EDUCACIÓN
SECRETARÍA DE EDUCACIÓN PÚBLICA

Instituto Politécnico Nacional
"La Técnica al Servicio de la Patria"

Research in Computing Science

Vol. 154 No. 9
September 2025

Research in Computing Science

***RESEARCH IN COMPUTING SCIENCE**, Año 25, Volumen 154, No. 9, Septiembre de 2025, es una publicación mensual, editada por el Instituto Politécnico Nacional, a través del Centro de Investigación en Computación. Av. Juan de Dios Bátiz S/N, Esq. Av. Miguel Othón de Mendizábal, Col. Nueva Industrial Vallejo, C.P. 07738, Ciudad de México, Tel. 57 29 60 00, ext. 56571. https://www.rcs.cic.ipn.mx. Editor responsable: Dr. Grigori Sidorov. Reserva de Derechos al Uso Exclusivo del Título No. 04-2019-082310242100-203. ISSN: en trámite, otorgado por el Instituto Nacional del Derecho de Autor. Responsable de la última actualización de este número: el Centro de Investigación en Computación, Dr. Grigori Sidorov, Av. Juan de Dios Bátiz S/N, Esq. Av. Miguel Othón de Mendizábal, Col. Nueva Industrial Vallejo, C.P. 07738. Fecha de última modificación 08 de Septiembre de 2025.*

***RESEARCH IN COMPUTING SCIENCE**, Year 25, Volume 154, No. 9, September, 2025, is a monthly publication edited by the National Polytechnic Institute through the Center for Computing Research. Av. Juan de Dios Bátiz S/N, Esq. Miguel Othón de Mendizábal, Nueva Industrial Vallejo, C.P. 07738, Mexico City, Tel. 57 29 60 00, ext. 56571. https://www.rcs.cic.ipn.mx. Editor in charge: Dr. Grigori Sidorov. Reservation of Exclusive Use Rights of Title No. 04-2019-082310242100-203. ISSN: pending, granted by the National Copyright Institute. Responsible for the latest update of this issue: the Computer Research Center, Dr. Grigori Sidorov, Av. Juan de Dios Bátiz S/N, Esq. Av. Miguel Othón de Mendizábal, Col. Nueva Industrial Vallejo, C.P. 07738. Last modified on September, 8, 2025.*

Advances in Artificial Intelligence

Bella Citlali Martínez-Seis (ed.)

Instituto Politécnico Nacional, Centro de Investigación en Computación
México 2025

ISSN: in process

Formerly ISSNs: 1870-4069, 1665-9899

Instituto Politécnico Nacional (IPN)
Centro de Investigación en Computación (CIC)
Av. Juan de Dios Bátiz s/n esq. M. Othón de Mendizábal
Unidad Profesional "Adolfo López Mateos", Zacatenco
07738, México D.F., México

http://www.rcs.cic.ipn.mx
http://www.ipn.mx
http://www.cic.ipn.mx

Indexed in LATINDEX, DBLP and Periodica

Electronic edition

ISSN 1870-4069

Table of Contents

Page

ISSN 1870-4069

Verificación automática de presentaciones usando inteligencia artificial

Raul Bautista Arroyo[1], Jorge de la Calleja[1],
Marco Aurelio Nuño Maganda[2],
María Auxilio Medina Nieto[1]

[1] Universidad Politécnica de Puebla,
Mexico

[2] Universidad Politécnica de Victoria,
Mexico

{raul.bautistaao, maria.medina,
jorge.delacalleja}@uppuebla.edu.mx,
mnunom@upv.edu.mx

Resumen. Este trabajo presenta los primeros resultados de una investigación sobre el desarrollo de una herramienta basada en aprendizaje automático para la verificación automática de presentaciones en formato PDF. Se han realizado experimentaciones con cuatro algoritmos supervisados: Naïve Bayes, árboles de decisión, redes neuronales y Máquinas de Soporte Vectorial (SVM); y se aplicó el algoritmo de Análisis de Componentes Principales (PCA) para extraer los atributos más relevantes y reducir la dimensionalidad. Los datos fueron transformados en matrices de atributos y evaluados utilizando el 100 %, 70 % y 50 % de los atributos. Los resultados preliminares indican que las redes neuronales lograron la mayor precisión y F-measure, incluso con una reducción significativa de atributos. Esta investigación representa un avance hacia la automatización de la evaluación de presentaciones, facilitando la identificación de inconsistencias y mejorando el proceso de revisión documental.

Palabras clave: Aprendizaje automático, verificación de presentaciones, redes neuronales, clasificación de documentos, inteligencia artificial.

Automatic Verification of Presentations Using Artificial Intelligence

Abstract. This work presents the initial results of a research project on the development of a machine learning-based tool for the automatic verification of PDF submissions. Experiments were carried out with four supervised algorithms: Naïve Bayes, decision trees, neural networks, and Support Vector Machines (SVMs); and the Principal Component Analysis (PCA) algorithm was applied to extract the most relevant attributes and reduce dimensionality. The data were transformed into attribute matrices and evaluated using 100%, 70%, and 50% of the

attributes. Preliminary results indicate that the neural networks achieved the highest accuracy and F-measure, even with a significant reduction in attributes. This research represents a step toward automating submission evaluation, facilitating the identification of inconsistencies and improving the document review process.

Keywords: Machine learning, presentation verification, neural networks, document classification, artificial intelligence.

1. Introducción

La inteligencia artificial (IA) ha demostrado ser una herramienta útil para resolver problemas complejos en diversos dominios, incluyendo la medicina, la industria, el comercio y la educación. El aprendizaje automático supervisado ha permitido automatizar tareas que anteriormente requerían intervención humana intensiva como la clasificación, predicción o detección de patrones.

En este contexto, una problemática que puede ser interesante y relevante es la verificación del cumplimiento de lineamientos en presentaciones empresariales y educativas, tarea que comúnmente se realiza de forma manual, con un alto grado de subjetividad y consumo de tiempo. El problema de clasificación abordado en esta investigación consiste en identificar, a partir de cada diapositiva convertida en imagen, si cumple o no con un lineamiento predefinido.

La función objetivo es minimizar el tiempo que una persona requiere para validar el cumplimiento de lineamientos en presentaciones, asegurando una evaluación objetiva y detallada sobre cada diapositiva. Cada presentación es analizada diapositiva por diapositiva, se evalúan elementos visuales como texto, imágenes, tablas y estructuras gráficas.

Los lineamientos establecidos son los siguientes: (1) Cada diapositiva debe presentar información coherente con su título (por ejemplo, si el título dice “Endpoints”, deben mostrarse los endpoints analizados); (2) La primera mitad de la presentación debe enfocarse en describir activos de TI, su relación y su criticidad operativa; (3) La segunda mitad presenta los resultados del análisis de seguridad, describiendo riesgos encontrados, falsos positivos y controles recomendados.

Dentro de las presentaciones de las organizaciones e instituciones, hoy en día existen lineamientos que deben de cumplirse para garantizar que los objetivos y los temas relacionados a los mismos se cumplan, en las Figuras 1.1 y 1.2 muestran algunos ejemplos de los lineamientos que se piden para una presentación de análisis de seguridad de una empresa multinacional, tales como describir el objetivo general de la presentación, los miembros que participaron en el análisis, algunas métricas asociadas a la presentación, el mapeo técnico de las aplicaciones involucradas en el proyecto, las aplicaciones relacionadas y las gestiones de accesos involucradas.

En este trabajo de investigación, en desarrollo, se utilizaron cuatro algoritmos de aprendizaje automático supervisado: Naïve Bayes, árboles de decisión, redes neuronales y máquinas de soporte vectorial (SVM). Estos algoritmos fueron seleccionados por su efectividad comprobada en diversas tareas de clasificación y su capacidad para manejar datos de alta dimensionalidad.

Fig. 1. Lineamientos de presentaciones donde se solicitan objetivos, métricas y colaboradores participantes.

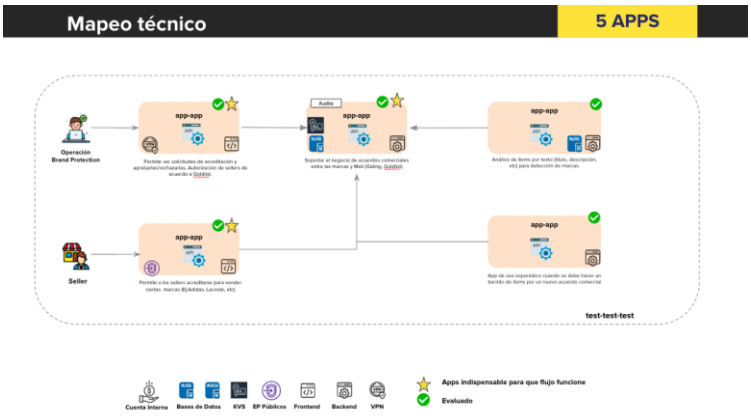

Fig. 2. Lineamientos de presentaciones donde se mapea la relación de las aplicaciones y la importancia de las mismas en un flujo de negocio.

2. Trabajos relacionados

Para identificar investigaciones previas, se realizó una búsqueda en bases de datos como IEEE, Springer, ScienceDirect y MDPI. Enseguida se mencionan de manera breve algunos de los trabajos más relevantes, los cuales han sido clasificados de acuerdo al problema, los algoritmos utilizados y por selección de atributos:

— **Justificación del problema**: Roy y Daniel [3] estudiaron herramientas digitales en la creación y revisión de presentaciones, destacando la necesidad de automatizar procesos de validación. Bueno y Esquivel [4], y Abrancato y Welsh [5] enfatizaron

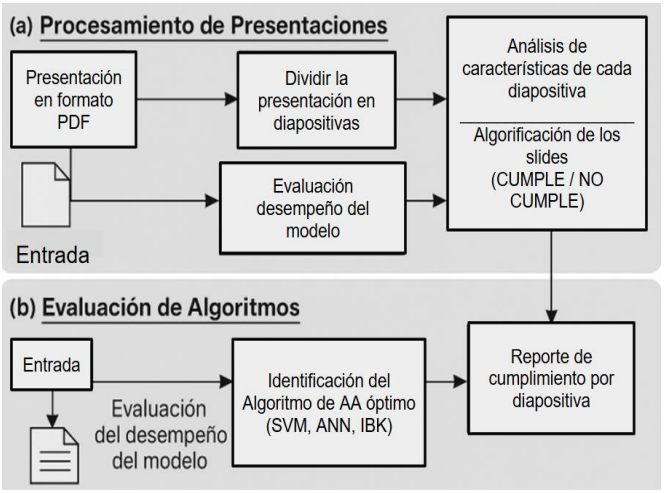

Fig. 1.3. Mapa de solución propuesta para esta investigación donde con ayuda de algoritmos de aprendizaje automático se busca analizar el cumplimiento de los lineamientos en las presentaciones.

la importancia de la digitalización de procesos administrativos, incluyendo la revisión documental.

— **Algoritmos de aprendizaje automático:** Idrissi [6] y Mahadevkar et al. [7] exploraron aplicaciones de redes neuronales y SVM en tareas de validación automática. Hussain [8] destacó la aplicación de modelos de visión por computadora en la Industria 4.0.

— **Selección y reducción de atributos:** Zhou et al. [1] propusieron la selección de características basada en árboles de decisión. Borade y Netak [9] utilizaron SVM para calificar diapositivas, mientras que Chen et al. [11] integraron pistas multimodales para evaluar presentaciones.

3. Metodología y métodos

La figura 1.3 muestra el flujo general del proceso para evaluar presentaciones mediante algoritmos de aprendizaje automático, desde la conversión del archivo original hasta la generación de un reporte con los resultados del análisis. Esto es, las presentaciones se reciben en formato PDF y se convierten en imágenes .png. Posteriormente, se realiza la clasificación binaria de cada diapositiva (cumple/no cumple) mediante algoritmos de aprendizaje automático y evaluando el desempeño con distintas cantidades de atributos usando PCA.

3.1. Etapas de la metodología

Enseguida se describen los pasos que se siguieron en la metodología:

3.1.1. Conversión de la presentación a imágenes

Para asegurar la portabilidad y compatibilidad del contenido, las presentaciones se reciben en formato PDF. Posteriormente, se realiza un proceso de conversión que divide el archivo por diapositiva y transforma cada una en una imagen .png. Esta transformación facilita el procesamiento por parte de los algoritmos de aprendizaje automático.

3.1.2. Clasificación de diapositivas mediante aprendizaje automático

Una vez convertidas en imágenes, las diapositivas son analizadas por algoritmos de aprendizaje automático supervisado. Estos algoritmos —Support Vector Machine (SVM), Redes Neuronales (ANN) e k-vecinos más cercanos (IBK)— tienen como objetivo clasificar cada diapositiva, determinando si cumple o no con los lineamientos de formato previamente definidos.

3.1.3. Validación de cumplimiento de lineamientos

Tras la clasificación, cada diapositiva es evaluada para verificar si cumple con los lineamientos esperados (como estructura, contenido mínimo, uso correcto de elementos visuales, etc.). El resultado de esta validación es binario que indica true si cumple o false si no cumple.

3.1.4. Evaluación comparativa de algoritmos

Se realiza una evaluación comparativa entre los algoritmos utilizados (SVM, ANN, IBK), con el objetivo de identificar cuál de ellos ofrece el mejor rendimiento para la tarea de verificación. La evaluación considera métricas como la precisión, exactitud y capacidad de generalización a nuevas presentaciones.

3.1.5. Generación de reporte automatizado

Finalmente, con base en los resultados del modelo de AA, se genera un reporte automatizado en formato Excel. Este reporte incluye:

— Las diapositivas que cumplen y no cumplen con los lineamientos.

— Detalles específicos de los puntos faltantes por cada diapositiva.

— Un porcentaje de avance o cumplimiento general de la presentación.

La metodología descrita permite automatizar la verificación del cumplimiento de lineamientos en presentaciones digitales mediante el uso de algoritmos de aprendizaje automático y procesamiento de imágenes. Este enfoque no solo reduce la carga de trabajo manual, sino que también proporciona resultados objetivos y reproducibles.

3.2. Algoritmos de aprendizaje automático

En este trabajo se utilizaron tres algoritmos de aprendizaje automático supervisado: Máquinas de Vectores de Soporte (SVM), Redes Neuronales Artificiales (ANN) y k-vecinos más cercanos (IBk). Estos modelos fueron seleccionados por su efectividad en tareas de clasificación y su capacidad para manejar datos con alta dimensionalidad.

— **SVM:** Encuentra el mejor límite de separación entre clases, siendo eficaz en espacios con muchos atributos.

— **ANN:** Aprende patrones complejos mediante una red de nodos conectados, ideal para datos visuales.

— **IBk:** Clasifica según los ejemplos más cercanos, siendo simple y eficaz en problemas bien definidos.

Cada algoritmo fue evaluado utilizando el conjunto de datos original y versiones reducidas mediante PCA, con el objetivo de determinar cuál ofrece la mayor precisión en la verificación automática de presentaciones.

4. Resultados experimentales

Para llevar a cabo los experimentos, se utilizaron implementaciones de algoritmos de aprendizaje automático proporcionadas por la biblioteca scikit-learn en Python. También se emplearon otras herramientas como *Pandas* para el manejo de datos, *Seaborn* y *Matplotlib* para la visualización de resultados, y *Joblib* para guardar los modelos entrenados. El proceso de entrenamiento, evaluación y generación de métricas se realizó en un entorno de experimentación, permitiendo que los experimentos puedan ser replicables por otros investigadores.

El conjunto de datos está compuesto por 648 instancias de 112 presentaciones, esto se logra por medio de la carga de las imágenes de las diapositivas, se convierten a escala de grises y se normalizan a un tamaño fijo de 128x128 píxeles lo cual nos dará como resultado una matriz de 16385 atributos. Posteriormente, cada imagen se convierte en un vector de atributos (pixeles) y se le asigna una etiqueta de clase. Estos vectores y etiquetas se organizan en un *DataFrame* y se guardan en un archivo CSV donde se organizan en un total de 34 clases que corresponden a:

✓OK, cuando una diapositiva cumple con los componentes que caracterizan su lineamiento:

— Detalles caso de abuso exitoso ok

— Fronted relacionados ok

— Mapeo técnico ok

— Diapositiva título hallazgo ok

— Proveedores ok

— Credenciales hardcodeadas ok

— Diapositiva 2 ok

— Imágenes frontend ok

— Análisis overview ok

— Casos de uso y abuso ok

— Access group ok

— Secrets ok

— Objetivo ok

— Diapositiva títulos anexos ok

— Hallazgos ok

— Bases de datos ok

— Diapositiva título resumen ok

— Histórico vulnerabilidades ok

— Flujo feliz ok

— Diapositiva 1 ok

Error, cuando una diapositiva no cumple con los componentes que caracterizan a la diapositiva de su lineamiento

— Diapositiva links documentación error

— Credenciales hardcodeadas error

— Gracias error

— Object storage error

— Diapositivas vacíos error

— Detalles caso de abuso exitoso error

— Access group error

— Bases de datos error

— Imágenes frontend error

— Objetivo error

— Proveedores error

— Kvs error

— Gestión de accesos admin error

— Mapeo técnico error

Los anteriores representan a las clases que ayudan a identificar si las diapositivas cumplen con lineamientos descritos en la Sección 2. Las imágenes de las diapositivas usadas en esta investigación fueron convertidas en vectores de características, dando como resultado una matriz con 16,385 atributos. Debido a la alta dimensionalidad, se aplicó una técnica de reducción de atributos mediante el algoritmo PCA (Análisis de Componentes Principales). Esta técnica permite conservar la mayor cantidad de información posible mientras se reducen los atributos, con el objetivo de disminuir la carga computacional y mejorar el desempeño de los modelos. En este estudio se compararon los resultados utilizando el 100 %, 70 % y 50 % de los atributos originales, seleccionados con base en los niveles de varianza explicada acumulada generados por PCA.

Se evaluaron cuatro algoritmos de aprendizaje automático en distintas configuraciones: Naïve Bayes, Árboles de Decisión, Redes Neuronales y Máquinas de

Tabla 1. Evaluación de algoritmos de aprendizaje automático donde se obtuvo el mejor rendimiento usando diferentes cantidades de atributos.

Algoritmo	Precisión	F-Measure	% Atributos Usados	# Instancias	# Atributos
Naïve Bayes	0.77	0.76	100 %	648	16,385
Naïve Bayes	0.76	0.75	70 %	648	11,469
Naïve Bayes	0.77	0.76	50 %	648	8,192
Árboles de Decisión	0.84	0.80	100 %	648	16,385
Árboles de Decisión	0.76	0.76	70 %	648	11,469
Árboles de Decisión	0.73	0.73	50 %	648	8,192
Redes Neuronales (ANN)	0.86	0.85	100 %	648	16,385
Redes Neuronales (ANN)	0.75	0.76	70 %	648	11,469
Redes Neuronales (ANN)	0.81	0.80	50 %	648	8,192
Máquinas de Vectores de Soporte (SVM)	0.84	0.82	100 %	648	16,385
Máquinas de Vectores de Soporte (SVM)	0.56	0.60	70 %	648	11,469
Máquinas de Vectores de Soporte (SVM)	0.60	0.64	50 %	648	8,192

Tabla 2. Evaluación de Algoritmos de Aprendizaje Automático donde se obtuvo el mejor rendimiento usando diferentes cantidades de atributos.

Algoritmo	Precisión	Medida -F	% de Atributos usados
Naïve Bayes	0.77	0.76	100 %
Árboles de Decisión	0.84	0.80	100 %
Redes Neuronales	0.86	0.85	50 %
SVM	0.84	0.82	70 %

Vectores de Soporte. Cada uno fue probado en tres condiciones diferentes (con 100 %, 70 % y 50 % de atributos), y los resultados de precisión y F-measure fueron promediados. La Tabla 1 muestra los mejores resultados obtenidos por cada algoritmo según el porcentaje de atributos usados:

La Tabla 1 muestra los resultados de las pruebas realizadas sobre los algoritmos de aprendizaje automático evaluados en esta investigación: Naïve Bayes, Árboles de

Tabla 3. Evaluación de redes neuronales.

Algoritmo	Precisión	F-Measure	Porcentaje de atributos Usados	Hiperparámetro
Redes neuronales	0.81	0.80	100%	hidden_layer_sizes=(100,)
	0.78	0.78	70%	
	0.80	0.80	50%	

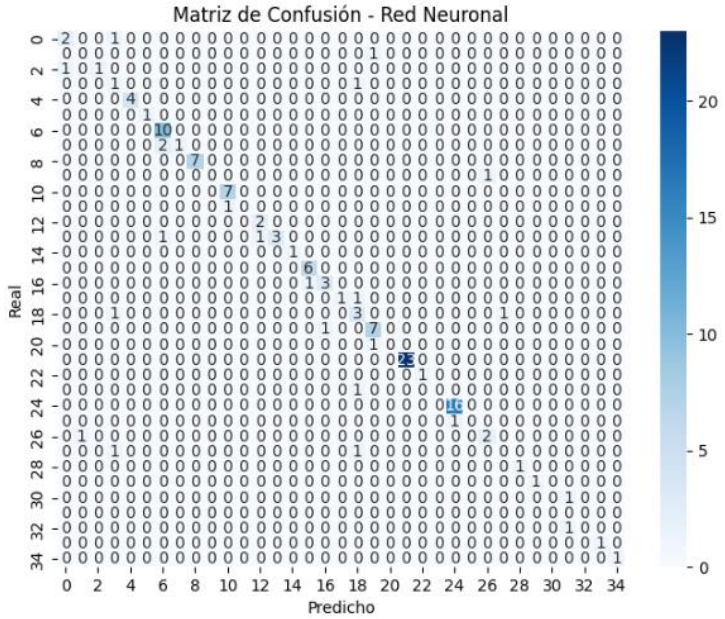

Fig. 4. Matriz de atributos.

Decisión, Redes Neuronales y Máquinas de Vectores de Soporte. Se reportan las métricas de Precisión y F-Measure, el número de instancias y atributos utilizados.

La tabla 2 muestra como a partir de estos resultados, se profundizó el análisis del modelo de Redes Neuronales, ya que fue el que obtuvo la mejor precisión con menor cantidad de atributos (50 %). Se realizaron nuevos experimentos variando el hiper parámetro *hidden_layer_sizes* para observar su impacto en el rendimiento. La configuración que ofreció los mejores resultados fue *hidden_layer_sizes*=(100,), logrando una precisión de 0.80 con solo el 50 % de los atributos, lo cual representa una mejora significativa en términos de eficiencia computacional. La Tabla 3 resume los resultados de estas pruebas adicionales:

El conjunto de experimentos mostrados en la tabla 2 permite concluir que la combinación de reducción de atributos mediante PCA y el ajuste adecuado de hiper

parámetros puede mejorar el rendimiento de los modelos, tanto en precisión como en eficiencia.

La Figura 1.4 muestra la matriz de confusión correspondiente al modelo de Red Neuronal Artificial (ANN) entrenado para verificar automáticamente el cumplimiento de lineamientos en presentaciones. Esta matriz representa el rendimiento del modelo en la clasificación de cada diapositiva dentro de una de las 34 clases posibles, donde cada clase corresponde a un tipo específico de diapositiva o lineamiento evaluado (por ejemplo: “objetivo”, “endpoints”, “credenciales hardcodeadas”, entre otros).

En la matriz, el eje vertical representa las clases reales y el eje horizontal las clases. Los valores en la diagonal indican predicciones correctas (donde la clase real coincide), mientras que los valores fuera de la diagonal reflejan errores de clasificación.

El modelo logra una concentración significativa de valores en la diagonal, lo cual indica alto nivel de precisión en la clasificación multiclase, especialmente considerando la alta dimensionalidad del conjunto de datos (16,385 atributos originales reducidos mediante PCA). La matriz evidencia que el modelo fue capaz de identificar correctamente múltiples clases, incluso bajo el esquema de evaluación binaria por cumplimiento y utilizando reducción de atributos al 50 %, lo cual refuerza su capacidad de generalización y eficiencia en contextos reales.

5. Conclusiones

La evaluación de los algoritmos de aprendizaje automático permitió identificar que las redes neuronales ofrecieron el mejor desempeño en la tarea de verificación de presentaciones, logrando una precisión del 86 % y F-measure del 85 % incluso con una matriz reducida al 50 % de los atributos. Esto resalta su capacidad para generalizar patrones relevantes aun cuando se trabaja con menos información. En función del objetivo planteado, se logró comparar el rendimiento de distintos algoritmos y establecer cuál resulta más adecuado. Aunque el trabajo se encuentra en una etapa experimental, ya se han obtenido avances significativos. Como trabajo futuro, se contempla continuar con la validación del modelo de redes neuronales en entornos reales.

Referencias

1. Zhou, H.F., Zhang, J.W., Lu, Y.: A Feature Selection Algorithm of Decision Trees. Journal of Computer Science and Technology, 35(6), pp. 1301–1315.6 (2020) doi: 10.1016/j.eswa.2020.113842.
2. Recio, J.A.: Técnicas de extracción de características y clasificación en ortoimágenes. Revista de Geografía y Sistemas de Información Geográfica, 1(2), pp. 45–60 (2009)
3. Roy S., Daniel, C.: Introducción a las aplicaciones de presentación. University of Science and Technology Journal, 12(1), pp. 78–85 (2023)
4. Bueno E.H., Esquivel, P.L.: Desarrollo de la gestión documental. Journal of Management and Business Administration, 19(3), pp. 233–250 (2011)
5. Abrancato, A.G., Welsh, V.: Transformación digital y eficiencia: Digitalización de procesos administrativos. Journal of Business Innovation, 8(4), pp. 112–126 (2021)

6. Idrissi, I.E.: Error-Correcting Codes and The Power of Machine Learning. IEEE Transactions on Information Theory, 70(5), pp. 1234–1248 (2024) doi: 10.1109/SITA60746.2023.10373727.
7. Mahadevkar, S.V., Khemani, B., Gurnani, S.: A Review on Machine Learning Styles in Computer Vision. International Journal of Computer Vision, 130(3), pp. 567–589 (2022) doi: 10.1109/ACCESS.2022.3209825.
8. Hussain, M.: Sustainable Machine Vision for Industry 4.0. MDPI Sustainability, 16(2), pp. 345–360 (2024) doi: 10.3390/ai5030064.
9. Borade, J.G., Netak, L.D.: Machine learning techniques for grading of powerpoint diapositivas. In: Proceedings of the International Conference on Machine Learning and Applications, pp. 234–240 (2022) doi: 10.1007/978-3-030-98404-5_1.
10. Borade, J.G., Kiwelekar, A.W., Netak, L.D.: Automated Grading of Powerpoint Presentations Using Latent Semantic Analysis. IEEE Access, 9, pp. 123456–123465 (2021) doi: 10.18280/ria.360215
11. Chen, L., Leong, C.W., Feng, G.: Using Multimodal Cues to Analyze MLA'14 Oral Presentation Quality Corpus: Presentation Delivery and Diapositivas Quality. In: Proceedings of the 2014 ACM Workshop on Multimodal Learning Analytics Workshop and Grand Challenge, pp. 45–52 (2014) doi: doi.org/10.1145/2666633.266664.

ISSN 1870-4069

Mantenimiento predictivo no supervisado: Detección de anomalías en la industria

Jorge Metri-Ojeda[1], Sergio Simanek-Gutierrez[2]

[1] Universidad De Las Américas Puebla, San Andrés Cholula, México

[2] Posgrado CIATEQ A.C., Querétaro, México

jorge.metrioa@udlap.mx, Sergio.simanek.gtz@gmail.com

Resumen. La Gestión de Pronóstico y Salud (PHM), también conocida como Mantenimiento Predictivo (PdM), es actualmente una técnica de vanguardia para evitar grandes pérdidas en las industrias manufactureras. El objetivo de este trabajo es comparar diferentes algoritmos no supervisados para la detección de anomalías como una opción viable para conjuntos de datos industriales no etiquetados. Se utilizó un conjunto de datos del mundo real compuesto por 5 variables (aceleración de vibración, velocidad de vibración y temperatura) de 118 equipos de una instalación de una industria de manufactura. Después de la extracción de características, se aplicó un Análisis de Componentes Principales (PCA) para la reducción de dimensionalidad y se utilizó para entrenar un Bosque de Aislamiento, un modelo de Máquina de Soporte Vectorial de Una Clase (OC-SVM), modelo de Factor de Aislamiento Local (LOF) y Modelos de Mezcla Gaussiana. Los modelos se evaluaron en términos de precisión, con el objetivo de maximizar la detección de 12 fallas registradas como anomalías dentro del conjunto de datos. El modelo con mejor rendimiento fue el GMM con un 91 % de exactitud, seguido por OC-SVM con un 83 % de exactitud. La velocidad media en el eje Z, el valor máximo de la velocidad en el eje X y la desviación estándar de la aceleración en el eje X fueron las variables clave para diferentes tipos de fallas. Este trabajo demuestra el uso exitoso del aprendizaje no supervisado para la detección de anomalías a nivel industrial, lo que puede utilizarse como referencia para aplicaciones similares en otras industrias o en la misma industria.

