conocimiento y variantes relacionadas [2, 12]. Asimismo, se han estudiado riesgos y defensas ante ataques de envenenamiento [8, 11]. Revisiones generales de AF contextualizan estos retos y oportunidades [16].

4 Contexto de aplicación

Mantener los datos en el dispositivo es prioritario en aplicaciones de la salud, monitoreo en el hogar y prevención de lesiones; el AF habilita colaboración entre múltiples usuarios sin consolidar datos centralizados, reduciendo riesgos asociados a la centralización masiva de datos [16].

5 Metodología

En esta investigación se seguirán los siguientes pasos para conseguir los objetivos y encontrar respuesta a la pregunta de investigación.

- **Colección de datos:**. Buscar conjuntos de datos públicos como UCI-HAR, PAMAP2 y KU-HAR. Evaluar la posibilidad de capturar un conjunto de datos nuevo.
- **Preprocesamiento de datos:**. Aplicar técnicas de preprocesamiento estándar para tratar con datos faltantes, escalamiento requerido y transformación de datos para su introducción a modelos de aprendizaje automático.
- **Partición de datos:**. Dividir el conjunto de datos por sujeto, considerando cada persona como cliente federado; separación por sujetos entre entrenamiento, validación y prueba.
- **Modelo y federación:** Usar modelos de aprendizaje profundo, como redes convolucionales o recurrentes sobre datos multimodales. Aplicar AF usando incialmente el método de combinación de parámetros *Federated Average*. Explorar otras técnicas de combinación de parámetros para adaptarse a las variaciones de los conjuntos de datos en RAH.
- **Evaluación:** Usar métricas para problemas de clasificación, como exactitud y F1-score, reportadas a nivel de cliente y global. Asimismo, comparación contra esquemas centralizados.
- **Comunicación de resultados:**. Participar en eventos de divulgación, en congresos científicos y compartir el conjunto de datos en dado caso que se genere.

References

1. Anguita, D., Ghio, A., Oneto, L., Parra, X., Reyes-Ortiz, J.L.: A public domain dataset for human activity recognition using smartphones. In: European Symposium on Artificial Neural Networks (ESANN) (2013)
2. Gad, G., Farrag, A., Fadlullah, Z.M., Fouda, M.M.: Communication-efficient federated learning in drone-assisted iot networks: Path planning and enhanced knowledge distillation techniques. In: IEEE PIMRC. pp. 1–7 (2023)
3. Haggenmüller, S., Schmitt, M., Krieghoff-Henning, E., Hekler, A., Maron, R.C., Wies, C., Utikal, J.S., Meier, F., Hobelsberger, S., Gellrich, F.F., et al.: Federated learning for decentralized artificial intelligence in melanoma diagnostics. JAMA Dermatology **160**(3), 303–311 (2024)
4. Hard, A., Rao, K., Mathews, R., Ramaswamy, S., Beaufays, F., Augenstein, S., Eichner, H.: Federated learning for mobile keyboard prediction. arXiv preprint arXiv:1811.03604 (2018)
5. Li, C., Niu, D., Jiang, B., Zuo, X., Yang, J.: Meta-har: Federated representation learning for human activity recognition. In: Proceedings of the Web Conference. pp. 912–922 (2021)
6. McMahan, B., Moore, E., Ramage, D., Hampson, S., Arcas, B.A.Y.: Communication-efficient learning of deep networks from decentralized data. In: AISTATS. pp. 1273–1282 (2017)
7. Ouyang, X., Xie, Z., Zhou, J., Huang, J., Xing, G.: Clusterfl: a similarity-aware federated learning system for human activity recognition. In: MobiSys. pp. 54–66 (2021)
8. Raza, A., Li, S., Tran, K.P., Koehl, L.: Using anomaly detection to detect poisoning attacks in federated learning applications. arXiv preprint arXiv:2207.08486 (2022)
9. Reiss, A., Stricker, D.: Introducing a new benchmarked dataset for activity monitoring. Proceedings of the 16th International Symposium on Wearable Computers pp. 108–109 (2012)
10. Ronao, C.A., Cho, S.B.: Human activity recognition with smartphone sensors using deep learning neural networks. Expert Systems with Applications **59**, 235–244 (2016)
11. Shahid, A.R., Imteaj, A., Badsha, S., Hossain, M.Z.: Assessing wearable human activity recognition systems against data poisoning attacks in differentially-private federated learning. In: IEEE SMARTCOMP. pp. 355–360 (2023)
12. Shen, Q., Feng, H., Song, R., Song, D., Xu, H.: Federated meta-learning with attention for diversity-aware human activity recognition. Sensors **23**(3), 1083 (2023)
13. Sikder, I.U., Nahid, A.A.: Ku-har: A dataset for human activity recognition using smartphones. Data in Brief **36**, 107107 (2021)
14. Trotta, A., Montori, F., Ciabattini, L., Billi, G., Bononi, L., Felice, M.D.: Edge human activity recognition using federated learning on constrained devices. Pervasive and Mobile Computing **104**, 101972 (2024)
15. Truhn, D., Arasteh, S.T., Saldanha, O.L., Müller-Franzes, G., Khader, F., Quirke, P., West, N.P., Gray, R., Hutchins, G.G.A., James, J.A., et al.: Encrypted federated learning for secure decentralized collaboration in cancer image analysis. Medical Image Analysis **92**, 103059 (2024)
16. Zhang, C., Xie, Y., Bai, H., Yu, B., Li, W., Gao, Y.: A survey on federated learning. Knowledge-Based Systems **216**, 106775 (2021)

Hand_IA, Sistema Inteligente para el Control de Prótesis de Mano basado en Señales de Electromiografía (EMG)

Daniela Padilla-Cuautle, Luis Villaseñor-Pineda, Alejandro Gutierrez-Giles, and Alejandro A. Torres-García

Instituto Nacional de Astrofísica, Óptica y Electrónica,
Luis Enrique Erro #1, Tonanzintla, Puebla, México
{daniela.pcuautle, villasen, alejandro.giles, alejandro.torres}@inaoe.mx

Abstract. La pérdida de una extremidad superior afecta gravemente la calidad de vida de las personas, ya que, además del daño físico, conlleva repercusiones psicológicas y socioeconómicas significativas. Este proyecto plantea la hipótesis de que, mediante la obtención y el procesamiento de señales electromiográficas de superficie (sEMG) de personas con y sin amputación transradial, es posible desarrollar un sistema de control basado en aprendizaje automático capaz de decodificar en tiempo real la actividad muscular del antebrazo para controlar una prótesis de mano. En una etapa preliminar, se implementaron dos modelos de redes neuronales: una red neuronal convolucional (CNN) y una red neuronal recurrente con memoria a corto y largo plazo (LSTM). Además, se aplicó el método de Análisis de Componentes Principales (PCA) para identificar los canales más representativos y optimizar el procesamiento de señales. Los resultados iniciales mostraron que el modelo CNN obtuvo un mejor desempeño, alcanzando una exactitud del 64.59% con cinco canales y del 79.15% al utilizar todos los canales en pruebas inter-sesión. Asimismo, la aplicación de PCA permitió reducir significativamente el tiempo de procesamiento sin comprometer el rendimiento del modelo.

Keywords: sEMG · Procesamiento de señales · Redes neuronales · Inter-sesión · PCA.

1 Introducción

La electromiografía (EMG) es una técnica que permite medir la respuesta muscular o la actividad eléctrica generada por un músculo en respuesta a la estimulación nerviosa. En los campos de la biomedicina, la biomecatrónica y las prótesis, la EMG se ha empleado ampliamente para el control de dispositivos de rehabilitación y el desarrollo de interfaces humano-computadora.

Su aplicación se basa principalmente en la clasificación de patrones de las señales electromiográficas, a partir de la extracción de un conjunto de características que representan la actividad registrada, lo que permite identificar y clasificar diferentes movimientos musculares.

El presente documento describe el trabajo de tesis orientado a la implementación de un sistema de control con aplicación en la rehabilitación clínica, específicamente en una prótesis de miembro superior, mediante el uso de señales electromiográficas y modelos de inteligencia artificial para la clasificación precisa de dichas señales.

1.1 Planteamiento del problema

El desarrollo de prótesis de brazo avanzadas es complejo. Aunque el uso de prótesis inteligentes controladas mediante señales electromiográficas (EMG) y modelos de inteligencia artificial (IA) representa actualmente una solución prometedora, aún enfrenta un desafío principal: la variabilidad de las señales del usuario a lo largo del tiempo. Esta variabilidad, es conocida como inter-sesión, que ocurre cuando las características de la señal cambian entre diferentes días o momentos de uso (debido a factores como la fatiga, re-colocación del sensor, etc), degradando el rendimiento de los modelos de IA entrenados previamente. Para solucionar este problema se propone utilizar técnicas de IA, como la transferencia de aprendizaje (Transfer learning), que permita adaptar los modelos de una sesión inicial a las nuevas condiciones de una sesión posterior.

1.2 Justificación

De acuerdo a la Academia Nacional de Medicina de México (ANMM), en 2024 se reportaron en promedio 75 amputaciones diarias, lo que equivale aproximadamente a 27,300 nuevos casos al año. Sin embargo, no reportan exactamente cuántas de estas amputaciones son de extremidades superiores. En este contexto, el desarrollo de prótesis mioeléctricas controladas por señales electromiográficas (EMG) constituye una alternativa prometedora. Sin embargo, su adopción, especialmente en países como México, aún enfrenta barreras tanto tecnológicas como de usabilidad. Siendo una de las barreras de usabilidad más críticas la de la variabilidad de las señales, ya que cambian significativamente de un día a otro (o incluso tras unas horas) debido a factores como el reposicionamiento de los electrodos, la fatiga muscular y cambios en la piel. Esta variabilidad provoca que un modelo de aprendizaje automático, entrenado con datos de una sesión, pierda su precisión y robustez al ser utilizado en una sesión posterior. Este proyecto propone el diseño de una estrategia de control para prótesis de mano mioeléctricas, empleando una pulsera de bajo costo (por ejemplo, *MYO Armband)* para la captación de señales EMG. Se aplicarán técnicas de procesamiento de señales y algoritmos de aprendizaje automático para clasificar eficazmente las características de la señal y traducirlas en comandos precisos para el movimiento de la prótesis. Para abordar el desafío de la variabilidad entre sesiones y la robustez a largo plazo, este proyecto empleará técnicas de aprendizaje por transferencia. Con el objetivo de que un modelo, ya entrenado con una sesión previa, pueda adaptarse a las características de una nueva sesión. Esto permitirá que el sistema se ajuste rápidamente a los cambios en la señal del usuario sin necesidad de un reentrenamiento completo.

1.3 Objetivos

Desarrollar un sistema de control para una mano protésica basado en técnicas de aprendizaje computacional para la identificación y clasificación en tiempo real de gestos y movimientos, a partir de señales EMG adquiridas de personas con amputación transradial y sujetos sanos.

- Diseñar e implementar una estrategia de identificación de biomarcadores en las señales de EMG que permita caracterizar movimientos funcionales esenciales en la vida diaria.
- Diseñar e implementar un modelo de aprendizaje automático —incluyendo estrategias de transferencia de aprendizaje— robusto frente a variabilidad inter-sesión, capaz de reconocer y clasificar con precisión gestos y movimientos de la mano en tiempo real.
- Evaluar un esquema de transferencia de aprendizaje para acelerar el entrenamiento de usuarios amputados utilizando modelos preentrenados con datos de sujetos sanos.
- Construir y probar el funcionamiento de la mano protésica.
- Integrar al sistema de control el modelo de clasificación y reconocimiento.
- Validar el sistema mediante la realización de pruebas funcionales.

2 Contribuciones esperadas

En el estado del arte existen varias investigaciones relacionadas con el desarrollo y la implementación de modelos de IA para el control de prótesis de extremidades superiores. La mayoría utiliza redes neuronales convolucionales, además, de explorar diferentes esquemas para la transferencia de aprendizaje. Por ejemplo, en [1] los autores proponen una estrategia de transferencia de aprendizaje combinada con una CNN de larga exposición (LECNN, por sus siglas en inglés) con el objetivo de reconocer la intención de movimiento, donde mencionan que las señales EMG se generan entre 30-150ms antes de que se realice el movimiento. Fratti R., et al., menciona la adaptación del modelo a usuarios que nunca antes han visto; la base de datos que emplean contiene 12 sujetos sanos y 3 sujetos amputados.
Por otro lado, al trabajar con sistemas en tiempo real, y bajo el supuesto que no todos los canales son de utilidad, se podría aplicar alguna estrategia para seleccionar sólo las señales útiles. En el presente trabajo se analiza la variabilidad de las señales inter-sesión, y el método de análisis de componentes principales para seleccionar los canales más representativos de las señales.
Otra contribución, no menos importante, es la base de datos que está en proceso de recopilación, tanto de personas sanas como amputadas, la cual se pondrá a disposición de la comunidad académica.

Finalmente, este proyecto integra conocimientos de biomedicina, robótica y rehabilitación con el objetivo de desarrollar un sistema de control inteligente para prótesis de mano basado en señales electromiográficas (EMG). Así la contribución final radica en ofrecer a las personas con amputación de miembro

superior una alternativa funcional que mejore su autonomía en las actividades diarias. Además, el sistema propuesto podría complementar los procesos de rehabilitación clínica, proporcionando herramientas tecnológicas para el monitoreo y evaluación del progreso motor tanto de pacientes como de profesionales de la salud.

3 Evaluación del los resultados

Para evaluar los resultados obtenidos en este proyecto, se pretende medir diversas métricas, las cuales están divididas en el desempeño del modelo de clasificación de señales y el desempeño del sistema de control de la mano. Para evaluar el modelo se va a medir el porcentaje de veces que el algoritmo interpreta correctamente la intención de movimiento del usuario. Para evaluar el desempeño de la mano protésica se va a medir el tiempo de latencia, que es el tiempo que transcurre desde que el usuario contrae el músculo hasta que la mano ejecuta el movimiento. En esta misma línea también se va a evaluar la tasa de éxito en la realización de tareas, que va a ser el porcentaje de veces que el sistema realizó el gesto o movimiento correcto.

Para presentar los resultados a la comunidad se va a publicar un artículo científico en una revista de ingeniería biomédica o de robótica. También, se creará contenido que se publicará en páginas del instituto.

4 Resultados alcanzados

En esta sección se muestran de forma resumida los resultados alcanzados hasta ahora en este trabajo.

4.1 Base de datos

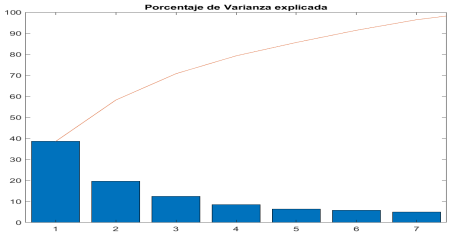
Se utilizó la base de datos titulada "EMG Data for Gestures" con 36 sujetos sanos, que contiene señales electromiográficas sin procesar capturadas por la pulsera MYO Armband de Thalmic de 8 sensores, y registradas a una frecuencia de muestreo de 200 Hz. El protocolo de adquisición de las señales fue el siguiente: cada sujeto realizó dos series, cada una compuesta por hasta siete gestos básicos realizados durante tres segundos con una pausa de tres segundos entre cada gesto. Finalmente, el número de registros fue de entre 40 000 y 50 000 en cada columna. Los gestos registrados son los siguientes: mano en reposo, mano apretada en puño, flexión de muñeca, extensión de muñeca, desviaciones radiales, desviaciones cubitales y palma extendida (No todos los sujetos lo realizaron).

La clase 0, corresponde al descanso o como lo mencionan los autores, datos sin marcar. Para este trabajo se seleccionaron los datos de solo 15 sujetos. Excluyendo las clases 0 y 7, ya que la primera no corresponde a ningún gesto y la segunda no la realizaron todos los sujetos, quedando con un total de 6 clases.

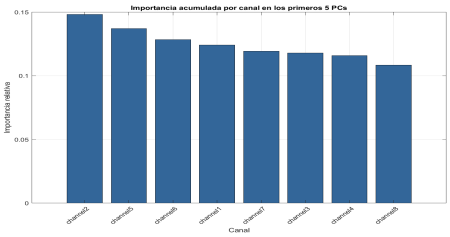
4.2 Preprocesamiento y Aplicación de PCA

El preprocesamiento involucra el filtrado, suavizado, rectificación y normalización de las señales. Para el filtrado se aplicó un filtro Butterworth pasa-bandas de cuarto orden de 10 a 95 Hz, esto debido a que la pulsera registra las señales a una frecuencia de muestreo de 200 Hz, también se aplicó un filtro Notch a 50 Hz para eliminar la interferencia de la línea eléctrica. Se aplicó un suavizado por media móvil, aplicando una ventana deslizante de valor n = 25, la cual promedia los valores de la señal dentro de esa ventana mejorando la calidad de la información. También se ajustaron los valores de la señal en un rango de [-1, 1] y finalmente, se aplicó una rectificación para obtener el valor absoluto de la señal.

Cuando se tienen muchos canales, el análisis de las señales puede volverse complejo, redundante y computacionalmente costoso. En este trabajo se empleó PCA para analizar la importancia relativa de cada canal en la construcción de los primeros componentes principales, este proceso se aplicó en un grupo de 21 de los 36 sujetos, los cuales no fueron utilizados para entrenar el modelo. En la figura 1 se muestra una gráfica del análisis de PCA, del lado izquierdo se representa el porcentaje de varianza explicada por cada componente principal, siendo los primeros 4 componentes los que explican la mayor parte de la varianza, al menos un 80%. Del lado derecho se muestra una gráfica de barras, que representa la relevancia de cada canal, determinando así los cinco canales a usar (canal: 2, 5, 6, 1 y 7).



(a) Porcentaje de la varianza explicada por cada componente principal.



(b) Importancia acumulada por canal.

Fig. 1: Análisis con PCA

4.3 Estructura del modelo

En ambos modelos (CNN y LSTM), previo al entrenamiento se aplicó un ventaneo, método que se refiere al procesamiento de datos en segmentos o ventanas, lo cual permite analizar y generar información a partir de datos secuenciales. Al examinar los datos en fragmentos más pequeños los modelos pueden captar patrones relevantes dentro de los datos con mayor facilidad. Las ventanas se

crearon con un tamaño de 150 milisegundos lo que resulta en un valor de 30 muestras, con un solapamiento entre ellas de 35 milisegundos (7 muestras), lo cual asegura que no se pierda información entre cada ventana.
Se realizó una prueba inter-sesión la cual consta de entrenar y probar el modelo con sesiones diferentes de los mismos sujetos, entonces los datos se dividieron en dos carpetas TRAIN y TEST respectivamente para evitar que se mezclara información. Para visualizar mejor los resultados por cada sujeto se entrenó con todos los sujetos y se probó con la segunda sesión de cada uno de ellos.