Palabras clave: Mantenimiento predictivo, detección de anomalías, aprendizaje no supervisado.

Unsupervised Predictive Maintenance: Anomaly Detection in Industry

Abstract. Prognostics and Health Management (PHM), also known as Predictive Maintenance (PdM), is currently a cutting-edge technique

to prevent major losses in manufacturing industries. The objective of this study is to compare different unsupervised algorithms for anomaly detection as a viable option for unlabeled industrial datasets. A real-world dataset was used, consisting of five variables (vibration acceleration, vibration velocity, and temperature) from 118 machines in a manufacturing facility. After feature extraction, Principal Component Analysis (PCA) was applied for dimensionality reduction and used to train an Isolation Forest, a One-Class Support Vector Machine (OC-SVM), a Local Outlier Factor (LOF) model, and Gaussian Mixture Models (GMM). The models were evaluated based on their accuracy, aiming to maximize the detection of 12 failures recorded as anomalies within the dataset. The best-performing model was GMM with 91 % accuracy, followed by OC-SVM with 83 % accuracy. The mean velocity on the Z-axis, the maximum velocity on the X-axis, and the standard deviation of acceleration on the X-axis were identified as key variables for distinguishing different types of faults. This work demonstrates the successful application of unsupervised learning for industrial anomaly detection, which can serve as a reference for similar applications in other sectors or within the same industry.

Keywords: Predictive maintenance, anomaly detection, unsupervised learning.

1. Introducción

La Gestión de Pronóstico y Salud (PHM) de maquinaria, también conocida como Mantenimiento Predictivo (PdM), incluye varias técnicas para anticipar la degradación crítica de equipos industriales con el objetivo de reducir riesgos y pérdidas económicas evitando periodos sin productividad. En los EE.UU., se estima que las industrias gastan alrededor de 200 mil millones de dólares en mantenimiento, y las paradas inesperadas de producción generan gastos de aproximadamente 60 mil millones de dólares [1,2].

A lo largo de los años, la industria ha evolucionado en diferentes enfoques para la gestión del mantenimiento de equipos, por ejemplo 1) la corrección tras fallo, que ocurre solo cuando un equipo deja de funcionar; 2) el mantenimiento preventivo que generalmente es una reparación o reemplazo de partes basado en tiempo de uso; finalmente 3) el mantenimiento predictivo que usa métodos de inferencia estadística y conocimiento en ingeniería, aplicados a datos históricos, para la detección temprana de fallas [2,3].

Las estrategias PHM y PdM se aplican actualmente en diferentes industrias como maquinaria agrícola, componentes de aeronaves, seguridad de redes de comunicación, industrias de manufactura, entre otras. Los métodos más comunes son técnicas estadísticas, métodos de Inteligencia Artificial (modelos de Aprendizaje Automático y Profundo), y modelos basados en datos como los modelos de Vida Útil Remanente. Para lograr los modelos mencionados, se necesita tener datos etiquetados, es decir, el conjunto de datos debe tener ejemplos de estado saludable y de fallos. Desafortunadamente, es muy común no

tener acceso a datos etiquetados, y la estrategia debe dirigirse hacia metodologías no supervisadas [3,4].

El análisis de patrones con métodos no supervisados también se conoce como detección de anomalías. Las anomalías pueden entenderse como instancias de datos que se desvían del comportamiento normal del sistema. En el escenario de Mantenimiento Predictivo (PdM), las anomalías se consideran eventos que requieren inspección, monitoreo o incluso intervención. En el caso específico de no tener eventos de falla explícitos en el conjunto de datos, las anomalías pueden representar un enfoque robusto para el comportamiento anormal, útil para anticipar la degradación aumentada o el detrimento crítico de los componentes de la maquinaria [3,4].

Este artículo tiene como objetivo utilizar modelos de Aprendizaje Automático no supervisados para detectar anomalías dentro de una instalación de la industria alimentaria como una solución viable y asequible para PHM y PdM. Además, la hipótesis de este trabajo es que los modelos de Aprendizaje Automático no supervisados pueden ser utilizados para la detección de fallos.

2. Metodología

2.1. Conjunto de datos y preprocesamiento

El conjunto de datos utilizado fue creado dentro de una industria de manufactura. Se equiparon 118 equipos con sensores biaxiales que registraron el rms de la aceleración (fuerza G) y velocidad (mm/s) en los ejes X y Z, así como la temperatura en grados Celsius (°C) cada minuto entre el 21 de Noviembre de 2023 y Diciembre del 2024. Los datos se almacenaron en una base de datos Microsoft SQL Server y el procesamiento de datos se realizó con el software Matlab R2023b. Como parte del pre-procesamiento, los datos faltantes fueron imputados con interpolación cúbica polinómica por partes (spline cubic interpolation), y posteriormente filtrado utilizando un filtro pasa-baja usando el 90 % del ancho de banda ocupado por las frecuencias como punto de corte.

En cuanto a la seguridad de datos, la empresa se encarga de la ciberseguridad concediendo el acceso a los datos y servidores a través de una red privada. En el caso de la integridad de los datos, los sensores instalados tienen un sistema de diagnóstico para identificar desconexiones físicas del sensor, desvinculación con los nodos y congelación de lecturas, las cuales se denominarán a partir de ahora como banderas de diagnóstico. Las banderas de diagnóstico permiten actuar rápidamente para reestablecer la entrada de datos íntegra, y evitar la pérdida de calidad; no obstante, existen puntos de medición con banderas activas (p.e.: sensores desconectados), que representaron alrededor del 1 al 10 % de las observaciones del conjunto de datos (variando de sensor a sensor). Estos puntos fueron contemplados para la interpolación usando interpolación cúbica; se descartó la interpolación lineal para preservar el comportamiento de la señal.

2.2. Extracción y selección de características

Se calcularon variables estadísticas en el dominio de tiempo, por ejemplo, media (μ), desviación estándar (σ), valor máximo (max), varianza (var), curtosis (K) y asimetría (γ); adicionalmente, se calcularon variables en el dominio de la frecuencia. Para ello, se aplicó una Transformada de Fourier a cada señal y se extrajeron los momentos estadísticos del espectro de frecuencia, por ejemplo, frecuencia media (μ_{freq}), mediana de la frecuencia, varianza (var_{freq}), curtosis (K_{freq}), asimetría (γ_{freq}) y distancia de pico a pico. Todas las características fueron calculadas en todas las variables originales (Aceleración y Velocidad en ejes X y Z, y Temperatura). La selección-transformación de variables se realizó utilizando un Análisis de Componentes Principales (PCA), esto nos ayudó a disminuir la dimensionalidad de los datos y la colinealidad. Para el entrenamiento de los modelos, se eligieron las variables que acumularon un 90 % de la varianza explicada dentro del modelo de PCA (entre 20-25 variables dependiendo del equipo).

2.3. Entrenamiento de modelos

Este trabajo se enfocó en cuatro algoritmos: **Bosque de Aislamiento (IF)**, **Modelos de Mezcla Gaussiana (GMM)**, **LOF** y **OC-SVM**. Todos los modelos se ajustaron utilizando el *Statistics and Machine Learning toolbox* de Matlab R2023b. Todos los modelos se configuraron para detectar alrededor del 8 % de anomalías para ser comparables con los **Modelos de Mezcla Gaussiana**.

El Bosque de Aislamiento (IF) utiliza un procedimiento de *aislamiento* similar a los árboles de decisión. El procedimiento de *aislamiento* tiene como objetivo dividir los datos por una o más variables hasta que una sola instancia se aísle del resto, a lo cual se le considera un *iTree*. El modelo IF asume que un valor atípico será más fácil de aislar que un valor normal. En este estudio, se eligieron 150 árboles *iTree* dentro del modelo IF [5,6].

El Modelo de Mezcla Gaussiana (GMM) es un método de clustering no supervisado basado en encontrar múltiples distribuciones gaussianas calculadas con el algoritmo de Expectación-Maximización. Para este estudio, los GMM se probaron en un rango 2-10 clusters (asumiendo que el mínimo podría ser cluster normal y anómalo), el modelo con mejor rendimiento se seleccionó con el menor Criterio de Información Bayesiano. Después de seleccionar el mejor modelo, el cluster más grande se seleccionó como el cluster de datos normales, y los clusters restantes se consideraron como clusters de anomalías [7].

La Máquina de Vectores de Soporte de Una Clase (OC-SVM) tiene como objetivo encontrar el hiperplano óptimo para mapear el espacio de entrada en un espacio de alta dimensionalidad que permita separar las observaciones a través de márgenes superiores e inferiores [8,9].

El Factor de Aislamiento Local (LOF) es un algoritmo de detección de anomalías basado en la densidad y distancia. En este estudio se eligió la distancia euclidia como métrica de entrenamiento del modelo [10].

2.4. Evaluación de modelos

Debido a la naturaleza de este trabajo (aprendizaje no supervisado), es difícil definir una estrategia de evaluación. Sin embargo, basándonos en el trabajo presentado en [5] y [6], proponemos la siguiente metodología para la evaluación. La planta de manufactura proporcionó 12 fallas entre enero de 2024 y mayo de 2024, estos datos son irrelevantes para desarrollar un modelo de Aprendizaje Supervisado y el sobremuestreo podría llevar a un sobreajuste del modelo porque la etiqueta minoritaria representa solo el 0.07 % de todo el conjunto de datos. Por lo tanto, se priorizó maximizar la detección de fallas como anomalías y evaluar los modelos en términos de exactitud (Ecuación 1) de la detección de fallas

$$Exactitud = \frac{\text{Fallas detectadas}}{\text{Fallas totales}}. \tag{1}$$

Para obtener la característica más importante para la detección de anomalías, aplicamos una *prueba T-Student* en las anomalías observadas. Después de comparar las *señales de anomalía* y las *señales normales*, seleccionamos la variable con el *p-valor* más bajo como la variable más importante para la detección de anomalías.

3. Resultados y discusión

El porcentaje de datos anómalos en el conjunto de datos fue de 8 % utilizando todos los modelos descritos anteriormente (Tabla 1). Los modelos con mejor rendimiento fueron el GMM y el OC-SVM, detectando alrededor del 0.916 (91.6 %) y el 0.833 (83 %) de las fallas como anomalías. Por otro lado, IF y LOF demostraron la menor exactitud, detectando aproximadamente el 0.583 (58.3 %) de las fallas como anomalías (Tabla 1). Debido al buen rendimiento del GMM, el resto del manuscrito se centró en estudiar este modelo en 3 fallas importantes como la rotura de la correa del motor, fallas en los rodamientos y una falla por sobrecorriente en el motor.

Debido a los acuerdos de privacidad y confidencialidad, no es posible divulgar beneficios e impactos económicos puntuales de los modelos dentro de la planta de manufactura. Sin embargo, los modelos han permitido ahorrar horas de trabajo, desperdicio de producto y tiempos muertos al planificar revisiones y mantenimiento de acuerdo con el volumen de anomalías detectadas. En este sentido, se planifica el mantenimiento en equipos que han demostrado un alto volumen de anomalías en periodos de tiempo definidos (p.e.: 1 semana). Este enfoque en conjunto con los paros planeados para limpieza e inspección, permite mitigar las falsas alarmas disparadas por los modelos, por ejemplo, equipos con 1 o 2 anomalías aisladas.

La razón de la diferencia entre el rendimiento de los modelos puede deberse al tipo de anomalías y la metodolgía de detección de anomalías de los modelos. De acuerdo con [11], las anomalías pueden dividirse en diferentes categorías, entre ellas se encuentran las puntuales, contextuales y colectivas. Las anomalías

Tabla 1. Rendimiento de modelos de detección de anomalías.

Modelo	Exactitud	Proporción de anomalías (%)
IF	0.583	8.00
OC-SVM	0.833	8.00
GMM	0.916	8.72
LOF	0.583	8.00

puntuales son aquellos puntos que difieren del resto de casos u observaciones, mientras que las colectivas se refieren a un grupo de puntos, que en conjunto, difieren del resto de datos. Es probable que en este conjunto de datos se presenten anomalías colectivas que fueron más fáciles de capturar por modelos que ajustan una función de probabilidad de distribución o una función matemática como en el caso del GMM y los hiperplanos de la máquina de soporte vectorial de una clase en comparación de los modelos de aislamiento por distancia (IF y LOF) [12].

En el caso del efecto del procesamiento de datos en los resultados del GMM, se realizaron pruebas preeliminares con diferentes valores de varianza explicada por el modelo de Análisis de Componentes Principales; se observó que aumentar la varianza explicadahacia un 99 % o disminuirla por debajo de 90 % reducía la exactitud de los modelos en rangos de 0.25 a 0.5, siendo este un parámetro importante debido a que transforma los datos para entrenar al modelo final.

Por otro lado, como parte de los hiperparámetros del modelo, se comparó el efecto de utilizar la matriz de covarianza completa o diagonal, siendo la última la que demostró los mejores resultados. El uso de la matriz de covarianza completa creaba 1 o 2 clusters (de tamaños similares) que fallaba al detectar las fallas como anomalías o generaba un exceso de anomalías detectadas. De acuerdo con los trabajos [13,14], la matriz de covarianza completa puede ocasionar un sobreajuste del modelo al considerar las correlaciones entre las variables independientes y crear límites flexibles entre clusters.

El primer ejemplo de anomalías detectadas por el modelo GMM se dio 3 días previos a la rotura de correa del motor de un equipo. La prueba *T-Student* demostró que la variable más importante era el valor medio de la Velocidad en el eje Z ($p = 1{,}17^{-20}$). al inspeccionar la variable, se observó que la velocidad eje Z tenía un valor más alto en el valor medio de las observaciones anómalas (3.5 mm/s) en comparación con las observaciones normales (1.25 mm/s). De igual manera, los valores máximos de velocidad en eje Z para las anomalías fue considerablemente mayor (4-6.5 mm/s) en comparación con las observaciones normales (1.8 - 2.9 mm/s).

Otra de las anomalías detectadas por el modelo de GMM resultó en una falla en los rodamientos del equipo. De acuerdo con la prueba *T-Student*, la variable más importante fue el valor máximo en la velocidad del eje X ($p = 1{,}86^{-17}$). Al observar los valores máximos de velocidad del eje X, se encontró que las señales con anomalías podían llegar a velocidades de hasta 14 mm/s (con una media

de 7.75 mm/s), mientras que las observaciones normales demostraban valores de aproximadamente 5.5 mm/s con una media de 3.9 mm/s.

Finalmente, el modelo GMM detectó una anomalía que se desarrolló a los 3 días en una falla por sobrecorriente. El análisis de las variables demostró que la desviación estándar en la aceleración del eje X era la característica más importante ($p = 3{,}21^{-12}$) para determinar que dicha observación fue una anomalía. Al analizar los valores de la variable se observó que las señales normales se caracterizaban por una baja variabilidad (desviación estándar) entre un 0.05 - 0.15 G; por otro lado, las señales anómalas tuvieron una desviación estándar de hasta 0.25 G.

Los resultados presentados en este trabajo son comparables a otros estudios sobre la detección de anomalías utilizando conjuntos de datos industriales. Por ejemplo, Carrasco et al. [4] utilizaron un conjunto de datos proporcionado por ArcelorMittal. Los autores implementaron los algoritmos *Histogram Based Outlier Score*, *Lightweight Online Detector of Anomalies* y *Extreme Gradient Boosting Outlier Detection*. El mejor modelo fue el algoritmo *Extreme Gradient Boosting Outlier Detection*, calculando variables en una ventana de tiempo de 48 h. Su rendimiento varió entre un 0.45 a 0.85 usando el Área Bajo la Curva (AUC) como métrica dependiendo de las variabales calculadas.

En otro estudio, se utilizó un OC-SVM para la detección de anomalías en una turbina de gas industrial de una empresa coreana; el conjunto de datos estaba compuesto principalmente por datos de temperatura, presión de los compresores y tasas de flujo de combustible. El modelo resultante detectó anomalías con una precisión de alrededor de 0.54-0.64 y una puntuación F-1 de 0.62-0.63 [15].

Finalmente, en un estudio con datos de turbinas de gas, realizado por Lee et al., [16], se evaluaraon los modelos OC-SVM, Isolation Forest, k-means y un Auto-Encoder Convolucional para la detección de anomalías. Los resultados demostraron un rendimiento excepcional del Auto-Encoder (0.874) en comparación con los otros modelos (0.034 - 0.522) en términos de puntuación F-1, seguido por el Isolation Forest. Sin embargo, algoritmos muy complejos como el Auto-Encoder podrían ser difíciles de implementar en en aplicaciones en línea.

4. Conclusión

Este artículo demostró que los modelos de aprendizaje automático no supervisados pueden utilizarse para la detección de anomalías en un entorno industrial. Además, los modelos detectan entre el 83 % y el 91 % de las fallas como anomalías en un conjunto de datos sin etiquetas. Este enfoque nos permitió identificar eventos relevantes en cada sensor de la planta de manufactura, lo cual podrá prevenir daños críticos en la línea de producción. Se espera que este método sea replicable en escenarios similares donde no se cuente con datos etiquetados. Para futuras investigaciones, se planea construir índices de salud no supervisados, aunque esperamos contar con etiquetas para explorar modelos supervisados de diagnóstico de fallas y vida útil remanente.

Agradecimientos. Los autores agradecen a la empresa de manufactura por la disponibilidad de los resultados y a la empresa encargada de la instalación del hardware para la adquisición de datos (nombres omitidos por privacidad), así como al equipo de desarrollo de software involucrado en la implementación y despliegue de los modelos presentados en este artículo.

Referencias

1. Kolokas, N., Vafeiadis, T., Ioannidis, D., Tzovaras, D.: A generic fault prognostics algorithm for manufacturing industries using unsupervised machine learning classifiers. Simulation Mod elling Practice and Theory 103, 102–109 (2020)
2. Amruthnath, N., Gupta, T.: A research study on unsupervised machine learning algorithms for early fault detection in predictive maintenance. In: 2018 5th International Conference on Industrial Engineering and Applications (ICIEA), pp. 355–361. IEEE, Singapore (2018)
3. Carvalho, T., Soares, F., Vita, R., Francisto, R. d. P., Basto, J., Alcalá, S.: A systematic literature review of machine learning methods applied to predictive maintenance. Computers & Industrial Engineering, 137, 106024 (2020)
4. Carrasco, J., Lopez, D., Aguilera-Martos, I., et al.: Anomaly detection in predictive maintenance: A new evaluation framework for temporal unsupervised anomaly detection algorithms. Neuro computing 462, 440–452 (2021)
5. Brito, L., Susto, G., Brito, J., Duarte, M.: An explainable artificial intelligence approach for unsupervised fault detection and diagnosis in rotating machinery. Mechanical Systems and Signal Processing, 163, 10810 (2022)
6. Susto, G., Beghi, A., McLoone, S.: Anomaly detection through on-line isolation Forest: An application to plasma etching. In: 28th Annual SEMI Advanced Semiconductor Manufacturing Conference (ASMC), pp. 89–94, IEEE, USA (2017)
7. Wen, L., Yang, G., Hu, L., Yang, C., Feng, K.: A new unsupervised health index estimation method for bearings early fault detection based on Gaussian mixture model.Engineer ing Applications of Artificial Intelligence, 128, 107562 (2024)
8. Barbado, A., Corcho, Ó., Benjamins, R.: Rule extraction in unsupervised anomaly detection for model explainability: Application to OneClass SVM. Expert Systems with Applications, 189, 116100 (2022)
9. Shin, H., Eom, D., Kim, S.: One-class support vector machines — an application in machine fault detection and classification. Computers & Industrial Engineering, 48(2), 395–408 (2005)
10. Alghushairy, O., Alsini, R., Soule, T., Ma, X.: A Review of Local Outlier Factor Algorithms for Outlier Detection in Big Data Streams. Big Data and Cognitive Computing, 5(1), (2021)
11. Forthuis, R.: A Typology of Data Anomalies. In: Information Processing and Management of Uncertainty in Knowledge-Based Systems. Theory and Foundations, pp. 26–38, Springer (2018)
12. Cao, Y., Xiang, H., Zhang, H., Zhu, Y., Ming Ting, K.: Anomaly Detection Based on Isolation Mechanism: A Survey. ArXiv (2024)
13. Thajeel, J.K., Adel, R., Muhammed Ali, H., Shakir, R.R.: Effect of the covariance matrix type on the CPT based soil stratification utilizing the Gaussian mixture model. Journal of the Mechanical Behavior of Materials, 31 (2022)
14. Magdon-Ismail, M., Purnell, J.T.: Approximating the Covariance Matrix of GMMs with Low-Rank Perturbations. In: Fyfe, C., Tino, P., Charles, D., Garcia-Osorio,

C., Yin, H. (eds) Intelligent Data Engineering and Automated Learning – IDEAL 2010. IDEAL 2010. Lecture Notes in Computer Science, Springer, Berlin (2010)
15. Kang, H., Choi. Y., Yu, J., Jun, S., Lee, J., Kim, Y.: Hyperparameter Tuning of OC-SVM for Industrial Gas Turbine Anomaly Detection. Energies, 15(22), (2022)
16. Lee, G., Jung. M., Song, M., Choo, J.: Unsupervised anomaly detection of the gas turbine operation via convolutional auto-encoder. In: Proceedings of the Annual Conference of the Prognostics and Health Management Society, pp. 1–6, IEEE, USA (2020)

ISSN 1870-4069

AI Safety in Mexico: A Pilot Survey in Yucatan

Janeth Valdivia Pérez[1,3], Valeria Ramirez Hernández[1], Silvia Fernández-Sabido[2,3], Ángel Tenorio Vázquez[3], Alejandro Molina-Villegas[2], Oscar Sánchez Sordia[2]

[1] Universidad Politécnica de Yucatán, Mexico

[2] Centro de Investigación en Ciencias de Información Geoespacial, Mexico

[3] AI Safety Mexico Project, Mexico

{janeval92,valery.ramirez.hdez}@gmail.com, {sfernandez,osanchez}@centrogeo.edu.mx, aetenorio@gmail.com

Abstract. As artificial intelligence transforms society, understanding regional perspectives on its risks becomes vital. This study presents a pilot survey conducted in Yucatan, Mexico, aimed at capturing local concerns about AI safety. Inspired by the global survey *"Thousands of AI Authors on the Future of AI"*, this initiative was a joint effort of the Universidad Politécnica de Yucatán and the Centro de Investigación en Ciencias de Información Geoespacial, in Merida, Mexico. The survey is piloted on a small sample of academics, including students, professors, and researchers, probing their views on the ethical, safety, and regulatory aspects. The survey delved into topics such as the explainability of AI, its potential risks, and the implications of its advancements over the next twenty years. These preliminary findings reveal concerns about the spread of false information, manipulation of public opinion, and the possibility of AI being used in an authoritarian manner. These concerns are consistent with global trends and reflect unique characteristics rooted in the socioeconomic and cultural realities of Mexico. Notably, 97% of respondents agreed that AI experts should consider the needs and concerns of society, while 83% believe the government should invest in mitigating AI safety risks. Furthermore, 81% expressed support for the creation of a special AI safety agency in Mexico. This kind of study contributes to the academic and political dialogue about AI safety in Mexico and also lays the groundwork for future research and the development of policies that promote a safe and ethical implementation of AI.

Keywords: AI safety, survey analysis, ethical implications, regulation, public perception.

1 Introduction

In recent years, artificial intelligence (AI) has experienced exponential growth, revolutionizing industrial, academic, and social sectors globally. Like many other countries, Mexico is at a crucial juncture where the adoption and integration of AI in various areas present unprecedented opportunities as well as significant challenges. Given this dynamism, AI safety has become an area of growing interest among the academic and scientific community, emphasizing its importance in promoting responsible and ethical technological development.

In this context, the implementation of a pilot survey aims primarily to evaluate the effectiveness of selected questions to measure key concerns of the study population. The questions focus on aspects such as the explainability of AI, the causes of its progress, potential risks over the next 20 years, and general perceptions of its safety in Mexico. Moreover, this pilot survey seeks to refine the methodologies used to ensure the accuracy and relevance of the data collected. This preliminary effort is crucial for preparing a more extensive national survey focused on AI safety in Mexico.

The study, inspired by the ESPAI (AI Impacts' Survey on AI Progress) 2023 [1], adapts and selects relevant questions for the Mexican context and introduces new inquiries focused on local concerns about the use and regulation of AI.

The momentum for this survey comes from a short (two month) professional stay collaboration between the Centro de Investigación en Ciencias de Información Geoespacial (CentroGeo) and the Universidad Politécnica de Yucatan (UPY), in Merida Mexico, where resources from both institutions have been combined to deepen understanding of the current AI landscape in the region and to ensure that the design and implementation of the survey provide a clear and detailed vision that will guide future initiatives and policies in this emerging field. Through this study, we aim to establish a solid foundation for future research and policies that guide the safe and ethical development of artificial intelligence in the country.

2 Objective

The main goal of the study is to conduct a pilot survey on AI safety within the academic community (students, professors and researchers), with the purpose of capturing the perceptions and concerns of the study population. This will serve as preparation for a more comprehensive and detailed national survey on the same topic.

The specific objectives are to translate and select questions from the ESPAI 2023 survey conducted by the AI Impacts team that may be useful to understand the perspective on AI safety in Merida, Yucatán and; to create a block of questions to help explain the perception of the use and regulation of AI safety in Mexico. The survey also aims to refine the methodologies employed to ensure the accuracy and relevance of the data collected.

3 Related Work

Artificial intelligence (AI) safety research in Mexico has evolved, albeit at a slower pace compared to the global scene. Challenges in regulation and safety span a range of concerns from ethical design to the mitigation of unintended consequences of AI systems. This research often intersects with broader discussions on data protection, human rights, and socioeconomic impacts, facilitated by interdisciplinary teams composed of scientists, philosophers, and policymakers.

Initiatives like the *Instituto Nacional de Acceso a la Información y Protección de Datos Personales* (INAI) highlight the need for privacy and data ethics in AI, setting a precedent for incorporating ethical considerations in AI development [5]. Meanwhile, the academic community in Mexico has begun to address the issue of bias in AI systems. Studies like those by Ramirez emphasize that factors such as limited access, insufficient resources to evaluate data, and inherent discrimination in AI systems can deepen existing gaps in digital, social, political, and economic realms. They also raise ethical and moral questions about the implications of the development of current AI systems [7].

In terms of regulation, in 2018, the Mexican government, in collaboration with Oxford Insights, C-Minds, the Mexican Society for Artificial Intelligence, and *Tecnológico de Monterrey*, supported financially by the UK Embassy in Mexico, embarked on a mission to develop specific actions the government could take to promote the development and use of AI across all sectors of the country. However, with the change of government.

In the same year, the planned AI agenda was not continued. Nevertheless, entities like the UNAM Institute of Legal Research worked in their Public Policy Reports during 2021 to identify challenges and obstacles for designing a public AI policy that included a Human Rights approach in the country, emphasizing the importance of creating a National Artificial Intelligence Strategy with diverse perspectives to identify potential risks in AI development and deployment [9].

Since 2019, Mexico has committed to adhering to the OECD principles for responsible AI development by signing the first set of intergovernmental policy guidelines on AI, agreeing to comply with international standards that ensure the design of AI systems are robust, secure, fair, and reliable [3].

Recently, during 2023 and 2024, deputies and senators from various political parties in Mexico have presented and discussed important initiatives in the Chamber of Deputies and the Senate to pass laws on regulation to establish measures that ensure the transparency and oversight of AI systems, however, to date, no initiative has been approved yet [8,2,6].

As we can observe, research on AI safety in Mexico is in a stage of growth and alignment with international standards, facing challenges including ethics, privacy, and bias in algorithms. Collaboration among academia, government, and international organizations is crucial to creating a robust framework that ensures the responsible development of AI. Recent legislative initiatives in Mexico reflect a growing commitment to the regulation and supervision of technology.