Red Neuronal Convolucional (CNN). La red CNN se implementó utilizando la biblioteca Keras de TensorFlow (TF), es una CNN unidimensional que utiliza 64 filtros con un tamaño de kernel de 12, incluye una capa de MaxPooling que reduce la dimensionalidad de los datos a la mitad.La función de activación definida es Softmax.

Red Neuronal Recurrente de memoria a corto, largo plazo (LSTM). Para la red LSTM, también se implementó la biblioteca Keras de TF. Este modelo se implementó ya que este tipo de red se especializa en el procesamiento y clasificación de datos secuenciales o series temporales, ya que pueden aprender dependencias a largo plazo y retener información relevante a lo largo de la secuencia.

Modelo	Sujetos	Cinco canales	Todos los canales
CNN	15	Exactitud: 64.59 ±8.9	Exactitud: 79.15 ±7.8
LSTM	15	Exactitud: 62.96 ±8.3	Exactitud: 77.34 ±8.6

Table 1: Resultados obtenidos

En este punto reducir el número de canales significó una reducción de al menos 10 minutos del tiempo de procesamiento.

5 Conclusión y Discusión

Este proyecto se basa en la adquisición y el procesamiento de señales musculares, extraídas de personas tanto sanas como de personas con amputación transradial. Para un primer acercamiento se aplicó la técnica de reducción de dimensionalidad con PCA para identificar y seleccionar el conjunto de canales que aportan mayor relevancia a la varianza de los datos. Esto mostró una reducción en el tiempo de procesamiento de alrededor de 10 minutos. En cuanto a los modelos utilizados, ambos obtuvieron resultados bastante similares, siendo la red CNN ligeramente superior a la LSTM.
Sin embargo, este enfoque tuvo como consecuencia una disminución significativa en cuanto a la exactitud de ambos modelos. Lo que nos permite concluir que,

aunque la reducción de canales permite agilizar el procesamiento de varios datos, es necesario considerar el impacto en el desempeño del sistema de clasificación. Como trabajo futuro se plantea explorar estrategias complementarias que permitan mitigar la pérdida de precisión asociada a la reducción de canales, así como el uso de modelos clásicos como lo son las Máquinas de Soporte Vectorial (SVM) y RandomForest, acompañadas con extracción de características. Finalmente, también se pretende probar con diferentes conjuntos de canales para observar si existen diferencias significativas en los resultados.

References

1. Lin, C., Niu, X., Zhang, J., et al. Improving Motion Intention Recognition for Trans-Radial Amputees Based on sEMG and Transfer Learning.(2023). https://doi.org/https://doi.org/10.3390/ app131911071
2. Fratti, R.; Marini, N.; Atzori, M.; Müller, H.; Tiengo, C.; Bassetto, F. A Multi-Scale CNN for Transfer Learning in sEMG-Based Hand Gesture Recognition for Prosthetic Devices. Sensors 2024, 24, 7147. https://doi.org/https://doi.org/10.3390/s24227147
3. Wang S, Zheng J, Huang Z, Zhang X, Prado da Fonseca V, Zheng B and Jiang X (2022), Integrating computer vision to prosthetic hand control with sEMG: Preliminary results in grasp classification. Front. Robot. AI 9:948238. https://doi.org/10.3389/frobt.2022.948238
4. Li W, Shi P and Yu H. Gesture Recognition Using Surface Electromyography and Deep Learning for Prostheses Hand: State-of-the-Art, Challenges, and Future. Front. Neurosci. (2021) 15:621885. https://doi.org/10.3389/fnins.2021.621885
5. Zhang X, Zhang M. (2023) Study on the methods of feature extraction based on electromyographic signal classification. https://doi.org/https://doi.org/10.1007/s11517-023-02812-3
6. Zhang B. The Application of Machine Learning in Gesture Prediction on EMG Dataset. (2020), 131-135. https://doi.org/10.1109/CDS49703.2020.00032.
7. Krilova, N., Kastalskiy, I., Kazantsev, V., et al. EMG Data for Gestures [Dataset]. (2018). UCI Machine Learning Repository. https://doi.org/https://doi.org/10.24432/C5ZP5C
   Delgado S. Unidad de Investigación en Órtesis y Prótesis, única en México y AL. (2022). [En línea] https://goo.su/zP3alDR

Diagnóstico de Cardiopatía Isquémica Crónica en DMT2 mediante TabTransformer: Contraste de Variables Predictivas con Factores de Riesgo Tradicionales.

Orlando Flores-Custodio[1 [0009−0004−4492−6478]] y Pablo Pancardo-García[1 [0000−0002−5482−6372]]

División Académica de Ciencias y Tecnologías de la Información, Universidad Juárez Autónoma de Tabasco, México,
florescustodioo@gmail.com, pablo.pancardo@ujat.mx

Resumen La cardiopatía isquémica crónica (CIC) es una complicación frecuente en pacientes con diabetes mellitus tipo 2 (DMT2), y su diagnóstico temprano es crucial para mejorar el tratamiento y el pronóstico del paciente. Sin embargo, el diagnóstico se complica porque los síntomas se manifiestan de forma inusual y los factores de riesgo son diversos. En esta investigación, se desarrolla, implementa, optimiza y evalúa un modelo de red TabTransformer capaz de diagnosticar CIC a partir del historial clínico de pacientes con DMT2. Se utilizó un conjunto de datos con información longitudinal de pacientes mexicanos, alcanzando una precisión de 87.72 %, un recall de 90.16 % y un valor de 0.9434 para el área bajo la curva ROC. El modelo TabTransformer superó a los enfoques tradicionales de *machine learning*(ML), demostrando su capacidad para captar relaciones complejas entre variables. Los resultados confirman la eficacia del enfoque propuesto para diagnosticar la CIC en pacientes diabéticos.

Palabras clave: Diagnóstico médico, cardiopatía isquémica crónica, TabTransformer, aprendizaje profundo, diabetes mellitus tipo 2

1 Introducción

La DMT2 es un desafío global, siendo la CIC una de sus complicaciones más comunes y letales [5]. El diagnóstico oportuno de la CIC en pacientes con DMT2 es complejo debido a la presentación atípica de síntomas y la influencia de múltiples factores de riesgo [9]. Este trabajo es una continuación natural de un trabajo previo de selección de características [7], centrándose en la implementación y evaluación de un modelo de *deep learning* para datos tabulares, el TabTransformer.

1.1 Hipótesis y objetivos

Hipótesis: Una red neuronal artificial Transformer puede diagnosticar la cardiopatía isquémica crónica en pacientes con DMT2 con al menos un 80 % de precisión.

Objetivo general: Implementar y evaluar un modelo de red neuronal artificial Transformer para diagnosticar la cardiopatía isquémica crónica en pacientes con DMT2.

Objetivos específicos:

- Desarrollar un modelo TabTransformer que diagnostique la CIC en pacientes con DMT2.
- Optimizar los hiperparámetros del modelo mediante algoritmo genetico (AG).
- Evaluar el rendimiento del modelo mediante métricas clínicamente relevantes.
- Comparar el rendimiento con modelos tradicionales de ML.
- Contrastar las variables identificadas computacionalmente con los factores de riesgo. reconocidos en la literatura médica

1.2 Contribuciones

- **Modelo TabTransformer en contexto mexicano:** Primera implementación para diagnóstico de CIC en población mexicana con DMT2, utilizando datos reales.
- **Rendimiento superior comprobado:** 87.72 % de precisión y 90.16 % de *recall*, superando ampliamente la hipótesis planteada.
- **Análisis comparativo de variables computacionales vs. clínicas:** Identificación de diferencias entre variables detectadas por el modelo y factores de riesgo tradicionales.
- **Interpretabilidad mediante SHAP:** Análisis de importancia de variables que facilita la comprensión clínica del modelo.

2 Antecedentes y Trabajos Relacionados

El diagnóstico de CIC ha utilizado ML tradicional (regresión, bosques aleatorios [15], Máquinas de Vectores de Soporte(SVM) [2]). Las redes neuronales [16] han demostrado ventaja con datos heterogéneos. La arquitectura Transformer [18], ha sido adaptada al ámbito médico para diversas tareas, como el análisis de imágenes [13], notas clínicas [10] e historiales clínicos electrónicos (EHR) [12], e incluso para la predicción de enfermedades cardíacas usando datos tabulares [17]. El TabTransformer [11], el modelo usado en este trabajo, es una adaptación específica para datos tabulares, utilizando mecanismos de atención para capturar interacciones complejas entre características, lo que lo hace idóneo para datos clínicos donde estas relaciones son fundamentales.

3 Metodología

Se utilizó el dataset hk_database del repositorio DiabetIA [4], con información longitudinal (2005-2020) de 137,131 pacientes con DMT2 del IMSS en Michoacán, México. Mediante técnicas de selección de características (ranking de

correlación, chi-cuadrada, CFS, Pearson) y stepwise regression con AIC [7], identificamos 18 variables predictoras [1].

El preprocesamiento incluyó: eliminación de variables con >25 % valores faltantes [14], imputación multivariada iterativa por regresión lineal, y balanceo por submuestreo (16,456 observaciones: 8,228 por clase). El TabTransformer fue implementado en PyTorch y optimizado mediante AG[8], cuyo cromosoma codificó hiperparámetros clave (dimensión del *embedding*, capas/cabezales del Transformer, tasa de aprendizaje). El entrenamiento se configuró con validación cruzada (CV-5), *early stopping*, optimizador Adam y regularización (*dropout, weight decay*). La arquitectura final incluye: *embedding* con dimensión de 60 para categóricas, normalización para continuas, 3 capas Transformer con 15 cabezas de atención, y MLP de 3 capas para clasificación.

4 Experimentos y Resultados

Los experimentos se realizaron con una división 64 %/16 %/20 % (entrenamiento/ validación/ prueba) y se evaluaron con *accuracy*, *recall*, F1-score y AUC. El modelo final, entrenado con los hiperparámetros optimizados (*learning rate* 0.0038, *batch* 36, 200 épocas máx., paciencia de 50 para *early stopping*), superó notablemente a los modelos base (Cuadro 1). La Figura 1 muestra la curva ROC y evolución del entrenamiento, con mejora sostenida hasta estabilización, sin sobreajuste.

Cuadro 1. Resultados de evaluación del TabTransformer y comparación con modelos base.

Modelo	**Accuracy**	**Recall**	**F1-score**	**AUC**
TabTransformer	87.72 %	90.16 %	88.02 %	0.9434
Regresión Logística	83.74 %	83.54 %	83.71 %	–
KNN	80.67 %	78.19 %	80.19 %	–
Árboles de decisión	83.26 %	83.78 %	83.35 %	–

El análisis SHAP (Figura 2) identificó como variables más influyentes: in_heart_rate_count, complications_and_ill_defined_descriptions_of_heart_disease y antidiabetics_count, validando la selección previa y subrayando su importancia diagnóstica.

[1] complications_and_il_defined_descriptions_of_heart_disease, in_heart_rate_count, dx_age_e11, in_respiratory_frequency_count, age_at_wx, hypertensive_heart_disease, antidiabetics_count, antiplatelets_mean/sum/count, heart_failure, circulatory_system_diseases, in_glucose_count, lipid_lowering_count/mean, in_heart_rate_max/std, chronic_ischemic_heart_disease

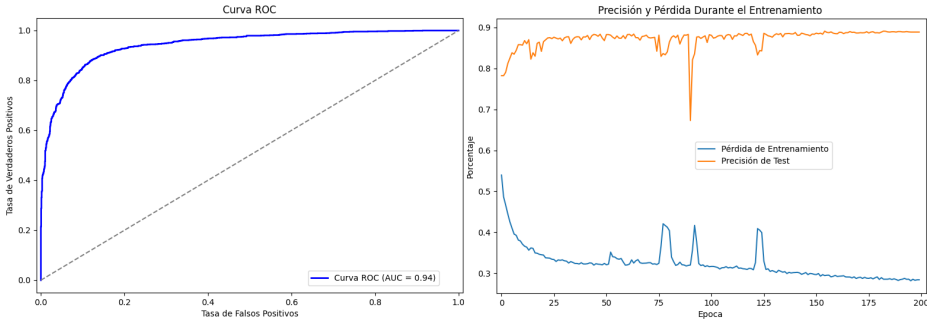


Figura 1. Curva ROC y evolución de la precisión y pérdida durante el entrenamiento del modelo.

5 Discusión

El TabTransformer mostró un rendimiento superior (87.72 % de *accuracy*, 90.16 % de *recall*), atribuible a su capacidad para capturar relaciones complejas mediante atención multi-cabeza, integrar variables categóricas y numéricas, y apoyarse en una selección previa robusta. El alto *recall* minimiza falsos negativos, aspecto clínicamente crítico para la detección oportuna.

5.1 Contraste entre variables identificadas y factores de riesgo

La literatura médica reconoce factores de riesgo tradicionales para CIC [1]: obesidad, sedentarismo, dislipidemia, hipertensión, tabaquismo y edad avanzada. El Cuadro 2 compara estos con las variables identificadas por nuestro modelo.

Cuadro 2. Comparación entre factores de riesgo tradicionales (izquierda) y variables identificadas (derecha).

Factor/Variable	Lit.	Mod.	Tipo
Obesidad/sobrepeso	✓		Causal
Sedentarismo	✓		Causal
Dislipidemia	✓		Causal
Hipertensión art.	✓	✓	Causal
Edad avanzada	✓	✓	Demog.
Tabaquismo	✓		Causal

Factor/Variable	Lit.	Mod.	Tipo
Hipertensión art.	✓	✓	Causal
Edad avanzada	✓	✓	Demog.
Conteo antiplaq.		✓	Tratam.
Media antiplaq.		✓	Tratam.
Conteo antidiab.		✓	Tratam.
Desv. est. frec. card.		✓	Medición
Insuf. cardíaca		✓	Compl.

Mientras la literatura se centra en factores causales, el modelo detectó también patrones de tratamiento (conteo de antiplaquetarios y antidiabéticos), complicaciones (insuficiencia cardíaca) y mediciones clínicas específicas (frecuencia cardíaca). Esto refleja un perfil de riesgo emergente derivado de la práctica clínica y la respuesta terapéutica

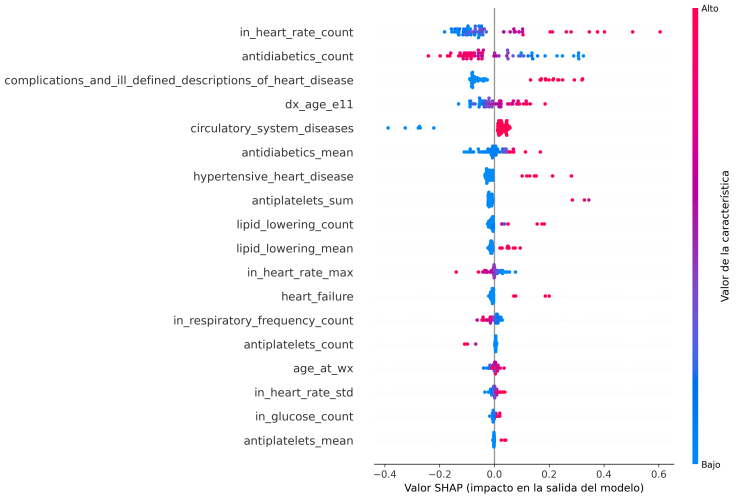


Figura 2. Importancia de las características según método SHAP.

5.2 Evaluación de resultados y divulgación

La hipótesis (80 % de precisión) fue ampliamente confirmada (87.72 %). Se cumplieron los cinco objetivos planteados: (1) desarrollo del TabTransformer con arquitectura optimizada, (2) evaluación con métricas múltiples, (3) superioridad frente a modelos base, y (4) contraste con factores tradicionales. La divulgación incluye: Articulo publicado en Springer Nature, Tesis de maestría

5.3 Implicaciones clínicas y limitaciones

El modelo podría apoyar la detección temprana de pacientes de alto riesgo y su integración en sistemas de apoyo clínico contribuiría a mejorar pronóstico y reducir morbimortalidad. El análisis SHAP facilita la interpretación y posible adopción por profesionales de la salud. Limitaciones: (1) resultados específicos a la base DiabetIA (población mexicana) y (2) exclusión de variables tradicionales con >25 % de valores faltantes.

6 Conclusiones y Trabajo Futuro

Este estudio demostró la efectividad del TabTransformer para el diagnóstico de CIC en pacientes con DMT2, superando la hipótesis inicial (87.72 % de *accuracy*, 90.16 % de *recall*) y los enfoques tradicionales. Las principales contribuciones son: (1) integración de una selección robusta de variables clínicas, (2) validación interpretativa con SHAP, y (3) análisis comparativo que evidencia que el modelo capta patrones de tratamiento y seguimiento complementarios a los factores causales clásicos.

Como trabajo futuro, se plantean tres líneas principales: (1) la validación externa del modelo en distintos conjuntos de datos para probar su generalización; (2) la validación cualitativa de las variables predictivas identificadas (p. ej., patrones de tratamiento) por expertos clínicos, para interpretar su causalidad y real valor pronóstico; y (3) la aplicación de pruebas estadísticas no paramétricas (p. ej., test de Friedman o Wilcoxon) para cuantificar la diferencia en el rendimiento entre el TabTransformer y otros modelos de ML.

Referencias

1. American Heart Association: Cardiovascular Disease and Diabetes. Circulation Research, 124, 1160–1162 (2019)
2. Antikainen, E., Linnosmaa, J., Umer, A., Oksala, N., Eskola, M., van Gils, M., Hernesniemi, J., Gabbouj, M.: Transformers for cardiac patient mortality risk prediction from heterogeneous electronic health records. Scientific Reports, 13(1), 3517 (2023)
3. Clavijo Rodríguez, J., García Torres, A., Pardo Rodríguez, J., Rodríguez Díaz, J.C.: Sistema inteligente de reconocimiento de enfermedades cardiovasculares basado en electrocardiogramas. Revista Cubana de Información en Ciencias de la Salud, 23(2), 147–157 (2012)
4. CONAHCYT: DiabetIA: Conjunto de Datos para el Estudio de la Diabetes Mellitus Tipo 2 en México. https://repositorio-salud.conacyt.mx/jspui/handle/1000/296
5. Dirección General de Epidemiología, Secretaría de Salud: Informe del Sistema de Vigilancia Epidemiológica Hospitalaria de Diabetes Mellitus Tipo 2, corte al 1 de abril de 2024. Secretaría de Salud, México (2024)
6. Esquivel-Prados, E., et al.: Influencia de la adherencia al tratamiento antidiabético en el control de la Diabetes Mellitus tipo 2 en Farmacia Comunitaria: Estudio Adh-DM2. Universidad de Granada, Tesis Doctoral (2025)
7. Flores Custodio, O., Pancardo García, P.: Aplicación de CRISP-DM para detectar variables relevantes de la cardiopatía isquémica crónica. Investigación Aplicada, un Enfoque en la Tecnología, No. 18, diciembre 2024, pp. 372. https://www.investigacionaplicadarevista.com/_files/ugd/5a3be4_75e15166787e4441b35e8a8d29ecfdf0.pdf#page=372
8. Han, Shifen, Xiao, Li: An improved adaptive genetic algorithm. SHS web of conferences, 140, 01044 (2022)
9. Hodelín Maynard, E.H., Maynard Bermúdez, R.E., Maynard Bermúdez, G.I., Hodelín Carballo, H.: Complicaciones crónicas de la diabetes mellitus tipo II en adultos mayores. Revista Información Científica, 97(3), 528–537 (2018)
10. Huang, K., Altosaar, J., Ranganath, R.: ClinicalBERT: Modeling Clinical Notes and Predicting Hospital Readmission. In: Proceedings of the 2020 Machine Learning for Health Workshop, NeurIPS, 158–169 (2020)
11. Huang, X., Khetan, A., Cvitkovic, M., and Karnin, Z.: TabTransformer: Tabular Data Modeling Using Contextual Embeddings. arXiv preprint arXiv:2012.06678 (2020)
12. Li, Y., Mamouei, M., Salimi-Khorshidi, G., Rao, S., Hassaine, A., Canoy, D., Lukasiewicz, T., and Rahimi, K.: Hi-BEHRT: Hierarchical Transformer-Based Model for Accurate Prediction of Clinical Events Using Multimodal Longitudinal Electronic Health Records. IEEE Journal of Biomedical and Health Informatics, 27(2), 1106–1117 (2022)

13. Matas González, J.A.: Clasificación de imágenes mediante Redes Neuronales Convolucionales y Transformers. Universidad de Valladolid, Trabajo de Fin de Grado (2021)
14. Medina, Fernando, Galván, Marco: Imputación de datos: teoría y práctica. Cepal (2007)
15. Shehzadi, S., Hassan, M.A., Rizwan, M., Kryvinska, N., Vincent, K.: Diagnosis of chronic ischemic heart disease using machine learning techniques. Computational Intelligence and Neuroscience, 2022(1), 3823350 (2022)
16. Silveri, G., Merlo, M., Restivo, L., Sinagra, G., Accardo, A.: Novel Classification of Ischemic Heart Disease Using Artificial Neural Network. In: 2020 Computing in Cardiology, pp. 1–4. IEEE (2020)
17. Sumon, M.S.I., Islam, M.S.B., Rahman, M.S., Hossain, M.S.A., Khandakar, A., Hasan, A., Murugappan, M., and Chowdhury, M.E.H.: CardioTabNet: A Novel Hybrid Transformer Model for Heart Disease Prediction Using Tabular Medical Data. Health Information Science and Systems, 13(1), 44 (2025)
18. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I.: Attention is all you need. Advances in Neural Information Processing Systems, 30 (2017)

Un Modelo Híbrido para la Predicción de Resistencia Antimicrobiana mediante Redes Neuronales Gráficas y PLN

Sergio Antonio de Jesús Saldaña Pacheco[1][0009−0005−4975−4596] and Pablo Pancardo García[2,3][0000−0002−5482−6372]

[1] Universidad Juárez Autónoma de Tabasco
División Académica de Ciencias y Tecnologías de la Información
Carretera Cunduacán-Jalpa KM. 1 Col. La Esmeralda CP. 86690
Cunduacán, Tabasco, México.
252h23009@alummo.ujat.mx,sergiosaldana@apliware.com
[2] pablo.pancardo@ujat.mx
http://www.ujat.mx/dacyti

Abstract. La resistencia antimicrobiana (AMR, por sus siglas en inglés) es una amenaza creciente para la salud pública global, y su predicción precisa resulta clave para el desarrollo de terapias efectivas. Actualmente, las redes neuronales gráficas (GNN, por sus siglas en inglés) y las técnicas de procesamiento de lenguaje natural (PLN) se destacan como enfoques prometedores para esta tarea. No obstante, su uso por separado presenta limitaciones: las GNN dependen de grafos biológicos que pueden estar incompletos o desactualizados, mientras que los modelos PLN, aunque eficaces para extraer conocimiento semántico, no capturan explícitamente relaciones topológicas entre entidades biológicas.

Este proyecto propone un modelo híbrido que combina representaciones gráficas y semánticas para predecir resistencia antimicrobiana en *Escherichia coli*. Se construye un grafo genómico a partir de bases de datos biológicas (CARD, BV-BRC y NCBI) y se extraen entidades relevantes desde literatura científica mediante BioBERT. Ambos tipos de datos se integran en una arquitectura basada en GNN, capaz de representar tanto la estructura biológica como el contexto semántico.

La evaluación se realizará mediante métricas como exactitud, F1-score y AUC-ROC, así como estudios de ablación para cada tipo de dato y los resultados serán validados por expertos biomédicos. Se prevé la divulgación de los resultados en conferencias y revistas especializadas, y la liberación de recursos en plataformas abiertas. Este enfoque busca avanzar el estado del arte mediante una integración computacional innovadora que mejora la predicción de AMR en cepas de E. Coli en $\approx 5\%$, en comparación con modelos que utilizan exclusivamente GNN o PLN.

Keywords: GNN · PLN · Resistencia antimicrobiana.

1 Planteamiento del problema

La resistencia antimicrobiana (AMR) representa una amenaza creciente para la salud pública [1], y su predicción precisa es clave para el desarrollo de terapias efectivas. Entre los enfoques recientes más prometedores para predecir AMR, destacan las redes neuronales gráficas (GNN) [2] y las técnicas de procesamiento de lenguaje natural (PLN), utilizadas para extraer conocimiento de la literatura biomédica [6].

Sin embargo, estos enfoques presentan limitaciones cuando se utilizan de forma aislada. En particular, las GNN dependen de grafos biológicos cuya calidad puede verse afectada por datos incompletos, dispersos o ruidosos. Según [2], muchas redes biomoleculares empleadas en bioinformática presentan estructuras poco densas, lo que compromete tanto el rendimiento como la interpretabilidad de los modelos.

Por otro lado, los modelos de PLN, aunque eficaces para extraer información semántica [6, 7], no representan explícitamente las relaciones estructurales entre entidades biológicas, lo que limita su capacidad para inferir relaciones topológicas complejas. Integrar datos estructurados y semánticos ha demostrado mejorar la precisión en tareas predictivas, por lo que la ausencia de esta integración podría afectar la precisión de los modelos de AMR [12].

2 Estado del arte

Diversos estudios han aplicado GNNs para predecir resistencia antimicrobiana. Por ejemplo, GCGACNN es un modelo que combina GNNs y bosques aleatorios para predecir asociaciones entre microbios y fármacos [5]; otro modelo se ha utilizado para predecir las concentraciones mínimas inhibitorias (MIC, por sus siglas en inglés) en estudios de resistencia antimicrobiana [3]; y también se ha desarrollado un modelo que aborda la AMR utilizando registros electrónicos de salud (EHR, por sus siglas en inglés) [4]. Asimismo, existen trabajos que utilizan PLN, como un sistema automatizado para extraer genes relacionados con la resistencia a los antibióticos a partir de la literatura biomédica [7]; y modelos de lenguaje especializados como BioBERT, que han mostrado resultados sobresalientes en tareas de minería de texto biomédico [6].

3 Objetivo General y específicos

Objetivo General: Desarrollar y evaluar un modelo híbrido basado en Redes Neuronales Gráficas y Procesamiento de Lenguaje Natural, para predecir la resistencia antimicrobiana en cepas de E. coli.
Objetivos específicos: (i) construir un grafo genómico a partir de datos de NCBI [10], CARD [8] y BV-BRC [9]; (ii) extraer entidades biomédicas relacionadas con AMR desde literatura científica utilizando BioBERT; (iii) integrar los datos estructurados y semánticos en un modelo híbrido basado en GNN; y (iv) evaluar el desempeño del modelo híbrido en la predicción de AMR en *E. coli*, comparándolo contra enfoques unimodales (solo GNN o solo PLN).

4 Hipótesis

Se espera que la integración de datos genómicos estructurados y conocimiento biomédico extraído mediante PLN en un modelo híbrido basado en GNN mejore la precisión en la predicción de resistencia antimicrobiana en *Escherichia coli* en $\approx 5\%$, en comparación con enfoques que utilizan únicamente GNN o PLN.

5 Principales contribuciones

- Proponemos un modelo híbrido GNN–PLN que integra información estructural y conocimiento semántico biomédico para predecir AMR.
- Construimos un dataset multimodal reproducible a partir de NCBI, CARD, BV-BRC y literatura biomédica para experimentación abierta.
- La demostración empírica de que el enfoque híbrido mejora la precisión en $\approx 5\%$ frente a métodos basados exclusivamente en GNN o PLN.

6 Metodología y evaluación de resultados

La metodología consiste en integrar datos del grafo genómico con embeddings de BioBERT para la predicción de AMR en E. coli, siguiendo el flujo ilustrado en la Figura 1. Este pipeline consiste en la recolección, normalización y estandarización de entidades provenientes de CARD, NCBI, BV-BRC y literatura científica, para construir un grafo con nodos y relaciones biológicas relevantes. Cada nodo recibe un embedding semántico asignado por su identificador estandarizado, el cual es proyectado a la misma dimensionalidad que el embedding estructural. Ambos embeddings se fusionan en el espacio latente [11] mediante un mecanismo de softmax gating, que calcula la contribución relativa de cada modalidad:

$$g = softmax(W[h; s] + b) \tag{1}$$

Donde h y s son los embeddings estructural y semántico respectivamente y $[h; s]$ representa su concatenación, proyectada linealmente mediante $W[h; s] + b$, produciendo:

$$g = [g_h, g_s] \tag{2}$$

g_h y g_s son escalares. La representación multimodal final se calcula como:

$$h' = g_h \odot h \; + \; g_s \odot s \tag{3}$$

Aquí, $\odot$ representa multiplicación elemento a elemento y h' conserva dimensión original de los embeddings y constituye la entrada para el proceso de message passing de la GNN. Se realizará un experimento piloto para verificar la factibilidad y capacidad predictiva del modelo en cepas de *E. coli*, usando criterios de comparación y métodos de referencia basados en trabajos previos. El modelo híbrido se evaluará con métricas estándar (exactitud, F1-score, AUC-ROC) y mediante estudios de ablación para medir el aporte relativo de los datos estructurados y semánticos, comparándolo con modelos base de solo GNN o PLN. Los resultados se validarán por expertos y se difundirán en conferencias, revistas y repositorios abiertos para garantizar reproducibilidad y transparencia.

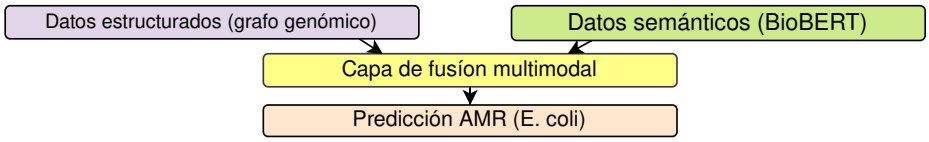


Fig. 1. Esquema conceptual del modelo híbrido GNN-PLN

Conflicto de intereses. Los autores declaran que no existen conflictos de intereses relevantes para el contenido de este artículo.

References

1. WHO. (2023). Antimicrobial resistance. Obtenido el 26 de septiembre de 2025 de https://www.who.int/news-room/fact-sheets/detail/antimicrobial-resistance.
2. Zhang, X.M., Liang, L., Liu, L., & Tang, M.J. (2021). Graph neural networks and their current applications in bioinformatics. Frontiers in Genetics, 12, 690049. https://doi.org/10.3389/fgene.2021.690049
3. Zhang, Z., Veerapaneni, R., Ayoola, M., Das, A. R., Chen, Z., Nanduri, B., & Ramkumar, M. (2024). Leveraging Graph Neural Networks for MIC prediction in antimicrobial resistance studies. En 2024 46th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). https://doi.org/10.1109/EMBC53108.2024.10782350
4. Gao, P., Chen, Z., Liu, X., Chen, P., Matsubara, Y., & Sakurai, Y. (2025). Antimicrobial resistance recommendations via electronic health records with graph representation and patient population modeling. Computer Methods and Programs in Biomedicine, 261, 108616. https://doi.org/10.1016/j.cmpb.2025.108616
5. Su, S., Liu, M., Zhou, J., & Zhang, J. (2024). GCGACNN: A graph neural network and random forest for predicting microbe–drug associations. Biomolecules, 14(8), 946. https://doi.org/10.3390/biom14080946
6. Lee, J., Yoon, W., Kim, S., Kim, D., Kim, S., So, C. H., & Kang, J. (2020). BioBERT: a pre-trained biomedical language representation model for biomedical text mining. Bioinformatics, 36(4), 1234–1240. https://doi.org/10.1093/bioinformatics/btz682
7. Brincat, A., & Hofmann, M. (2022). Automated extraction of genes associated with antibiotic resistance from the biomedical literature. Database, 2022, baab077. https://doi.org/10.1093/database/baab077
8. CARD. (2025). Obtenido el 24 de septiembre de 2025 de https://card.mcmaster.ca/
9. BV-BRC. (2025). Bacterial and Viral Bioinformatics Resource Center. Obtenido el 26 de septiembre de 2025 de de https://www.bv-brc.org/
10. NCBI. (2025). Pathogen detection. National Center for Biotechnology Information. Obtenido el 24 de septiembre de 2025 de https://www.ncbi.nlm.nih.gov/pathogens/
11. Duoyi Zhang, Md Abul Bashar, Richi Nayak, Leila Cuttle. (2025). A Novel Modality Contribution Confidence-enhanced Multimodal Deep Learning Framework for Multiomics Data. *BMC Bioinformatics, 26*(1), 271. https://doi.org/10.1186/s12859-025-06219-9
12. Alshahrani, M., Almansour, A., Alkhaldi, A., Thafar, M. A., Uludag, M., Essack, M., & Hoehndorf, R. (2022). Combining biomedical knowledge graphs and text to improve predictions for drug-target interactions and drug-indications. *PeerJ, 10*, e13061. https://doi.org/10.7717/peerj.13061

Optimización de modelos para identificar arritmia cardiaca con aprendizaje automático

Santiago Arias-García[1][0000-0003-3578-7707], José Hernández-Torruco[1]*[0000-0003-3146-9349], Betania Hernández-Ocaña[1][0000-0001-5700-7615], Oscar Chávez-Bosquez[1][0000-0002-0324-9886]

[1]Universidad Juárez Autónoma de Tabasco, México

231H18005@alumno.ujat.mx

Resumen. La detección temprana y precisa de arritmias cardíacas a través de señales de electrocardiograma (ECG) es fundamental para prevenir complicaciones graves de salud. Sin embargo, la alta dimensionalidad de los datos del ECG, que pueden incluir cientos de atributos, presenta un desafío significativo para los algoritmos de clasificación, afectando tanto la pureza como su eficiencia computacional. Este estudio propone un enfoque híbrido para optimizar la clasificación de arritmias, combinando un método de filtro (Correlation-based Feature Selection - CFS) con algoritmos metaheurísticos (Búsqueda Tabú y Recocido Simulado) para la selección de características. Utilizando el conjunto de datos de arritmias de la UCI, se demuestra que este enfoque reduce drásticamente el número de características necesarias, mejorando al mismo tiempo la pureza de la clasificación. El modelo final, que integra Recocido Simulado con un clasificador Random Forest, logra un promedio de accuracy balanceado del 83.89% utilizando solo 8 de las 279 características originales, superando en rendimiento y eficiencia a los modelos que utilizan únicamente métodos de filtro. Con Búsqueda Tabú y Random Forest logra un desempeño del 83.34% del promedio de accuracy balanceado utilizando solo 14 características. No se utilizó balanceo de datos en este trabajo.

Palabras clave: Arritmia. Clasificación. Selección de características. Búsqueda Tabú. Recocido Simulado. ECG.

1 Problema y justificación

La arritmia cardiaca, una alteración en el ritmo normal del corazón, es una de las principales causas de morbilidad y mortalidad a nivel mundial [1], [2]. Su identificación a tiempo a través de registros de señales ECG es muy importante para el diagnóstico y tratamiento clínico oportuno. La interpretación manual de estos registros es una tarea que consume tiempo, es propensa a errores y requiere de un médico especializado, sobre todo en análisis de larga duración.

Para hacer frente a estas limitaciones, se han desarrollado sistemas de diagnóstico asistido por ordenador (CAD) que emplean técnicas de aprendizaje automático para clasificar automáticamente las señales de ECG. Estos sistemas tienen el potencial de optimizar el diagnóstico médico y reducir los costos asociados [3]. Sin embargo, los principales desafíos en el procesamiento de este conjunto de datos son el número limitado de ejemplos de entrenamiento en comparación con el alto número de características, un alto sesgo hacia el caso de ECG normal (sin arritmia), y la presencia de valores faltantes (NA).

El conjunto de datos de Arritmia de la UCI (University of California, Irvine), un referente en este campo de estudio consta de 452 casos y 279 variables [1], [3]–[6]. Este elevado número de características, muchas de las cuales pueden ser redundantes o irrelevantes, no solo aumenta la complejidad y el costo computacional de los modelos de clasificación, sino que también puede afectar negativamente su rendimiento [2], [6]. Por lo tanto, se recurre a la selección de variables relevantes, ya que los algoritmos de clasificación funcionan de manera más efectiva cuando existen atributos relevantes dentro del conjunto de datos. El objetivo es encontrar un subconjunto óptimo de características que permita lograr una mejor tasa de clasificación.

2 Objetivos de la investigación

El objetivo principal de este trabajo es desarrollar y evaluar un enfoque de selección de características híbrido para mejorar el rendimiento del modelo en la clasificación de arritmia cardiaca, utilizando el conjunto de datos de la UCI. Para ello, se plantean los siguientes objetivos específicos:

1. Identificar un método filtro eficiente como línea base para una reducción inicial de variables, evaluando su desempeño con distintos clasificadores para encontrar el modelo con el accuracy balanceado promedio más alto.
2. Utilizar el subconjunto de datos obtenido en el paso anterior como punto de partida para una segunda fase de selección de características, aplicando las metaheurísticas Búsqueda Tabú y Recocido Simulado.
3. Evaluar el rendimiento de los modelos resultantes (Filtro, Búsqueda Tabú y Recocido Simulado) en términos de reducción de variables y la pureza en la clasificación, tomando como métrica principal el accuracy balanceado.
4. Demostrar que un enfoque híbrido secuencial puede alcanzar un rendimiento superior o comparable con un subconjunto seleccionado pequeño en comparación de los métodos tradicionales.

3 Contribuciones esperadas y trabajos relacionados

La principal contribución de este estudio es un enfoque híbrido secuencial que combina la simplicidad y rapidez de los métodos filtro con la capacidad de optimización de las metaheurísticas para la selección de variables relevantes.

En la literatura, diversos investigadores han abordado el problema de la clasificación de arritmias. Algunos trabajos se han enfocado en aplicar distintos algoritmos de clasificación sobre el conjunto de datos completo. Por ejemplo, Al-Saffar et al. (2023) lograron un 92.47% de accuracy utilizando todas las 279 variables con una Red Neuronal Artificial (ANN), aunque esto implica un alto costo computacional [7]. Otros estudios reconocen la importancia de la selección de características para mejorar el rendimiento del modelo .