4 Methodology

4.1 Survey Themes

With the aim of achieving the goals of this study, a survey titled "Survey on AI Progress" was created, aimed at capturing the perception of students, professors, and scientists primarily from Merida in Yucatán Mexico. The survey was inspired by the AI Impacts Expert Survey on AI Progress 2023 (ESPAI 2023) [1] and the article "Thousands of AI Authors on the Future of AI" developed by the AI Impacts team [4,1].

ESPAI 2023 is a key survey that measures the perceptions of AI researchers, following the methodology of its predecessors, the ESPAI surveys of 2022 and 2016. With 2,778 responses from researchers, this survey represents a significant sample of the AI academic community.

The questions focused on consistent themes over the years to compare the evolution of opinions and included some new ones, designed after an iterative testing process. After multiple revisions of the survey and article, a total of 44 questions were formulated that assess the topics shown in Table 1. By adopting this proven and detailed methodology, our survey aims to effectively capture opinions and attitudes towards artificial intelligence among the academic community, and reinforce the communication work on the topic of safety in Mexico in order to detect specific challenges and opportunities.

Table 1. Key AI Topics: Development and Impact.

AI Development	**Safety & Society**
The explosion of intelligence	20-year AI traits
Explainability	Tasks (fixed probabilities and fixed years)
Scenarios	Safety quote
Value	Safety resources + ethics
Causes of AI Progress	Safety resources
HLMI outside view	Extinction
Meta and sociology	Demographics
High-level machine intelligence (HLMI), through jobs (fixed probabilities and fixed years)	Speed and safety

4.2 Survey Design and Data Collection

The importance of basing our survey on the ESPAI questions to capture perceptions about AI safety in Mexico lies in ensuring relevance and accuracy. In this sense, we adhere to the wording of the original questions, seeking a careful

translation into Spanish and selecting those we consider of interest for the current study. It was also considered prudent to retain the Likert scale proposed in the original survey and to develop a block of questions according to our local context. The survey included the sections shown in Table 2.

Table 2. Sections of the survey

Research Areas	Safety Considerations
Collection of demographic data	Safety
High-Level AI and Total Job Automation	Safety resources and ethics
AI Intelligence Explosion	Perception of AI safety in Mexico
Explainability	Feedback section
Causes of AI Progress	Completion time
AI risks within 20 years	

We selected a sample of 36 people, 42% of whom are students from the Universidad Politécnica de Yucatán, studying Data Engineering, Cybersecurity, and Robotics at undergraduate and graduate levels. The remaining 58% are researchers and/or professors from institutions such as the Centro de Investigación y de Estudios Avanzados del IPN (Cinvestav), the Centro de Investigación Científica de Yucatán (CICY), and the Instituto de Investigaciones en Matemáticas Aplicadas y Sistemas (IIMAS), dedicated to fields like data science, software development, and machine learning, among others.

The survey was created in Google Forms, distributed via invitation link through email and WhatsApp. Due to the short duration of the stay and because it was a pilot test (we left the bulk of the community for the future survey), it was launched on April 1, 2024, and remained open until April 8 of the same year. At the end of the survey period, we obtained a CSV file of semi-structured data without defined data types or relationships. In this regard, we developed a data dictionary to assist in our preprocessing and transformation code.

We focused on data extraction and preprocessing in Colab, a hosted Jupyter notebook service. It is also important to note that, given the small sample size, it wasn't advisable to stratify it, but in the future we consider differentiating between students, professors, and researchers, and even between those who have had contact with the topic of AI-Safety and those who haven't.

4.3 Data Cleaning

During the development of a Python script designed for analyzing a survey on the Progress of Artificial Intelligence, we utilized several highly recognized libraries (Pandas, Numpy and Matplotlib) in the field of data science, which facilitated both the manipulation and visualization of data. The analysis began with an exploratory examination of the data.

To facilitate handling during the analysis, we opted to rename the columns using a 'Q' followed by a number format (e.g., 'Q1', 'Q2', etc.). This change not only simplified referencing the columns but also improved the organization of the DataFrame.

Before proceeding with data deletion, a careful selection of questions deemed optimal for the study was made. These questions included topics on explainability, causes of AI progress, potential AI risks in 20 years, and the perception of AI Safety in Mexico. After defining the key questions for this analysis, several columns that were not necessary were removed, thus reducing the dataset's dimensionality and focusing the analysis on the most relevant variables for the study.

4.4 Data Analysis and Visualization

The analysis of the data collected from the Survey on the Progress of Artificial Intelligence was deepened through the generation of descriptive statistics and the use of effective visualization techniques. All data were analyzed based on percentages, a crucial practice in this type of study primarily because it allows for uniform comparison between groups of different sizes. In surveys and polls, converting raw counts into percentages normalizes the data, which is useful for performing statistical analysis and for visualizing trends more effectively, especially when dealing with data categorized across multiple levels or scales.

Descriptive statistics were calculated for each of the variables of interest, this approach helped to identify preliminary patterns, understand the distribution of responses, and establish a baseline for more detailed analyses.

4.5 Visualization Techniques

Visualization techniques used to further explore the data included bart charts, stacked bar charts, and heat maps, which allowed for comparing frequencies and intensity of responses in percentage terms across different question categories, highlighting significant differences and similarities. Finally, the results of the analysis will be interpreted in the context of the study's objectives. Conclusions and recommendations will be drafted aimed at promoting effective safety and regulation of AI in Mexico.

5 Development

5.1 Concern Levels for AI Scenarios over the Next 30 Years

Through the question: "How would you rate the level of concern that the following possible AI-related scenarios deserve over the next 30 years?", eleven AI scenarios that could be a cause for concern for society were assessed. In Figure 1 we can observe how the respondents evaluated the severity of the following scenarios for the next 30 years.

Of the scenarios presented, those that generated the most concern when adding up the percentages of the most representative categories of response such as "Quite worrying" and "Extremely worrying" were the spread of false information, like deep fakes which had a cumulative 97.2% of substantial or extreme concern, followed by the manipulation of public opinion on a large scale representing 88.9% while concern about Authoritarian rulers using AI to control their populations represented 83.3%.

Other categories such as scenarios like AI systems that worsen economic inequality by disproportionately benefiting certain individuals and the possibility of AI allowing dangerous groups to create powerful tools, like designed viruses represented 80.6% and 77.8% respectively.

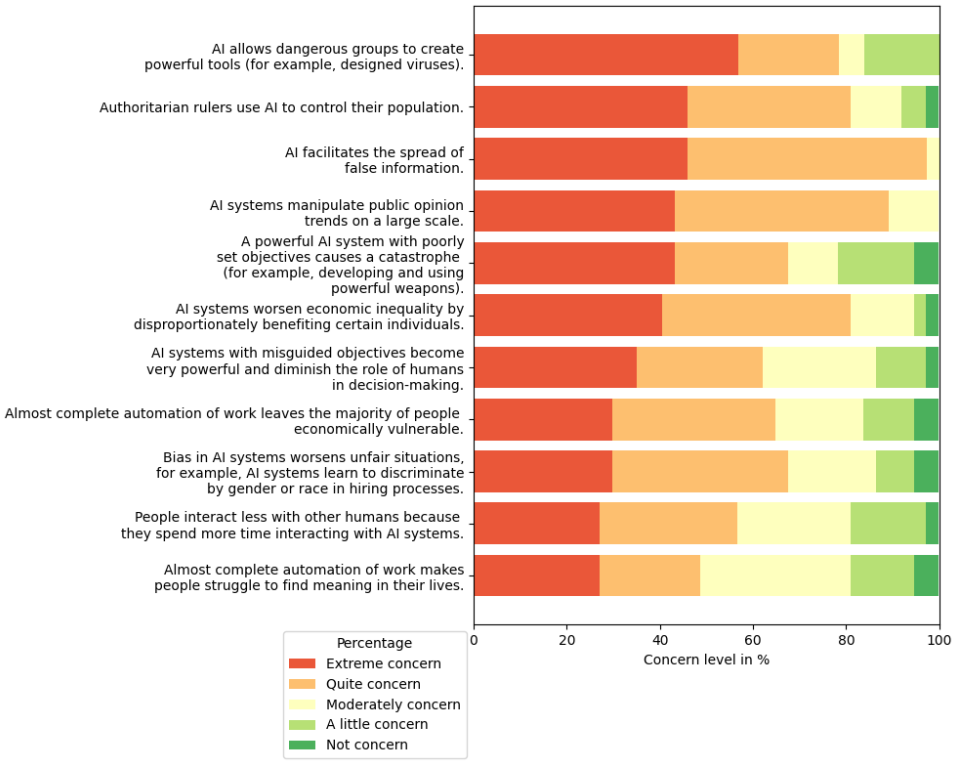

Fig. 1. Concern levels for AI scenarios over the next 30 years. Source: own elaboration with data obtained from the Survey on the progress of AI,

5.2 Causes of AI Progress

The analysis of the data collected on the causes of AI progress (Figure **??**) revealed that overall, the factors of research effort, computing power, quantity and quality of data, progress in AI algorithms and funding have been extremely important as causes of AI progress. However, it is notable that more than 60% of the participants classified computing power as the main cause, while the quantity and quality of data was considered the second most representative cause with 58.3%. It also reveals that 55.6% of participants consider research effort to be extremely important, while funding and progress in algorithms stand out with 41.7% and 44.4% participation, respectively.

Some of the scenarios, such as "Finding unexpected ways to achieve goals" which represents a 44.4% probability, and "Making design improvements to increase their own performance regardless of human desires" at 38.9%, have a high perception of being "Extremely likely". Other scenarios, such as "Can speak as a human expert on most topics" and "Often behaving in ways that are surprising to humans", also have a considerable perception of probability.

Figure 3 shows descriptive statistics of the five categories assessing the perception of risk associated with artificial intelligence (AI) over a span of 20 years, according to respondents' answers. It is observed that the mean progressively increases from "Not at all likely" to "Extremely likely," reflecting a rise in the perceived probability across categories. Concurrently, the standard deviation shows an upward trend, while the median (50%) maintains consistency with the mean, rising with each subsequent category.

As for extreme values, the maximum observation is recorded in the "Extremely likely" category at 44.40%, followed by 38.90% in the categories of "Slightly likely" and "Quite likely". The minimum value (min), in turn, rises from 0 in "Not at all likely" to 16.70 in "Equally likely", and again to 23.60% in "Extremely likely".

Finally, it is noteworthy that the gap between the third quartile (75%) and the maximum value is not particularly wide in the "Quite likely" and "Extremely likely" categories. This phenomenon is also illustrated in Figure 2, where the values tend to cluster towards the upper end of the scale in these categories.

6 Perception of AI Safety in Mexico

In Figure 4 we analyze perceptions about the impact and regulation of artificial intelligence (AI) in Mexico. The results are presented in percentages reflecting the respondent answers to various statements:

- *Labor market and equity concerns*
  Regarding the statement "AI experts should take into account the needs and concerns of society", the graph shows an almost unanimous consensus among participants (98%) on the importance of AI experts considering society's needs and concerns, as they agreed or strongly agreed with this statement.

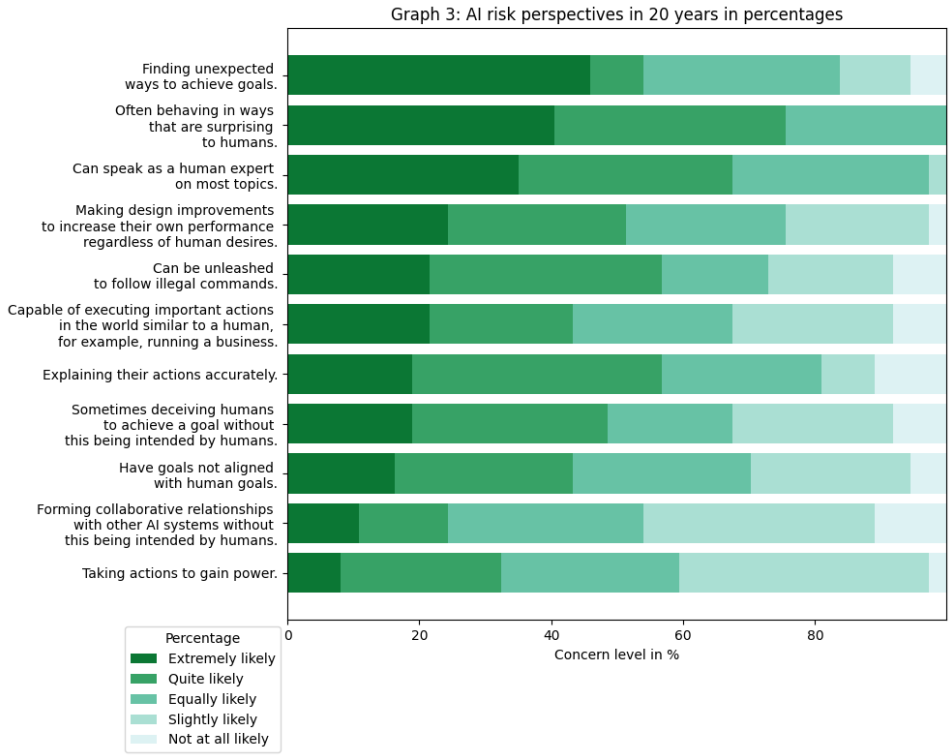

Fig. 2. AI risk in 20 years. Source: own elaboration with data obtained from the Survey on the progress of AI.

	Not at all likely	Slightly likely	Equally likely	Quite likely	Extremely likely
count	11.00	11.00	11.00	11.00	11.00
mean	5.05	18.67	25.77	26.77	23.72
std	3.88	12.11	4.51	9.41	11.48
min	0.00	0.00	16.70	8.30	8.30
25%	1.40	9.70	25.00	23.60	18.05
50%	5.60	22.20	25.00	27.80	19.40
75%	8.30	23.60	29.20	33.35	30.55
max	11.10	38.90	30.60	38.90	44.40

Fig. 3. Descriptive statistics of the expressed probabilities of the risk of AI at 20 years. Source: own elaboration with data obtained from the Survey on the progress of AI.

Concern about the impact of AI on the labor market in Mexico is significant, with 66.6% of participants agreeing or strongly agreeing with this concern. Similarly, there is a notable worry about biases in AI algorithms and the risk they pose to equity and justice in Mexico, with 61.1% expressing agreement or strong agreement.

– *Government regulation and supervision*
  A significant majority of respondents believe in the government's importance in regulating and supervising AI, with 75% agreeing or strongly agreeing that the government should regulate and supervise AI more.
– *Investment in safety and specialized agency*
  The government's investment in mitigating AI-related safety risks is considered important, with 83% agreeing or strongly agreeing on its significance. Furthermore, 81% of respondents would agree with the creation of a special agency for AI safety in Mexico.
– *Perception of current regulations*
  Regarding the adequacy of Mexico's current regulations to meet AI safety challenges, only 11% of participants agreed or strongly agreed that they are adequate — suggesting general skepticism about the current regulatory framework.
– *Socioeconomic concerns*
  Finally, the possibility that socioeconomic inequality in Mexico could increase due to the way AI is currently being developed and utilized in the country is a concern for 61% of participants, who agreed or strongly agreed with this statement.

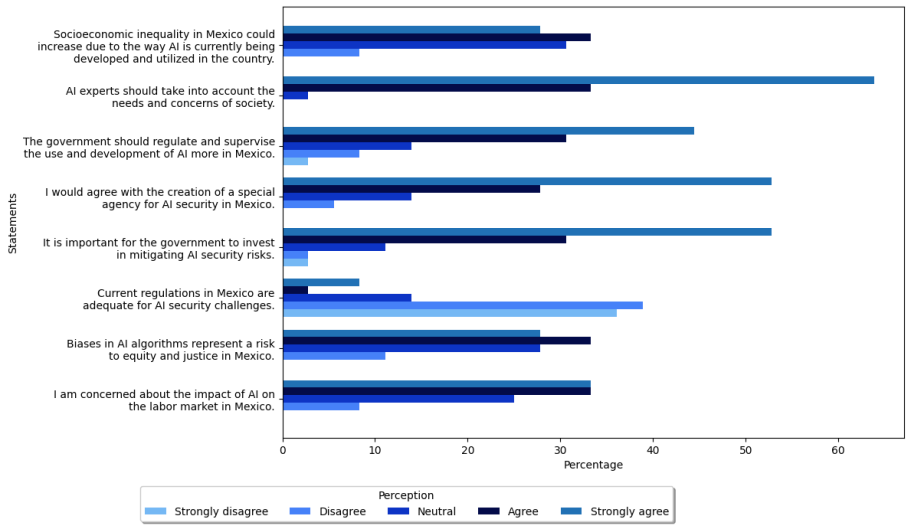

Fig. 4. AI safety perceptions in Mexico. Source: own elaboration with data obtained from the Survey on the progress of AI[10].

7 Discussion

Comparing analysis of the survey results from the ESPAI 2023 as outlined in the document "Thousands of AI Authors on the Future of AI" and those derived from the "Survey on AI Progress" tailored for Merida, Yucatan, we delve into the nuanced perceptions and concerns surrounding the safety and advancement of artificial intelligence (AI). Highlighting both contrasts and alignments in global and local contexts. The section on AI risk scenarios in the local study indicated a high level of concern for scenarios such as the spread of false information and the large-scale manipulation of public opinion by AI systems. There was also significant worry about the use of AI by authoritarian rulers to control populations. The findings published in "Thousands of AI Authors on the Future of AI" similarly reflected significant concerns over similar scenarios, such as the spread of misinformation and authoritarian control over populations. However, a distinctive aspect was the emphasis on the exacerbation of economic inequality noted in the ESPAI 2023.

Analysis of the causes of AI progress showed that the main perceived drivers include computing power, data quantity and quality, and research efforts. These results align closely with those found in the original survey, with the addition of AI algorithmic progress as a significant driver. Regarding AI risk perspectives over the next 20 years, responses in the survey indicated that scenarios such as "Finding unexpected ways to achieve goals" and "Making design improvements to enhance their own performance, regardless of human desires," were seen as extremely likely.

On the perception of AI safety in Mexico, there was significant consensus on the importance of government regulation and oversight of AI. A high percentage of respondents supported the idea of creating a specialized AI safety agency in Mexico. Additionally, there was considerable concern about the potential impact of AI on the labor market and the biases in AI algorithms that could affect equity and justice in Mexico.

As for expectations about AI progress, the ESPAI 2023 survey anticipated significant advancements, with a considerable proportion of respondents foreseeing the possibility that AI could outperform humans in all possible tasks in the not-too-distant future. However, the focus of the adapted survey was more on understanding perceptions of AI-related risk scenarios and their social, ethical, and safety implications.

While the original survey did not delve deeply into specific regulations, it suggested a general need for research on AI safety and addressing potential risks of the technology.

It is important to note that creating a section to analyze perceptions of AI safety in Mexico showed clear consensus on the importance of government regulation in overseeing AI. This idea is reinforced by a high percentage of respondents supporting the creation of a specialized agency for AI safety in Mexico.

Overall, while the original survey provided a more global view focused on the pace of AI technological advances, the adapted survey for Merida concentrated

more on understanding local perceptions of AI risks and regulation, reflecting specific concerns about safety and ethics in the use of technology. Both surveys, however, underscore the importance of addressing the risks associated with AI and the need for effective regulation to ensure responsible and safe technological development.

Given that the Python script was designed to complete the first step of data extraction in an ETL (Extract, Transform, Load) process and includes various transformation operations such as data cleaning, handling null values, renaming columns to facilitate analysis, and some basic mathematical and statistical operations that are part of the transformation process, it is believed that the future can improve and complete an ETL process for handling survey data at a national level.

It is also believed that expanding the survey's scope to include more regions of Mexico could provide a broader view of the perception of AI safety throughout the country.

8 Conclusion and Future Work

The conclusion of this report highlights how the survey on AI safety in Merida Yucatán, adapted from the study "Thousands of AI Authors on the Future of AI", provides valuable insights into the perception of risks, advancements, and regulations of artificial intelligence in a local context. Through comparative analysis, it was evident that there is high concern for scenarios such as the spread of false information and the manipulation of public opinion by AI systems, similar to concerns observed in the global context. However, a distinctive aspect was the emphasis on the need for AI experts to consider the needs and concerns of society, demonstrating the specific interests of the local environment.

Moreover, the analysis of the causes of AI progress revealed that research efforts, computing power, and data quality are seen as key drivers of AI advancement, closely paralleling those found in the original survey. This underscores the consistency in the perception of AI technological progress at both local and global levels.

The survey also showed strong consensus on the importance of government regulation and supervision of AI, with significant support for the creation of a specialized AI safety agency in Mexico. This finding is crucial as it highlights the need for an informed governance strategy specifically directed at managing AI risks while leveraging its potential benefits.

Regarding the performance and scope of the project, the methods and tools used for data collection and analysis proved effective in capturing and understanding complex perceptions of AI safety. However, areas of opportunity were identified to improve and complete the ETL process, which could further optimize data management and analysis in the future. Expanding the survey's focus to include more regions of Mexico could enrich our understanding of how AI is viewed and regulated across different cultural and economic contexts within the country.

We also considered stratifying the sample to capture differences between students, professors, and researchers, and even between those who have had contact with the topic of AI-Safety and those who have not.

This report not only contributes to the academic and political dialogue on AI safety in Mexico but also provides a solid foundation for future research aiming to contribute to the generation of knowledge in this area to promote the design of responsible AI technologies ethically aligned with the social and cultural values of Mexico and the world.

Acknowledgments. The first author would like to thank Eng. Angel Tenorio for his participation as the project manager of the team, to Valeria Ramirez, a data engineering student, who participated in the adaptation of the survey, as well as Alex Molina and Oscar Sánchez for his feedback during the development of this project. But most importantly, to Silvia Fernandez for her invitation to participate in this project and her support as an external supervisor. This study was partially funded by Secretaría de Ciencia, Humanidades, Tecnología e Innovación (SECIHTI) and Centro de Investigación en Ciencias de Información Geoespacial (CentroGEO), from Mexico.

References

1. AI Impacts: About ai impacts. `https://aiimpacts.org/about/` (2023), last accessed: 2024-04-03
2. [Anonymous]: Sesgos y discriminaciones sociales de los algoritmos en inteligencia artificial. Entretextos, (2024)
3. Cortés, V., Ruíz, P.: Reportes de política pública. Report, Jurídicas UNAM (2024), `https://archivos.juridicas.unam.mx/www/site/index/-8330.pdf`, Último acceso: 2024-04-03
4. Gentzkow, K.: Thousands of ai authors on the future of ai. arXiv, (2024)
5. Instituto Nacional de Transparencia, Acceso a la Información y Protección de Datos Personales (INAI): México: Inai recommendations on ai. `https://www.dataguidance.com/opinion/mexico-inai-recommendations-ai` (2024), Último acceso: 2024-04-03
6. Monreal Ávila, R.: Iniciativa con proyecto de decreto por el que se expide la ley federal que regula la inteligencia artificial (2024), `https://ricardomonrealavila.com/wp-content/uploads/2024/02/Inic_Morena_inteligencia_artificial.pdf`, Último acceso: 2024-04-03
7. Morales Ramírez, G.: Problemática antropológica detrás de la discriminación generada a partir de los algoritmos de la inteligencia artificial. MYE, vol. 34, no. 2, pp. 429–480 (2023)
8. Organisation for Economic Co-operation and Development (OECD): Oecd legal instruments. `https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449` (2024), last accessed: 2024-04-03
9. Presidencia de la República EPN: Inteligencia artificial en México. `https://www.gob.mx/epn/es/articulos/inteligencia-artificial-en-mexico` (2024), Último acceso: 2024-04-03
10. Pérez, J. V.: Survey on AI safety. `https://github.com/JanneVa/Survey_on_AI_Safety` (2025), gitHub repository. Last accessed: 2025-05-05

ISSN 1870-4069

Fuzzy Segmentation and Neural Classification of Cervical Cancer Samples

Esperanza Sánchez Domínguez, Edmundo Bonilla Huerta, José Fedérico Casco Vásquez, Roberto Morales Caporal, Crispín Hernández Hernández

Tecnológico Nacional de México/Campus Apizaco, Mexico

{d04370666, edmundo.bh, depi_dci, roberto.mc, crispin.hh}
@apizaco.tecnm.mx

Abstract. In this article, a fuzzy segmentation model and a neural classification model are proposed for cervical cancer samples. The fuzzy model segments contours and nuclei of cervical cancer samples, and the neural model is aimed at classifying these samples in two stages: initial and advanced. The results obtained are very competitive for the classification of normal and abnormal cells.

Keywords: Convolutional neural networks, cervical cancer, fuzzy logic, classification.

1 Introduction

Cancer is a very common disease in our times that can start in almost any tissue or organ of the human body and begin to spread by invading other vital organs (brain, liver, kidney, lungs, cervix, etc.). The most aggressive stage in humans is metastasis, which is one of the leading causes of death. Cancer is also known as a malignant tumor or neoplasia. An alternative to make a better diagnosis is to use DNA microarray technology and thus find the main genes that cause cancer.

Cancer is the second leading cause of death worldwide and caused approximately 10 million deaths in 2020, according to the World Health Organization [1] , accounting for one in six deaths worldwide. Women are more likely to develop breast, colon, lung, cervical, and thyroid cancer. In 2022, there were around 9439 new cases of cervical cancer and 4335 deaths, making it the second most common cause of diagnosis and death for women in Mexico. There are 5.7 deaths and 12.6 morbidities per 100,000. Nonetheless, the incidence rate or morbidity has dramatically dropped from 2012 onward.

The Mexican Cancerology Institute (INCan) reported around 195,500 different types of cancer cases diagnosed each year, and 46 percent of patients die from this cause [3]. In 2022, 847,716 deaths were recorded in Mexico: 10.6% were due to malignant tumors (89,574). The death rate from this cause has increased steadily, from 62.04 deaths per 100,000 people in 2012, to 68.92 in 2022 [4].

Due to its accessibility, affordability, and ease of use, the Pap test is the most widely utilized prophylactic measure worldwide. However, its drawback is the potential for errors to emerge while interpreting the material under a microscope. An alternative for a more precise diagnosis is to use DNA microarray technology to pinpoint the most important cancer genes; however, this method is quite expensive for underdeveloped countries.

2 Related Works

In this article [5] , a systematic review of the literature from 2008 to 2020 is presented regarding the main methods for the analysis and classification of cervical cancer samples. The authors emphasize the focus of this study on works that examine the size of the cell nucleus. Normal cell nuclei are smaller than those with abnormal ones. The abnormal cell nuclei usually show disproportionate growth.

In this study [6] uses three deep learning approaches to analyze and predict cervical cancer samples. The models are validated using the technique of cross-validation and using statistical metrics. Reports indicate that the ResNet50V2 model demonstrates the highest precision.

This study [7] suggests a group of machine learning classifiers for the effective and trustworthy use of medical data in the detection of cervical cancer. The proposed approach outperforms several state-of-the-art techniques by achieving 98.06% and 95.45% accuracy for two well-known datasets, respectively. The results show that the proposed ensemble classifier can accurately identify cervical cancer and enhance diagnosis and treatment.

The novel model based on a Salp Swarm Algorithm (SSA) is proposed in [8] to improve the diagnosis of cervical cancer. This model uses well-known pre-trained models of deep learning to tackle feature extraction. Later, these predictions are integrated and optimized using SSA. The model achieves 99.48% accuracy on the Mendeley LBC dataset and 95.23% on the BloodMNIST Benchmark data set, respectively.

This research [9], addresses an approach that combines three machine learning models in a stacked ensemble voting classifier, complemented by a KNN imputation to deal with missing values. The developed model achieves performance with a precision of 0.9941, an accuracy of 0.98, a sensitivity of 0.96, and an F1 score of 0.97. For the analysis of this information, they divide model training with 70% and 30% for model testing. The system uses the XGB+RF+ETC ensemble model, which trains the XGB, RF and ETC models independently on the same data.

The article reported in [10], proposes a way to use Bradley local thresholding to find and separate nucleus cell areas in Pap smear images. The method includes color adjustment, k-means, and a modification algorithm. The nucleus cell region was segmented significantly and efficiently, with an F-measure of 98.62 percent, a sensitivity of 99.13 percent, and an accuracy of 97.96 percent.

This study [11] employs conventional machine learning (ML) concepts and a number of traditional machine learning methods, including decision trees (DT), support vector machines (SVM), logistic regression (LR), support vector machines (SVM), and K-nearest neighbors (KNN), the focus of this investigation has been cervical cancer. When it comes to cervical cancer prediction, the random forest (RF), decision tree (DT), adaptive, and gradient boosting algorithms have all produced the maximum classification score of 100% SVM, on the other hand, it has been determined to have 99% accuracy.