Los métodos filtro son un proceso inicial popular, ya que son los más sencillos y rápidos. Mitra & Samanta (2013) utilizaron el filtro CFS para seleccionar 18 características y alcanzaron un 87.71% de accuracy promedio con redes neuronales [2]. Yilmaz (2013) empleó Fisher Score para reducir el conjunto a 65 variables, obteniendo un accuracy del 82.09% con un clasificador LS-SVM [3]. Por su parte, Xu et al. (2017) usaron Fisher Discriminant Ratio (FDR) para seleccionar 236 variables y lograron un 80.64% de accuracy con una Deep Neural Network (DNN) [8].

Técnicas de extracción de características como el Análisis de Componentes Principales (PCA) también han sido exploradas. Elsayad (2009) redujo la dimensionalidad a 100 componentes, alcanzando un 74.12% de accuracy con redes LVQ [5]. Khare et al. (2012) combinaron la Correlación de Spearman con PCA para obtener 115 características y lograron un 85.98% de accuracy con SVM [6].

Las metaheurísticas, como los algoritmos genéticos (GA), son otra técnica utilizada para solucionar problemas de optimización. Kadam et al. (2020) emplearon un GA elitista para seleccionar 92 variables y obtuvieron un 87.83% de accuracy promedio con SVM [9]. Ayar & Sabamoniri (2018) usaron un GA para identificar 61 características, con las que alcanzaron un 86.96% de accuracy utilizando un árbol de decisión C4.5.

Nuestro enfoque base se distingue por proponer una metodología secuencial donde los métodos filtro y las metaheurísticas no son alternativas, sino fases complementarias. Primero, se utiliza un método filtro (CFS) para obtener un subconjunto de variables prometedor (37), reduciendo así el espacio de búsqueda. Posteriormente, se aplican las metaheurísticas (Búsqueda Tabú y Recocido Simulado) sobre este conjunto reducido para refinar la selección y encontrar un subconjunto de características óptimo.

4 Metodología y evaluación

Para la creación de los modelos predictivos, se utilizó el conjunto de datos de arritmia de la UCI [10]. Como parte del preprocesamiento, se procedió a imputar los datos faltantes por medio de la media del total de los datos de cada variable [11].

Dado que el conjunto de datos presenta un desbalanceo hacia la clase mayoritaria (casos con ECG normales), se utilizó como métrica principal el accuracy balanceado, ya que proporciona un balance entre la sensibilidad y especificidad del modelo. Esta métrica promedia la especificidad y sensibilidad del modelo, lo que la hace robusta en estas condiciones.

En un trabajo relacionado con el preprocesamiento del conjunto de datos de arritmia cardiaca [11], realizamos una imputación de los datos faltantes mediante la media, mediana y la moda, siendo la media la técnica más efectiva. Se eliminó la variable V14 porque presentaba datos faltantes arriba del 80%, las demás características con valores NA se imputaron con la media.

Todas las etapas, imputación de valores faltantes, selección de características y ajuste de hiperparámetros, se ejecutaron exclusivamente dentro de cada partición de entrenamiento y posteriormente se aplicaron al conjunto de prueba correspondiente [11]. En la imagen 1 se observa el proceso general para la construcción de los modelos utilizando cada método filtro y cada metaheurística correspondiente.

4.1 Flujo del experimento

En el presente estudio se realizó un diseño estructurado en tres etapas con el propósito de reducir la dimensionalidad del conjunto de datos y optimizar el rendimiento de los modelos de clasificación.

1. Fase 1 (línea base): Para evaluar el rendimiento de la optimización con los métodos propuestos, se construyeron modelos de clasificación con todas las variables del conjunto de datos imputado (278) [11]. Se utilizaron los clasificadores C4.5, Rpart, PART, K-Nearest Neighbor (KNN), Support Vector Machine con kernel lineal (SVMLin) y Random Forest (RF). Con esta línea de referencia de rendimiento, se compararon con los métodos de selección de características;
2. Fase 2 (métodos filtro): Se implementó los métodos filtro del paquete FSelector [12] CFS, Consistency, Chi-Square, Information Gain, Gain Ratio, Symmetrical Uncertainty, y OneR. Los métodos CFS y Consistency construyen subconjuntos con las variables más relevantes, se aplicaron los clasificadores correspondientes de cada subconjunto. Los demás métodos filtro, generan una lista de las variables con un orden de importancia, pero para encontrar el mejor subconjunto, se construyeron pequeños subconjuntos incrementados que fueron desde los primeros dos, hasta incluir todas las variables. Una vez identificado el valor máximo de accuracy balanceado promedio de los subconjuntos incrementados, se realizó la implementación del clasificador correspondiente;
3. Fase 3 (algoritmos metaheurísticos): Una vez identificado el mejor subconjunto del método filtro, se utilizó como entrada para la optimización de selección de características mediante los algoritmos metaheurísticos recocido simulado y búsqueda tabú. Cada metaheurística realizó la búsqueda del mejor subconjunto, maximizando el accuracy balanceado promedio y minimizando el número de características seleccionadas.

4.2 Configuración experimental

4.2.1 Entorno y validación

El software estadístico utilizado para la implementación de los experimentos fue RStudio. Se empleó el paquete *FSelector* [12] para los métodos filtro; *Caret* [13], *RWeka* [14], *kknn* [15], *Rpart* [16], *e1071* [17] *y randomForest* [18] para los

clasificadores correspondientes. Para las metaheurísticas se implementaron con las librerías *tabuSearch* [19] para búsqueda tabú y *GenSA* [20] para recocido simulado.

En la implementación de los clasificadores, se utilizó la técnica de validación repetida hold-out [21] mediante *createDatapartition()* del paquete caret [13], donde se dividió aleatoriamente las proporciones de 2/3 para el entrenamiento del modelo y 1/3 para la prueba. Este procedimiento se ejecutó 30 veces con las semillas reproducibles (*set.seed(1), …, set.seed(30)*) para crear diferentes particiones y poder medir la estabilidad con la deviación estándar. Ver imagen 1.

Para las configuraciones de los clasificadores kNN, SVMLin y Random Forest, se aplicaron procedimientos de *grid search* para identificar los hiperparámetros óptimos antes de crear los modelos predictivos: número óptimo de vecinos (*k*), parámetro de penalización (*cost*) y las combinaciones de *mtry-ntree*.

- kNN: k = 1–20, distance = 1–2, kernel = {optimal, epanechnikov, triangular, rectangular, gaussian}; validación cruzada (trainControl(method = "cv"));
- SVMLin: cost ∈ {0.001, 0.01, 0.1, 1, 10, 50, 80, 100}; tune(svm, kernel = "linear");
- Random Forest: Grid search: ntree ∈ {100, 200, 300, 400, 500}, mtry ∈ {1, …, p}; validación cruzada repetida (repeatedcv, 10 folds × 3 repeticiones).

Las configuraciones de los algoritmos metaheurísticos se realizaron con los óptimos globales dependiendo de parámetros de las funciones. La función objetivo era aumentar el accuracy balanceado promedio y disminuir el tamaño del subconjunto de características.

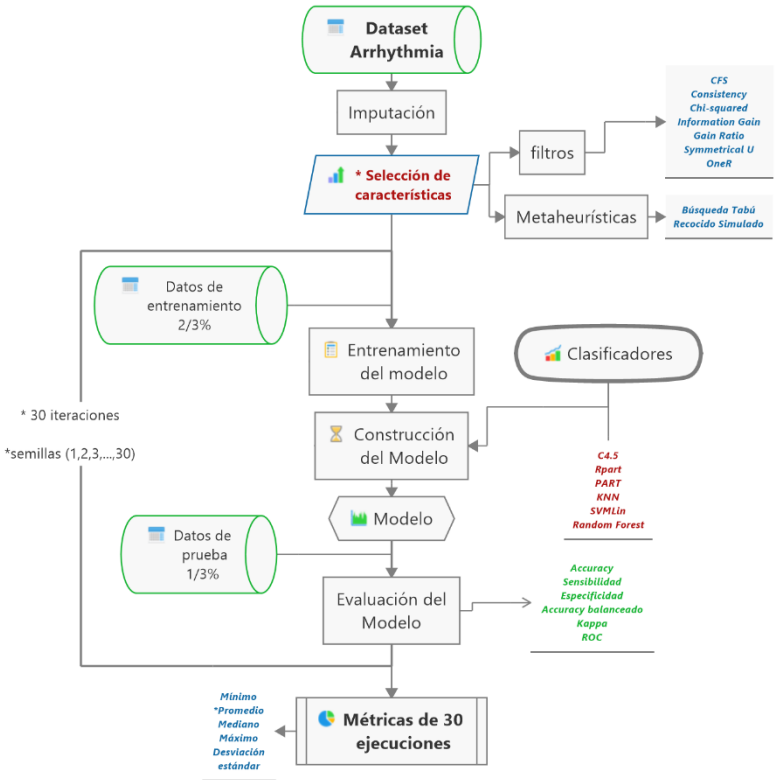


Imagen 1. Proceso de construcción de los modelos utilizando selección de características y clasificadores.

Tabla 1. Resultados de la implementación de los filtros y clasificadores. Los valores son los promedios de 30 ejecuciones y la desviación estándar.

Filtros	Clasificadores	C4.5	Rpart	PART	KNN	SVMLin	Random Forest
CFS	*Variables encontradas*	37	37	37	37	37	**37**
	Accuracy Balanceado	0.7723	0.7557	0.7718	0.7269	0.7510	**0.8293**
	+ SD	+ 0.0308	+ 0.0429	+ 0.0365	+ 0.0223	+ 0.0245	+ 0.0283
	Sensibilidad	0.7454	0.7329	0.7382	0.5527	0.5773	0.8068
	Especificidad	0.7992	0.7786	0.8053	0.9012	0.9247	0.8519
	AUC-ROC	0.7756	0.7590	0.7769	0.7664	0.7960	0.8323
Consistency	*Variables encontradas*	20	20	20	20	20	20
	Accuracy balanceado	0.7565	0.7684	0.7568	0.6824	0.7488	0.8198
	+ SD	+ 0.0387	+ 0.0338	+ 0.0391	+ 0.0284	+ 0.0260	+ 0.0274
	Sensibilidad	0.7155	0.7459	0.7309	0.4348	0.6338	0.8116
	Especificidad	0.7975	0.7909	0.7827	0.9259	0.8638	0.8395
	AUC-ROC	0.7615	0.7713	0.7593	0.7416	0.7690	0.8255
OneR	*Variables encontradas*	53	42	84	20	62	80
	Accuracy balanceado	0.7603	0.7610	0.7536	0.7155	0.7588	0.8275
	+ SD	+ 0.0286	+ 0.0405	+ 0.0338	+ 0.0354	+ 0.0269	+ 0.0259
	Sensibilidad	0.7285	0.7401	0.7343	0.4758	0.6159	0.8053
	Especificidad	0.7922	0.7819	0.7728	0.9551	0.9016	0.8490
	AUC-ROC	0.7638	0.7642	0.7555	0.7926	0.7898	0.8302
Chi-squared	*Variables encontradas*	18	20	35	17	49	78
	Accuracy balanceado	0.7694	0.7674	0.7682	0.7290	0.7595	0.8261
	+ SD	+ 0.0359	+ 0.0324	+ 0.0332	+ 0.0309	+ 0.0264	+ 0.0267
	Sensibilidad	0.7290	0.7401	0.7348	0.5198	0.5928	0.8048
	Especificidad	0.8099	0.7947	0.8016	0.9383	0.9263	0.8473
	AUC-ROC	0.7740	0.7709	0.7736	0.7889	0.8015	0.8291
Information Gain	*Variables encontradas*	15	21	48	30	62	64
	Accuracy balanceado	0.7666	0.7664	0.7700	0.7415	0.7606	0.8295
	+ SD	+ 0.0344	+ 0.0333	+ 0.0321	+ 0.0305	+ 0.0255	+ 0.0234
	Sensibilidad	0.7213	0.7406	0.7333	0.5072	0.6126	0.8126
	Especificidad	0.8119	0.7922	0.8066	0.9514	0.9086	0.8465
	AUC-ROC	0.7724	0.7697	0.7750	0.7983	0.7944	0.8319
Symmetrical Uncertainty	*Variables encontradas*	35	37	34	7	62	59
	Accuracy balanceado	0.7706	0.7660	0.7687	0.7525	0.7537	0.8292
	+ SD	+ 0.0360	+ 0.0412	+ 0.0333	+ 0.0295	+ 0.0243	+ 0.0240
	Sensibilidad	0.7304	0.7435	0.7246	0.5536	0.6005	0.8169
	Especificidad	0.8107	0.7885	0.8128	0.9420	0.9070	0.8416
	AUC-ROC	0.7754	0.7685	0.7761	0.8030	0.7891	0.8306
Gain Ratio	*Variables encontradas*	44	41	66	42	48	61
	Accuracy balanceado	0.7612	0.7779	0.7613	0.7372	0.7549	0.8268
	+ SD	+ 0.0324	+ 0.0299	+ 0.0356	+ 0.0279	+ 0.0259	+ 0.0259
	Sensibilidad	0.7150	0.7464	0.7382	0.5111	0.5942	0.8179
	Especificidad	0.8074	0.8095	0.7844	0.9576	0.9156	0.8358
	AUC-ROC	0.7669	0.7820	0.7642	0.8057	0.7934	0.8286

Adicionalmente, se realizó una prueba de hipótesis de Wilcoxon pareado entre los tres enfoques evaluados (CFS, Búsqueda Tabú y Recocido Simulado) para determinar diferencias significativas en el rendimiento promedio.

4.2.2 Métricas de evaluación

Para reflejar el desbalance de clases, la métrica principal fue el Balanced Accuracy. Asimismo, se calcularon las métricas complementarias: sensibilidad (recall or clase), especificidad, ROC. Los valores se reportan como media ± desviación estándar. Además, se generaron matrices de confusión para los modelos evaluados.

5 Resultados

En el enfoque base utilizando todas (278) variables del dataset, se obtuvo el mejor resultado utilizando el clasificador Random Forest. Alcanzó un accuracy balanceado promedio de 81.47% y una desviación estándar de 0.0250. Este rendimiento se obtuvo con una configuración de los parámetros óptimos de mtry = 50 y ntree = 100.

Una vez identificados los diferentes subconjuntos de variables relevantes, se procedió a crear los modelos predictivos. Los resultados se resumen a continuación:

1. Modelos con Métodos Filtro: Se evaluaron 7 métodos filtro del paquete FSelector con 8 algoritmos de clasificación. El mejor modelo de esta fase fue el que utilizó el filtro Correlation-based Feature Selection (CFS) junto con el clasificador Random Forest. Este rendimiento se obtuvo con una configuración de los parámetros óptimos de mtry = 11 y ntree = 300. Este modelo seleccionó 37 variables y obtuvo un promedio de accuracy balanceado de 82.93% y una desviación estándar (SD) de 0.0283 . Ver Tabla 1.

2. Modelos con Metaheurísticas: El subconjunto de 37 características encontrado por CFS se utilizó como entrada para las metaheurísticas.

- Búsqueda Tabú: Esta metaheurística redujo el conjunto de datos a 14 variables, con las que se alcanzó un promedio de accuracy balanceado de 83.34% + 0.0290 (SD) utilizando el clasificador Random Forest. Este rendimiento se obtuvo con una configuración de los parámetros óptimos de *mtry = 8* y *ntree = 400*. Las características seleccionadas fueron: V15, V30, V69, V76, V91, V112, V197, V211, V217, V224, V228, V234, V250, V279.
- Recocido Simulado: Este enfoque encontró el subconjunto óptimo, seleccionando solo 8 variables. Con este conjunto, el clasificador Random Forest logró el mayor rendimiento, con un promedio de accuracy balanceado de 83.89% + 0.0249 (SD). Este rendimiento se obtuvo con una configuración de los parámetros óptimos de *mtry = 2* y *ntree = 400*. Las características seleccionadas fueron: V15, V76, V112, V224, V234, V277, V69, V169.

Los resultados demuestran que la selección de características con los métodos filtros mejora el rendimiento de los clasificadores. Además, al reducir las variables con ayuda de los métodos filtros, el costo computacional se reduce, haciendo posible tener un buen rendimiento al utilizar las metaheurísticas para una selección posterior. La combinación de CFS con Recocido Simulado y Random Forest fue el modelo con mayor rendimiento, logrando una reducción de características del 97% (de 279 a 8) y superando a los modelos base. La Tabla 2 muestra la comparación de los mejores resultados de nuestro estudio con los de otras investigaciones.

Con el propósito de determinar si existían diferencias estadísticamente significativas en el rendimiento de los modelos obtenidos con el filtro CFS y las metaheurísticas Recocido Simulado y Búsqueda Tabú, se aplicó la prueba de rangos con signo de Wilcoxon considerando las 30 ejecuciones independientes realizadas con cada método.

En la comparación CFS vs Búsqueda Tabú, se obtuvo un valor de W = 174.5 y un valor de Z = -0.649, con un valor p = 0.5157. De manera similar, para la comparación CFS vs Recocido Simulado, los resultados fueron W = 153.5, Z = -1.6249 y p = 0.1052. En ambos casos, los valores p fueron superiores al nivel de significancia

convencional ($p < 0.05$), por lo que no se detectaron diferencias estadísticamente significativas entre los métodos comparados.

Estos resultados indican que el rendimiento medio obtenido con el filtro CFS es estadísticamente equivalente al logrado mediante las metaheurísticas evaluadas. En consecuencia, aunque los enfoques de optimización presentan mecanismos distintos de búsqueda y selección de características, su efectividad en términos de desempeño predictivo no difirió significativamente bajo las condiciones experimentales establecidas.

Tabla 2. Comparación del rendimiento promedio con otros estudios sobre el dataset de arritmia de la UCI.