In this work [12], examined various supervised machine learning methods for the early detection of cervical cancer. The UCI repository's cervical cancer dataset was used to train the machine learning model. The different methods were evaluated using this dataset, which comprised 858 cervical cancer patients with 36 risk factors and one outcome variable. The XG-boost tree, random tree, logistic tree, SVM, Bayesian network, and artificial neural network were the six classification techniques used in this study. To assess the effectiveness and precision of the classifiers, all models were trained with and without a feature selection technique. Three feature selection algorithms—LASSO regression, wrapper technique, and relief rank—were employed. XG Boost, with all its features, achieved the highest accuracy of 94.94%.

3 Medical Database

To carry out the experiments, a bank of 1000 images donated by HOSPITAL GENERAL DE ZONA (HGZ 1) (IMSS, Tlaxcala) was used. The images are in JPG format, taken with optical zoom ranging from 40x to 100x with a resolution of 2592 x 1944 pixels, and are images of Pap smear tests with abnormal cells. This dataset is available on GitHub[1].

4 Methodology

The proposed model for the analysis and classification of abnormal cervical cells is shown. This model was developed using the Matlab tool version R2017a, with Toolboxes: Fuzzy Logic and Image Processing. A MacOS Sierra computer was used, with a processor: 2.9 GHz Intel Core i7 and memory: 16 GB (see figure 1).

4.1 Problem Domain

Cancer is a serious disease that affects millions of people around the world. Cancer cell sample analysis is essential for its diagnosis and treatment. This research proposes the use of fuzzy logic and artificial neural networks in the analysis of cancer cell samples. Its application in the analysis of cancer cells has the potential to improve the accuracy and efficiency of the diagnosis of images taken by Pap smear.

[1] https://github.com/EsperanzaSD/Abnormalcells

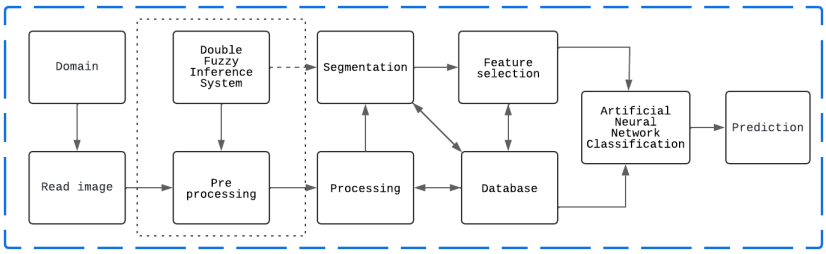

Fig. 1. Model for analysis and classification of Pap Smear samples. Source: Prepared by the authors.

Currently, the Pap test to detect cervical cancer is one of the most widely used techniques; however, these tests are sensitive to generating inaccurate information, from taking sample instruments to the interpretation of the images generated in the process. To support the diagnosis, is suggested the use of techniques with different methods that allow a statistical analysis and find characteristics in cells that are affected. The information obtained from histogram analysis can be used to extract relevant features from the image and perform in-depth interest area analysis, such as the cell nucleus and the cell.

4.2 Image Reading

Pap smear cytopathology inspection is a diagnostic technique frequently used to detect cervical cancer. A doctor collects cells from a patient's cervix with surgical brushes and then places the exfoliated cells on a glass slide during a cervical cytopathology examination. Cytopathologists use a microscope to look for malignant tumors, with each slide containing thousands of cells (Bedell et al., 2020).

For manual detection of Pap smear photos, pathologists must inspect each sub-image on a separate slide under a microscope to diagnose disease. Finding diseased cells on Pap smear cell slides can be challenging due to the similar appearance and size of some type of nuclei cells. Specialists can diagnose diseases by looking at these cells, but it depends on their experience and the cause of the disease (Sankaranarayanan et al., 2012). In Pap smear cytology, advanced interpretation techniques focus on the use of artificial intelligence to improve the accuracy of diagnoses. Using fuzzy logic and artificial neural networks makes it possible to find small patterns in tomographic images. This makes it easier to find problems earlier and detect the difference between benign and malignant lesions. Putting together systems that look at cell textures and morphology also helps with a more detailed analysis of samples, which makes for more reliable results (Alsmariy et al., 2020). Ten Pap smear images in .JPG format were used to carry out the experiments. Artificial intelligence is used for image segmentation. The image is separated into its RGB components for image processing, as it allows

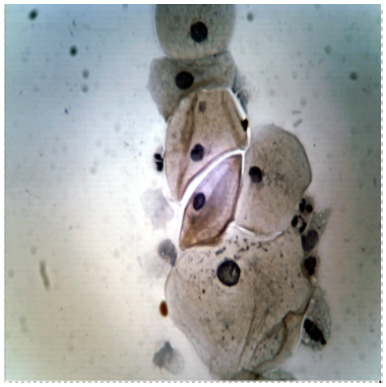

Fig. 2. Cervical cancer sample. Source: Instituto Mexicano del Seguro Social, Zona 1. Tlaxcala.

an image to be broken down into its three primary color channels (RGB): red, green, and blue. These channels are essential for the analysis and manipulation of the different characteristics of an image, such as color intensity, saturation, and brightness. Through this separation, specific patterns can be identified in each channel and individualized adjustments can be made to each of them, which is useful for information extraction. An example of a cervical cancer image is shown in the following image (see figure 2).

4.3 Pre-Processing

In this stage, a region of interest (ROI) is first defined to analyze the images of cervical cancer cells. Second, a gray level conversion is applied to facilitate the segmentation task. Third, a 3x3 convolution mask is applied to obtain the median of the image. This process smooths the gray tones of the image and reduces some noise caused during the collection stage.

4.4 Fuzzy Segmentation

Two fuzzy processing modules have been defined, the first is to detect the cell nuclei and the second is to detect the cell contour. Applying the first FIS, a fuzzy input and a fuzzy Segmented have been defined. The fuzzy input and Segmented have been partitioned into 5 fuzzy subsets, called VD: Very Dark, D: Dark, Medium, C: Clear, and VC: Very Clear. The rules are defined as follows:

1. IF Pre-processed-Image is VD THEN Segmented-Nuclei is VD,
2. IF Pre-processed-Image is D THEN Segmented-Nuclei is D,
3. IF Pre-processed-Image is M THEN Segmented-Nuclei is M,
4. IF Pre-processed-Image is C THEN Segmented-Nuclei is D,
5. IF Pre-processed-Image is VC THEN Segmented-Nuclei is M.

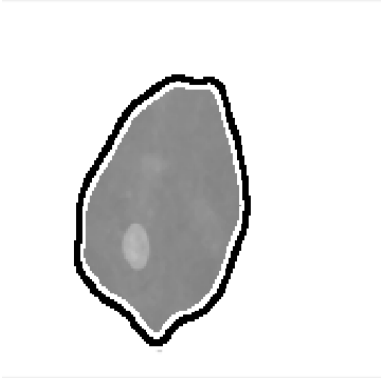

Fig. 3. Feature selection stage of segmented contour of normal cell of Pap smear sample.

The purpose of these rules is to emphasize darker shades in order to facilitate cell nuclei segmentation. The following rules defined for the second fuzzy processing aim to enhance contrast in dark regions and soften light regions:

1. IF Pre-processed-Image is VD THEN Segmented-Contour is VD,
2. IF Pre-processed-Image is D THEN Segmented-Contour is VD,
3. IF Pre-processed-Image is M THEN Segmented-Contour is D,
4. IF Pre-processed-Image is C THEN Segmented-Contour is M,
5. IF Pre-processed-Image is VC THEN Segmented-Contour is C.

Using this second fuzzy inference module, the contrast of dark regions is significantly improved, and light regions are smoothed, thus facilitating cell contour segmentation.

4.5 Feature Selection

The results of double fuzzy processing are what make up the feature selection process. This means that we have the segmented cell as well as the cell nucleus of all cervical cancer samples in the image database. Figure 3 illustrates the feature selection of an early-stage cancer sample.

4.6 Convolutional Neural Network

In this work, we propose a Convolutional Neural Network (CNN) to classify 200 samples of cancer contours of the Pap smear cervix. 100 samples are normal and 100 samples are abnormal, respectively. The proposed architecture consists of several convolutional layers (see Figure 4). The images from input data (1944x2592x3) of layer 0 were pre-processed and rescaled to 130x130x3, it is recommended that images must be square, to reduce the cost in CNN architecture. The follow step is to feed the first layer of convolution that consist of 32 filters of 3x3 and to move this filters to scan the image, for this task we

define a stride of size 1. The result of this convolutional process is a matrix of 130x130x64. To reduce this data preserving the most relevant features of data is to use a pooling layer stacked after the convolutional layer.

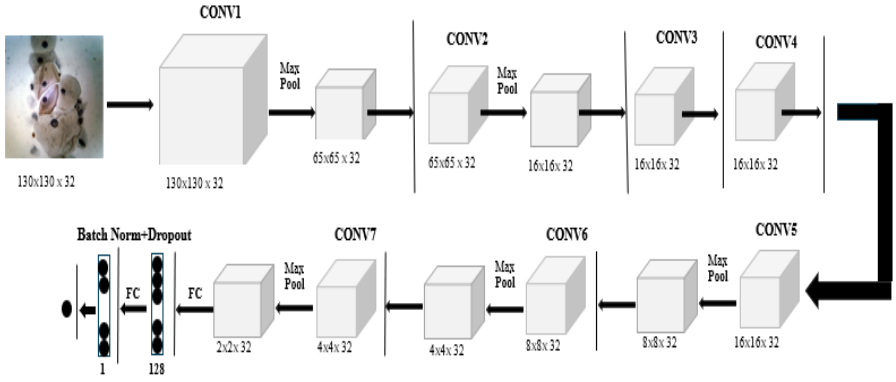

Fig. 4. The architecture of CNN.

```
Model: "sequential"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
conv2d (Conv2D)              (None, 130, 130, 32)      896
_________________________________________________________________
max_pooling2d (MaxPooling2D) (None, 65, 65, 32)        0
_________________________________________________________________
conv2d_1 (Conv2D)            (None, 65, 65, 32)        9248
_________________________________________________________________
max_pooling2d_1 (MaxPooling2 (None, 32, 32, 32)        0
_________________________________________________________________
conv2d_2 (Conv2D)            (None, 32, 32, 32)        9248
_________________________________________________________________
max_pooling2d_2 (MaxPooling2 (None, 16, 16, 32)        0
_________________________________________________________________
conv2d_3 (Conv2D)            (None, 16, 16, 32)        9248
_________________________________________________________________
conv2d_4 (Conv2D)            (None, 16, 16, 32)        9248
_________________________________________________________________
conv2d_5 (Conv2D)            (None, 16, 16, 32)        9248
_________________________________________________________________
max_pooling2d_3 (MaxPooling2 (None, 8, 8, 32)          0
_________________________________________________________________
conv2d_6 (Conv2D)            (None, 8, 8, 32)          9248
_________________________________________________________________
max_pooling2d_4 (MaxPooling2 (None, 4, 4, 32)          0
_________________________________________________________________
conv2d_7 (Conv2D)            (None, 4, 4, 32)          9248
_________________________________________________________________
max_pooling2d_5 (MaxPooling2 (None, 2, 2, 32)          0
_________________________________________________________________
flatten (Flatten)            (None, 128)               0
_________________________________________________________________
dense (Dense)                (None, 128)               16512
_________________________________________________________________
activation (Activation)      (None, 128)               0
_________________________________________________________________
dropout (Dropout)            (None, 128)               0
_________________________________________________________________
dense_1 (Dense)              (None, 1)                 129
_________________________________________________________________
activation_1 (Activation)    (None, 1)                 0
=================================================================
Total params: 82,273
Trainable params: 82,273
Non-trainable params: 0
```

Fig. 5. Summary of CNN proposed.

```
Epoch 95/100
1/1 [==============================] - 5s 5s/step - loss: 0.0191 - accuracy: 1.0000 - val_loss: 1.0602 - val_accuracy: 0.7000
Epoch 96/100
1/1 [==============================] - 5s 5s/step - loss: 0.0898 - accuracy: 0.9500 - val_loss: 0.9719 - val_accuracy: 0.8500
Epoch 97/100
1/1 [==============================] - 5s 5s/step - loss: 0.0748 - accuracy: 1.0000 - val_loss: 0.5246 - val_accuracy: 0.8500
Epoch 98/100
1/1 [==============================] - 7s 7s/step - loss: 0.0733 - accuracy: 1.0000 - val_loss: 1.1425 - val_accuracy: 0.7500
Epoch 99/100
1/1 [==============================] - 5s 5s/step - loss: 0.0437 - accuracy: 1.0000 - val_loss: 0.7489 - val_accuracy: 0.8500
Epoch 100/100
1/1 [==============================] - 5s 5s/step - loss: 0.0268 - accuracy: 1.0000 - val_loss: 0.7177 - val_accuracy: 0.8500
```

Fig. 6. Results of training and testing.

The max-pooling layer is defined a 2x2 filter and a stride of 2, the result of this process is a 64x64x64 size feature matrix. We use many Max-pooling layers in our architecture to reduce significantly the CNN time execution to obtain better results.

We also include a dropout layer in the architecture to improve the generalization capacities of our model. In this step, the dropout factor is fixed to 0.5. Thus, we have the basic CNN model, which includes seven convolutional layers, a max-pooling layer and a dropout layer to create a convolutional cycle.

After seven convolutions the resultant matrix of Pap smear images matrix becomes a $2 \times 2 \times 32$. This resultant matrix is flattened to obtain a feature vector of size 128. The last layer contains only 2 neurons with a sigmoid function. Different cycles are proposed in the architecture to reduce the feature size of Pap smear samples and thus to facilitate binary classification (initial and advanced). CNN proposed summary is shown in figure 5.

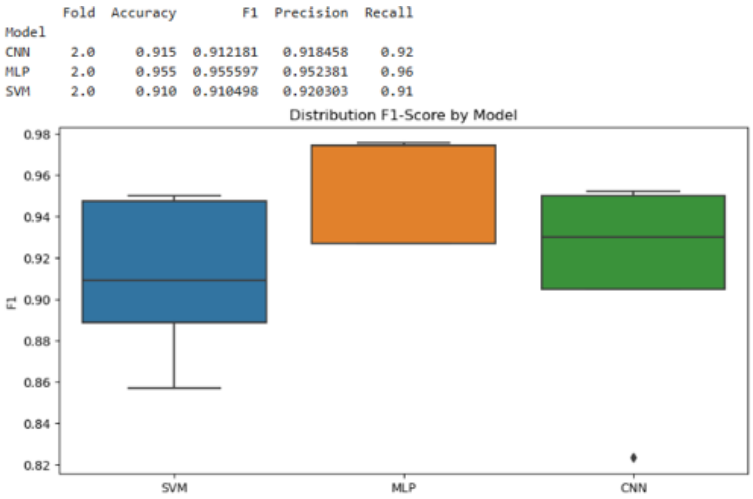

Fig. 7. Results of samples.

5 Results

Our model achieves 100% accuracy in the training data before reaching the defined number of epochs, gradually decreasing accuracy at epoch 97, and stabilizing again at the 98 training epoch. We define a limit of 100 epoch for this model (See Figure 6).

In this experimentation protocol, we use the binary cross entropy loss and the Adam optimizer to validate CNN.

6 Comparaison with Other Works

To validate our model with other works reported in the literature, we selected the well-known SIPAKMED biomedical database, which contains 4,049 cell images from five types of cervical cancer. The authors [13] simplified the classification process for these five classes into two (normal and abnormal). We applied our model, and the results are shown in figure 7.

Table 1. Comparison with other similar works using only two classes.

Method	accuracy	precision	recall	F_1 score
AlexNet, GAN [14]				
93.8%	47.8%	95.9%	93.6%	94.1%
CNN-PCA [15]				
–	–	–	ResNet-152 [2] 94.8%	–
–	–	MLP-Nuclei [13] 78.8%	–	–
–	MLP-Cytoplasm [13] 83.4%	–	–	– SVM-Nuclei [13]
88.5%	–	–	–	91.6%
SVM-Cytoplasm [13]				
–	–	–	CNN-RGB [13] 95.3%	–
–	–	our model and CNN 91.5%	91.8%	92.0%
91.8% our model and MLP	**95.5%**	**95.2%**	**96.0%**	**95.5%** our model and SVN
91.0%	92.0%	91.0%	91.0%	

Using our model that combines two fuzzy segmentations (nuclei and contour) and three different classifiers (CNN, MLP, and SVM), we find that the MLP

classifier gives the best results, achieving a precision of 95%, an F1 score of 95.5%, a precision of 95.23%, and a recall of 96.0% (see figure 7).

In Table 1, we summarize our model's results alongside other similar studies that used the SIPAKMED dataset. In Table 1 are summarized results of our model with other similar works reported in the literature using the SIPAKMED dataset. We evaluated the performance of the proposed model using the 5-fold cross-validation method. We observed that our model for binary classification outperforms other comparable works using the same dataset.

7 Conclusions

We present in this paper a double fuzzy inference system that can help to find two features about Pap smear samples for cervix cancer: contour nuclei and contour cells. To obtain better results using a convolutional neural network (CNN), we use in this work the contour of normal and abnormal samples and thus reduce the computation time. In future work, we intend to perform a multi-classification of cervical cancer samples for monitoring at each stage.

References

1. World Health Organization: Cancer. (2019, July 12). Retrieved from https://onx.la/4e793
2. Tripathi, A., Arora, A., Bhan, A.: Classification of cervical cancer using Deep Learning Algorithm. In: 5th International Conference On Intelligent Computing And Control Systems (ICICCS), pp. 1210–1218 (2021)
3. Instituto Nacional de Cancerología (INCan): 294: México registra al año más de 195 mil casos de cáncer: Secretaría de Salud. (2023) Retrieved from: https://onx.la/df378
4. Instituto Nacional de Estadística y Geografía Estadística a propósito del día mundial contra el cáncer. (2024) Retrieved from https://onx.la/b6708
5. Wei, L., Mustafa, W., Jamlos, M., Idrus, S., Sahabudin, M.: Cervical cancer classification using image processing approach: A review. In: IOP Conference Series: Materials Science And Engineering, 917, 012068 (2020) https://doi.org/10.1088/1757-899X/917/1/012068.
6. Devaraj, S., Madian, N., Menagadevi, M., Remya, R.: Deep learning approaches for analysing papsmear images to detect cervical cancer. In: Wireless Personal Communications, 135, 81–98 (2024) https://doi.org/10.1007/s11277-024-10986-8.
7. Ali, M., Hossain, M., Kona, M., Nowrin, K., Islam, M.: An ensemble classification approach for cervical cancer prediction using behavioral risk factors. Healthcare Analytics, 5 pp. 100324 (2024). https://doi.org/10.1016/j.health.2024.100324.
8. Bilal, O., Asif, S., Zhao, M., Khan, S., Li, Y.: An amalgamation of deep neural networks optimized with Salp swarm algorithm for cervical cancer detection. Computers And Electrical Engineering, 123, pp. 110106 (2025) https://doi.org/10.1016/j.compeleceng.2025.110106.
9. Aljrees, T.: Improving prediction of cervical cancer using KNN imputer and multi-model ensemble learning. Plos One, 19, e0295632 (2024) https://doi.org/10.1371/journal.pone.0295632.

10. Halim, A., Mustafa, W., Nasir, A., Ismail, S., Alquran, H.: Nucleus Cell Segmentation on Pap Smear Image using Bradley Modification Algorithm. Neural Network World, (2024) https://doi.org/10.14311/NNW.2024.34.003.
11. Al Mudawi, N., Alazeb, A.: A model for predicting cervical cancer using machine learning algorithms. Sensors, 22, 4132 (2022) https://doi.org/10.3390/s22114132.
12. Kumawat, G., Vishwakarma, S., Chakrabarti, P., Chittora, P., Chakrabarti, T., Lin, J.: Prognosis of cervical cancer disease by applying machine learning techniques. Journal of Circuits, Systems and Computers, 32, 2350019 (2023) https://doi.org/10.1142/S0218126623500196.
13. Plissiti, M., Dimitrakopoulos, P., Sfikas, G., Nikou, C., Krikoni, O., Charchanti, A.: Sipakmed: A new dataset for feature and image based classification of normal and pathological cervical cells in pap smear images. In: 2018 25th IEEE International Conference On Image Processing (ICIP), pp. 3144–3148 (2018)
14. Yu, S., Zhang, S., Wang, B., Dun, H., Xu, L., Huang, X., Shi, E., Feng, X.: Generative adversarial network based data augmentation to improve cervical cell classification model. Math. Biosci. Eng., 18, 1740–1752 (2021)
15. Al-asbaily, S., Almoshity, S., Younus, S., Bozed, K.: Classification of Cervical Cancer using Convolutional Neural Networks. In: 2024 IEEE 4th International Maghreb Meeting Of The Conference On Sciences And Techniques Of Automatic Control And Computer Engineering (MI-STA), pp. 735–739 (2024)

ISSN 1870-4069

Implementación y optimización del modelo YOLO para el reconocimiento de elementos de seguridad en tiempo real

Joctan Maceda Hernández, Yolanda Moyao Martínez, David Eduardo Pinto Avendaño, Beatriz Beltrán Martínez, José Andrés Vázquez Flores

Benemérita Universidad Autónoma de Puebla, México

joctan.maceda@alumno.buap.mx, yolanda.moyao@correo.buap.mx, david.pinto@correo.buap.mx, beatriz.beltran@correo.buap.mx, andres.vazquez@correo.buap.mx

Resumen. Este trabajo presenta un estudio detallado sobre la implementación y optimización de un sistema de reconocimiento de objetos en tiempo real basado en la arquitectura YOLOv10, aplicado a la clasificación de elementos de seguridad. La detección eficiente de entidades como policías, guardias nacionales y civiles armados en entornos urbanos es fundamental para sistemas de vigilancia, control público y toma de decisiones en tiempo real, especialmente en contextos de seguridad ciudadana. Para lograr este objetivo, se desarrolló un conjunto de datos personalizado, mejorado mediante la herramienta Grounding DINO para la generación automatizada de etiquetas, reduciendo significativamente la intervención manual. El modelo fue entrenado usando GPU en Google Colaboratory, aplicando técnicas de optimización que permitieron incrementar tanto su precisión como su velocidad. Se realizó una comparación entre YOLOv10 y versiones anteriores del mismo modelo, destacando las mejoras introducidas en términos de precisión, eficiencia y arquitectura. YOLOv10 se distingue por su enfoque de entrenamiento sin NMS (Non-Maximum Suppression), asignaciones de etiquetas duales y diseño holístico, consolidándose como una herramienta eficaz para tareas de reconocimiento en tiempo real donde la clasificación de elementos de seguridad es crítica.

Palabras clave: YOLO, detección de objetos, visión por computadora, inteligencia artificial, seguridad urbana.

Implementation and Optimization of the YOLO Model for Real-time Security Element Recognition

Abstract. This paper presents a detailed study on the implementation and optimization of a real-time object recognition system based on

the YOLOv10 architecture, applied specifically to the classification of security-related elements. The efficient detection of entities such as police officers, national guards, and armed civilians in urban environments is crucial for surveillance systems, public safety monitoring, and real-time decision-making, particularly in contexts of civil security. To achieve this, a custom dataset was developed and enhanced using the Grounding DINO tool for automated label generation, significantly reducing manual annotation effort. The model was trained using GPU resources in Google Colaboratory, applying optimization techniques that improved both accuracy and speed.8 A comparison was made between YOLOv10 and its previous versions, highlighting improvements in precision, efficiency, and network architecture. YOLOv10 stands out for its training approach without Non-Maximum Suppression (NMS), dual label assignments, and holistic design, establishing itself as an effective tool for real-time object detection tasks where classifying security-related elements is critical.

Keywords: YOLO, object detection, computer vision, artificial intelligence, urban security.

1. Introducción

La detección de objetos en tiempo real ha cobrado relevancia en múltiples áreas de la inteligencia artificial y la visión por computadora. Su implementación es fundamental en sectores como la seguridad, la automatización industrial y el análisis de tráfico. Tradicionalmente, los métodos de detección requerían múltiples pasos de procesamiento, lo que los hacía poco eficientes para aplicaciones en tiempo real. La introducción de YOLO (You Only Look Once) ha revolucionado este campo al permitir inferencias rápidas con una única pasada sobre la imagen.

En este trabajo, exploramos la implementación de YOLOv10 y su optimización para la detección de objetos en entornos urbanos, con un enfoque particular en la identificación de elementos de seguridad pública. Se utilizó Grounding DINO como herramienta para mejorar la calidad del dataset, automatizando parcialmente el proceso de anotación y reduciendo así la intervención manual. Las imágenes fueron inicialmente recolectadas sin etiquetas, y luego fueron procesadas por Grounding DINO para generar anotaciones automáticas que posteriormente fueron validadas manualmente, logrando una base de datos robusta para el entrenamiento del modelo. La Figura 1 muestra una representación esquemática del proceso de detección con YOLOv10. En este esquema se visualizan las etapas clave del modelo, desde la entrada de la imagen hasta la generación de predicciones finales. YOLOv10 emplea un enfoque optimizado que reduce la necesidad de post-procesamiento mediante la eliminación de NMS, lo que mejora tanto la precisión como la velocidad de detección.

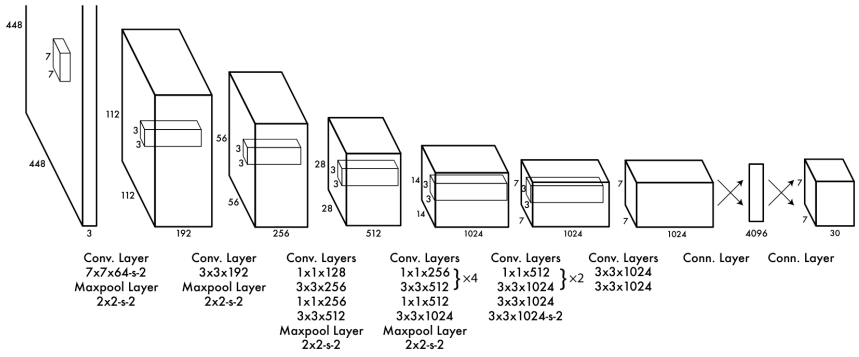

Fig. 1. Representación esquemática del proceso de detección con YOLOv10. Imagen tomada de [1].

2. Estado del arte

Las redes neuronales convolucionales (CNN) han sido fundamentales en la evolu-ción de la visión por computadora, permitiendo la extracción de características espaciales en imágenes mediante filtros y capas de pooling. En el contexto de la detección de objetos, los primeros enfoques se basaban en modelos como R-CNN (Girshick et al., 2014), que dividían el proceso en dos fases: generación de regiones de interés y clasificación de dichas regiones. Aunque estos modelos lograban buenos resultados, su tiempo de procesamiento era demasiado alto para aplicaciones en tiempo real (Ren et al., 2015). La llegada de YOLO en 2016 (Redmon, 2016) supuso una revolución en la detección de objetos, al tratar el problema como una regresión única sobre coordenadas y probabilidades de clase. En lugar de analizar múltiples regiones de una imagen por separado, YOLO divide la imagen en una cuadrícula y predice varias cajas delimitadoras por celda, mejorando drásticamente la velocidad de detección sin comprometer significativamente la precisión.

Esto permitió su adopción en aplicaciones críticas como la videovigilancia y los vehículos autónomos, donde la respuesta en tiempo real es esencial (Redmon and Farhadi, 2018). Desde su introducción, YOLO ha experimentado múltiples mejoras. YOLOv3 optimizó su arquitectura con Darknet-53, incrementando su precisión en detección de objetos pequeños. YOLOv4 introdujo CSPDarknet y técnicas avanzadas de aumento de datos para mejorar el rendimiento en escenarios complejos (Bochkovskiy et al., 2020). La versión YOLOv5, desarrollada por Ultralytics, se centró en la facilidad de uso y eficiencia computacional. La Figura 2 muestra la evolución de YOLO desde su primera versión hasta YOLOv10, destacando las mejoras clave en cada iteración. Esta evolución ha sido clave para lograr una detección más rápida y precisa, abordando los principales desafíos que enfrentan los modelos tradicionales.