Estudio	**Método de Selección / Extracción**	**Nº de Vars.**	**Clasificador**	**Métrica (%)**
Elsayad (2009) [5]	PCA	100	LVQ2.1	Accuracy: 74.12
Yilmaz (2013) [3]	Fisher Score	65	LS-SVM	Accuracy (promedio): 82.09
Mitra & Samanta (2013) [2]	CFS	18	LM (Red Neuronal)	Accuracy (promedio): 87.71
Xu et al. (2017) [8]	FDR	236	DNN	Accuracy (promedio): 80.64
Ayar & Sabamoniri (2018)[22]	Genetic Algorithm	61	C4.5	Accuracy: 86.96
Kadam et al. (2020) [9]	Elitist GA	92	SVM	Accuracy (promedio): 87.83
Al-Saffar et al. (2023) [7]	Ninguno (solo preproc.)	279	ANN	Accuracy: 92.47
Este estudio (CFS)	**CFS**	**37**	**Random Forest**	***Accuracy* Balanceado (promedio): 82.93**
Este estudio (B. Tabú)	**CFS + Búsqueda Tabú**	**14**	**Random Forest**	***Accuracy* Balanceado (promedio): 83.34**
Este estudio (R. Simulado)	**CFS + Recocido Simulado**	**8**	**Random Forest**	***Accuracy* Balanceado (promedio): 83.89**

A continuación, se muestran la matriz de confusión de los mejores resultados de cada método de selección de características (CFS – Búsqueda Tabú – Recocido Simulado).

```
Confusion Matrix and Statistics

          Reference
Prediction  1  2
         1 72  8
         2  9 61

               Accuracy : 0.8867
                 95% CI : (0.8248, 0.9326)
    No Information Rate : 0.54
    P-Value [Acc > NIR] : <2e-16

                  Kappa : 0.7721

 Mcnemar's Test P-Value : 1

            Sensitivity : 0.8841
            Specificity : 0.8889
         Pos Pred Value : 0.8714
         Neg Pred Value : 0.9000
             Prevalence : 0.4600
         Detection Rate : 0.4067
   Detection Prevalence : 0.4667
      Balanced Accuracy : 0.8865

       'Positive' Class : 2
```

(**a**)

```
Confusion Matrix and Statistics

          Reference
Prediction  1  2
         1 72 10
         2  9 59

               Accuracy : 0.8733
                 95% CI : (0.8093, 0.922)
    No Information Rate : 0.54
    P-Value [Acc > NIR] : <2e-16

                  Kappa : 0.7448

 Mcnemar's Test P-Value : 1

            Sensitivity : 0.8551
            Specificity : 0.8889
         Pos Pred Value : 0.8676
         Neg Pred Value : 0.8780
             Prevalence : 0.4600
         Detection Rate : 0.3933
   Detection Prevalence : 0.4533
      Balanced Accuracy : 0.8720

       'Positive' Class : 2
```

(**b**)

```
Confusion Matrix and Statistics

          Reference
Prediction  1  2
         1 71  8
         2 10 61

               Accuracy : 0.88
                 95% CI : (0.817, 0.9273)
    No Information Rate : 0.54
    P-Value [Acc > NIR] : <2e-16

                  Kappa : 0.759

 Mcnemar's Test P-Value : 0.8137

            Sensitivity : 0.8841
            Specificity : 0.8765
         Pos Pred Value : 0.8592
         Neg Pred Value : 0.8987
             Prevalence : 0.4600
         Detection Rate : 0.4067
   Detection Prevalence : 0.4733
      Balanced Accuracy : 0.8803

       'Positive' Class : 2
```

(c)

Imagen 2. En esta imagen se muestra la matriz de confusión de los mejores valores de cada subconjunto óptimo de los métodos de selección de características: (a) CFS – Random Forest; (b) Búsqueda Tabú – Random Forest; (c) Recocido Simulado – Random Forest.

6 Conclusiones y trabajo futuro

En este trabajo se ha presentado un enfoque híbrido y secuencial para la selección de variables relevantes en la identificación de arritmia cardiaca. La combinación del método filtro CFS con las metaheurísticas Búsqueda Tabú y Recocido Simulado fue eficaz para reducir la dimensionalidad del conjunto de datos y mejorar el rendimiento de los modelos predictivos. El mejor modelo, que integra Recocido Simulado con un clasificador Random Forest, obtuvo un accuracy balanceado promedio de 83.89% con solo 8 características.

Este estudio contribuye a la importancia de implementar metaheurísticas para seleccionar características relevantes. Se encontró que el clasificador Random Forest produjo los mejores resultados de manera consistente. Este hallazgo no sustituye a los médicos, sino que

El sistema propuesto está orientado a apoyar la toma de decisiones clínicas y no a sustituir el diagnóstico médico. Dado que el conjunto de datos proviene de un único repositorio (UCI), se requiere validación externa en contextos clínicos reales.

Como trabajo a futuro, se propone aplicar otras metaheurísticas para buscar patrones de interés, así como explorar técnicas de balanceo de clases para incrementar estadísticamente el rendimiento de los modelos.

Referencia

[1] K. Polat, S. Şahan, and S. Güneş, "A new method to medical diagnosis: Artificial immune recognition system (AIRS) with fuzzy weighted pre-processing and application to ECG arrhythmia," *Expert Syst. Appl.*, vol. 31, no. 2, pp. 264–269, Aug. 2006, doi: 10.1016/j.eswa.2005.09.019.

[2] M. Mitra and R. K. Samanta, "Cardiac Arrhythmia Classification Using Neural Networks with Selected Features," *Procedia Technol.*, vol. 10, pp. 76–84, Jan. 2013, doi: 10.1016/j.protcy.2013.12.339.

[3] E. Yilmaz, "An expert system based on fisher score and LS-SVM for cardiac arrhythmia diagnosis," *Comput. Math. Methods Med.*, vol. 2013, 2013, doi: 10.1155/2013/849674.

[4] A. Uyar and F. Gürgen, "Arrhythmia classification using serial fusion of support vector machines and logistic regression," *2007 4th IEEE Work. Intell. Data Acquis. Adv. Comput. Syst. Technol. Appl. IDAACS*, pp. 560–565, 2007, doi: 10.1109/IDAACS.2007.4488483.

[5] A. M. Elsayad, "Classification of ECG arrhythmia using learning vector quantization neural networks," in *Proceedings - The 2009 International Conference on Computer Engineering and Systems, ICCES'09*, 2009, pp. 139–144. doi: 10.1109/ICCES.2009.5383295.

[6] S. Khare, A. Bhandari, S. Singh, and A. Arora, "ECG Arrhythmia Classification Using Spearman Rank Correlation and Support Vector Machine," in *Advances in Intelligent and Soft Computing*, vol. 131 AISC, no. VOL. 2, 2012, pp. 591–598. doi: 10.1007/978-81-322-0491-6_54.

[7] B. Al-Saffar, Y. H. Ali, A. M. Muslim, and H. A. Ali, "ECG Signal Classification Based on Neural Network," in *Lecture Notes in Networks and Systems*, 2023, vol. 573 LNNS, pp. 3–11. doi: 10.1007/978-3-031-20429-6_1.

[8] Sean shensheng Xu, Man-Wai Mak, and Chi-Chung Cheung, "Deep neural networks versus support vector machines for ECG arrhythmia classification," in *2017 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*, Jul. 2017, pp. 127–132. doi: 10.1109/ICMEW.2017.8026250.

[9] V. J. Kadam, S. S. Yadav, and S. M. Jadhav, "Soft-Margin SVM Incorporating Feature Selection Using Improved Elitist GA for Arrhythmia Classification," in *Advances in Intelligent Systems and Computing*, 2020, vol. 941, pp. 965–976. doi: 10.1007/978-3-030-16660-1_94.

[10] Sang-Hong Lee, Jung-Kwon Uhm, and J. S. Lim, "Extracting input features and fuzzy rules for detecting ECG arrhythmia based on NEWFM," in *2007 International Conference on Intelligent and Advanced Systems*, Nov. 2007, pp. 22–25. doi: 10.1109/ICIAS.2007.4658341.

[11] S. Arias-García, J. Hernández-Torruco, and B. Hernández-Ocaña, "Imputación de datos de pacientes con arritmia cardiaca," *Investig. Apl. Un Enfoque en la Tecnol.*, vol. 6, no. 2021, pp. 351–359, Dec. 2021, Accessed: May 30, 2023. [Online]. Available: https://scholar.google.com/citations?view_op=view_citation&hl=es&user=ABDCkcAAAAAJ&cstart=100&pagesize=100&citation_for_view=ABDCkcAAAAAJ:qxL8FJ1GzNcC

[12] P. Romanski, L. K. Maintainer, and L. Kotthoff, “Package ‘FSelector’ Type Package Title Selecting attributes,” 2015.
[13] M. Kuhn, “Building predictive models in R using the caret package,” *J. Stat. Softw.*, vol. 28, no. 5, pp. 1–26, Nov. 2008, doi: 10.18637/jss.v028.i05.
[14] K. Hornik, “RWeka Odds and Ends,” pp. 1–4, 2012, Accessed: Jun. 16, 2022. [Online]. Available: http://ftp.arklinux.org/pub/cran/web/packages/RWeka/vignettes/RWeka.pdf
[15] K. Schliep and K. Hechenbichler, “Weighted k-Nearest Neighbors [R package kknn version 1.4.1],” *CRAN Contrib. Packag.*, May 2025, doi: 10.32614/CRAN.PACKAGE.KKNN.
[16] Terry Therneau, Beth Atkinson, and Brian Ripley, “Package ‘rpart,’” p. 34, Jan. 2022, Accessed: Aug. 24, 2022. [Online]. Available: https://cran.r-project.org/package=rpart
[17] D. Meyer, E. Dimitriadou, K. Hornik, A. Weingessel, and F. Leisch, “Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien [R package e1071 version 1.7-16],” *CRAN Contrib. Packag.*, Sep. 2024, doi: 10.32614/CRAN.PACKAGE.E1071.
[18] L. Breiman, “Random forests,” *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, Oct. 2001, doi: 10.1023/A:1010933404324.
[19] K. Domijan, “Tabu Search Algorithm for Binary Configurations,” p. 6, Mar. 2018, doi: 10.1111/j.1751-5823.2002.tb00174.x.
[20] S. Gubian, Y. Xiang, B. Suomela, J. Hoeng, and S. A. Maintainer, “Package ‘GenSA’ Title R Functions for Generalized Simulated Annealing,” 2025.
[21] S. Arias-García, J. Hernández-Torruco, B. Hernández-Ocaña, and O. Chávez-Bosquez, “Cardiac Arrhythmia Identification Using Feature Selection and Rule-Based Classifiers,” *IFMBE Proc.*, vol. 110, pp. 120–128, 2024, doi: 10.1007/978-3-031-62520-6_14.
[22] M. Ayar and S. Sabamoniri, “An ECG-based feature selection and heartbeat classification model using a hybrid heuristic algorithm,” *Informatics Med. Unlocked*, vol. 13, pp. 167–175, Jan. 2018, doi: 10.1016/j.imu.2018.06.002.

Neuroevolution of Spiking Neuron Networks

Carlos-Alberto López-Herrera[0009−0009−5603−7690],
Héctor-Gabriel Acosta-Mesa[0000−0002−0935−7642], and
Efrén Mezura-Montes[0000−0002−1565−5267]

Artificial Intelligence Research Institute, University of Veracruz, Mexico
zs23000650@estudiantes.uv.mx, {heacosta,emezura}@uv.mx
https://www.uv.mx/iiia/

Abstract. Spiking Neural Networks (SNNs) are well-suited for processing temporal data; however, their performance depends strongly on the architectural design. Liquid State Machines (LSMs), a reservoir computing model, typically rely on random or heuristic constructions that limit consistency and efficiency. This research project introduces a neuroevolutionary algorithm that jointly optimizes neuron parameters and spatial properties (positions and connectivity). Preliminary results indicate that the combined encoding and configuration-only approach yields more robust reservoirs than position-only schemes, and that the evolutionary process enhances generalization over particle swarm optimization while requiring fewer neurons. Future work will expand to larger event-based benchmarks and analyze structural metrics to inform the design of compact and practical LSMs.

Keywords: Spiking Neural Network · Liquid State Machine · Neuroevolution.

1 Introduction

Spiking Neural Networks (SNNs) represent the third generation of artificial neural models [1], characterized by their ability to process information through discrete spike events rather than continuous activations. This attribute makes them particularly suitable for tasks involving time-varying or event-based data, such as auditory [2, 3] and visual streams [4, 5], where conventional Artificial Neural Networks (ANNs) often struggle to capture fine-grained temporal dynamics. Among the various neuron models proposed for SNNs, the Leaky Integrate-and-Fire (LIF) model [6] is one of the most widely adopted due to its balance between biological plausibility and computational manageability. The LIF models operate under the principle that a neuron integrates incoming currents into its membrane potential, which gradually decays ("leaks") over time. When this potential surpasses a defined threshold, the neuron emits a spike and enters a refractory period during which it cannot fire again (Fig. 1). Parameters such as the membrane potential, membrane time constant, refractory period, and the firing threshold control these dynamics, shaping how individual neurons respond to stimuli and interact within a network [6].

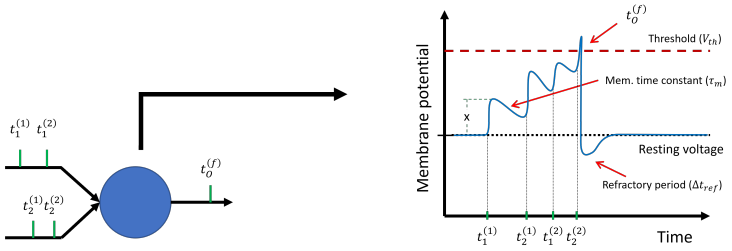


Fig. 1. Illustration of a spiking neuron model. On the left, incoming presynaptic spikes are integrated, incrementally modifying the neuron's membrane potential. On the right, the membrane potential trace illustrates how inputs accumulate and gradually decay over time, following the membrane time constant. Once the threshold is crossed, the neuron emits a spike, after which the potential is reset and followed by a refractory period.

Within the broader family of SNNs, Liquid State Machines (LSMs) [7] have emerged as a prominent architecture under the reservoir computing paradigm. LSMs consist of a randomly connected recurrent network of spiking neurons, known as liquid, which transform input streams of spike trains into high-dimensional states. These states can then be linearly separated by a simple readout mechanism, i.e., a classifier, enabling sequence processing with minimal training requirements.

Despite their potential, the performance of LSMs is highly sensitive to architectural design choices. Factors such as synaptic connectivity and spatial organization of the liquid can significantly affect their dynamics. Traditional approaches rely on random instantiation [7] or reinforcement learning methods [8], resulting in inconsistent results and suboptimal performance. In this context, Neuroevolution (NE) has arisen as a promising alternative. Derived from a broader field of Evolutionary Computation (EC), NE optimizes neural architecture through biologically inspired mechanisms such as mutation, crossover, and selection. Recent research has begun to explore these strategies for systematically tuning the liquid hyperparameters [9–12]. Yet, all efforts have been mainly focused on the connectivity aspect of the model.

This work addresses this gap by exploring NE applied to LMSs. In particular, we propose a Genetic Algorithm (GA) based approach that evolves the liquid by encoding neuron configurations (e.g., membrane time constant, refractory period, and spike threshold) together with their spatial positions. By systematically analyzing the impact of these design factors, our study aims to provide deeper insight into the properties governing effective LSM architectures.

The rest of this work is divided as follows: Sections 2 and 3 present the research hypothesis and objectives, respectively. Section 4 discusses the expected contributions of this study and their relation to previous work. Section 5 describes the strategy for evaluating the proposed approach and communicating

the outcomes. Finally, Section 6 outlines preliminary findings and the directions for future work.

2 Hypothesis

We hypothesize that an NE algorithm that considers both structural properties, such as neuron positions within the liquid, as well as functional properties, such as membrane time constant, refractory period, and spike threshold, can evolve LSM that achieve statistically higher classification accuracy while maintaining more compact network sizes, compared to existing state-of-the-art (SOTA) methods, across both both simple and event-driven classification benchmarks.

3 Objectives

The main objective of this research is to advance the state of the art in evolving LSMs by incorporating neuron configurations as explicit hyperparameters into the evolutionary process, thereby enhancing the efficiency and capabilities of these models.

To achieve this, the specific objectives are:

1. Review SOTA approaches in EC, and NE applied to SNNs and LSMs. Particular attention will be given to encoding strategies, network representations, and fitness functions.
2. Establish a comprehensible review of benchmarks and metrics used in SOTA literature to establish an experiential design that enables a fair comparison.
3. Develop a NE algorithm tailored to LSMs that can explore both structural configurations, synaptic factors, and neuron parameters. The LIF model will serve as the computational unit.
4. Assess the performance of the proposed algorithm against existing SOTA methods using selected benchmarks, such as event-driven and straightforward classification tasks.
5. Investigate the impact of evolving different neuron configurations on overall performance and on the behavior of the evolutionary process.
6. Explore potential applications of the proposed framework in domains that benefit from efficient temporal processing.
7. Identify limitations, challenges, and areas of opportunity of the proposed algorithm, providing recommendations for future research directions.

4 Contributions

This research introduces an NE algorithm that explicitly integrates individual neuron configurations (i.e., functional properties) and spatial information in the optimization of LSMs. By doing so, it provides an innovative perspective that expands the design possibilities of reservoir computing models and contributes to a deeper understanding of their structural and dynamical characteristics.

4.1 Related work contrast

Early approaches to LSM design generated random reservoirs [7], which were similar to Random Search (RS) and often resulted in inconsistent performance. Later studies introduced heuristic connectivity rules [8], but fixed both spatial and functional configurations of neurons within the liquid. Recent research has implemented EC and NE approaches. Tian et. al. [10] introduce a multi-liquid framework optimized with Simulated Annealing, achieving competitive results but focusing only on liquid-level adjustments rather than neuron configurations or connection patterns. Zhou et. al. [11] proposed the Generative LSM using cooperative co-evolution with a surrogate model, but their design was restricted to fixed subliquids, but their design was restricted to fixed subliquids structures. Finally, Álvarez-Canchila et al. [12] applied Particle Swarm Optimization (PSO) to optimize synaptic weights, although this approach limited the exploration of complex architectures and excluded neuron-level adaptations. In contrast, our approach directly integrates functional and spatial properties within the GA framework, enabling a more comprehensive exploration of the search space.

5 Evaluation and Communication of Results

The proposed framework will be evaluated through a combination of classification accuracy, separability metrics, and structural complexity measures. Benchmark tasks will include synthetic datasets such as the Frequency Recognition (FR) and Pattern Recognition (PR) tasks [13], as well as selected event-based benchmarks, such as the Neuromorphic MNIST (N-MNIST) [14], the Free Spoken Digit Dataset (FSDD) [15], and the DVS Gesture Dataset [16]. Comparative analyses and statistical tests will be carried out against SOTA methods, ensuring a fair and robust experimental design.

Results will be disseminated through specialized conferences with a rigorous peer review and indexed proceedings. To date, one publication has already been accepted [17] that fulfills these criteria, and another is currently under review. In addition, manuscripts will be submitted to peer-reviewed journals, with one article already under revision. Finally, science communication articles are also planned to engage a broader audience.

6 Preliminary Results and Future Work

In the aforementioned work [17], we introduced the proposed encoding scheme. The genotype was represented with two parallel chromosomes: per-neuron configurations and 2D spatial positions, with connection probabilities and synaptic weights scaled by the distance between neurons. The search space is, therefore, constrained by specific physiological ranges for neuron parameters and spatial limitations, as well as rules to maintain stability and biological plausibility. Three variants were tested: E (full scheme), E1 (only configurations), and E2 (only positions). Results showed that the full scheme matched or outperformed E1, while E2 offered no measurable gains (Fig. 2).

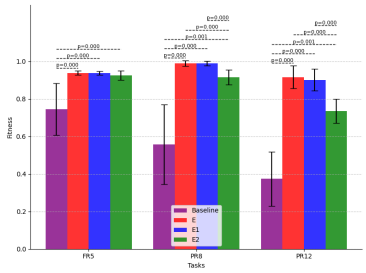


Fig. 2. Encoding comparison

In a subsequent study, now under review, we extended the evaluation to the GA itself. The method was tested against the proposals from [13] and [12]. Preliminary analysis suggests significant improvements in accuracy and generalization on FR and PR synthetic data, while achieving competitive results on real datasets such as N-MNIST and FSDD using smaller liquids.

Future efforts will focus on extending the evaluation of the proposed framework in four specific directions. First, we need to improve the encoding proposal based on the conclusions obtained in [17], while also scaling experimentation to other consider datasets, such as the DVS Gesture. Second, we plan to expand the structural and dynamical analysis of evolved liquids, taking into account their characteristics. Third, we want to include complexity and resource efficiency metrics to provide a more comprehensive characterization of the evolved LSMs beyond accuracy-based evaluations. Fourth, we intend to explore potential applications in domains that benefit from efficient temporal processing.

Finally, Fig. 3 shows the research schedule. The project is currently in the fifth semester of the PhD program, with goals achieved as planned. A pre-doctoral examination is scheduled for the end of this semester, and the dissertation defense is projected for August 2027.

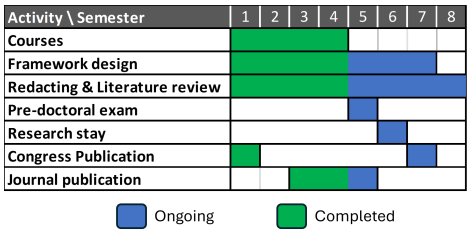


Fig. 3. Research schedule

Acknowledgments. This research project is supported by the Mexican Ministry of Science, Humanities, Technology, and Innovation (SECIHTI for Secretaría de Ciencia,

Humanidades, Tecnología e Innovación) in the form of a scholarship to pursue graduate studies at the University of Veracruz.

References

1. Maass, W. (1997). Networks of spiking neurons: the third generation of neural network models. Neural networks, 10(9), 1659-1671. https://doi.org/10.1016/S0893-6080(97)00011-7
2. Deng, B., Fan, Y., Wang, J., & Yang, S. (2023). Auditory perception architecture with spiking neural network and implementation on FPGA. Neural Networks, 165, 31-42. https://doi.org/10.1016/j.neunet.2023.05.026
3. Baek, S., & Lee, J. (2024). SNN and sound: a comprehensive review of spiking neural networks in sound. Biomedical Engineering Letters, 14(5), 981-991. https://doi.org/10.1007/s13534-024-00406-y
4. Aydin, A., Gehrig, M., Gehrig, D., & Scaramuzza, D. (2024). A hybrid ANN-SNN architecture for low-power and low-latency visual perception. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 5701-5711). https://doi.org/10.1109/CVPRW63382.2024.00579
5. Hussaini, S., Milford, M., & Fischer, T. (2024). Applications of spiking neural networks in visual place recognition. IEEE Transactions on Robotics. https://doi.org/10.1109/TRO.2024.3508053
6. Stein, R. B. (1967). Some models of neuronal variability. Biophysical journal, 7(1), 37-68. https://doi.org/10.1016/S0006-3495(67)86574-3
7. Maass, W., Natschläger, T., & Markram, H. (2002). Real-time computing without stable states: A new framework for neural computation based on perturbations. Neural computation, 14(11), 2531-2560. https://doi.org/10.1162/089976602760407955
8. Norton, D., & Ventura, D. (2006). Preparing more effective liquid state machines using hebbian learning. In The 2006 IEEE International Joint Conference on Neural Network Proceedings (pp. 4243-4248). IEEE. https://doi.org/10.1109/IJCNN.2006.246996
9. Yaqoob, M., & Wróbel, B. (2017). Very small spiking neural networks evolved to recognize a pattern in a continuous input stream. In 2017 IEEE Symposium series on computational intelligence (SSCI) (pp. 1-8). IEEE. https://doi.org/10.1109/SSCI.2017.8285420
10. Tian, S., Qu, L., Wang, L., Hu, K., Li, N., & Xu, W. (2021). A neural architecture search based framework for liquid state machine design. Neurocomputing, 443, 174-182. https://doi.org/10.1016/j.neucom.2021.02.076
11. Zhou, Y., Jin, Y., Sun, Y., & Ding, J. (2023). Surrogate-assisted cooperative co-evolutionary reservoir architecture search for liquid state machines. IEEE Transactions on Emerging Topics in Computational Intelligence, 7(5), 1484-1498. https://doi.org/10.1109/TETCI.2022.3228538
12. Alvarez-Canchila, O. I., Espinal, A., Sotelo-Figueroa, M. A., Soria-Alcaraz, J. A., & Rostro-Gonzalez, H. (2024). Enhancing liquid state machine classification through reservoir separability optimization using swarm intelligence and multitask learning. IEEE Access. https://doi.org/10.1109/ACCESS.2024.3510459
13. Norton, D., & Ventura, D. (2010). Improving liquid state machines through iterative refinement of the reservoir. Neurocomputing, 73(16-18), 2893-2904. https://doi.org/10.1016/j.neucom.2010.08.005

14. Orchard, G., Jayawant, A., Cohen, G. K., & Thakor, N. (2015). Converting static image datasets to spiking neuromorphic datasets using saccades. Frontiers in neuroscience, 9, 437.
15. Jackson, Z. Free Spoken Digit Dataset (FSDD). https://github.com/Jakobovski/free-spoken-digit-dataset, 2018. Accessed: 2025-06-02.
16. Amir, A., Taba, B., Berg, D., Melano, T., McKinstry, J., Di Nolfo, C., ... & Modha, D. (2017). A low power, fully event-based gesture recognition system. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 7243-7252). https://doi.org/10.1109/CVPR.2017.781
17. López-Herrera, C. A., Acosta-Mesa, H. G., & Mezura-Montes, E. (2025). Evaluating Encoding of Neuron Configuration and Position in Neuroevolution of Liquid State Machines. In Proceedings of the Genetic and Evolutionary Computation Conference Companion (pp. 2384-2388). https://doi.org/10.1145/3712255.3734330

Multi-objective Evolutionary Neural Architecture Search for the design of Convolutional Neural Networks considering explainability mechanisms

Jesús-Arnulfo Barradas-Palmeros, Efrén Mezura-Montes, and Héctor-Gabriel Acosta-Mesa

Artificial Intelligence Research Institute, Universidad Veracruzana, Xalapa, Mexico.
zs23000652@estudiantes.uv.mx, emezura@uv.mx, heacosta@uv.mx

Abstract. This research project aims to automate the design process of intrinsically interpretable Convolutional Neural Networks (CNNs). The Proto-Concepts model is considered the basis architecture to be evolved, and it provides insights into the decision process of CNNs following a 'this looks like those' rationale. The Deep Genetic Algorithm (DeepGA) is adapted for the design of Proto-Concepts, resulting in Proto-DeepGA, which follows a multi-objective optimization process of Evolutionary Neural Architecture Search (ENAS). Traditionally, Proto-Concepts employs standard architectures from the literature, whereas using Proto-DeepGA allows the creation of one tailored for the application. Therefore, specific design criteria, such as network performance and complexity, can be incorporated in the search process. Proto-DeepGA is currently in its pilot testing phase, utilizing a reduced dataset, and extensive experimentation is expected. In addition, studies on the effects of using cost reduction mechanisms and varying the parent selection strategies in ENAS have been reported in one journal paper and two conference papers. The work is currently in the fifth semester of the doctoral program.

Keywords: Evolutionary Neural Architecture Search · Explainable Artificial Intelligence · Convolutional Neural Networks

1 Introduction

Convolutional Neural Networks (CNNs) are one of the most successful Deep Learning (DL) models for computer vision tasks [7,9]. Nonetheless, two of the problems that CNNs face are their opacity and the complexity of their design process, which requires in-depth expertise. Considered as black-box models, CNNs only produce an output, and the decision process remains hidden. In this regard, various techniques have been proposed to illuminate the internal workings of the model, aligning with the trend towards explainable Artificial Intelligence. As presented in [1], efforts to add interpretability to a model can be classified into post-hoc and intrinsic. A post-hoc explainer is an additional model that is added to the main model to generate an explanation. On the contrary, intrinsic interpretability is achieved through the model's inherent characteristics, which

enable the generation of explanations. Efforts in the latter are presented in [6], where the Prototypical-Part Network (ProtoPNet) is presented, involving the addition of a prototype layer to the CNN architecture that allows the generation of explanations in the form of "this looks like that". Later, Proto-Concepts [12] extended the work of the ProtoPNet following a "this looks like those" approach.

Regarding CNNs design, Neural Architecture Search (NAS) has emerged as an approach to automate the process of defining a CNN architecture and its hyperparameters [17]. As seen in [20], NAS is based on three elements: search space, search strategy, and validation strategy. The first includes all the possible network architectures and depends on the network encoding. The second is how the search space is traversed. Finally, the third aspect is the mechanism used for estimating an architecture's quality, which has a direct impact on the computational cost. NAS typically requires training and evaluating several architectures. In the search strategy, Evolutionary Computation methods have been successfully applied, consolidating the area of Evolutionary NAS (ENAS) [10]. The Deep Genetic Algorithm (DeepGA) was proposed in [21] for the automatic design of CNNs. DeepGA considers two design criteria: the network's accuracy and its complexity, measured as the number of trainable parameters.

In this research project, the adaptation of DeepGA into Proto-DeepGA is proposed for the design of the Proto-Concepts architecture, an intrinsically explainable CNN model. By considering the network's performance and complexity as design criteria, it is expected to automate the process of creating intrinsically explainable CNNs tailored to a specific application. Currently, the classical approach for utilizing the ProtoPNet and Proto-Concepts models involves testing a series of handcrafted architectures and selecting the one that achieves the highest performance. Therefore, the use of ENAS is a promising area for exploration. Additionally, the automated design of intrinsically interpretable CNNs is crucial for their use in sensitive applications, such as healthcare, where enhancing trust in the model is expected [16].

2 Hypothesis

Using Multi-objective ENAS, which considers both CNN accuracy and complexity, it is possible to evolve prototypical part network architectures with competitive classification accuracy and reduced complexity, as measured by the number of network parameters, compared to state-of-the-art methods. The network is expected to provide users with an interpretable model that offers insights into its decision-making process through class prototypes.

3 Research Objectives

3.1 Main objective

To develop a multi-objective evolutionary method for the automatic design of intrinsically explainable CNN architectures for the image classification task. The architecture's performance and complexity are considered as objectives.

3.2 Specific objectives

The following specific objectives are considered to achieve the stated main objective.

- Review the state-of-the-art approaches in explainable AI for CNNs, with emphasis on the Prototypical Part Networks proposals and applications. Additionally, review the state-of-the-art NAS approaches for CNN design. The review will provide the conceptual foundations to bridge the gap between those two concepts.
- Extend DeepGA to automatically evolve CNN architectures with intrinsic explainability mechanisms, considering the network's accuracy performance and complexity as objectives in a multi-objective ENAS framework.
- Develop and implement strategies to reduce the computational cost of DeepGA without compromising the quality of the solutions. The number of evaluations and computational time of the method will be used as indicators.
- Analyze, and implement techniques to select representative solutions from the Pareto front obtained with the multi-objective ENAS algorithm, balancing the trade-offs between the objectives.
- Evaluate the performance of the proposed ENAS algorithm using multi-objective performance indicators, such as the hypervolume indicator, to assess the Pareto front quality.
- Compare the evolved architecture of the intrinsically explainable CNN with the state-of-the-art manually designed models in terms of classification accuracy and network complexity.

4 Expected contributions and related work

The primary expected contribution of this research project is the design of Proto-DeepGA, an ENAS algorithm that constructs intrinsically explainable CNNs tailored to specific applications. DeepGA has been applied in various domains, including medical [11], automotive [22,23], and agricultural [13] applications. On the other hand, variants of the ProtoPNet [6], such as ProtoPool [18], ProtoTree [14], ProtoPShare [19], and Proto-Concepts [12], all share the design pipeline of testing a variety of CNN architectures for their backbone. Therefore, a gap in the literature is revealed, where the automation of designing prototypical-part networks has not been explored.

5 Evaluation and communication of findings

The Prototypical-part networks have been commonly evaluated in the literature using the Caltech-UCSD Birds-200-2011 dataset [24], which contains 11,788 images of 200 bird breeds. Consequently, Proto-DeepGA will be evaluated using the aforementioned dataset. Additionally, statistical tests will be employed to evaluate the significance of the results in comparison with state-of-the-art manually

designed networks. The metrics considered for comparison include accuracy and network complexity. Regarding the process of DeepGA adaptation, the CIFAR-100 and CIFAR-10 datasets [8] will be used, as they are considered benchmarks in image classification tasks. Additionally, the Fashion-MNIST dataset [25] is also considered.

For the dissemination of results, findings will be presented at specialized conferences with a rigorous peer-review process and indexed proceedings, as has been done with the published papers ([2] and [3]) derived from the research proposal. Similarly, papers will be submitted to peer-reviewed journals, as in the case of [4]. More details of the aforementioned papers will be provided in the following subsection. In addition, science communication articles are planned targeting a broader audience. An example of the previous is the article entitled 'A Treasure Map Without an X: Evolving Neural Network Architecture for Computer Vision' (original Spanish title 'Un mapa del tesoro sin equis: evolucionando arquitecturas de redes neuronales para visión por computadora') [5] published on the Komputer Sapiens magazine. Komputer Sapiens is auspicated by the Mexican Society on Artificial Intelligence.

6 Results to date and future directions

This section summarizes the results obtained to date and outlines the next steps planned for the research project. An initial step was to reduce the computational cost associated with DeepGA, one of the main drawbacks of NAS. In [2], population memory and early stopping were added to DeepGA. The former avoids repeated evaluations, and the latter reduces the number of training epochs required for evaluation. The results were extended in [4], achieving competitive performance of the algorithm with a time reduction ranging from 46.0% to 64.5% when five epochs are used for evaluation during the search process instead of using 10. The results are presented in Fig. 1. Additionally, the population memory mechanism avoided 45% of the evaluations on average.

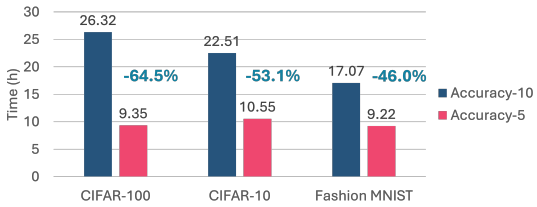


Fig. 1. Results of the computational cost reduction in DeepGA presented in [4]. The accuracy-10 and accuracy-5 configurations represent the number of epochs used for an evaluation during search.

Another study was conducted in [3] regarding the effects of parent selection in DeepGA since early convergence was detected in some of the algorithm runs. The

selection pressure was varied in relation to the tournament size in the Stochastic Tournament (ST) mechanism used in DeepGA, and the Controlled Selection (CS) from [15] was tested. The results showed that CS avoided early convergence, improved population diversity, and allowed the method to select less complex networks with similar performance to those obtained with the ST configuration.

Proto-DeepGA is currently undergoing pilot tests, where the Proto-Concepts architecture is automatically designed for a fraction of the Birds dataset, following a single-objective optimization process. In this case, 10 of the classes were selected for a proof of concept. Proto-DeepGA requires an adaptation of the encoding and operators of the genetic algorithm, as it only considers convolutional layers. The fully connected layers blocks from the individuals are substituted for the prototypical layer used in Proto-Concepts. In addition, the training process was modified since the prototypical-part networks require three stages: warm-up, joint, and fine-tuning. Considering the same fraction of the data, the classical version of DeepGA achieved 62.14% of accuracy, whereas Proto-DeepGA achieved 68.31%.

Future work includes extending the pilot tests of Proto-DeepGA to the complete Birds dataset and comparing the findings with the state-of-the-art manually designed architectures. Additionally, the multi-objective approach will be adopted. A potential long-term plan is the application of Proto-DeepGA in various domains. Other research directions include a deeper study of the effects of the mechanisms from the genetic algorithm, such as replacement, which is conducted in an elitist form. The analysis of the effects of using data augmentation in the search process of the ENAS algorithm is a proposed direction for future research. Fig. 2 presents the research timeline with the remaining stages of the project. The project has now reached the fifth semester of the doctoral studies, and the PhD defense is expected in August 2027.

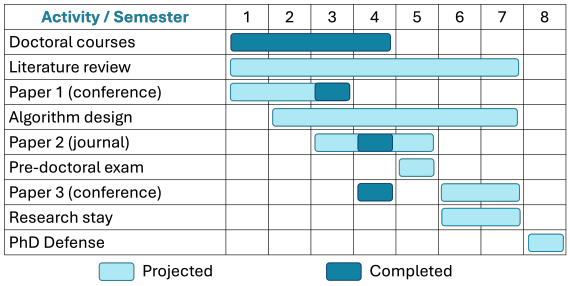


Fig. 2. Research timeline.

Acknowledgments. This research project is supported by the Mexican Ministry of Science, Humanities, Technology, and Innovation (SECIHTI for Secretaría de Ciencia, Humanidades, Tecnología e Innovación) in the form of a scholarship to pursue graduate studies at the University of Veracruz.