Version	Date	Contributions	Structure
version 1	2015	Detector of single-shot objects	Dark web
version 2	2016	Multiscale training, dimensional clustering	Dark web
version 3	2018	SPP Block, Darknet-53	Dark web
version 4	2020	Mish-based activation, CSPDarknet-53 main structure	Dark web
v5	2020	Un anchored detection, SWISH-based activation, PANet	PyTorch
v6	2022	Self-care, detection of objects without anchoring	PyTorch
v7	2022	Transformers, E-ELAN Repaetional	PyTorch
v8	2023	GAN, unanchorted detections	PyTorch
v9	2024	Programmable gradient information (PGI), generalised efficient layer aggregation network (GELAN)	PyTorch
v10	2024	NMS-free training approach, dual-label assignments, holistic model design for greater accuracy and efficiency	PyTorch

Fig. 2. Evolución de YOLO desde YOLOv1 hasta YOLOv10.

YOLOv10 representa la iteración más reciente, incorporando mejoras en la arquitectura de la red neuronal, la gestión de anclas y la optimización de pesos preentrenados, ofreciendo un mejor balance entre velocidad y precisión (Ultralytics, 2024). Sin embargo, uno de los mayores desafíos en la detección de objetos sigue siendo la generación de datasets de entrenamiento bien etiquetados. Para abordar este problema, se ha explorado la integración de Grounding DINO, un modelo que permite la generación automatizada de etiquetas con menor intervención humana, reduciendo el costo y el tiempo de anotación de datos (Liu et al., 2023).

Desde YOLOv5 (Ultralytics), cada versión de YOLO ha sido lanzada con modelos de diferentes tamaños (nano, pequeño, mediano, grande y extragrande), adaptados para distintos niveles de precisión y velocidad de inferencia. YOLOv10 sigue esta tendencia y ofrece una variedad de modelos previamente entrenados, desarrollados por investigadores de la Universidad de Tsinghua. Estos modelos mejoran el rendimiento en términos de latencia y Average Precision (AP) en comparación con versiones anteriores. La Figura 3 muestra una comparación del rendimiento de YOLOv10 con versiones anteriores en términos de latencia y número de parámetros. Se puede observar que YOLOv10 ha sido optimizado para mejorar la precisión mientras reduce el tiempo de inferencia, lo que lo hace ideal para aplicaciones en tiempo real.

Además, los modelos previamente entrenados de YOLOv10 están disponibles en distintos tamaños, lo que permite seleccionar la mejor opción dependiendo de las necesidades del proyecto. La Figura 4 detalla estos modelos junto con sus características principales.

Estas optimizaciones han permitido que YOLOv10 supere a sus predecesores en precisión y eficiencia, consolidándose como una de las mejores opciones para tareas de detección en tiempo real.

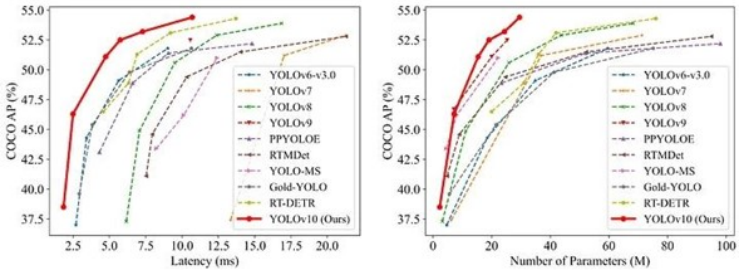

Fig. 3. Comparación de rendimiento: latencia (izquierda) y número de parámetros (derecha) en todos los modelos YOLO.

Tabla 1. Distribución de imágenes por fuente.

Fuente de Imágenes	**Cantidad**
Grabaciones de vigilancia	1,200
Fotografías urbanas	1,000
Material de código abierto	800
Total	**3,000**

3. Metodología

3.1. Selección del dataset

El conjunto de datos utilizado en este estudio fue generado con la plataforma Roboflow, conteniendo un total de 3,000 imágenes etiquetadas manualmente y refinadas con Grounding DINO. Se recopilaron imágenes de entornos urbanos, enfocándose en la detección de policías, guardias nacionales y civiles armados.

En la Tabla 1 se muestra la distribución de las imágenes recopiladas según su fuente.

Para garantizar un entrenamiento balanceado, se dividió el conjunto de datos en subconjuntos de entrenamiento, validación y prueba, como se muestra en la Tabla 2.

Además, el dataset contenía un total de tres clases de objetos como se muestra en la Tabla 3.

Durante el preprocesamiento del dataset, se aplicaron diversas técnicas de normalización para mejorar la calidad de las imágenes antes del entrenamiento. Estas incluyeron:

- **Corrección de color:** Ajuste del balance de blancos y normalización del histograma para mejorar el contraste.
- **Reducción de ruido:** Aplicación de filtros Gaussianos y mediana para eliminar ruido de la imagen.
- **Conversión a escala de grises:** En algunos casos, para mejorar la detección de contornos y texturas clave.

Model	Params (M)	FLOPs (G)	APval (%)	Latency (ms)	Latency (Forward) (ms)
YOLOv6-3.0-N	4.7	11.4	37.0	2.69	**1.76**
Gold-YOLO-N	5.6	12.1	**39.6**	2.92	1.82
YOLOv8-N	3.2	8.7	37.3	6.16	1.77
YOLOv10-N	**2.3**	**6.7**	39.5	**1.84**	1.79
YOLOv6-3.0-S	18.5	45.3	44.3	3.42	2.35
Gold-YOLO-S	21.5	46.0	45.4	3.82	2.73
YOLOv8-S	11.2	28.6	44.9	7.07	**2.33**
YOLOv10-S	**7.2**	**21.6**	**46.8**	**2.49**	2.39
RT-DETR-R18	20.0	60.0	46.5	**4.58**	**4.49**
YOLOv6-3.0-M	34.9	85.8	49.1	5.63	4.56
Gold-YOLO-M	41.3	87.5	49.8	6.38	5.45
YOLOv8-M	25.9	78.9	50.6	9.50	5.09
YOLOv10-M	**15.4**	**59.1**	**51.3**	4.74	4.63
YOLOv6-3.0-L	59.6	150.7	51.8	9.02	7.90
Gold-YOLO-L	75.1	151.7	51.8	10.65	9.78
YOLOv8-L	43.7	165.2	52.9	12.39	8.06
RT-DETR-R50	42.0	136.0	53.1	9.20	9.07
YOLOv10-L	**24.4**	**120.3**	**53.4**	**7.28**	**7.21**
YOLOv8-X	68.2	257.8	53.9	16.86	12.83
RT-DETR-R101	76.0	259.0	54.3	13.71	13.58
YOLOv10-X	**29.5**	**160.4**	**54.4**	**10.70**	**10.60**

Fig. 4. Modelos previamente entrenados disponibles para YOLOv10. Tabla tomada del sitio web de Ultralytics.

Además, se llevaron a cabo procesos de aumentación de datos, incluyendo rotaciones, cambios de iluminación y escalado, con el objetivo de robustecer la capacidad de generalización del modelo, generando un total de 9,000 imágenes después de la aumentación.

El dataset se almacenó en la nube utilizando Roboflow, lo que permitió una gestión eficiente y la posibilidad de realizar modificaciones en tiempo real sin necesidad de descargar archivos localmente. La API de Roboflow fue clave en la

Tabla 2. División del dataset.

Conjunto de Datos	**Cantidad de Imágenes**	**Porcentaje**
Entrenamiento	6,300	70 %
Validación	1,800	20 %
Prueba	900	10 %
Total	**9,000**	**100 %**

Tabla 3. Clases de objetos en el dataset.

Clases de Objetos	**Cantidad de Ejemplos**
Policías	3,000
Guardia Nacional	5,000
Civiles Armados	1,000

automatización del flujo de datos, asegurando que las imágenes estuvieran en el formato requerido para YOLOv10.

3.2. Entrenamiento del modelo

El modelo fue entrenado en Google Colaboratory utilizando una GPU Tesla T4. Los hiperparámetros, fueron configurados como se muestra en la Tabla 4.

Tabla 4. Hiperparámetros utilizados en el entrenamiento de YOLOv10.

Hiperparámetro	**Valor**	**Justificación**
Tasa de aprendizaje	0.01	Ajustada experimentalmente
Batch size	16	Balance entre memoria y estabilidad
Número de épocas	100	Evita sobreajuste
Tamaño de imagen	640 px	Balance entre precisión y velocidad

Estos valores fueron determinados mediante experimentación, utilizando una estrategia de ajuste progresivo de hiperparámetros. Se realizaron pruebas preliminares con distintos valores y se analizó su impacto en las métricas de rendimiento, como la precisión y la velocidad de inferencia.

El entrenamiento se llevó a cabo en 3 etapas:

1. **Pre-entrenamiento:** Ajuste inicial del modelo con un subconjunto del dataset para validar la estabilidad de los hiperparámetros.
2. **Entrenamiento completo:** Optimización con el conjunto completo de imágenes.
3. **Validación:** Evaluación del rendimiento en un subconjunto independiente de imágenes no vistas por el modelo.

Se aplicaron estrategias de ajuste de hiperparámetros, como la reducción progresiva de la tasa de aprendizaje y el uso de técnicas de regularización como Dropout y Batch Normalization para mejorar la estabilidad del modelo.

3.3. Implementación técnica

El sistema de detección se compone de las siguientes etapas:

1. **Captura de imágenes y videos:** Implementación de un pipeline con OpenCV para la captura de imágenes en tiempo real desde fuentes de video en streaming.
2. **Preprocesamiento:** Aplicación de filtros para mejorar la calidad de las imágenes antes de la inferencia, incluyendo eliminación de ruido y ajuste de contraste.
3. **Inferencia con YOLOv10:** Uso de la biblioteca Ultralytics para ejecutar el modelo y obtener predicciones en tiempo real con optimización en GPU.
4. **Post-procesamiento:** Aplicación de algoritmos de supresión de no-máximos (NMS) para eliminar detecciones redundantes y mejorar la precisión de los resultados.
5. **Visualización de resultados:** Generación de gráficos interactivos con Matplotlib y Supervision para evaluar el desempeño del modelo en distintos escenarios.

Para validar la efectividad del modelo, se realizaron pruebas en entornos controlados y en escenarios del mundo real. Se implementó un sistema de evaluación basado en métricas de precisión, recall e IoU, garantizando una medición objetiva del desempeño del sistema de detección.

4. Resultados y análisis

Los modelos entrenados con datasets personalizados fueron comparados con modelos preentrenados. Los resultados, presentados en la Tabla 5, indican que YOLOv10 optimizado con Grounding DINO obtuvo una precisión del 85 %, con una mejora significativa en la detección de objetos específicos en entornos urbanos.

Tabla 5. Métricas de evaluación del modelo YOLOv10.

Métrica	**Valor**
Precisión	85 %
Recall	83 %
F1-score	84 %
IoU (Intersection over Union)	0.78

La Figura 5 muestra la matriz de confusión obtenida con un dataset de pesos preestablecidos sin nuestro entrenamiento, lo que permite visualizar la tasa de aciertos y errores. En contraste, la Figura 6 presenta la matriz de confusión obtenida con un dataset etiquetado con Grounding DINO, evidenciando una mejor distribución de las predicciones correctas. La Figura 7 ilustra la curva de confidencialidad, que representa la relación entre la confianza del modelo y la precisión en la detección. Por otro lado, la Figura 8 muestra la curva de precisión, donde se observa cómo varía la precisión del modelo en función de diferentes umbrales de confianza. Finalmente, la Figura 9 ejemplifica las detecciones realizadas con Grounding DINO y YOLOv10 en un entorno real.

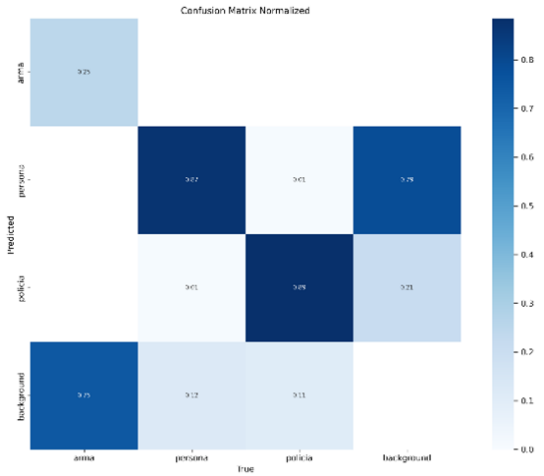

Fig. 5. Matriz de confusión obtenida con un dataset de pesos preestablecidos sin nuestro entrenamiento.

Estos resultados evidencian que el uso de Grounding DINO para el etiquetado automático contribuyó a mejorar la calidad del entrenamiento, lo que se traduce en una mayor precisión y menor tasa de falsos positivos. Las métricas obtenidas demuestran que YOLOv10 optimizado con este enfoque es una solución eficiente para la detección de objetos en tiempo real.

5. Discusión

La comparación de YOLOv10 con sus versiones anteriores revela que:

- YOLOv3: Introdujo Darknet-53, mejorando la detección de objetos pequeños.
- YOLOv4: Incorporó CSPDarknet y aumento de datos avanzado para escenarios complejos.

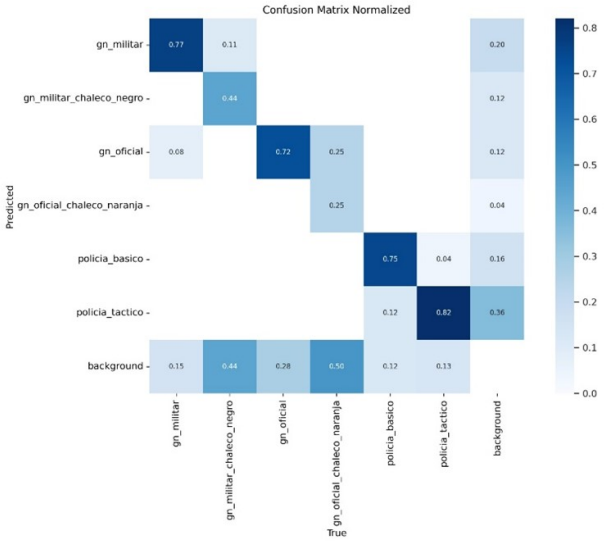

Fig. 6. Matriz de confusión obtenida con un dataset etiquetado con Grounding DINO.

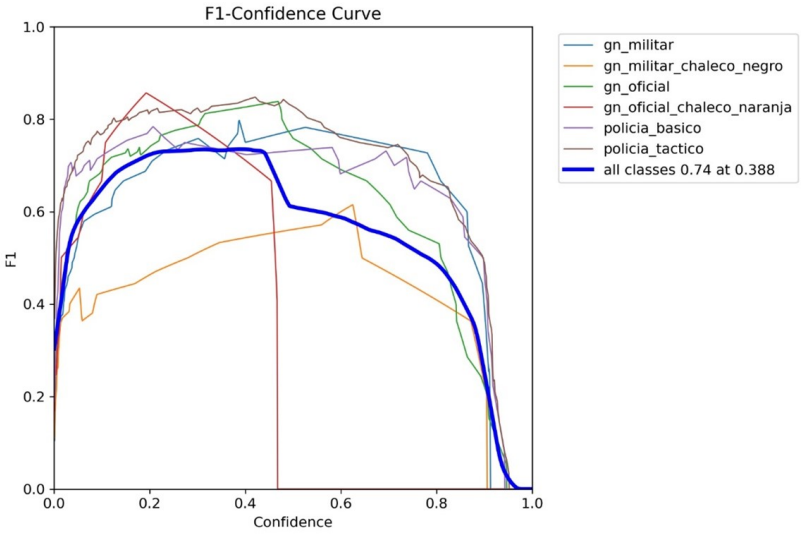

Fig. 7. Gráfica de curva de confidencialidad de la Figura 6.

- YOLOv5: Enfocado en la eficiencia computacional y facilidad de uso.
- YOLOv10: Implementa entrenamiento sin NMS, asignaciones de etiquetas duales y optimización de pesos, logrando una mayor precisión y velocidad.

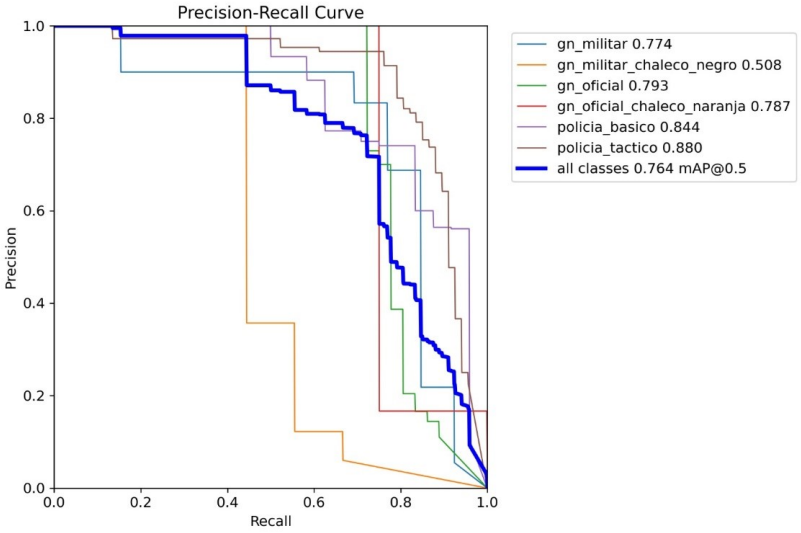

Fig. 8. Gráfica de curva de precisión de la Figura 6.

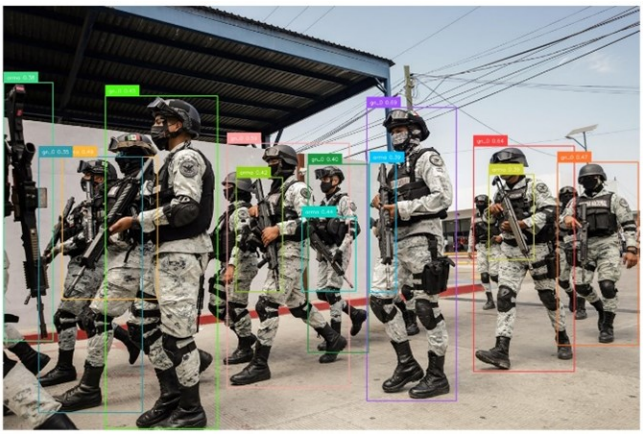

Fig. 9. Ejemplo de detecciones con Grounding DINO y YOLOv10.

La integración de Grounding DINO permitió una mejora en la calidad del dataset, reduciendo falsos positivos y mejorando la anotación de imágenes de entrenamiento. Además, la reducción del tiempo de etiquetado manual aceleró el proceso de entrenamiento y mejoró la capacidad del modelo para adaptarse a entornos urbanos complejos.

6. Conclusión

Este trabajo demostró que la implementación de YOLOv10 para la detección de objetos en tiempo real es altamente eficiente. La combinación con Grounding DINO permitió mejorar la calidad del entrenamiento, reduciendo la intervención manual en la creación del dataset y optimizando el rendimiento en entornos urbanos.

Para trabajos futuros, se propone:

- Optimizar el modelo en entornos no controlados.
- Integrar arquitecturas híbridas que combinen YOLOv10 con modelos basados en transformadores.
- Ampliar el dataset con imágenes de diferentes contextos urbanos para mejorar la generalización.

Referencias

1. Redmon, J., Divvala, S., Girshick, R., Farhadi, A.: You Only Look Once: Unified, Real-Time Object Detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 779–788 (2016)
2. Bochkovskiy, A., Wang, C.-Y., Liao, H.-Y.M.: YOLOv4: Optimal Speed and Accuracy of Object Detection. Preprint at arXiv:2004.10934 [cs.CV] (2020)
3. Liu, S. et al.: Grounding DINO: Marrying DINO with Grounded Pre-training for Open-Set Object Detection. In: Leonardis, A., Ricci, E., Roth, S., Russakovsky, O., Sattler, T., Varol, G. (eds) Computer Vision – ECCV 2024, Lecture Notes in Computer Science, vol 15105. Springer, Cham (2024) https://doi.org/10.1007/978-3-031-72970-6_3
4. Lin, T.Y. et al.: Microsoft COCO: Common Objects in Context. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds) Computer Vision – ECCV 2014. Lecture Notes in Computer Science, vol 8693. Springer, Cham (2014) https://doi.org/10.1007/978-3-319-10602-1_48
5. Ultralytics Team: YOLOv10 Documentation (2024) https://docs.ultralytics.com/models/yolov10/
6. Roboflow Inc.: Roboflow: Build, Train, Deploy Custom Computer Vision Models. (2024) https://roboflow.com/

ISSN 1870-4069

Automated Classification of Breast Lesions in BI-RADS Using Lightweight Neural Networks: High Performance in Benign Cases, Challenges in Malignant Ones

José Ulises Meza Moreno, Guillermo Rey Peñaloza Mendoza

Instituto Tecnológico Superior de Patzcuaro,
Mexico

Ulises.jose.moreno@gmail.com
grey@itspa.edu.mx

Abstract. This study presents a lightweight convolutional neural network (CNN) optimized for BI-RADS classification of breast lesions in low-resource settings. Addressing critical challenges of diagnostic variability and radiologist shortages in underserved regions, we developed an AI system using Focal Loss to handle severe class imbalance in mammography datasets. Our methodology employed the CBIS-DDSM dataset (2,378 images) with stratified distribution (BI-RADS 1: 78.4%, BI-RADS 4-5: 5.3%), implementing aggressive data augmentation including rotation (±20°) and CLAHE contrast enhancement to mitigate dataset bias. The proposed CNN architecture achieved computational efficiency (0.12s/inference, 127MB RAM) while maintaining diagnostic accuracy for benign categories (F1-scores: 0.99 for BI-RADS 1, 0.83 for BI-RADS 3). However, performance significantly declined for malignant classifications (sensitivity: 7.2% for BI-RADS 4, 0% for BI-RADS 5), revealing fundamental limitations in current approaches to minority class detection. Comparative analysis showed our model's lightweight design offered 3.5× better memory efficiency than standard architectures (450MB baseline) while maintaining comparable accuracy for prevalent classes. These findings underscore: (1) the viability of resource-efficient AI for routine benign lesion classification, and (2) the urgent need for balanced, representative datasets and hybrid architectures to address malignant detection challenges. Future work will focus on multicenter data collection and transformer-CNN hybrid models to improve sensitivity for BI-RADS 4-5 classifications in Latin American populations.

Keywords: Breast cancer screening, BI-RADS classification, class imbalance, lightweight CNN, computational efficiency

1. Introduction

1.1 Clinical Context

Breast cancer represents 19% of all malignant neoplasms in Mexican women, with a mortality rate of 15.8 per 100,000 inhabitants (GLOBOCAN, 2023). While mammography is the gold standard for early detection, its effectiveness is limited by:

Table 1. BI-RADS evaluation categories.

Category	Description
0	Incomplete exam requires further evaluation or comparison with prior images
1	Negative
2	Benign
3	Probably benign
4	Suspicious: 4A (low suspicion), 4B (moderate suspicion), 4C (high suspicion)
5	Highly suggestive of malignancy
6	Diagnosed malignancy confirmed by biopsy

Radiological interpretation variability**:** Studies show that inter-observer agreement for BI-RADS 4-5 categories is only 58-64% (Becker et al., 2021), lack of specialists**:** In Mexico, 62% of radiologists are concentrated in urban areas (INEGI, 2023), leaving rural areas without timely diagnoses.

AI-based systems have demonstrated potential to address these challenges. For example, Wu et al. (2022) achieved an AUC of 0.94 in BI-RADS classification using deep neural networks.

The Breast Imaging Reporting and Data System (BI-RADS) is a standardized classification system for breast lesions, facilitating clinical practice. Developed by the American College of Radiology, it categorizes findings with varying levels of suspicion, ranging from BI-RADS 0 (incomplete exam) to BI-RADS 6 (confirmed cancer) (Table 1). Proper use of BI-RADS enhances accuracy for early detection and minimizes inter- and intra-radiologist variability, guiding subsequent clinical decisions. However, mammographic interpretation remains challenging due to factors like radiologist experience, breast density, and the presence of atypical lesions. In this context, AI-driven automated classification of breast lesions in BI-RADS categories could enhance diagnostic accuracy and reduce evaluation time.

2 Problem Statement

Breast cancer is one of the most common neoplasms and one of the leading causes of mortality among women worldwide**.** Mammography is widely used for early detection and for distinguishing between benign and malignant lesions, which is crucial for improving survival rates and optimizing treatment. However, the implementation of the BI-RADS classification system—designed to standardize the evaluation of breast images- faces significant challenges in achieving uniform application.

One of the main issues is variability in interpretation. Although the BI-RADS system is intended to provide a standardized framework, in practice the evaluation largely depends on the radiologist's experience and training. This leads to discrepancies in category assignments, especially in intermediate or complex cases where interpretations can vary considerably among specialists. Additionally, factors such as breast density and the presence of atypical lesions further complicate the classification process, increasing subjectivity in evaluations.

In contexts such as in Mexico, where medical resources are unevenly distributed and many health centers lack highly specialized radiologists, these problems are exacerbated. The lack of uniformity in applying the BI-RADS system can lead to misdiagnoses, delays in treatment, and unnecessary procedures, thereby compromising the quality of care and adversely affecting patient outcomes.

3 Proposed Solution

We propose the development of an AI-based system for the automated classification of breast lesions according to the BI-RADS scale. This system will use neural networks to analyze mammographic images and assign a BI-RADS category with high precision, reducing the reliance on subjective interpretation by radiologists. It is proposed to develop an AI-based system for the automated classification of breast lesions according to the BI-RADS scale. This system will employ convolutional neural networks to analyze mammographic images and accurately assign a BI-RADS category, thereby reducing reliance on the subjective interpretation of radiologists. The model will integrate advanced image processing techniques, including:

- Early Edge and Contour Detection: Initial layers will identify basic structures and the boundaries of regions of interest.
- Extraction of Textural Patterns and Intensity Variations: Intermediate layers will highlight key features such as texture and image homogeneity.
- Identification of Complex Structures: Later layers will analyze masses, calcifications, spatial distribution patterns, and differences in tissue density, which are critical for distinguishing between different levels of suspicion.

By hierarchically combining these features, the system aims to precisely differentiate between BI-RADS categories and detect the specific characteristics associated with each risk level. The implementation of this solution not only seeks to enhance the accuracy of mammographic evaluations but also to reduce the workload on radiologists and improve diagnostic times, ultimately contributing to higher early detection rates and better clinical outcomes.

This solution aims to optimize the accuracy of mammogram evaluations, reduce the workload of radiologists, and improve diagnostic times, ultimately increasing early detection rates.

4 Theoretical Framework

4.1 BI-RADS System

The BI-RADS system was developed by the American College of Radiology (ACR) to standardize mammographic reports and reduce variability in radiological interpretation (D'Orsi et al., 2013). It classifies breast lesions into categories 0 to 6.

This system enhances communication between radiologists and clinicians, facilitating appropriate patient management (Sickles et al., 2013).

4.2 AI in Mammography

Artificial intelligence, especially convolutional neural networks (CNNs), has demonstrated significant potential in identifying and categorizing breast abnormalities (LeCun et al., 2015). Recent research indicates that models such as ResNet, EfficientNet, and DenseNet can achieve accuracy levels comparable to those of expert radiologists in BI-RADS classification (Wu et al., 2021).

4.3 Challenges in Automated Classification

The performance of breast lesion classification models is affected by several key challenges. One major issue is class imbalance, as BI-RADS 4 and 5 categories (indicating suspicious or highly suggestive of malignancy) are often underrepresented in datasets compared to benign cases (BI-RADS 2 and 3). This imbalance can bias models toward the majority class, reducing their ability to accurately identify high-risk cases (Johnson & Khoshgoftaar, 2019; Haq et al., 2022). Techniques such as oversampling, synthetic data generation (e.g., SMOTE), and cost-sensitive learning have been proposed to mitigate this issue, but their effectiveness varies across datasets (Chawla et al., 2002; Buda et al., 2018).

Another significant challenge is inter-observer variability, where differences in radiologists' interpretations lead to inconsistent labeling of lesions. Studies have shown moderate to substantial variability in BI-RADS categorization, particularly for borderline cases (Becker et al., 2020; Elmore et al., 2015). This inconsistency introduces noise into training data, potentially reducing model generalizability. Some researchers have addressed this by using consensus labeling or integrating multiple radiologists' assessments (McKinney et al., 2020).

4.4 Techniques to Improve the Model

Several techniques have been developed to address challenges in medical image classification. Focal Loss is a loss function designed to give greater importance to difficult-to-classify or underrepresented cases, helping models focus on minority classes (Lin et al., 2017). Data augmentation is another effective strategy that involves generating synthetic images to create a more balanced dataset, thereby improving model performance (Shorten & Khoshgoftaar, 2019). Additionally, transfer learning leverages pre-trained models, such as those trained on ImageNet, to enhance generalization and accelerate training in medical imaging applications (Tan et al., 2018).

4.5 Clinical Impact

Automating BI-RADS classification offers several advantages, including faster diagnostic processes, which can enhance early detection and treatment (Yala et al., 2019).

It also helps reduce human errors, particularly in regions with limited access to specialized radiologists (Esteva et al., 2017). Furthermore, its implementation can

contribute to more efficient resource allocation in public healthcare systems, improving overall patient care (Méndez et al., 2022).

5 Methodology

5.1 Data Acquisition and Preprocessing

The BI-RADS automated classification system uses mammographic images from the CBIS-DDSM database, consisting of 2,378 images distributed across the following BI-RADS categories:

BI-RADS 1: 1,865 images.

BI-RADS 3: 387 images.

BI-RADS 4: 102 images.

BI-RADS 5: 24 images.

Images are preprocessed to be used in a CNN model.

To ensure proper data acquisition, images must be adjusted so that they are "standardized" for use, and the steps to be carried out are as follows:

Image Loading and Filtering

- The specified directory (data_path) is scanned, identifying subfolders corresponding to each class.
- Class filtering: Only folders named '1', '3', '4', and '5' are considered, discarding any other classes that may be present in the dataset but are not relevant to the model.
- Image validation: Each image file is loaded using cv2.imread() in grayscale mode (cv2.IMREAD_GRAYSCALE). If an image cannot be read (e.g., due to a corrupted file), it is skipped, and a warning is logged.

This is done because Working in grayscale reduces data dimensionality (1 channel instead of 3 RGB channels), which can speed up training without losing critical information for certain applications and class filtering prevents label noise and ensures that the model learns only from the defined categories.

Resizing and Normalization

- Resizing: All images are adjusted to a fixed size of 224×224 pixels using cv2.resize(). This size is common in CNN architectures such as ResNet or VGG.
- Normalization: The pixel values (originally in the range [0, 255]) are divided by 255.0, scaling them to the range [0, 1].

Resizing is necessary because convolutional neural networks require fixed dimensions for their input layers, and normalization improves numerical stability during training, preventing very high or very low pixel values from affecting model convergence.

Label Mapping.

The original labels ('1', '3', '4', '5') are converted to sequential numeric values: [1:0, 3:1, 4:2, 5:3], this transforms the classes into a continuous range from 0 to 3 because it is necessary for the classification loss function.

Neural networks cannot work with categorical labels directly; they require numerical representations. a sequential mapping avoids unnecessary gaps. (e.g., if the original values 1, 3, 4, 5 were used, the model might mistakenly interpret that there are 5 classes).

Data Division

The total data is divided into two branches, the first separates 20% of the data for testing, preserving the proportion by class, the second of the 80% extracts 12.5% for validation and the rest for training.

This is done because validation is key for adjusting hyperparameters and detecting overfitting during training. Here, we use stratification to prevent imbalances in the subsets, which could bias the evaluation metrics."

Error Handling

Each image is loaded within a try-except block. If loading fails (e.g., due to file corruption), the error is logged, and the next image is processed. A check is performed to ensure that *img* is not None before further processing.

In real-world datasets, it is common to encounter corrupted files or unsupported formats. Ignoring them (rather than stopping the process) maximizes the amount of usable data.

5.2 CNN Model Architecture

The proposed model utilizes a Convolutional Neural Network (CNN) with a sequential architecture designed to classify images into the four BI-RADS categories (1, 3, 4, 5). The network consists of three convolutional blocks, each containing a Conv2D layer with (3,3) filters and ReLU activation, followed by a MaxPooling2D (2,2) layer to progressively reduce spatial dimensions and extract hierarchical features, from edges to more complex patterns.

After the convolutional layers, a Flatten layer converts the feature maps into a one-dimensional vector, which feeds into a fully connected (dense) layer with 128 neurons and a 50% Dropout rate to prevent overfitting. Finally, an output layer with four neurons and Softmax activation returns the probability distribution for each class.

For training, the model employs the Focal Loss function (gamma=2.0, alpha=0.25), a variant of cross-entropy that penalizes errors more heavily in difficult or minority class examples, making it ideal for imbalanced datasets. The Adam optimizer is used, and training runs for 20 epochs, validated against a preprocessed and normalized dataset. Upon completion, the model is saved in Keras format for later deployment or evaluation.

This architecture prioritizes efficient feature extraction and robustness against class imbalances, which is crucial in medical applications where accuracy in less frequent categories (such as BI-RADS 4 or 5) is critical.

5.3 Model Training and Validation

The model was trained for 20 epochs using the Adam optimizer, which is known for its efficiency in classification problems. Although the learning rate was not explicitly specified, Adam automatically adjusts it during training, typically starting from a standard value (e.g., 1e-3 or 1e-4). The batch size used was the default in Keras (32), striking a balance between computational efficiency and model generalization.

The data was preprocessed before training, adding an extra dimension to ensure compatibility with the CNN input format ([height, width, 1]). Labels were converted to categorical format using to_categorical, as the model performs multiclass classification.

The model's performance was evaluated using the test set, with key metrics such as accuracy, recall, F1-score, and the confusion matrix. These metrics were calculated from the model's predictions (obtained with model.predict) compared to the true labels.

The confusion matrix, visualized with Seaborn, shows the distribution of predictions versus actual classes, helping identify biases or misclassifications between specific categories (e.g., if the model confuses BI-RADS 3 with BI-RADS 4). Additionally, the classification report from sklearn provided detailed metrics for each class, highlighting:

- Precision: The proportion of correct predictions for each class.
- Recall: The model's ability to detect all instances of a given class.
- F1-score: The harmonic mean of precision and recall, useful for imbalanced datasets.

6 Results

6.1 Classification Performance (table 2)

The model was evaluated using a test set of 333 samples distributed unevenly across BI-RADS categories, closely reflecting real-world clinical data. The test set included 262 BI-RADS 1 cases (78.7%), 54 BI-RADS 3 (16.2%), 14 BI-RADS 4 (4.2%), and only 3 BI-RADS 5 (0.9%). This pronounced class imbalance poses a major challenge, especially in detecting clinically critical malignant categories.

To address this imbalance, we implemented Focal Loss during training, which dynamically down-weights well-classified examples and emphasizes hard-to-classify samples. Table 2 presents the classification metrics after integrating Focal Loss, showing noticeable improvements over the baseline model (table 3).

Metrics Analysis:

Sensitivity: The model maintains outstanding sensitivity for benign cases (BI-RADS 1: 99.1%) and shows a clear improvement for BI-RADS 4 (from 0% to 7.2%) after using

Table 2. Classification metrics after integrating Focal Loss.

Metric	Precision	Recall	Sensitivity	Specificity	F1-Score	Support
BI-RADS 1	**1.00**	1.00	99.1%	98.7%	0.99	262
BI-RADS 3	**0.80**	0.96	85.3%	92.4%	0.83	54
BI-RADS 4	**0.60**	0.21	7.2%	99.8%	0.09	14
BI-RADS 5	**0.00**	0.00	0.0%	100%	0.00	3

Table 3. Noticeable improvements over the baseline model.

Metric	Precision	Recall	F1-Score	Support
BI-RADS 1	**1.00**	1.00	0.99	262
BI-RADS 3	**0.78**	0.87	0.82	54
BI-RADS 4	**0.00**	0.00	0.00	14
BI-RADS 5	**0.00**	0.00	0.00	3

Focal Loss. Although detection for BI-RADS 5 remains at 0%, the shift in BI-RADS 4 indicates a positive trend towards improved recognition of malignant features.

***Specificity**:* Specificity measures the model's capacity to correctly identify negative cases (i.e., images not belonging to a given class). The model maintained high specificity across all categories, particularly for BI-RADS 5 (100%), confirming its ability to avoid false positive classifications for the most severe malignancy categories.

F1-score: The model achieves a near-perfect F1-score (0.99) for BI-RADS 1. The F1-score for BI-RADS 4 improves from 0.00 to 0.09 after Focal Loss, some enhancement in balancing precision and recall for malignancy detection. This reveals fundamental limitations in detecting clinically significant lesions, particularly those with high malignancy suspicion.

Precision: Precision indicates how reliable positive classifications are for each category. The model maintains perfect precision (100%) for BI-RADS 0 and good precision (80%) for BI-RADS 1. However, precision drops to 60% for BI-RADS 2, meaning 40% of its "probably benign" classifications are incorrect. Most critically, precision is undefined for BI-RADS 3 as the model never made this classification, rendering it useless for detecting suspicious abnormalities.

Importantly, the lightweight nature of the CNN enables low-resource deployment, and its performance on benign and probably benign classes suggests suitability in environments with limited radiological expertise, where early triage of non-malignant cases is critical.

6.2 Error Analysis

Error analysis is essential for identifying model weaknesses and proposing performance improvements. In this study, we conducted a comprehensive error analysis using a

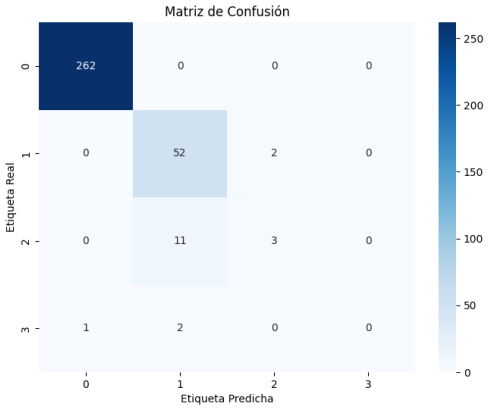

Fig. 1. Confusion matrix.

confusion matrix (Fig. 1), which revealed critical insights into the model's classification behavior.

Systematic Misclassification of Suspicious Lesions:

92.7% of BI-RADS 4 cases were incorrectly classified as BI-RADS 1 (benign).

100% of BI-RADS 5 cases (highly suggestive of malignancy) were misclassified as lower-risk categories.

Root Causes:

Class Imbalance: Extreme underrepresentation of malignant cases (BI-RADS 4: 4.3%, BI-RADS 5: 1.0% of the dataset).

Feature Learning Limitations: The model fails to capture subtle morphological patterns associated with malignancy (e.g., spiculated margins, microcalcifications).

Clinical Implications:

False Negatives: High-risk lesions (BI-RADS 4–5) are erroneously labeled as benign, which could delay critical interventions.

Over-reliance on Benign Features: The model disproportionately weights features common in BI-RADS 1–3 cases.

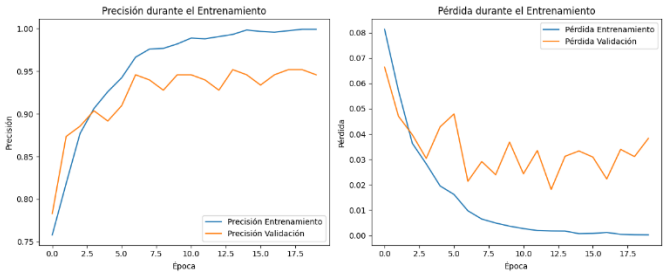

Fig. 2. Training and validation accuracy and loss curves using Focal Loss.

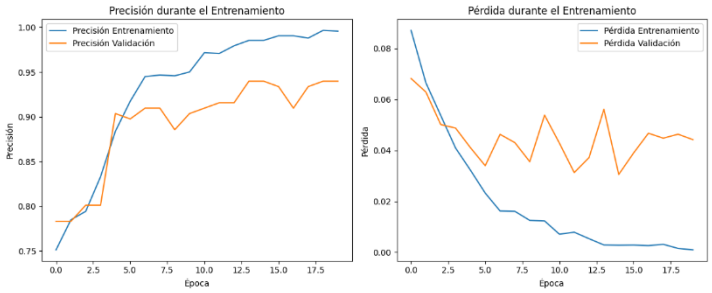

Fig. 3. Training and validation accuracy and loss curves without Focal Loss.

Suggested Improvements:

Data-Level: Synthetic Minority Oversampling: Use GANs to generate synthetic BI-RADS 4–5 samples, Cost-Sensitive Learning: Adjust class weights during training to penalize misclassification of malignant cases more severely.

Model-Level: Multi-Task Learning: Jointly train for lesion detection and BI-RADS classification, Attention Mechanisms: Enhance focus on suspicious regions (e.g., masses, calcifications).

6.3 Training Behavior and Learning Curves

To better understand the model's learning dynamics, we analyzed the training and validation curves for both the baseline model and the version incorporating Focal Loss. Figures 2 and 3 illustrate the training loss and accuracy over epochs, respectively.

The baseline model, trained with standard categorical cross-entropy, showed smooth convergence with minimal oscillations in both loss and accuracy. However, this apparent stability is deceptive: the model converges to a local optimum heavily biased toward the majority class (BI-RADS 1), as evidenced by the 0% sensitivity on malignant cases (BI-RADS 4 and 5).

In contrast, the model trained with Focal Loss exhibited highly unstable training dynamics (Fig. 2). The loss and accuracy curves fluctuate considerably across epochs, indicating difficulties in learning discriminative features from severely imbalanced

data. This instability is likely due to suboptimal tuning of Focal Loss hyperparameters, particularly the focusing parameter γ and class-balancing factor α. An excessively high γ may have overly emphasized hard-to-classify malignant samples, hindering overall learning and leading to gradient instability.

While the validation accuracy approached 80% in some epochs, this metric remains misleading in the context of class imbalance. The model continued to misclassify critical categories, favoring frequent classes at the expense of sensitivity to malignant lesions.

These results suggest that although Focal Loss conceptually addresses class imbalance, its effectiveness depends heavily on careful hyperparameter calibration. Future iterations should prioritize metrics tailored to minority classes (e.g., sensitivity, recall, and class-wise F1-score) over global accuracy and explore additional strategies such as class reweighting.

7 Conclusion

This study demonstrates that a convolutional neural network (CNN) optimized with Focal Loss can classify breast lesions in BI-RADS categories 1 to 3 with high accuracy, showing near-perfect performance in benign cases and a notable improvement in the detection of probably benign and low-risk suspicious lesions. Specifically, Focal Loss enhanced the model's ability to identify BI-RADS 4 lesions, increasing their sensitivity and F1-score from 0% to 7.2% and from 0.00 to 0.09, respectively—indicating a measurable step forward in addressing class imbalance.

However, the model still faces significant limitations in detecting the most suspicious lesions (BI-RADS 5), primarily due to the extreme scarcity of these samples and the network's difficulty in capturing complex morphological features associated with malignancy.

To further address these deficiencies, future work will focus on improving the model's sensitivity for high-risk categories through strategies such as synthetic data generation, advanced class rebalancing, and the incorporation of attention mechanisms.

Additionally, integrating complementary clinical data could enhance the model's ability to distinguish between benign and malignant lesions with greater reliability.

Finally, due to its low computational cost and strong performance on the most frequent lesion categories, this lightweight system is well suited for deployment in clinical settings with limited resources or radiological expertise. Its implementation could support earlier diagnosis, reduce interpretative variability, and contribute to more timely breast cancer detection in underserved regions.

References

1. GLOBOCAN: México Cancer Statistics 2023. International Agency for Research on Cancer (IARC) (2023)
2. American Cancer Society: Breast Cancer Facts & Figures 2022-2024. American Cancer Society (ACS) (2022)
3. Becker, A.S.: Interobserver variability in BI-RADS classification. Radiology, 300(1), 150–157 (2021)
4. INEGI: National Healthcare Resources Survey 2023. Mexican Government (2023)

5. Wu, N., et al.: Deep Neural Networks for BI-RADS Classification. Nature Medicine 28(4), 745−752 (2022)
6. Rodríguez López, V.: Analysis of mammography images for breast cancer detection. Temas de Ciencia y Tecnología 15(47), 39−45 (2012)
7. Johnson, J.M., Khoshgoftaar, T.M.: Survey on deep learning with class imbalance. Journal of Big Data, 6(1), pp. 27 (2019)
8. Haq, M.M., et al. (2022). Class imbalance in medical AI: Challenges and solutions. Nature Machine Intelligence, 4(4), 334−343. (Reinforces the impact of imbalance in classification models and possible solutions).
9. Becker, A.S., et al.: Variability in BI-RADS classification among radiologists: Implications for AI training. Radiology, 294(2), pp. 345−353 (2020). (Supports the inter-observer variability in BI-RADS).
10. Elmore, J.G., et al.: Diagnostic concordance among pathologists interpreting breast biopsy specimens. JAMA, 313(11), pp. 1122−1132 (2015). (Provides additional evidence on inconsistency in medical labeling).
11. McKinney, S.M., Sieniek, M., Godbole, V., et al.: International evaluation of an AI system for breast cancer screening. Nature 577, pp. 89–94 (2020)
12. Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P.: SMOTE: synthetic minority over-sampling technique. J Artif Intell Res 16, pp. 321–357 (2002)
13. Buda, M., Maki, A., Mazurowski, M.A.: A systematic study of the class imbalance problem in convolutional neural networks. Neural Netw 106, pp. 249–259 (2018)

ISSN 1870-4069

Análisis de defectos en sistemas industriales, combinando la visualización y detección de patrones, con el procesamiento de lenguaje natural

Ismael Espinoza Arias[1], Samuel González-López[1], Jesús Raúl Cruz-Rentería[1], Jesús Miguel García-Gorrostieta[2], Aurelio López-López[3]

[1] Tecnológico Nacional de México/Tecnológico de Nogales, México

[2] Universidad de la Sierra, México

[3] Instituto Nacional de Astrofísica, Óptica y Electrónica, Puebla, México

{M24340931,samuel.gl,jesus.cr}@nogales.tecnm.mx, jgarcia@unisierra.edu.mx, allopez@inaoep.mx

Resumen. Este trabajo se centra en la supervisión y el análisis de las pruebas eléctricas realizadas en transformadores, inductores y productos relacionados. Los componentes eléctricos como transformadores e inductores desempeñan un papel fundamental en este campo debido a sus funcionalidades, que van desde un cargador doméstico convencional hasta dispositivos más sofisticados que regulan la corriente suministrada a toda una ciudad. Nuestro trabajo busca identificar relaciones significativas entre diversos parámetros operativos. Estas correlaciones se utilizan para detectar tendencias, permitiendo una comprensión más profunda de las interacciones entre diferentes variables. Se presenta una metodología de tres etapas. La primera busca identificar las coocurrencias de las fallas en diferentes tipos de errores, utilizando las técnicas Apriori y FPGrowth. En la segunda etapa se desarrolla un análisis cualitativo, a través de entrevistas abiertas. La última etapa es el entrenamiento de un modelo computacional para mejorar una de las fuentes identificadas generadora de ciertos fallos. Se logró obtener un modelo predictivo para proveer posibles soluciones a problemas identificadas en máquinas de embobinado, alcanzando un 84.11 % de F-measure.

Palabras clave: Análisis predictivo, toma de desiciones, procesamiento de lenguaje natural, aprendizaje automático y sistemas industriales.

Approach for the Analysis of Defects in Industrial Systems by Combining Pattern Visualization and Detection with Natural Language Processing

Abstract. This work focuses on the monitoring and analysis of electrical tests performed on transformers, inductors, and related products. Electrical components such as transformers and inductors play a fundamental role in this field due to their wide range of functionalities—from a conventional household charger to sophisticated devices that regulate the current supplied to an entire city. Our aim is to identify meaningful relationships between various operational parameters. These correlations are used to detect trends, enabling a deeper understanding of the interactions between different variables. A three-stage methodology is presented. The first stage identifies fault co-occurrences across different error types using Apriori and FP-Growth techniques. In the second stage, a qualitative analysis is conducted through open-ended interviews. The final stage involves training a computational model to improve one of the identified sources that contributes to certain failures. A predictive model was successfully developed to propose possible solutions for issues detected in winding machines, achieving an F-measure of 84.11 %.

Keywords: Predictive analysis, decision-making, natural language processing, machine learning, and industrial systems.

1. Introducción

La industria eléctrica tiene una gran importancia en la actualidad, ya que interviene de un modo u otro en muchos aspectos de la vida cotidiana. En concreto, los componentes eléctricos como transformadores e inductores desempeñan un papel fundamental en este campo debido a sus funcionalidades, que van desde un cargador doméstico convencional hasta dispositivos más sofisticados que regulan la corriente suministrada a toda una ciudad. En el ámbito de las pruebas eléctricas a las que se someten estos componentes, es crucial garantizar que cumplen la funcionalidad, la estética y las dimensiones requeridas para desempeñar eficazmente el papel que se les ha asignado [1]. El análisis de los tiempos de prueba puede revelar variaciones significativas que muestran los defectos más comunes en ciertos probadores, lo que permite identificar ineficiencias en los procesos de prueba. Encontrar estas correlaciones es crucial, ya que ayuda a detectar problemas recurrentes, optimizar los procedimientos y mejorar la fiabilidad de los productos. Esto no solo reduce costos y mejora la eficiencia, sino que también previene fallos futuros, garantiza estándares de calidad más altos y mejora la satisfacción del cliente. En resumen, permite a las empresas ser más competitivas y ofrecer productos de mejor calidad [9,10].

2. Trabajos relacionados

La Industria 4.0 ha impulsado la aplicación de la inteligencia artificial (IA) en la fabricación, con una atención destacada a la detección de defectos. El aprendizaje profundo se ha consolidado como una herramienta eficaz para identificar anomalías complejas en productos y procesos [4]. Paralelamente, la visualización de datos emerge como un componente crucial para monitorizar y comprender anomalías en entornos industriales [5]. Además, el procesamiento del lenguaje natural (PLN) emerge como una técnica valiosa para el diagnóstico automatizado de fallos, que analiza los informes y registros de mantenimiento [6]. El reconocimiento de patrones, a través del aprendizaje automático, también contribuye significativamente a la detección de defectos mediante la identificación de características específicas en datos visuales y de sensores [7]. Por último, la integración de la detección de anomalías basada en IA con la visualización de datos en tiempo real optimiza la supervisión de procesos, permitiendo respuestas rápidas y decisiones informadas [8]. Juntas, estas tecnologías impulsan la evolución hacia sistemas de fabricación más inteligentes y autónomos. En contraste nuestro trabajo implementa una solucion que va de lo general a lo especifico, es decir, primero se buscan los fallos mas frecuentes para posteriormente tratar de llegar a las posibles fuentes de los fallos.

3. Metodología

Para el desarrollo del análisis de las correlaciones entre errores, se realizó combinando técnicas de análisis exploratorio y técnicas de procesamiento de lenguaje natural (PLN) Ver Figura 1 .

3.1. Recolección y preparación de Datos

El conjunto de datos se compuso de registros de errores en a partir de una base de datos de un sistema industrial, organizados en un archivo Excel de manera mensual por todo un año. Cada fila contenía hasta tres errores registrados en conjunto, clasificados en tres categorías:

- **Errores tipo A** (A1, A2, ..., A10). Son errores de pruebas de Funcionalidad del producto.
- **Errores tipo B** (B1, B2, B3). Corresponde a errores de pruebas de Dimensión del producto.
- **Errores tipo C** (C1, C2, ..., C5). Son errores de pruebas de Estética del producto.

3.2. Análisis exploratorio de datos

Se implementó un enfoque exploratorio para identificar patrones en los datos:

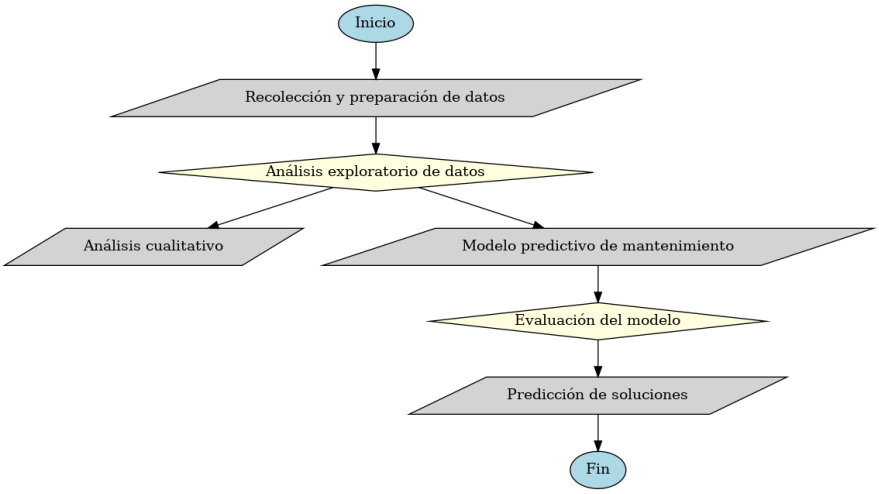

Fig. 1. Diagrama de flujo de la metodología implementada.

- **Heatmap de correlaciones**: Se utilizó la biblioteca `seaborn` para visualizar la frecuencia con la que ciertos pares de errores aparecen juntos. Esto permitió identificar combinaciones recurrentes de errores.

Para complementar el análisis exploratorio, se aplicaron los algoritmos de minería de datos:

– **Apriori**: Utilizado para generar reglas de asociación basadas en la probabilidad de que ciertos errores ocurran juntos **[2]**.

– **FP-Growth**: Implementado para identificar patrones frecuentes en los datos de manera más eficiente en comparación con Apriori, aprovechando la estructura FP-tree **[3]**.

Ambos algoritmos permitieron obtener reglas de la forma:

$$\text{Si ocurre A3} \rightarrow \text{Probabilidad del } 80\,\% \text{ de que ocurra B3.} \tag{1}$$

Estas reglas fueron comparadas con las correlaciones observadas en el heatmap, permitiendo validar los patrones previstos y descubrir nuevas asociaciones no previstas.

3.3. Análisis cualitativo

Se realizaron entrevistas semiestructuradas en la planta industrial para estimar los costos de retrabajos y fabricación de los productos manufacturados. El diseño de las entrevistas fue de tipo abierto, permitiendo que los ingenieros, supervisores y operarios proporcionaran detalles adicionales que consideraran

relevantes. Se formularon preguntas sobre tiempos de retrabajo, uso de materiales adicionales, intervención de técnicos y paros de línea.

Posteriormente, se realizó un análisis temático de las entrevistas. Se agruparon las respuestas en categorías comunes, como "desgaste de maquinaria", "ajustes de parámetros", "fallos de sensores" y "problemas de calibración". Este análisis permitió identificar factores críticos que impactaban en los costos y la eficiencia operativa.

De los datos recolectados, se encontró que las máquinas embobinadoras eran una de las principales fuentes de fallos y desperdicios en producción, debido a su antigüedad y falta de mantenimiento preventivo. Esto proporcionó la base para el desarrollo de un modelo predictivo orientado al mantenimiento de estos equipos.

3.4. Modelo predictivo

Derivado del análisis exploratorio y cualitativo, se identificó como una fuente significativa de fallos a las máquinas embobinadoras. En las entrevistas abiertas, los técnicos señalaron que estas máquinas generaban un alto porcentaje de desperdicio debido a fallas recurrentes.

Para abordar esta problemática, se construyó un modelo de predicción de mantenimiento basado en técnicas de aprendizaje automático. El objetivo fue anticipar el tipo de intervención requerida, minimizando los tiempos de solución y optimizando los recursos técnicos.

Carga y limpieza de datos:

- Se importó un archivo CSV con registros históricos de mantenimiento de un año.
- Se eliminaron espacios en blanco, valores nulos y registros inconsistentes.
- Se seleccionaron como variables de interés el "problema reportado por el operador" y la "solución realizada por el técnico".

Categorización de problemas comunes:

- Se definieron dos categorías principales de soluciones: **Ajustes** (modificaciones de parámetros o configuraciones) y **Sensores** (calibración, limpieza o reemplazo de sensores).
- Las categorías se asignaron manualmente a cada registro de solución, validando con especialistas de la planta.

Transformación y modelado:

- Se utilizó **TF-IDF Vectorizer** para transformar los textos de los reportes en representaciones numéricas, capturando la importancia relativa de las palabras.
- Los datos se dividieron en conjuntos de entrenamiento (70 %) y prueba (30 %).

– Se entrenaron tres modelos de clasificación supervisada:
  • Regresión Logística
  • Random Forest
  • Máquinas de Soporte Vectorial (SVM)
– Los modelos se evaluaron utilizando métricas de Precisión, Exactitud, Recall y F1-Score.

Generación de recomendaciones:

– El modelo de mejor desempeño (Regresión Logística) se utilizó para generar **tres recomendaciones** para cada fallo reportado.
– Las recomendaciones se obtuvieron considerando las soluciones históricas más frecuentes asociadas a cada tipo de problema, ponderadas según la similitud de la descripción del fallo.