References

1. Ali, S., Abuhmed, T., El-Sappagh, S., Muhammad, K., Alonso-Moral, J.M., Confalonieri, R., Guidotti, R., Del Ser, J., Díaz-Rodríguez, N., Herrera, F.: Explainable artificial intelligence (xai): What we know and what is left to attain trustworthy artificial intelligence. Information Fusion **99**, 101805 (2023). https://doi.org/https://doi.org/10.1016/j.inffus.2023.101805
2. Barradas-Palmeros, J.A., López-Herrera, C.A., Acosta-Mesa, H.G., Mezura-Montes, E.: Efficient neural architecture search: Computational cost reduction mechanisms in deepga. In: Martínez-Villaseñor, L., Ochoa-Ruiz, G., Montes Rivera, M., Barrón-Estrada, M.L., Acosta-Mesa, H.G. (eds.) Advances in Computational Intelligence. MICAI 2024 International Workshops. pp. 125–134. Springer Nature Switzerland, Cham (2025). https://doi.org/10.1007/978-3-031-83882-8_12
3. Barradas-Palmeros, J.A., López-Herrera, C.A., Quiroz-Castellanos, M., Mezura-Montes, E., Acosta-Mesa, H.G.: Comparative analysis of parent selection strategies in deepga for neural architecture search (accepted). In: 2023 IEEE Latin American Conference on Computational Intelligence (LA-CCI). pp. 1–6 (2025)
4. Barradas-Palmeros, J.A., López-Herrera, C.A., Mezura-Montes, E., Acosta-Mesa, H.G., López-Lobato, A.L.: Testing neural architecture search efficient evaluation methods in deepga. Mathematical and Computational Applications **30**(4) (2025). https://doi.org/10.3390/mca30040074
5. Barradas-Palmeros Jesús-Arnulfo, Mezura-Montes Efrén, A.M.H.G.: Un mapa del tesoro sin equis: evolucionando arquitecturas de redes neuronales para visión por computadora. Komputer Sapiens **2** (17), mayo-agosto 2025
6. Chen, C., Li, O., Tao, C., Barnett, A.J., Su, J., Rudin, C.: This Looks like That: Deep Learning for Interpretable Image Recognition. Curran Associates Inc., Red Hook, NY, USA (2019)
7. Krichen, M.: Convolutional neural networks: A survey. Computers **12**(8) (2023). https://doi.org/10.3390/computers12080151
8. Krizhevsky, A.: Learning multiple layers of features from tiny images. Tech. rep., Computer Science Department, University of Toronto (2009), https://www.cs.toronto.edu/~kriz/cifar.html
9. Li, Z., Liu, F., Yang, W., Peng, S., Zhou, J.: A survey of convolutional neural networks: Analysis, applications, and prospects. IEEE Transactions on Neural Networks and Learning Systems **33**(12), 6999–7019 (2022). https://doi.org/10.1109/TNNLS.2021.3084827
10. Liu, Y., Sun, Y., Xue, B., Zhang, M., Yen, G.G., Tan, K.C.: A survey on evolutionary neural architecture search. IEEE Transactions on Neural Networks and Learning Systems **34**(2), 550–570 (2023). https://doi.org/10.1109/TNNLS.2021.3100554
11. Llaguno-Roque, J.L., Barrientos-Martínez, R.E., Acosta-Mesa, H.G., Romo-González, T., Mezura-Montes, E.: Neuroevolution of convolutional neural networks for breast cancer diagnosis using western blot strips. Mathematical and Computational Applications **28**(3) (2023). https://doi.org/10.3390/mca28030072
12. Ma, C., Zhao, B., Chen, C., Rudin, C.: This looks like those: Illuminating prototypical concepts using multiple visualizations. In: Oh, A., Naumann, T., Globerson, A., Saenko, K., Hardt, M., Levine, S. (eds.) Advances in Neural Information Processing Systems. vol. 36, pp. 39212–39235. Curran Associates, Inc. (2023)
13. Morales-Reyes, J.L., Aquino-Bolaños, E.N., Acosta-Mesa, H.G., Márquez-Grajales, A.: Estimation of anthocyanins in heterogeneous and homogeneous bean landraces using probabilistic colorimetric representation with a neuroevolutionary approach.

Mathematical and Computational Applications **29**(4) (2024). https://doi.org/10.3390/mca29040068
14. Nauta, M., van Bree, R., Seifert, C.: Neural prototype trees for interpretable fine-grained image recognition. In: 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 14928–14938 (2021). https://doi.org/10.1109/CVPR46437.2021.01469
15. Quiroz-Castellanos, M., Cruz-Reyes, L., Torres-Jimenez, J., Gómez S., C., Huacuja, H.J.F., Alvim, A.C.: A grouping genetic algorithm with controlled gene transmission for the bin packing problem. Computers & Operations Research **55**, 52–64 (2015). https://doi.org/https://doi.org/10.1016/j.cor.2014.10.010
16. Rawal, A., McCoy, J., Rawat, D.B., Sadler, B.M., Amant, R.S.: Recent advances in trustworthy explainable artificial intelligence: Status, challenges, and perspectives. IEEE Transactions on Artificial Intelligence **3**(6), 852–866 (2022). https://doi.org/10.1109/TAI.2021.3133846
17. Ren, P., Xiao, Y., Chang, X., Huang, P.y., Li, Z., Chen, X., Wang, X.: A comprehensive survey of neural architecture search: Challenges and solutions. ACM Comput. Surv. **54**(4) (may 2021). https://doi.org/10.1145/3447582
18. Rymarczyk, D., Struski, Ł., Górszczak, M., Lewandowska, K., Tabor, J., Zieliński, B.: Interpretable image classification with differentiable prototypes assignment. In: Avidan, S., Brostow, G., Cissé, M., Farinella, G.M., Hassner, T. (eds.) Computer Vision – ECCV 2022. pp. 351–368. Springer Nature Switzerland, Cham (2022)
19. Rymarczyk, D., Struski, L., Tabor, J., Zieliński, B.: Protopshare: Prototypical parts sharing for similarity discovery in interpretable image classification. In: Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining. p. 1420–1430. KDD '21, Association for Computing Machinery, New York, NY, USA (2021). https://doi.org/10.1145/3447548.3467245
20. Salmani Pour Avval, S., Eskue, N.D., Groves, R.M., Yaghoubi, V.: Systematic review on neural architecture search. Artificial Intelligence Review **58**(3), 73 (Jan 2025). https://doi.org/10.1007/s10462-024-11058-w
21. Vargas-Hákim, G.A., Mezura-Montes, E., Acosta-Mesa, H.G.: Hybrid encodings for neuroevolution of convolutional neural networks: a case study. In: Proceedings of the Genetic and Evolutionary Computation Conference Companion. p. 1762–1770. GECCO '21, Association for Computing Machinery, New York, NY, USA (2021). https://doi.org/10.1145/3449726.3463133
22. Velazco-Muñoz, J.D., Acosta-Mesa, H.G., Mezura-Montes, E.: Reducing parameters by neuroevolution in cnn for steering angle estimation. In: Mezura-Montes, E., Acosta-Mesa, H.G., Carrasco-Ochoa, J.A., Martínez-Trinidad, J.F., Olvera-López, J.A. (eds.) Pattern Recognition. pp. 377–386. Springer Nature Switzerland, Cham (2024)
23. Vázquez-Santiago, D.I., Acosta-Mesa, H.G., Mezura-Montes, E.: Vehicle make and model recognition as an open-set recognition problem and new class discovery. Mathematical and Computational Applications **28**(4) (2023). https://doi.org/10.3390/mca28040080
24. Wah, C., Branson, S., Welinder, P., Perona, P., Belongie, S.: The caltech-ucsd birds-200-2011 dataset. Tech. Rep. CNS-TR-2011-001, California Institute of Technology (2011)
25. Xiao, H., Rasul, K., Vollgraf, R.: Fashion-mnist: a novel image dataset for benchmarking machine learning algorithms (2017), https://arxiv.org/abs/1708.07747

Identificación de biomarcadores en señales de electroencefalogramas para el apoyo al diagnóstico del trastorno depresivo mediante inteligencia artificial

Alejandro Antonio Torres García[1], Juan Pablo Tiburcio Pérez[1], and Luis Villaseñor Pineda[1]

Instituto Nacional de Astrofísica Óptica y Electrónica, MX
juanp.tiburciop@inaoe.mx

Abstract. El Trastorno Depresivo Mayor (TDM) representa una de las principales causas de discapacidad a nivel mundial, y su diagnóstico clínico actual depende de métodos subjetivos que presentan baja precisión. En este estudio se propone un enfoque de apoyo al diagnóstico utilizando señales EEG en reposo, caracterizadas mediante la Transformada Wavelet Continua y el análisis espectral por bandas de frecuencia. Se extrajeron métricas estadísticas relevantes y se evaluaron 34 modelos de clasificación supervisada en MATLAB, destacando Linear Discriminant y Binary GLM Logistic Regression por su capacidad de generalización (89.0% de precisión en datos no vistos). Aunque la clasificación se realizó a nivel de instancia, se exploró su extrapolación a nivel paciente mediante umbrales de confianza. Los resultados sugieren que el uso combinado de EEG e inteligencia artificial constituye una herramienta prometedora para el desarrollo de sistemas objetivos y eficientes en el diagnóstico de la depresión.

Keywords: Electroencefalografía · Depresión · Aprendizaje automático · clasificación supervisada.

1 Problema a resolver

1.1 Planteamiento del problema

La depresión, o *trastorno depresivo mayor (TDM)*, afecta al 5% de los adultos a nivel mundial [6]. Se caracteriza por tristeza persistente, pérdida de interés o anhedonia, y puede llevar a ideación o conductas suicidas, siendo proyectada como la principal causa de carga de enfermedad en 2030 [5][6].

El diagnóstico actual se basa en entrevistas clínicas y cuestionarios como HAMD y BDI [4][6], los cuales son subjetivos y presentan baja precisión diagnóstica (47.3% según metaanálisis) [6]. Además, la superposición de síntomas con otros trastornos, como la esquizofrenia, genera diagnósticos erróneos [4].

Estudios con EEG han mostrado anomalías en TDM:

- Alteraciones en bandas delta, theta, alfa y beta, con asimetría alfa frontal destacada [6][8][3].
- Alteraciones topológicas en conectividad cerebral mediante teoría de grafos [2][3].
- Cambios en PREs, como FRN y Rew-P, asociados con sensibilidad al castigo y anhedonia [1][4].

Para esta investigación se utilizó la base de datos pública *PREDICT*, proporcionada por la Universidad de Nuevo México, con registros EEG de pacientes clasificados según su puntaje en el cuestionario BDI-II. De esta base se seleccionaron 5 pacientes control (promedio BDI-II: 1.4, desviación estándar: 1.35), 5 pacientes con depresión (promedio BDI-II: 20.4, desviación estándar: 5.68), y 2 pacientes adicionales para prueba: uno de control (puntaje: 5) y uno con depresión (puntaje: 19).

1.2 Justificación

Las limitaciones de los métodos tradicionales evidencian la necesidad de biomarcadores objetivos para un diagnóstico más fiable [6]. Incluso síntomas subclínicos se asocian con deterioro cognitivo y déficits en el control ejecutivo [1].

La EEG es una técnica no invasiva, de bajo costo y con alta resolución temporal [2], que ha demostrado detectar alteraciones relacionadas con la depresión y es candidata para biomarcadores clínicos [4]. Sin embargo, el EEG es ruidoso y de alta dimensionalidad, lo que motiva el uso de inteligencia artificial, especialmente aprendizaje profundo, para extraer patrones complejos sin necesidad de características manuales [6].

Entre los enfoques más destacados se encuentran:

- CNN, que extraen características espacio-temporales [8].
- GNN/GCN, que modelan la conectividad cerebral en forma de grafos [2][4][6].
- Transformers, que capturan dependencias de largo alcance mediante autoatención [7][8].

Finalmente, la combinación de características temporales, espaciales y frecuenciales mejora la precisión diagnóstica, destacando la relevancia de los enfoques multimodales en la investigación actual.

2 Objetivos

- **Objetivo general:** Desarrollar y evaluar un sistema de clasificación automática basado en señales EEG en reposo y aprendizaje automático, para contribuir al diagnóstico objetivo del Trastorno Depresivo Mayor.
- **Objetivos específicos:**
  - Preprocesar las señales EEG del conjunto de datos PREDICT mediante técnicas estándar, asegurando su limpieza y segmentación para el análisis.

- Caracterizar las señales EEG a través de la Transformada Wavelet Continua y el análisis del espectro de potencia, extrayendo métricas estadísticas en bandas de frecuencia relevantes.
- Entrenar y validar distintos modelos de clasificación supervisada sobre las características extraídas, utilizando validación cruzada y evaluación con datos no vistos.
- Explorar la viabilidad de la clasificación a nivel de paciente a partir de umbrales de instancias correctamente clasificadas, como aproximación hacia aplicaciones clínicas.

3 Contribuciones esperadas y trabajo relacionado

Las principales contribuciones esperadas de esta investigación son:

– La implementación de un *pipeline* reproducible de preprocesamiento, caracterización multibanda mediante CWT y análisis espectral, aplicable a señales EEG para el diagnóstico de depresión.
– La demostración de que modelos clásicos de aprendizaje automático, como *Linear Discriminant* y *Binary GLM*, alcanzan altos niveles de generalización en datos no vistos, con un rendimiento competitivo frente a enfoques más complejos.
– La exploración de un criterio preliminar de clasificación a nivel paciente, acercando el análisis a una aplicación clínica real.

En relación con el trabajo previo, estudios basados en EEG han mostrado la relevancia de la asimetría alfa frontal, alteraciones en bandas de baja frecuencia y cambios en la conectividad cerebral en pacientes con TDM. Asimismo, enfoques recientes con CNN, GCN y Transformers han demostrado gran potencial en la clasificación automática, aunque con retos en la necesidad de grandes volúmenes de datos y en la interpretabilidad clínica. Frente a ello, esta investigación se posiciona como un aporte intermedio que combina interpretabilidad, eficiencia computacional y precisión diagnóstica.

4 Evaluación y divulgación de resultados

La evaluación de los resultados se realizará en dos etapas principales. En primer lugar, se empleará validación cruzada *k-fold* (con $k = 5$) durante el entrenamiento de los modelos, lo que permitirá estimar su capacidad de generalización a partir de los subconjuntos de entrenamiento y validación. Adicionalmente, se utilizará un conjunto de prueba independiente, compuesto por pacientes no vistos en el entrenamiento, con el fin de valorar de manera objetiva el rendimiento de los clasificadores en condiciones reales.

Las métricas empleadas para la evaluación incluyen la exactitud global, matrices de confusión, curvas ROC y el área bajo la curva (AUC), lo que permitirá analizar tanto el desempeño global como el equilibrio entre sensibilidad y especificidad. Asimismo, se explorará un criterio de clasificación a nivel paciente

basado en la proporción de épocas correctamente clasificadas, con el objetivo de aproximarse a un escenario clínico real.

En cuanto a la divulgación, los resultados serán presentados mediante figuras y tablas que resuman el desempeño de los modelos. Se contempla la difusión de este trabajo en congresos de bioingeniería y neurociencia, así como en revistas académicas enfocadas en neurociencia computacional y aplicaciones de inteligencia artificial en salud. De igual forma, se buscará poner a disposición de la comunidad los códigos de procesamiento y entrenamiento empleados, favoreciendo la reproducibilidad y la comparación con futuros estudios.

5 Resultados alcanzados y líneas de trabajo futuro

Los resultados obtenidos muestran que los modelos *Linear Discriminant* y *Binary GLM Logistic Regression* alcanzaron una precisión del 89.0% en datos no vistos, lo que evidencia su capacidad de generalización y los posiciona como los clasificadores más prometedores en este estudio. El modelo *Cosine KNN*, aunque mostró un desempeño alto en validación cruzada, redujo significativamente su exactitud en el conjunto de prueba, lo que sugiere sobreajuste.

Si bien la clasificación se realizó a nivel de instancia, se exploró un criterio preliminar para clasificación por paciente, basado en un umbral del 90% de épocas correctamente clasificadas. Los resultados obtenidos quedaron ligeramente por debajo de este límite, lo que indica potencial clínico, pero también la necesidad de optimizar el método para escenarios reales.

Desde una perspectiva neurofisiológica, los resultados obtenidos concuerdan con hallazgos previos que reportan alteraciones en la actividad de las bandas alfa y theta en pacientes con depresión, asociadas a disfunciones en la corteza prefrontal y cingulada anterior. Estas regiones están implicadas en el procesamiento emocional y la autorregulación afectiva, lo que respalda la relevancia de las características espectrales empleadas en este estudio.

En cuanto a líneas de trabajo futuro, se plantean las siguientes:

- Incluir a todos los pacientes del conjunto de datos, en especial aquellos con puntajes BDI-II cercanos al umbral clínico, con el fin de mejorar la robustez y generalización del modelo.
- Implementar técnicas de selección de canales y análisis por bandas de frecuencia para identificar las regiones espacio-frecuenciales más relevantes.
- Comparar el rendimiento de modelos clásicos con arquitecturas de aprendizaje profundo como CNN, GCN y Transformers, evaluando tanto precisión como interpretabilidad.
- Validar de manera más amplia la clasificación a nivel paciente, empleando criterios clínicamente relevantes y conjuntos de datos más diversos.

En conjunto, estos resultados refuerzan el potencial de la EEG y el aprendizaje automático como herramientas para el desarrollo de biomarcadores objetivos en el diagnóstico del Trastorno Depresivo Mayor, a la vez que abren oportunidades de mejora y expansión en investigaciones futuras.

6 Anexo

Para consultar el pipeline de pre procesamiento y caracterización CWT+Pwelch disponible en GitHub Pre-processing-and-CWT-Pwelch-Caractherization

References

1. Cavanagh, J. F., Bismark, A. W., Frank, M. J., & Allen, J. J. (2019). Multiple dissociations between comorbid depression and anxiety on reward and punishment processing: Evidence from computationally informed EEG. Computational Psychiatry (Cambridge, Mass.), 3, 1.
2. Chen, T., Guo, Y., Hao, S., & Hong, R. (2022). Exploring self-attention graph pooling with EEG-based topological structure and soft label for depression detection. IEEE transactions on affective computing, 13(4), 2106-2118.
3. Lu, H., You, Z., Guo, Y., & Hu, X. (2024). Mast-gcn: Multi-scale adaptive spatial-temporal graph convolutional network for eeg-based depression recognition. IEEE Transactions on Affective Computing.
4. Peng, D., Liu, W., Luo, Y., Mao, Z., Zheng, W. L., & Lu, B. L. (2023, July). Deep Depression detection with resting-state and cognitive-Task EEG. In 2023 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC) (pp. 1-4). IEEE.
5. Ravan, M., Noroozi, A., Sanchez, M. M., Borden, L., Alam, N., Flor-Henry, P., ... & Hasey, G. (2024). Diagnostic deep learning algorithms that use resting EEG to distinguish major depressive disorder, bipolar disorder, and schizophrenia from each other and from healthy volunteers. Journal of Affective Disorders, 346, 285-298.
6. Wang, Z., Feng, J., Jiang, R., Shi, Y., Li, X., Xue, R., ... & Deng, W. (2022). Automated rest EEG-based diagnosis of depression and schizophrenia using a deep convolutional neural network. IEEE Access, 10, 104472-104485.
7. Xi, Y., Chen, Y., Meng, T., Lan, Z., & Zhang, L. (2025). Depression detection based on the temporal-spatial-frequency feature fusion of EEG. Biomedical Signal Processing and Control, 100, 106930.
8. Yang, K., Hu, Y., Zeng, Y., Tong, L., Gao, Y., Pei, C., ... & Yan, B. (2023). EEG Network Analysis of Depressive Emotion Interference Spatial Cognition Based on a Simulated Robotic Arm Docking Task. Brain Sciences, 14(1), 44.

Modelo predictivo para clasificar la infección por hongo fusarium oxysporum en hojas de tomate por medio de redes neuronales profundas y visión transformers

Astrid A. Torres[1[0000-0003-1815-2953]], Yasmín Hernández [1[000-0002-8842-0899]],

Javier Ortíz Hernández [1[000-0002-6481-1537]]

[1] TecNM/Centro Nacional de Investigación y Desarrollo, Tecnológico, Cuernavaca, México
d24ce113@cenidet.tecnm.mx, *yasmin.hp@cenidet.tecnm.mx, javier.oh@cenidet.tecnm.mx

Abstract. El tomate es uno de los cultivos hortícolas más importantes de México, tanto por su importancia económica como por su papel en la seguridad alimentaria. Sin embargo, enfrenta amenazas fitosanitarias, entre las que destaca la marchitez vascular causada por el hongo *Fusarium oxysporum*. Tradicionalmente, la estimación de la severidad se basa en la inspección visual experta, lo que limita la precisión diagnóstica. El aprendizaje profundo mediante redes neuronales convolucionales y vision transformers, permiten extraer patrones visuales y analizar relaciones globales en imágenes, ofreciendo objetividad en la estimación de severidad foliar, se propone un modelo predictivo para estimar con mayor precisión el grado de severidad de la infección foliar causada por *Fusarium oxysporum* en hojas de tomate. Se construyó un conjunto de datos propio con imágenes recolectadas en condiciones reales de invernadero, etiquetadas según distintos niveles de severidad. Los resultados preliminares indican que el modelo supera las limitaciones de la inspección visual, logrando clasificaciones más consistentes y escalables, con potencial para mejorar el manejo fitosanitario y la sostenibilidad del cultivo de tomate.

Keywords: Fusarium oxysporum, Jitomate, Redes neuronales convolucionales, Transformadores

1 Introducción

El tomate es uno de los cultivos hortícolas más relevantes en México, tanto por su importancia económica como por su papel en la seguridad alimentaria. México, es uno de los principales exportadores de tomate a nivel mundial, (SIAP, 2023). Sin embargo, el cultivo enfrenta múltiples desafíos fitosanitarios, entre los cuales destacan las enfermedades fúngicas, debido a su frecuencia, impacto y dificultad de control (Larios, Pérez & Rodríguez, 2021).

Entre los patógenos se encuentra el hongo Fusarium oxysporum, agente que causa de la marchitez vascular de la planta de tomate. Esta enfermedad afecta las raíces y hojas de la planta, comprometiendo el sistema vascular, lo cual provoca síntomas como marchitez, clorosis, necrosis progresiva y, en casos graves, la muerte de la planta. La marchitez vascular tiene carácter endémico en diversas zonas productoras del país y puede ocasionar pérdidas económicas severas si no se detecta y trata de manera oportuna (Agrios, 2005).

Uno de los principales retos en el manejo de la marchitez vascular es la identificación precisa de la severidad del daño foliar. Los síntomas en las hojas pueden confundirse fácilmente con otras enfermedades o con deficiencias nutricionales. Los métodos tradicionales de evaluación dependen de la observación del experto, lo cual introduce subjetividad y limita el diagnóstico. Esta situación dificulta la aplicación oportuna de tratamientos adecuados, favoreciendo la propagación del patógeno y generando pérdidas económicas a lo largo de la cadena de producción. En los últimos años, diversos trabajos han explorado el uso de herramientas de visión computacional e inteligencia artificial para la detección y clasificación de enfermedades en cultivos.

1.1 Problema a resolver

El problema central que se busca resolver es la identificación precisa de su severidad en las hojas, ya que los síntomas foliares pueden confundirse con deficiencias nutricionales o con otras enfermedades fúngicas. Además, la progresión del daño es gradual y variable, por lo que los métodos tradicionales basados en la observación del experto son limitados. La falta de detección oportuna de severidad tiene consecuencias, no se aplican los tratamientos adecuados a tiempo, se propaga el patógeno en el cultivo y se generan pérdidas económicas que afectan a productores y a toda la cadena agroalimentaria. La importancia de resolver el problema radica en una estimación objetiva y estandarizada de la severidad, para que se apliquen insumos de manera racional, así como la rentabilidad para los productores al reducir pérdidas por la enfermedad, trazabilidad de la enfermedad en un modelo de predicción que apoye a la agricultura en este sentido, el uso de modelos basados en visión por computadora y aprendizaje profundo posibilita procesar grandes volúmenes de imágenes y eliminar el sesgo humano.

Por otro lado, esta investigación busca trabajar con imágenes recolectadas del invernadero ubicado en el Instituto Tecnológico del Altiplano de Tlaxcala, el dataset construido tiene la clasificación de la enfermedad en las hojas de tomate, la investigación contempla el potencial de integración temporal en el análisis del progreso de la enfermedad. De este modo, el modelo se acerca más a la complejidad del escenario agrícola real, aumentando su aplicabilidad en la práctica.

1.2 Objetivo

Desarrollar un modelo predictivo con redes neuronales convolucionales y transformers, que permita clasificar la severidad de la infección foliar causada por el hongo Fusarium oxysporum en hojas de tomate, para proporcionar trazabilidad en el proceso de diagnóstico y ofrecer escalabilidad en su implementación, a fin de apoyar la toma de decisiones oportunas en el manejo fitosanitario del cultivo.

2 Trabajo relacionado

Sharma et al. (2023) – Tomato leaf disease detection using machine learning- cuyo Objetivo es desarrollar un modelo automático para detectar diversas enfermedades en hojas de tomate, se implementaron redes neuronales convolucionales y ResNet50, obteniendo como resultado mejor precisión para la clasificación de la enfermedad.

Lightweight CNN-RNN model for tomato leaf disease detection, el objetivo es implementar un modelo eficiente en tiempo real para detección de enfermedades en tomate, se utiliza una metodología con redes neuronales y redes neuronales recurrentes con baja demanda de cómputo y optimizado para dispositivos de bajo costo, los resultados: Obtuvo resultados comparables a modelos más pesados, mostrando que es viable usar modelos compactos para aplicaciones prácticas en invernaderos.

Predicting yellow mosaic disease severity in yard-long bean using visible imaging coupled with machine learning model, en el artículo se observa un aporte que se diferencia de los enfoques basados en arquitecturas profundas como CNN o ResNet es el trabajo de Dubey et al. (2025), quien desarrollaron un modelo de predicción de la severidad de la enfermedad del mosaico amarillo (Yellow Mosaic Disease, YMD) en yard-long bean utilizando únicamente imágenes RGB en condiciones de campo. A diferencia de los estudios que se centran en la clasificación discreta de la enfermedad, este trabajo buscó estimar cuantitativamente los grados de severidad mediante el cálculo de 45 índices derivados del espectro visible (RGB), entre los que destacaron el RCC (Red Color Composite) y el ExR (Excess Red) por su alta correlación positiva (~0.87) con la severidad, mientras que el índice GRD (Green-Red Difference) mostró correlación negativa (~–0.88).

Para el modelado predictivo, los autores evaluaron nueve algoritmos de aprendizaje automático, entre ellos Random Forest, Cubist, XGBoost, KNN y GBM. Los resultados demostraron que estos métodos alcanzaron valores de ajuste muy sólidos, con R^2 superiores a 0.88 y un índice de concordancia d mayor a 0.96 en fase de validación, siendo Random Forest el modelo más robusto. El estudio concluye que el uso de índices simples de color combinados con técnicas de machine learning clásicas constituye una alternativa y de bajo costo para predecir la severidad de enfermedades foliares, ofreciendo una vía complementaria a los enfoques basados en deep learning.

El análisis de los trabajos mencionados identifica que los enfoques de aprendizaje profundo han mejorado la detección de enfermedades, particularmente mediante redes convolucionales, recientemente, con el uso de arquitecturas emergentes como los Transformers de visión (ViT). Sin embargo, los estudios se han concentrado principalmente en la clasificación binaria (sano/enfermo) o en la detección de diversas enfermedades, mientras que los modelos orientados a la estimación cuantitativa y estandarizada de la severidad en las hojas de tomate es una oportunidad de estudio.

Asimismo, se destaca que la construcción de datasets propios bajo condiciones reales de cultivo es importante para desarrollar modelos aplicables en la práctica agrícola, ya que la mayoría de los avances previos han trabajado con bases de datos comerciales.

En conclusión, los Transformers en esquemas híbridos aparecen como una línea de investigación para obtener modelos precisos, explicables y escalables, capaces de fortalecer el manejo fitosanitario del tomate. Esta línea de trabajo puede contribuir tanto a la reducción de pérdidas económicas como al fortalecimiento de la producción agrícola en la planta de tomate.

3 Modelo predictivo para clasificar la infección por hongo

La investigación aporta la identificación precisa de la severidad del daño foliar causado por Fusarium oxysporum. A diferencia de estudios previos centrados únicamente en la clasificación binaria de la enfermedad (sano/enfermo), este trabajo se enfoca en la estimación cuantitativa y estandarizada de la severidad, integrando técnicas avanzadas de visión por computadora y aprendizaje profundo.

Las contribuciones de este estudio no se limitan al ámbito metodológico, sino que abarcan también la generación de recursos propios, como la construcción de un dataset especializado, y la aplicabilidad práctica en escenarios reales de invernadero y campo abierto. De esta manera, se busca no solo fortalecer el conocimiento científico en el área, también ofrecer herramientas que mejoren la toma de decisiones, en el cultivo de tomate. A continuación, se describen las contribuciones de la investigación.

Construcción de un dataset propio: En este punto se genera un conjunto de datos con imágenes de hojas de tomate recolectadas en invernadero del Instituto Tecnológico del Altiplano de Tlaxcala. Por otro lado, el dataset está etiquetado con diversos grados de severidad de la infección por Fusarium oxysporum, lo cual representa un avance respecto a estudios previos que solo consideran clasificaciones binarias (sano/enfermo). A continuación, se muestran en las **figuras 1 y 2** el invernadero al que se hace referencia.

Fig 1. Plántula de tomate

Fig 2. Invernadero de tomate

Detección de la severidad, en este punto se propone un modelo predictivo capaz de estimar de los distintos grados de severidad (leve, moderado, avanzado, crítico; equivalentes a niveles 1, 2, 3, 4, etc.) del daño foliar causado por Fusarium oxysporum. Esta clasificación detallada permite no solo identificar la presencia de la enfermedad, sino también determinar el nivel de avance de la infección en cada hoja. Este enfoque supera los métodos tradicionales basados en la observación visual, que suelen ser subjetivos y variables entre expertos, y reduce significativamente el margen de error humano. Al incorporar una escala graduada de severidad, el modelo proporciona información más fina para la toma de decisiones.

Por un lado, la extracción de características emplea convoluciones profundas que capturan la morfología y textura de las lesiones en distintos niveles de severidad. Estas características se convierten en vectores representativos que alimentan la siguiente etapa. En paralelo, se integran variables de contexto —como condiciones ambientales o temporales— que refuerzan la capacidad del sistema para inferir relaciones entre factores externos y progresión de la enfermedad. Posteriormente, la capa de predicción se apoya en un esquema de aprendizaje supervisado capaz de estimar no solo la categoría de la enfermedad, sino también su tendencia de avance. En esta fase, se utilizan funciones de pérdida diseñadas para minimizar errores de clasificación y de proyección futura, lo que fortalece la precisión del modelo.

Finalmente, el modelo se valida a través de métricas como precisión, recall, F1-score y error cuadrático medio, asegurando que la predicción no sea únicamente exacta en el momento actual, sino también confiable a mediano plazo. De esta forma, el Capítulo 3 plantea un modelo predictivo robusto, escalable y adaptable, que no solo clasifica, sino que anticipa el impacto de la enfermedad en el cultivo, contribuyendo directamente a la toma de decisiones agrícolas más oportunas y basadas en evidencia. Técnicas de visión por computadora y aprendizaje profundo, aquí el modelo integrará estrategias como segmentación foliar, técnicas de aumento de datos y arquitecturas basadas en transformers de visión (ViT), lo que representa un aporte en el campo de la fitopatología computacional.

Contribución científica y práctica al sector agropecuario, ya que la investigación aporta una herramienta tecnológica a la ciencia agropecuaria mediante la generación de conocimiento y metodologías basado en inteligencia artificial. También se ofrece una herramienta aplicable en escenarios reales de campo, capaz de mejorar la toma de decisiones en la producción de tomate.

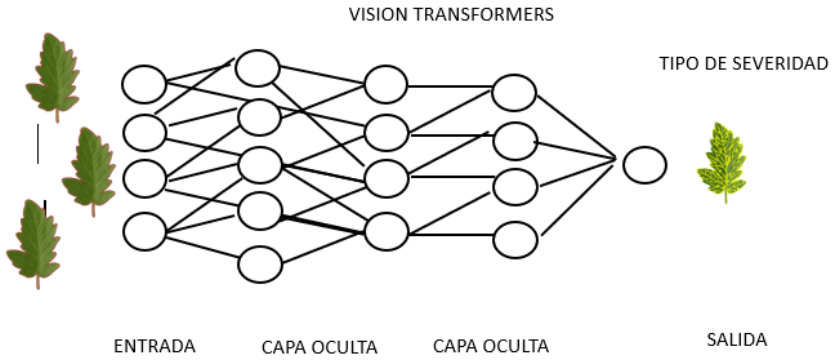


Imagen 3. Esquema de modelo predictivo

4 Resultados esperados

La validación de una investigación de carácter científico requiere que los resultados obtenidos sean evaluados con criterios claros y objetivos. En este caso, la evaluación del modelo propuesto para estimar los grados de severidad del daño foliar causado por Fusarium oxysporum en hojas de tomate debe considerar tanto la calidad de los datos empleados como el desempeño técnico del modelo, así como su aplicabilidad en escenarios reales de producción de tomate.

Dado que la propuesta de investigación integra elementos como la construcción de un dataset propio, la clasificación detallada en niveles de severidad, el uso de arquitecturas avanzadas de inteligencia artificial, la evaluación de resultados no puede limitarse a métricas de precisión; es necesario incorporar medidas que permitan verificar la confiabilidad del modelo y la utilidad práctica de la herramienta en la toma de decisiones agronómicas, a través de los expertos en el área.

En este contexto, los resultados de la investigación se evaluarán a partir de los siguientes puntos: el dataset construido, rendimiento del modelo, comparación con métodos tradicionales, desempeño en condiciones reales y explicabilidad de las predicciones, así como impacto práctico y científico. Estos puntos apoyaran a la evaluación del modelo predictivo ya los expertos en el área en la productividad del cultivo de tomate.

Se espera como resultado de la investigación un modelo predictivo basado en arquitecturas de aprendizaje profundo y Transformers de visión (ViT) que logre clasificar con precisión el nivel de severidad del daño foliar causado por *Fusarium oxysporum* en hojas de tomate. Este modelo apoya las limitaciones de la evaluación visual tradicional, proporcionando estimaciones objetivas, reproducibles y libres de sesgo humano.

Asimismo, se anticipa la construcción de un dataset especializado con imágenes obtenidas en condiciones reales del invernadero ubicado en el Instituto Tecnológico del Altiplano de Tlaxcala, etiquetadas según distintos grados de severidad (leve, moderado, avanzado y crítico). Este recurso no solo fortalecerá la validación del modelo, sino que representará un aporte a la comunidad científica al disponer de datos más cercanos al contexto productivo real, en contraste con bases de datos controladas como *PlantVillage*.

En términos de desempeño, se esperan altos valores de precisión, recall, finalmente, se prevé que la propuesta contribuya al manejo fitosanitario del cultivo mediante la optimización del uso de insumos, la reducción de pérdidas económicas y la toma de decisiones de manera oportuna favoreciendo la producción agrícola.

5 Conclusiones y trabajo a futuro

La marchitez vascular causada por Fusarium oxysporum representa un problema constante para la producción de tomate, no solo por su capacidad destructiva en el sistema vascular de la planta, sino también por las pérdidas económicas que ocasiona en toda la cadena agroalimentaria.

Los métodos tradicionales basados en la inspección visual de expertos resultan insuficientes frente a este problema, ya que están sujetos a la subjetividad y lentitud en la detección, lo que retrasa la aplicación de medidas de control. Este panorama hace evidente la necesidad de soluciones tecnológicas basadas en inteligencia artificial (IA), que permitan un diagnóstico objetivo, rápido y preciso.

Los modelos de aprendizaje profundo, como las redes neuronales convolucionales, identifican y clasifican enfermedades en hojas de tomate. Sin embargo, la mayoría de los estudios se han enfocado en clasificaciones binarias (sano/enfermo) o en la simple detección de patologías, sin avanzar lo suficiente hacia la estimación cuantitativa de la severidad, que es crucial para un manejo fitosanitario efectivo.

En este contexto, la aparición de arquitecturas como los Transformers de visión (ViT), al ofrecer un procesamiento más robusto, explicable y escalable de imágenes agrícolas. La integración de estas arquitecturas con redes neuronales convolucionales, en modelos híbridos constituye línea de investigación de este proyecto, el desarrollo el dataset propio permite entrenar y validar el modelo, superando las limitaciones de bases de datos como PlantVillage.

En conclusión, la construcción de un modelo predictivo con Transformers con un dataset propio permitirían la clasificación precisa de los grados de severidad del daño foliar, optimizando el uso de insumos, reduciendo pérdidas económicas y contribuyendo a una agricultura de la región, la sostenibilidad, eficiencia en la producción de tomate. Como trabajo a fututo se sugiere identificar el grado de severidad en la raíz, tallo y fruto de la planta.

Referencias

1. Agrios, G. N. (2005). Plant pathology (5th ed.). Elsevier Academic Press.
2. Alpaydin, E. (2020). *Introduction to machine learning* (4th ed.). MIT Press.
3. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., … Houlsby, N. (2020). An image is worth 16×16 words: Transformers for image recognition at scale. *International Conference on Learning Representations (ICLR)*.
4. Dubey, A. K., et al. (2025). Predicting yellow mosaic disease severity in yard-long bean using visible imaging coupled with machine learning model. *Scientific Reports*.

5. George Princess, T., & Poovammal, E. (2025). *An efficacious detecting tomato leaf disease using RCOA-based RNN method*. Journal of Advances in Information Technology, 16(1), 49-56.
6. Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.
7. Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems* (Vol. 25, pp. 1097–1105).
8. Larios, J., Pérez, M., & Rodríguez, H. (2021). Impacto de las enfermedades fúngicas en cultivos hortícolas. *Revista Mexicana de Fitopatología, 39*(2), 123–135.
9. Le, A. T. (2024). Tomato disease detection with lightweight recurrent and convolutional neural networks. *Frontiers in Sustainability*.
10. LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature, 521*(7553), 436–444.
11. Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature, 323*(6088), 533–536.
12. Sanida, T., Sideris, A., Sanida, M. V., & Dasygenis, M. (2023). Tomato leaf disease identification via two–stage transfer learning approach. *Smart Agricultural Technology, 5*, 100275.
13. Sharma, R., et al. (2023). Tomato leaf disease detection using machine learning. *CEUR Workshop Proceedings, 3283*
14. SIAP. (2023). *Avances de siembras y cosechas: Tomate rojo (tomate)*. Servicio de Información Agroalimentaria y Pesquera.
15. Szeliski, R. (2022). Computer vision: Algorithms and applications (2nd ed.). Springer.
16. Thanjaivadivel, M., Gobinath, C., Vellingiri, J., Kaliraj, S., & Josephin, J. S. F. (2025). EnConv: Enhanced CNN for leaf disease classification. *Journal of Plant Diseases and Protection, 132*, 32.
17. Zhou, Z.-H. (2021). *Machine learning* (2nd ed.). Springer.