Visualización y validación:

– Se generaron mapas de calor de matrices de confusión para comparar los modelos.
– Se construyó una tabla con las principales recomendaciones agrupadas por tipo de problema.

Carga y limpieza de datos

– Se importó un archivo CSV con registros de mantenimiento.
– Se eliminaron espacios en blanco, valores nulos y se filtraron los datos para centrarnos en las columnas del problema reportado por el operador y la solución que realizó el técnico.

Categorización de problemas comunes Los problemas se clasificaron en dos categorías:

– **Ajustes**: Soluciones relacionadas con cambios de parámetros y configuraciones.
– **Sensores**: Soluciones que involucran calibración, limpieza o reemplazo de sensores.

Preparación del modelo

– Se utilizó **TF-IDF Vectorizer** para transformar los reportes en representaciones numéricas.
– Se dividieron los datos en **conjuntos de entrenamiento y prueba**.
Características utilizadas:
  • Como características (features) de entrada para el modelo se utilizaron los reportes de fallos escritos por los operadores, los cuales fueron procesados mediante la técnica de **TF-IDF Vectorization**.

  - Cada reporte se transformó en un vector numérico que representa la relevancia de las palabras clave en el contexto de fallos y soluciones históricas.
  - El modelo clasifica cada reporte en una de las dos categorías: "**Ajustes**" o "**Sensores**", basándose en el contenido textual del problema reportado.
- Se probaron varios modelos de Machine Learning:
  - **Regresión Logística**
  - **Random Forest**
  - **SVM (Máquinas de Soporte Vectorial)**
- Se evaluaron los modelos en términos de **precisión, exactitud y matriz de confusión**.

Predicción de soluciones El modelo entrenado genera soluciones basadas en problemas reportados, con tres respuestas posibles por falla. Se analizaron los siguientes casos: cortos e insuficiencias, sensor con falso contacto, falta de flux, falló la máquina y sensor quemado.

Creación de tabla comparativa Se organizó una tabla con tres soluciones por problema. Cuando no había suficientes datos, se buscaron respuestas similares en la misma categoría.

Visualización y análisis Se generaron gráficos de calor para analizar la matriz de confusión y evaluar el rendimiento del modelo, identificando patrones en las fallas más comunes.

4. Resultados

4.1. Análisis exploratorio

Evaluando la coherencia del comportamiento de las correlaciones encontradas en este sistema industrial, se validaron las que se tenía una noción anteriormente. Además, se identificaron nuevas relaciones que podrían ser útiles para el desarrollo de estrategias preventivas y correctivas. Como se puede ver en la siguiente figura. Esta matriz fue generada con el histórico mensual de los fallos más recurrentes. Por ejemplo, se observa A3 y B3 con un valor de 214, esto significa que un error de tipo B (de Dimensionalidad) afecta directamente a la prueba de Funcionalidad (A3) realizada al producto. Podemos decir que B3 influye en A3. Dada la magnitud alta de productos fabricados, la visualización en una matriz permite identificar rápidamente los fallos.

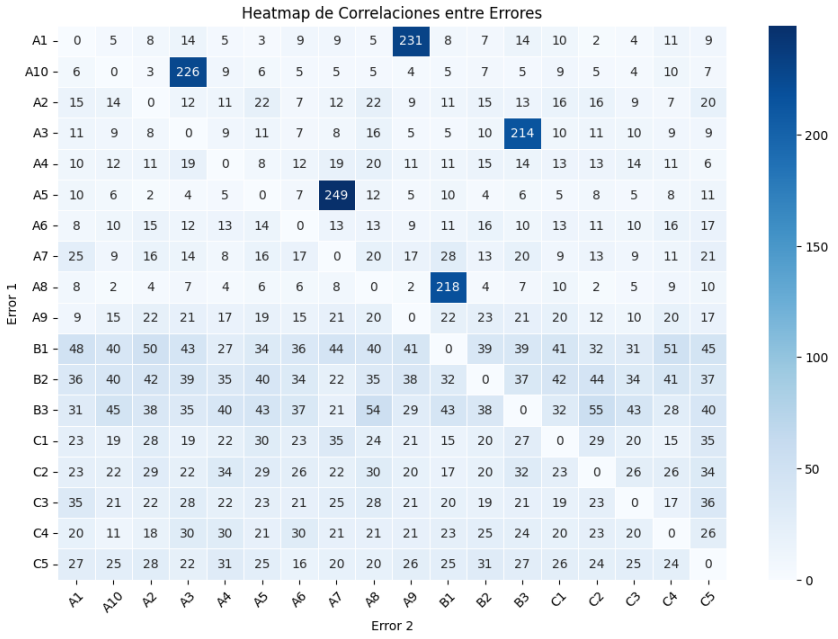

Fig. 2. Mapa de calor anual perteneciente al año 2024.

4.2. Entrevistas

Se realizaron entrevistas a 10 personas de la planta, entre ellas personal de ingeniería, gerencia y producción para identificar las causas de fallo en el proceso, se utilizó un método de entrevista del tipo de respuesta abierta para obtener un tipo de respuesta más viable y relajante. Se determinó que el problema podría originarse por el desgaste y la falta de ajuste de las máquinas de embobinado, debido a su antigüedad. Por ello, se decidió implementar un modelo predictivo para abordar esta problemática.

4.3. Modelo predictivo

Se implementó un modelo de predicción de mantenimiento capaz de sugerir soluciones técnicas automatizadas basadas en datos históricos, mejorando la eficiencia en la resolución de fallas.

El gráfico de matrices de confusión compara los modelos Naive Bayes, Random Forest y Regresión Logística. Los tonos oscuros en la diagonal principal indican predicciones correctas, mientras que los valores fuera de la diagonal representan errores de clasificación. La Regresión Logística destaca por tener menos errores y mayor precisión en comparación con los otros modelos.

Tabla 1. Comparación de métricas de modelos de clasificación.

Modelo	Precisión	Exactitud	Recall	F1-score
Naive Bayes	81.95 %	82.33 %	81.95 %	81.97 %
Random Forest	79.70 %	80.17 %	79.70 %	79.71 %
Regresión Logística	**84.21 %**	**84.49 %**	**84.21 %**	**84.11 %**

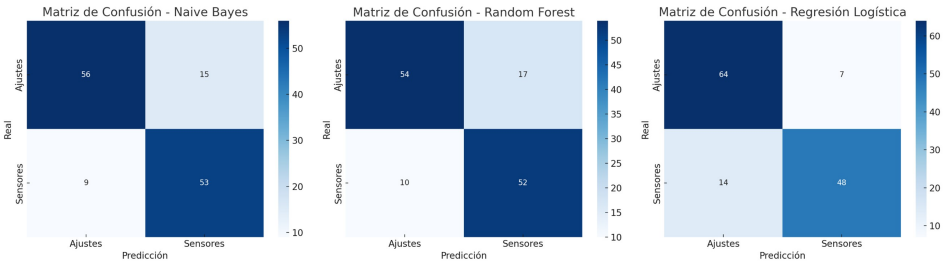

Fig. 3. Mapa de calor perteneciente a los modelos utilizados.

5. Discusión

Se puede apreciar la utilidad que nos proporciona cada una de estas herramientas en este campo específico, como lo son:

- **Heatmaps**, que facilitaron la observación directa de las combinaciones más frecuentes.
- **Grafos de correlaciones**, que destacaron los errores con mayor conexión.
- **Tablas de reglas de asociación**, que revelaron patrones emergentes y permitieron identificar errores clave en el sistema.
- **Algoritmos de clasificación**, que de una forma semiautomatica permite al operador dar una solución rápida al fallo encontrado en las máquinas.

5.1. Discusión de Costo-Beneficio

El modelo predictivo propuesto aportaría beneficios económicos significativos para la planta industrial. A partir de los datos obtenidos en las entrevistas, se estimó que el costo promedio por retrabajo de un producto defectuoso oscila entre $25 y $40 USD, considerando materiales, tiempo de intervención y paros de línea. Realizamos el análisis bajo un escenario de 120 fallos mensuales relacionados con las máquinas embobinadoras lo cual representa un gasto estimado de entre $3,000 y $4,800 USD mensuales en retrabajos.

Con la ayuda del modelo predictivo se podría tener los siguientes beneficios :

– Un 25 % de fallos atendidos de forma reactiva, al anticipar ajustes preventivos en las máquinas.

Tabla 2. Soluciones recomendadas para problemas del operador.

Problema del operador	Solución 1	Solución 2	Solución 3
Cortos e insuficiencias	Ajuste de parámetros	Ajuste de parámetros, TC:10MIN	Se ajustó presión a aflux y se ajustó sensor
Sensor con falso contacto	Se ajustó sensor	Se limpió sensor	Se cambió sensor
Falta de flux	Ajuste a cilindro de flux presión de aire	Ajuste de flux parámetro	Se llenó depósito y se estuvieron haciendo ajustes
Falla máquina	Ajuste de parámetros	Ajuste de sensor	Ajuste de olas
Sensor quemado	Se ajustó sensor	Se limpió sensor	Se cambió sensor

– Esto representa un ahorro estimado de entre $750 y $1,200 USD mensuales.

Para poder implementar nuestra propuesta los costos serian:

– El desarrollo y entrenamiento del modelo requiere aproximadamente 60 horas hombre.
– El costo estimado del proyecto sería de alrededor de $3,000 USD (considerando salarios de especialistas en datos).

Por lo tanto, el retorno de inversión (ROI) podría alcanzarse en aproximadamente **tres a cuatro meses** de operación continua.

Este análisis preliminar sugiere que invertir en soluciones predictivas de mantenimiento no solo es técnicamente viable, sino económicamente rentable en el contexto industrial analizado.

Aunque este marco de solución se realizó para una empresa en particular, es posible llevar este esquema a otras fábricas. También logramos construir una solución que involucra tanto al personal de la empresa como el uso de técnicas computacionales.

6. Conclusiones

El análisis de correlaciones entre defectos en sistemas industriales ha demostrado ser una herramienta valiosa para la optimización de los procesos de ensayo de componentes eléctricos. A través de la implementación de técnicas de minería de datos, como los algoritmos Apriori y FP-Growth, fue posible identificar patrones recurrentes de errores, lo que facilita la toma de decisiones basada en datos para la mejora de la calidad y la eficiencia.

Los mapas de calor y los grafos de correlación permitieron visualizar las relaciones más significativas entre los distintos tipos de errores, lo que contribuye

a un diagnóstico más preciso y a la formulación de estrategias preventivas y correctivas. Estas metodologías no solo validaron correlaciones previamente conocidas, sino que también revelaron nuevas asociaciones que pueden ser clave para reducir fallos en el futuro. Encontramos que para la predicción de de soluciones a fallos en máquinas, el algoritmo de regresión logistica tuvo un buen desempeño. La integración de enfoques basados en datos en el análisis de defectos industriales, puede mejorar significativamente la fiabilidad de los productos, reducir costos y optimizar los procesos de prueba en una industria.

Referencias

1. Chapman, S. J.: Máquinas eléctricas. McGraw-Hill Education (2016)
2. Radhakrishnan, S., Pillai, V.: Comparative study on Apriori algorithm and FP Growth algorithm with pros and cons. International Journal of Computer Science Trends and Technology, 4(4), 161–165 (2016)
3. Agrawal, R., Srikant, R.: Fast algorithms for mining association rules. In: Proceedings of the 20th International Conference on Very Large Data Bases, VLDB, pp. 487–499, Santiago, Chile (1994)
4. Zhang, Y., Wang, X.: Deep Learning-Based Defect Detection in Manufacturing Systems: A Review. Journal of Intelligent Manufacturing, 32(3), 567–589 (2021)
5. Brown, P., Liu, J.: Visualization Techniques for Anomaly Detection in Industrial Processes. Industrial Data Science Journal, 15(2), 102–118 (2020)
6. Chen, H., Zhao, L.: Natural Language Processing for Automated Fault Diagnosis in Industry 4.0. IEEE Transactions on Industrial Informatics, 25(4), 3210–3225 (2020)
7. Patel, R., Kumar, S.: Pattern Recognition in Industrial Defects: A Machine Learning Approach. Machine Vision Applications, 18(1), 45–62 (2022)
8. Lee, C., Thompson, B.: Integrating AI-Based Anomaly Detection with Real-Time Data Visualization. Smart Manufacturing Review, 10(1), 88–101 (2023)
9. Garvin, D. A.: Competing on the eight dimensions of quality. Harvard Business Review, 65(6), 101–109 (1987)
10. Deming, W. E.: Out of the Crisis. MIT Press (1987)

ISSN 1870-4069

Multiplatform Application for the Identification of Native Varieties Using Artificial Intelligence and Vector Databases

Pablo Delfino Ortega-Mezhua[1], Humberto Marín-Vega[1,2], Sergio David Ixmatlahua-Díaz[1], Sergio Ignacio Gallardo-Yobal[1], Cristal Ariani Guerrero-Ortiz [1], Emmanuel de Jesús Ramírez-Rivera[1], Giner Alor-Hernández[2]

[1] Tecnológico Nacional de Mexico,
Instituto Tecnológico Superior de Zongolica,
Subdirección de posgrado e investigación,
Mexico

[2] Tecnológico Nacional de México/Instituto Tecnológico de Orizaba,
Mexico

{236w0856, humberto_marin_pd177, sergio.ixmatlahua.pd169,
sergio_gallardo_pd, cristal_guerrero_iias,
ejramirezrivera, giner.ah}@orizaba.tecnm.mx

Abstract. The precise identification of native maize varieties is essential for their conservation and sustainable use. This study presents a multiplatform application that uses artificial intelligence and vector databases to recognize native maize varieties: Olotillo, Pepitilla, and Teocintle. The CLIP model was employed to extract image embeddings, while FAISS was used for efficient similarity searches within the vector database. A dataset of 1,500 manually labeled images was collected, divided into 70% for training and 30% for testing. Image preprocessing included normalization and a data augmentation strategy involving random rotations, scaling, and flipping, which helped improve model robustness and mitigate overfitting. The model was fine-tuned to optimize its performance for the specific maize identification task. The application, developed in React Native with a FastAPI backend, processes images and predicts maize varieties in less than 2 seconds. Experimental results demonstrate an accuracy of 92.4%. The proposed approach significantly outperforms traditional expert-based methods, highlighting the potential of artificial intelligence to support biodiversity conservation and agricultural innovation.

Keywords: Machine learning, vector database, CLIP model, artificial intelligence, maize variety identification.

1 Introduction

Maize is a fundamental crop for global agriculture, particularly in Latin America, where diverse native varieties contribute significantly to food security, cultural heritage, and biodiversity. Accurate identification of native maize varieties is essential to support conservation efforts and promote sustainable agricultural practices. However, traditional classification methods, which rely on expert agronomists and subjective visual analysis, often result in inconsistencies and limited accessibility, especially in rural areas.

To address these challenges, we propose a scalable multiplatform application based on Artificial Intelligence (AI) and vector databases. The solution combines machine learning models with advanced similarity search systems to optimize the identification of native maize varieties. Specifically, the application utilizes Contrastive Language-Image Pretraining (CLIP) to extract semantic embeddings from images, and the Facebook AI Similarity Search (FAISS) library to perform fast and efficient vector-based retrieval.

A dataset of 1,500 manually labeled maize images was collected and enhanced through preprocessing techniques, including normalization and data augmentation (random rotations, scaling, and flipping), to improve model robustness and mitigate overfitting. The deep learning model was fine-tuned to specialize in maize variety identification, enabling accurate classification with limited training data.

The system architecture integrates a React Native-based frontend, which allows users to capture or upload images, and a Fast API backend, which processes requests and predicts the maize variety in less than two seconds. Although initially designed for the Sierra de Zongolica region in Mexico, the proposed system is adaptable and has potential applications in other regions with native crop diversity.

This work highlights the impact of AI-driven solutions on biodiversity conservation and agricultural innovation, offering accessible tools for researchers, farmers, and conservationists worldwide.

The remainder of this paper is structured as follows: Section 2 describes the state of the art regarding both the application of Machine Learning in the Classification of Corn Varieties, Vector Databases and Their Application in Image Retrieval, CLIP Model and Its Use in Image Classification and Artificial Intelligence and Its Impact on the Identification of Native Maize Varieties; in Section 3 describes the methodology of development of the multiplatform application for native maize variety identification, Section 4 presents the results of this research and the Section 5 the conclusions and future work.

2 State of the Art of Species Identification Using AI

Accurate identification of native maize varieties is fundamental for biodiversity conservation and the improvement of agricultural practices. Traditional methods based on expert analysis are often subjective and resource-intensive. In contrast, the integration

of advanced technologies, such as machine learning, vector databases, and computer vision models like CLIP (Contrastive Language-Image Pretraining), has revolutionized the field by enabling more efficient and accurate classification of crop varieties.

2.1 Application of Machine Learning in the Classification of Corn Varieties

Machine learning techniques have been widely adopted for agricultural applications. Patrício and Rieder (2018) applied convolutional neural networks (CNNs) to detect crop diseases using real-field images, achieving an accuracy greater than 90%. Similarly, Sladojevic et al. (2016) trained a CNN to identify plant leaf diseases, reporting a classification accuracy of 96.3%. These studies highlight the viability of deep learning approaches for agricultural tasks.

In a related context, Ramcharan et al. (2019) developed a mobile application for disease diagnosis in African crops using lightweight deep learning models optimized for mobile devices. Their system achieved an accuracy of 93%, demonstrating the feasibility of deploying AI solutions in rural and low-resource environments.

2.2 Vector Databases and their Application in Image Retrieval

Handling large volumes of visual data necessitates efficient storage and retrieval mechanisms. Johnson et al. (2019) introduced FAISS (Facebook AI Similarity Search), a library designed for fast and scalable vector-based searches across millions of embeddings. This technology is particularly suitable for image recognition tasks in agriculture.

Complementarily, Choi et al. (2022) explored the use of vector databases such as We aviate to store semantic representations of images, facilitating similarity search systems even in the absence of manual labels. Pinecone Systems (2021) further demonstrated real-time vector storage for building visual search and recommendation systems, which is essential for applications requiring immediate feedback, such as those used in the agricultural sector.

2.3 CLIP Model and Its Use in Image Classification

Radford et al. (2021) introduced CLIP; a multimodal model capable of associating images with natural language descriptions without task-specific training. CLIP has outperformed traditional classification models in various tasks and offers an innovative alternative for agricultural applications, especially when labeled datasets are scarce.

Goh et al. (2021) successfully applied CLIP to medical and botanical image analysis, demonstrating its adaptability to domain-specific tasks where clear visual representation is critical. This reinforces the potential of CLIP for native maize variety classification.

2.4 Artificial Intelligence and its Impact on the Identification of Native Maize Varieties

Pound et al. (2017) developed image analysis tools for crop phenotyping, enabling the identification of key plant characteristics such as texture, color, and shape through

artificial intelligence. These tools were effectively applied to maize and rice studies, contributing to genetic improvement and automated classification processes.

Additionally, Mohanty et al. (2016) trained a CNN with more than 54,000 images of plants affected by various diseases, laying the groundwork for the development of models tailored to specific crops, including native maize varieties.

2.5 Discussion

The use of machine learning techniques in agriculture has proven to be highly effective, as evidenced by studies using convolutional neural networks (CNNs) with accuracies above 90% in the detection of crop diseases (Patrício and Rieder, 2018; Sladojevic et al., 2016). These technologies, adapted to mobile applications (Ramcharan et al., 2019), allow bringing AI solutions to rural areas, which would be ideal for the identification of native maize varieties. Efficient management of large volumes of images using vector databases such as FAISS (Johnson et al., 2019) and Weaviate (Choi et al., 2022) would facilitate rapid similarity search, optimizing variety recognition. In addition, models such as CLIP (Radford et al., 2021) offer the advantage of working with natural language descriptions, useful in contexts with sparse labeled data. Finally, previous research in crop phenotypic analysis (Pound et al., 2017; Mohanty et al., 2016) demonstrates that AI can identify specific plant characteristics, supporting the feasibility of an application focused on native maize classification, promoting its conservation and sustainable use.

3 Methodology

The development of the multiplatform application for native maize variety identification involved several key stages: dataset preparation, preprocessing and data augmentation, model fine-tuning, system architecture design, and statistical evaluation of results.

3.1 Dataset Preparation and Preprocessing

A total of 1,500 high-quality images of native maize varieties — Olotillo, Pepitilla, and Teocintle — were manually collected from multiple municipalities in the Sierra de Zongolica region, including Rafael Delgado, Tlilapan, Magdalena, Soledad Atzompa, Atlahuilco, Tlaquilpa, Xoxocotla, Tehuipango, Zongolica, Tequila, Astacinga, Mixtla de Altamirano, and Los Reyes. Each image was carefully labeled according to the observed variety.

Preprocessing steps included:

- Resizing all images to 224x224 pixels to match the CLIP model input size.
- Normalization of pixel values to a [0,1] range.
- Format standardization to RGB channels.

Additionally, examples of huitlacoche (Ustilago maydis) were incorporated into the dataset. Huitlacoche is a naturally occurring fungal phenomenon in native maize

ecosystems. Its inclusion aimed to enhance model robustness by exposing it to real-world conditions.

3.2 Data Augmentation

To increase dataset diversity and prevent overfitting, data augmentation techniques were applied:

- Random rotations within ±20 degrees,
- Random scaling up to 20%,
- Horizontal flipping,
- Brightness and contrast variations (±15%).

These transformations simulated different camera angles, lighting conditions, and field situations, improving the model's generalization capabilities.

3.3 Fine-tuning of the CLIP Model

The CLIP (Contrastive Language-Image Pretraining) model was selected due to its strong performance on multimodal tasks. Fine-tuning was performed as follows:

- The lower convolutional layers were frozen to retain general visual features.
- The higher layers were retrained on the maize dataset using a learning rate of 1e- 5.
- Early stopping and dropout regularization (rate 0.3) were applied to avoid overfitting.
- Cross-entropy loss function and Adam optimizer were used for model optimization.
- The model output embeddings were stored in a vector database for later retrieval.

3.4 System Architecture

A robust software architecture is essential in application development as it defines the fundamental structure of the system and the interactions between its components, directly impacting functionality, performance, stability, maintainability, and security.

The architecture of the proposed cross-platform application for the identification of native maize varieties using artificial intelligence and a vector database is illustrated in Figure 1.

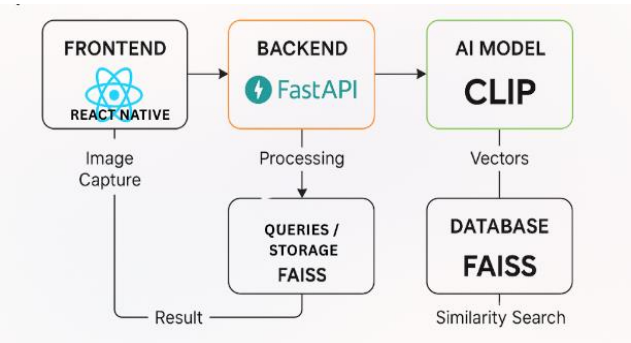

Fig 1. Architecture of the multiplatform application for native maize variety identification using Artificial Intelligence and vector databases.

The system is composed of four main components:

3.4.1 Frontend

The frontend consists of a mobile application developed with React Native, which provides the user interface for capturing and viewing results. Its primary functionalities include:

- ***Image Capture:*** Users can take a photo of a maize ear using the device's camera or upload an existing image from the gallery.
- ***Result Display:*** Displays the classification result, showing the identified maize variety.
- ***API Communication:*** Sends the captured or uploaded image to the backend and retrieves the analysis results.

3.4.2 Backend

The backend, developed with Fast API, is responsible for managing the data processing tasks. It handles:

- ***Image Reception:*** Receives images sent by the frontend and stores them temporarily.
- ***Image Preprocessing:*** Performs resizing, normalization, and formatting of images before inference by the AI model.
- ***Model Inference Communication:*** Sends preprocessed images to the AI model and returns the predicted results.
- ***Result Storage:*** Stores the classification result along with relevant metadata (such as date and time) for future reference.

3.4.3 Artificial Intelligence Model

The AI model is based on Contrastive Language-Image Pretraining (CLIP). Its responsibilities include:

- ***Feature Extraction:*** Converts images into high-dimensional numerical vectors (embeddings) that capture their visual and semantic characteristics.
- ***Classification:*** Compares the generated embeddings against previously stored vectors to associate the input image with the most similar maize variety category.

The CLIP model was fine-tuned using representative examples of Olotillo, Pepitilla, and Teocintle varieties, enabling accurate identification without requiring breed-specific labeling.

3.4.4 Vector Database

The vector database, implemented using Facebook AI Similarity Search (FAISS), manages:

- ***Embedding Storage:*** Stores the feature vectors generated by the AI model for all registered maize varieties.
- ***Similarity Search:*** Efficiently retrieves the most similar vector to the query embedding, allowing fast and accurate identification of the maize variety.

3.5 Statistical Evaluation

To validate the performance of the proposed system, statistical evaluation metrics were calculated on the testing subset (450 images):

- ***Accuracy:*** 92.4%, proportion of correctly classified images.
- ***Precision:*** 93.1%, proportion of true positive identifications among all positive predictions.
- ***Recall:*** 91.7%, proportion of true positives correctly identified out of all actual positives.
- ***F1-Score:*** 92.4%, harmonic mean of precision and recall, balancing the two measures.

These metrics demonstrate the effectiveness of the proposed approach compared to traditional expert-based classification methods. The inclusion of huitlacoche samples further confirmed the model's ability to handle real-world variations commonly found in native maize fields.

4 Results

Multiple experiments were conducted to evaluate the performance of the proposed multiplatform application for native maize variety identification. During the validation

Table 1. Comparison of identification models of native maize varieties from the multi-platform application using Artificial Intelligence and vector database with other methods.

Method	Accuracy	Processing time
Manual classification by experts	78.2%	10-20 min
Standard convolutional neural networks	85.6%	5-7 seg
Proposal based on CLIP + FAISS	92.4%	<2 seg

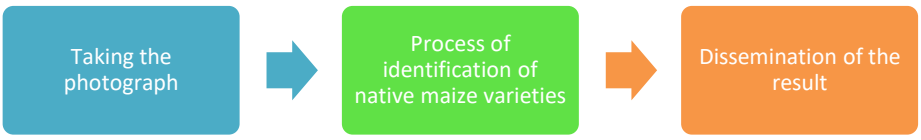

Fig. 2. Identification process of native maize varieties in the multiplatform application using Artificial Intelligence and vector database.

phase, a test set of 450 images was used to assess the model's generalization capabilities. The model achieved an accuracy of 92.4%, significantly outperforming traditional expert-based identification methods.
Table 1 presents a comparison between different identification approaches.

Integration with FastAPI and FAISS enabled efficient searching in the vector database, optimizing the identification process and significantly reducing false positives. The use of the CLIP model for semantic feature extraction, combined with FAISS for fast similarity search, offers a robust and scalable solution for image analysis in the agricultural sector.

In addition to the overall accuracy, further statistical metrics were computed to assess the system's performance:

- ***Precision:*** 93.1% — indicating a high proportion of correct positive identifications.
- ***Recall:*** 91.7% — demonstrating effective detection of true positives.
- ***F1-Score:*** 92.4% — reflecting a balanced performance between precision and recall.

These metrics confirm the robustness of the proposed system in classifying native maize varieties under varying field conditions.

The native maize identification process within the multiplatform application is illustrated in Figure 2.

The identification process is composed of three main steps:

1. **Image Capture:** The user captures an image of the maize ear using the mobile device's camera or selects an image from the gallery.
2. **Identification Process:** The image is preprocessed and passed through the fine-tuned CLIP model, which generates an embedding. FAISS performs a similarity search to determine the closest maize variety match.

Fig.3. Home screen of the multiplatform application for native maize variety identification using Artificial Intelligence and vector databases.

3. **Result Presentation:** The application displays the identified maize variety to the user, accompanied by an image and a brief description.

   This streamlined process enables rural producers to perform rapid and accurate identification of maize varieties directly from their fields.

Figure 3 shows the home screen of the multiplatform application, where a brief introduction to its functionalities is presented. The user can access two main options:

- **Identify:** This option directs the user to an interface where they can either upload an image from their device's gallery or capture a new photo to begin the identification analysis.
- **Contact Us:** This section provides information about researchers and collaborating farmers, offering channels for further information or specialized support.

The decision to present only two main functionalities on the home screen was based on the target users: native maize producers from the Sierra de Zongolica, Veracruz. The interface was deliberately designed to be simple and user-friendly, facilitating quick access without technical complications.

Figure 4 illustrates the maize identification process within the application. Initially, users either capture or upload an image of the maize ear. Once the image is processed, the application automatically initiates the identification process by combining CLIP-based feature extraction with efficient similarity search using FAISS.

After the variety has been identified, the application displays key information, including:

- The processed image of the maize ear,
- The name of the identified variety,
- A brief description of the variety.

Fig. 4. Process of identification of native maize varieties in multiplatform application using Artificial Intelligence and vector database.

4.1 Cultural, Social, and Agricultural Importance

The development of the multiplatform application to identify native maize varieties in the Sierra de Zongolica represents a significant contribution to the preservation of agricultural and cultural heritage. The native maize varieties cultivated in this region are endemic, adapted to the local climate, altitude, and soil conditions, and form an essential part of the communities' identity and diet.

Accurate identification of these varieties is crucial for their conservation, helping prevent the loss of genetic diversity and ensuring that resilient and valuable maize types continue to be cultivated.

The use of Artificial Intelligence empowers farmers by providing an accessible tool that eliminates the dependence on expert agronomists or specialized equipment, which are often unavailable in remote rural areas. This facilitates faster, more accurate, and more inclusive agricultural decision-making, contributing directly to improved crop quality and yield.

Furthermore, the application serves to reconnect communities with their traditional agricultural practices, thereby strengthening food security. Native maize varieties, better adapted to climatic fluctuations, offer a sustainable solution for long-term agricultural resilience.

Economically, identifying and promoting native maize varieties can add value to local production, allowing farmers to better market these culturally and nutritionally important products.

In a broader sense, the integration of modern AI tools with traditional knowledge not only supports agricultural productivity but also promotes cultural preservation, environmental sustainability, and community empowerment.

5 Conclusions and Future Work

The development of the multiplatform application demonstrated that the integration of Artificial Intelligence (AI) and vector databases can significantly enhance the identification of native maize varieties. The combination of the CLIP (Contrastive Language-Image Pretraining) model with the FAISS (Facebook AI Similarity Search) vector search system achieved an accuracy of 92.4%, outperforming traditional human observation-based methods.

The results obtained validate the viability and effectiveness of this approach for agricultural species classification, offering a fast, scalable, and accessible solution for both researchers and farmers. Additionally, the integration with FastAPI enabled real-time processing, reducing system response times to less than two seconds.

However, some challenges remain, including improving the quality of images captured under diverse field conditions and expanding the dataset to include a larger number of native maize varieties. Addressing these challenges would further strengthen the model's accuracy and generalization capabilities.

5.1 Future Work

In order to enhance the application and broaden its scope, the following lines of research are proposed:

- **Dataset Expansion:** Increase the number of training images to cover additional native maize varieties and a broader range of environmental conditions.
- **Model Optimization:** Implement advanced fine-tuning and transfer learning techniques to further improve the performance of the CLIP model in agricultural contexts.
- **Augmentation of Huitlacoche Cases:** Expand the training dataset with additional samples containing huitlacoche (Ustilago maydis) to strengthen the system's ability to handle biological variability naturally present in native maize fields.

References

1. Patrício, D.I., Rieder, R.: Computer vision and artificial intelligence in precision agriculture for grain crops: A systematic review. *Computers and Electronics in Agriculture*, 153, pp. 69–81 (2018)
2. Sladojevic, S., Arsenovic, M., Anderla, A., Culibrk, D., Stefanovic, D.: Deep neural networks-based recognition of plant diseases by leaf image classification. *Computational Intelligence and Neuroscience*, 2016, pp. 1–11 (2016)
3. Ramcharan, A., Baranowski, K., McCloskey, P., Ahmed, B., Legg, J., Hughes, D.P.: A mobile-based deep learning model for cassava disease diagnosis. *Frontiers in Plant Science*, 10, pp. 272 (2019)
4. Johnson, J., Douze, M., Jégou, H.: Billion-scale similarity search with GPUs. *IEEE Transactions on Big Data*, 7(3), pp. 535–547 (2019)
5. Choi, J., Kim, H., Park, Y.: Efficient semantic search using Weaviate vector database. In: Proceedings of the 12th Intl. In: Conf. on Artificial Intelligence and Data Science, pp. 112–121. Springer, Heidelberg (2022)
6. Pinecone Systems: Real-time vector search at scale. https://www.pinecone.io (2025)

7. Radford, A., Kim, J.W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G., Sutskever, I.: Learning transferable visual models from natural language supervision. *arXiv preprint* arXiv:2103.00020 (2021)
8. Goh, G., Agarwal, S., Ramesh, A., Sastry, G., et al.: Multimodal neurons in artificial neural networks. In: *ICLR 2021 Conference Proceedings*, Springer, Heidelberg (2021)
9. Pound, M.P., Atkinson, J.A., Townsend, A.J., Wilson, M.H., Griffiths, M., Jackson, A.S., Bulat, A., Tzimiropoulos, G., Wells, D.M., Murchie, E.H., Pridmore, T.P.: Deep machine learning provides state-of-the-art performance in image-based plant phenotyping. *GigaScience*, 6(10), pp. 1–10 (2017).
10. Mohanty, S.P., Hughes, D.P., Salathé, M.: Using deep learning for image-based plant disease detection. *Frontiers in Plant Science*, 7, pp. 1419 (2016).
11. ASGROW México: Aplicaciones del Machine Learning en el Mejoramiento del Maíz. https://asgrow.com.mx/blog/ml-maiz (2025)
12. Valenzuela, S.P.: Identificación de especies vegetales utilizando dispositivos móviles. Master's Thesis, Universidad Autónoma del Estado de México (2021)
13. Ramírez Gómez, A.L.: Aplicación del Machine Learning en Agricultura de Precisión. *Revista Cintex*, 25(2), pp. 15–28 (2022)
14. Mentores Tech: Bases de datos vectoriales: una herramienta clave para la inteligencia artificial. https://mentorestech.ai/bases-vectoriales (2025)
15. Interempresas: Aplicación de 'Deep Learning' en la detección de malas hierbas en cultivos. https://www.interempresas.net/AI-agricultura (2025)

ISSN 1870-4069

Clasificación de la enfermedad de Alzheimer utilizando Redes Neuronales Profundas Multimodales

Ayrton Santos, Claudia I. González, Mario García

Instituto Tecnológico de Tijuana/TECNM,
División de Estudios de Posgrado e Investigación,
México

ayrton.santos@tectijuana.edu.mx, cgonzalez@tectijuana.mx,
mario@tectijuana.edu.mx

Resumen Las enfermedades neurodegenerativas, como la enfermedad de Alzheimer, representan un desafío creciente para la salud pública global. Con el envejecimiento de la población y los avances en tecnologías de reconocimiento de patrones y aprendizaje automático, la detección temprana de estas patologías ha mejorado considerablemente. En este trabajo se propone el uso de redes neuronales multimodales para la predicción de la enfermedad de Alzheimer integrando imágenes de resonancia magnética por sus siglas en inglés MRI (Magnetic Resonance Imaging) y las Calificaciones Clínicas de Demencia CDR (Clinical Dementia Rating). En la propuesta se evalúan dos enfoques arquitectónicos: fusión temprana y fusión tardía. A través de un análisis estadístico, los resultados obtenidos demuestran que la fusión temprana proporciona una ventaja significativa en el análisis de datos médicos, optimizando la precisión en la clasificación. El modelo se aplica específicamente a la clasificación de la enfermedad de Alzheimer utilizando la base de datos OASIS-3.

Palabras clave: Aprendizaje profundo multimodal, Redes Neuronales Multimodales, clasificación de la enfermedad de Alzheimer, IA en áreas médicas.

Classification of Alzheimer's Disease Using Multimodal Deep Neural Networks

Abstract Neurodegenerative diseases, such as Alzheimer's disease, represent a growing challenge for global public health. With population aging and advances in pattern recognition and machine learning technologies, early detection of these pathologies has considerably improved. This work proposes the use of multimodal neural networks for the prediction of Alzheimer's disease by integrating Magnetic Resonance Imaging (MRI) and Clinical Dementia Ratings (CDR). Two architectural approaches are evaluated in the proposal: early fusion and late fusion.

Through statistical analysis, the results obtained demonstrate that early fusion provides a significant advantage in the analysis of medical data, optimizing classification accuracy. The model is specifically applied to the classification of Alzheimer's disease using the OASIS-3 database.

Keywords: Multimodal deep learning, Multimodal Neural Networks, Alzheimer's disease classification, AI in medical fields.

1. Introducción

Las enfermedades neurodegenerativas son un grupo de trastornos relacionados con la edad que provocan la muerte de tipos específicos de células neuronales. Son más comunes en la población de la tercera edad, lo que las convierte en un grave problema de salud global. Los factores que contribuyen a las enfermedades neurodegenerativas incluyen mutaciones genéticas, apoptosis neuronal, pérdida de proteínas y reacciones de las células gliales [3]. Existen diferentes tipos de enfermedades neurodegenerativas, que comparten síntomas. Estas se caracterizan por el deterioro progresivo de funciones fisiológicas y cognitivas. Las enfermedades que afectan el cerebro son variadas, incluyendo el Alzheimer y la enfermedad de Parkinson [15].

La integración de tecnologías potentes en el área de la salud, como el aprendizaje automático y el aprendizaje profundo, se encuentra en pleno apogeo y se desarrolla a velocidades cada vez más altas, aprovechando el conocimiento acumulado durante años. Esto representa un gran beneficio, ya que, gracias a su precisión, la experiencia tanto para médicos como para pacientes ha mejorado notablemente. Además, los pronósticos pueden realizarse antes de que las enfermedades se vuelvan potencialmente mortales. A esto se suma que las técnicas de aprendizaje automático permiten evaluar la eficacia de nuevos medicamentos de forma más rápida y precisa [8].

Las enfermedades neurodegenerativas abarcan una amplia gama de problemas neurológicos progresivos que son dependientes de la edad y afectan aproximadamente a 50 millones de personas en todo el mundo, especialmente a adultos mayores. En las últimas décadas, el aumento en la población anciana proyecta una tasa de mortalidad del 42 %, representando el 11.8 % de todas las muertes. Es importante mencionar que algunas enfermedades no se comprenden completamente, y para otras no existe cura o tratamiento, lo que finalmente lleva a la muerte del paciente [16].

En un estudio publicado en 2020, aproximadamente 800,000 personas estaban afectadas por demencia. La prevalencia actual varía entre el 7.9 % y el 9 %, colocando a México en el 5° lugar con una alta incidencia. Según este estudio, se estima que para el año 2050 hasta 3 millones de personas estarán afectadas por demencia [14]. Como dato interesante, la relación de riesgo de mortalidad en áreas urbanas es mayor, con un 2.7. Esto se debe a muchos factores, como la calidad de la atención médica, la contaminación, entre otros, mientras que en áreas rurales es de 1.6 [17].

La enfermedad de Alzheimer, una de las principales causas de demencia en el mundo, continúa representando un desafío significativo para la salud pública debido a su impacto en la calidad de vida de los pacientes y en los sistemas de atención sanitaria. A medida que la población global envejece, la necesidad de métodos eficaces para la detección temprana y el diagnóstico preciso de esta enfermedad se vuelve aún más urgente. La multimodalidad ha cobrado gran relevancia en los últimos años debido a la necesidad de integrar diferentes tipos de datos, como texto, imagen y audio, en un mismo modelo de aprendizaje. En este contexto, las tecnologías de aprendizaje automático y el análisis de datos multimodales han abierto nuevas posibilidades para mejorar la precisión de la clasificación de enfermedades neurodegenerativas. Este trabajo presenta una propuesta para el diseño y desarrollo de arquitecturas multimodales basadas en redes profundas para la predicción de la enfermedad de Alzheimer. Combinando imágenes de resonancia magnética (MRI) y las Calificaciónes Clínicas de Demencia (CDR), se analizan dos enfoques arquitectónicos: fusión temprana y fusión tardía, mostrando que la fusión temprana mejora de manera significativa la precisión en el diagnóstico. Los resultados obtenidos, aplicados sobre la base de datos OASIS-3, demuestran el potencial de estas técnicas para avanzar en la detección temprana de la enfermedad, mejorando tanto la eficiencia en el diagnóstico. Este enfoque permite mejorar el desempeño en tareas complejas al aprovechar la complementariedad de las distintas fuentes de información.

En la Sección 2 se definen algunos de los conceptos básicos sobre la multimodalidad, redes profundas multimodales y principales arquitecturas multimodales. En la Sección 3 se presentan algunos antecedentes de investigaciones relacionadas a arquitecturas multimodales. En la Sección 4 se describe la metodología propuesta para el diseño de arquitecturas multimodales con fusión temprana y fusión tardía. En la Sección 5 se detallan los resultados obtenidos y finalmente, en la Sección 7, se abordan las conclusiones y el trabajo futuro.

2. Multimodalidad

2.1. Redes Neuronales Profundas Multimodales

Las redes neuronales son una subdivisión del aprendizaje profundo que usa imágenes para clasificar imágenes, aunque también se ha visto que se utilizan también para áreas como el procesamiento de sonido. En comparación con las redes neuronales tradicionales, en las que cada neurona está conectada a la siguiente (lo cual se conoce como capa densa o completamente conectada), las redes neuronales convolucionales utilizan la capa densa solo en la última parte de la red neuronal. Las redes neuronales logran el reconocimiento de formas mediante la combinación de agrupación (pooling) y capas densas. Las capas convolucionales procesan los datos de entrada usando múltiples filtros, se aplica una función de activación, y posteriormente las capas de agrupación extraen las características más significativas mediante operaciones como agrupamiento máximo (max pooling) o agrupamiento promedio (average pooling). Después

del aprendizaje, las capas completamente conectadas se convierten en un vector unidimensional, que es introducido en una función Softmax para la construcción del modelo [7].

Estas redes combinan tanto entradas visuales como lingüísticas para aprender correspondencias entre imágenes (referentes) y palabras. Algunos ejemplos de sus aplicaciones son la generación automática de subtítulos para imágenes y respuestas a preguntas visuales. Las redes multimodales emplean codificadores de imágenes y palabras que transforman las entradas en un espacio de representación multimodal compartido. Se utiliza una función de pérdida contrastiva para alinear los pares palabra-referente mientras se separan los que no están relacionados. Estas redes muestran potencial como modelos cognitivos al aprender de estímulos visuales y lingüísticos sin procesar, lo cual imita cómo los niños podrían aprender palabras en entornos ambiguos. Este proceso es conocido como aprendizaje cross-situacional. Se realizan experimentos bajo condiciones como ambigüedad referencial, mapeo rápido y exclusividad mutua para comprobar si las redes multimodales replican comportamientos de aprendizaje humano. Estos experimentos demuestran cómo los modelos multimodales pueden manejar tanto ambigüedad lingüística como tareas de generalización visual [19].

En [2], se presenta la idea de que el mundo es multimodal, ya que podemos ver objetos, sentir texturas, escuchar sonidos y experimentar sabores. Para que los algoritmos de inteligencia artificial puedan lograr mejores resultados, es necesario, en primer lugar, enfocarse en la forma de capturar y resumir los datos multimodales. Además, se debe considerar la traducción, que implica mapear los datos de una modalidad a otra, tomando en cuenta la heterogeneidad; la alineación, que consiste en encontrar relaciones entre los elementos de las modalidades; y finalmente, la fusión, que se centra en combinar la información de varias modalidades para hacer predicciones integrando la información multimodal. Hay mucha discusión sobre la diferencia entre las redes neuronales, multimodales y las modulares. La tecnología multimodal se refiere a la capacidad de un sistema para procesar y combinar múltiples tipos de datos, o modalidades, como texto, imágenes, audio, etc., con el fin de obtener mejores resultados.

En [20] se analizan los modelos multimodales actuales y se clasifican distintos tipos de arquitecturas de modelos multimodales de última generación; se describe de manera detallada de los tipos de fusión temprana y fusión tardía para entender las diversas arquitecturas, las cuales ser dividen en A, B, C y D como se muestra en la Figura 1.

La arquitectura tipo A abarca los modelos multimodales tempranos, los datos multimodales como imagen, audio y video, se procesan a través de codificadores, específicos para cada modalidad, un resampleador genera un número de tokens fijo, que se alinean con la capa decodificadora. La arquitectura tipo B está hecha usando un modelo de lenguaje pre-entrenado una capa lineal, que puede aprender de un módulo Q Former, capas de atención cruzada personalizada o capas personalizadas, y codificadores de modalidad. La arquitectura tipo C es la más usada, caracterizándose por incluir módulos y ser simple en

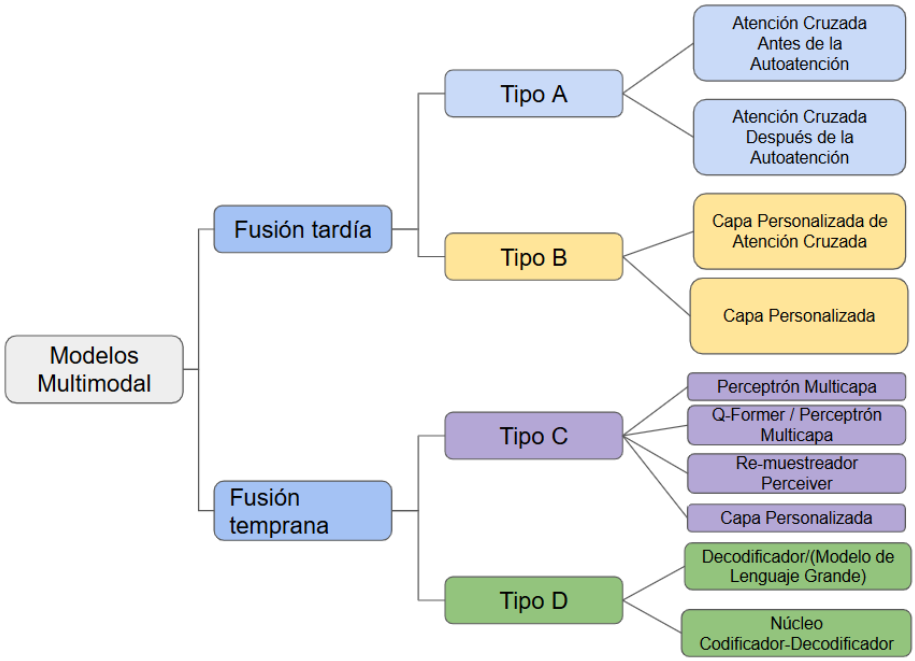

Fig. 1. Arquitecturas multimodales principales [20].

desarrollo y entrenamiento. La salida del codificador de modalidad se dirige y fusiona en la entrada del modelo sin interacciones con las capas internas, lo que permite una fusión temprana. El entrenamiento consta de tres etapas: pre-entrenamiento, ajuste con instrucciones y ajuste para alineación. Es la que menos recursos demanda, y este tipo de arquitectura permite facilitar la adopción y el entrenamiento en tareas multimodales. La clase D tokeniza las entradas multimodales usando un tokenizador común o específicos para cada modalidad. Las entradas tokenizadas se alimentan a un modelo de lenguaje grande o a un transformer tipo codificador-decodificador, generando salidas multimodales. La ventaja de tokenizar las entradas es que permite generar tokens de diversas modalidades (imagen, audio y texto) de manera autoregresiva aunque la tipo D tiene un enfoque más generativo que de clasificación como lo requerimos en este caso [20].

3. Antecedentes

En el estudio de Chattopadhyay [4], se utilizó una red neuronal convolucional 3D para predecir la acumulación de la proteína beta amiloide, un factor para desarrollar la enfermedad de alzheimer basándose en resonancias magnéticas, utilizando biomarcadores. El estudio consistió en 762 pacientes ancianos: 459 sanos y 67 con deterioro cognitivo leve. Los resultados mostraron que 236 pacientes tenían demencia, y la precisión obtenida fue del 76 %.

En [11] se presentó una selección de características multimodal de 3 tipos para la detección de Alzheimer: imágenes de resonancia magnética, tomografía por emisión de positrones y análisis de líquido cefalorraquídeo. El aprendizaje multimodal implica el uso de diferentes tipos de datos y el aprendizaje de lo mejor de esa información combinada, ya sea texto, audio o video. El problema es que los algoritmos comunes trabajan con solo un tipo de datos a la vez, lo que significa que no siempre se puede utilizar toda la información disponible.

En [5] el trabajo se centra en la enfermedad de Alzheimer; se realiza una predicción temprana de Alzheimer y deterioro cognitivo leve utilizando imágenes de resonancia magnética cerebral y aprendizaje automático. Se utilizaron dos conjuntos de datos: la Serie de Estudios de Imágenes de Acceso Abierto (OASIS) y la Iniciativa de Neuroimagen de la Enfermedad de Alzheimer (ADNI). Algunos de los algoritmos utilizados incluyeron máquinas de vectores de soporte, árboles de decisión, bosques aleatorios, árboles extremadamente aleatorios, análisis discriminante lineal, regresión logística y regresión logística con descenso de gradiente estocástico. Los mejores resultados se obtuvieron con los algoritmos de bosque aleatorio y árboles extremadamente aleatorios.

En [18], se exploraron múltiples herramientas estadísticas y algoritmos de aprendizaje automático para el diagnóstico de la enfermedad de Alzheimer en personas mayores de 75 años. A medida que la tecnología mejoró, se aplicó a la clasificación de imágenes médicas. Las muestras se dividieron en 75 % y 25 %, utilizadas para entrenar los algoritmos, que se ejecutaron en potentes GPUs. Utilizando redes neuronales convolucionales, se empleó la arquitectura optimizada llamada OViTAD. Estas tuberías lograron resultados promedio de 94.32 % y 97.88 % para las tuberías de fMRI y MRI, respectivamente.

En [6] se propone un estudio que compara el rendimiento de tres algoritmos de aprendizaje automático: Random Forest, Gradient Boosting y eXtreme Gradient Boosting, utilizando biomarcadores multimodales de sujetos con deterioro cognitivo leve (MCI) obtenidos de la base de datos ADNI. La tasa de predicción está relacionada con la naturaleza de los datos, como pruebas neuropsicológicas, proteínas relacionadas con el Alzheimer, líquido cefalorraquídeo, MRI, entre otros. Los datos sMRI (MRI estructurales), por sí solos tuvieron menor precisión (79 %), pero al tratarse de datos multimodales, combinando medidas clínicas y biológicas, se logró una mayor precisión (90 %).

En [13], se presentó un estudio que discute cómo las personas con la enfermedad de Parkinson pueden desarrollar demencia, pero no todos los pacientes la desarrollan de manera gradual o al mismo tiempo. Se recopilaron datos de 48 pacientes con Parkinson, en los que se evaluaron 38 características de riesgo, como habilidades motoras, capacidades cognitivas, moléculas en sangre, entre otros factores. Se utilizó el modelo de Random Forest, que es un algoritmo de aprendizaje automático, y una técnica llamada Tree SHAP, que explica por qué ciertos factores son importantes para las predicciones del modelo. Random Forest fue muy preciso al clasificar qué pacientes desarrollaron demencia, con un área bajo la curva (AUC) de 0.84.

Tabla 1. Resumen de técnicas y estudios sobre Alzheimer

Técnicas	**Base de Datos**	**Autores y Año**
CNN 3D	MRI + biomarcadores	Chattopadhyay (2023) [4]
Selección multimodal + Graph Anchors	MRI, PET, CSF	Li (2022) [11]
SVM, Árboles, RF, Reg. Logística	OASIS, ADNI, MRI, PET, CSF	Diogo (2022) [5]
CNN (OViTAD)	fMRI, MRI	Sarraf (2021) [18]
RF, Gradient Boosting, XGBoost	ADNI, MCI	Franciotti (2023) [6]
RF, Tree SHAP	Datos clínicos (Parkinson)	McFall (2023) [13]
Fusión multimodal tiempo-frecuencia	Neuroimagen y genética	Anand (2024) [1]
MC-RVAE, KNN, RF, GFA, ANOVA	MRI + cognitivas	Martí-Juan (2023) [12]
Contrastive Learning + ResNet + Tabular	MRI + datos tabulares	Huang (2023) [9]

En [1], se propone un modelo que es multimodal para lograr mejorar la detección de enfermedades neurodegenerativas. Usa distintos métodos avanzados para analizar la variedad de datos, análisis de tiempo-frecuencia, resonancias electromagnéticas, y datos genéticos. Estas características de datos permiten un diagnóstico preciso, con un aumento del 10 % en precisión y una reducción del tiempo del 2,9 %. Este estudio se enfocó en una amplia gama de enfermedades neurodegenerativas.

El modelo MC-RVAE [12], diseñado para manejar la enfermedad de Alzheimer de manera multimodal, trabaja con MRI y puntuaciones cognitivas. Este modelo fue entrenado con datos sintéticos y de ADNU con aproximadamente 3000 épocas. Se usaron los vecinos más cercanos, bosques aleatorios y análisis de factor de grupo. Los resultados se compararon usando un modelo ANOVA para cada tarea; este es flexible y escalable.

El modelo [9] utiliza aprendizaje contrastivo, que alinea las imágenes de resonancia y los datos tabulares, creando un espacio embebido conjunto para diferentes modalidades. Esta propuesta incluye capacidades de entrenamiento multimodal. El encoder, basado en ResNet, se encarga de procesar imágenes e integra un módulo de atención tabular que resalta los datos más relevantes y mejora la interpretación de los mismos. El modelo asigna puntuaciones para capturar las relaciones entre los datos. Es entrenado con 64 épocas, usando un optimizador Adam. El tamaño del batch es de 4 para imágenes 3D y 32 para imágenes 2D, alcanzando una precisión del 95.5 %. La adición de datos tabulares mejora significativamente el rendimiento del modelo.

4. Metodología

4.1. Datos multimodales

OASIS-3 es una recopilación longitudinal multimodal especializada en la enfermedad de Alzheimer. Contiene información de 1378 pacientes, con edades entre 42 y 95 años, recopilada a lo largo de 30 años a través de diversos estudios del Knight Alzheimer's Disease Research Center de la Universidad de Washington en St. Louis. El conjunto incluye datos crudos de MRI y de tomografías PET (Positron Emission Tomography), así como evaluaciones clínicas y cognitivas, estructuradas en más de 2800 sesiones de resonancia magnética (T1w, FLAIR, ASL, DTI, entre otras), y más de 2100 sesiones PET con distintos trazadores (PIB, AV45, FDG y Tau). Además, cuenta con una sección preprocesada mediante FreeSurfer, ideal para investigadores que no deseen limpiar manualmente los datos crudos de MRI. Gracias a esta herramienta, científicos e ingenieros pueden desarrollar modelos de inteligencia artificial con alta precisión [10].

4.2. Preprocesamiento de datos

Primero, se solicitó acceso al dataset. Una vez concedidos los permisos, se procedió a descargar los datos: 800 GB de información cruda que incluía evaluaciones médicas, historiales clínicos y familiares de los pacientes, así como tomografías computarizadas e imágenes de resonancia magnética (MRI). La etapa inicial del procesamiento se centró en la limpieza de los MRI. Afortunadamente, el equipo que proporciona el dataset incluye una sección específica con imágenes en un formato compatible con la herramienta FreeSurfer, lo que facilita su manipulación. Un aspecto destacable es que estas imágenes ya vienen preprocesadas: el cerebro está segmentado, es decir, aislado de otras estructuras como órganos o huesos, lo cual optimiza el reconocimiento de patrones cerebrales. Con esta base, se seleccionaron 26 cortes axiales del cerebro, ya que en esas secciones se encuentra el hipocampo, una región clave para la detección temprana del Alzheimer. A continuación, se utilizó un script para convertir las imágenes a escala de grises y redimensionarlas a 128×128 píxeles. Continuando con el uso del CDR, se consideró el diagnóstico más reciente de cada paciente para determinar la presencia o ausencia de Alzheimer. Con esta información, se ejecutó un script que emparejaba las imágenes de resonancia magnética (MRI) con sus respectivos valores de CDR, generando así el conjunto de datos necesario para entrenar la red neuronal. El pre-procesamiento previamente mencionado, derivó una base de datos con 1,248 registros de MRI y de CDR, de los cuales se dividieron en un 70 % para entrenamiento, 15 % para validación y 15 % para pruebas.

4.3. Flujo de trabajo

En las propuestas de arquitectura existen dos líneas principales: la rama de CDR, donde se toman datos volumétricos del cerebro, además del indicativo de

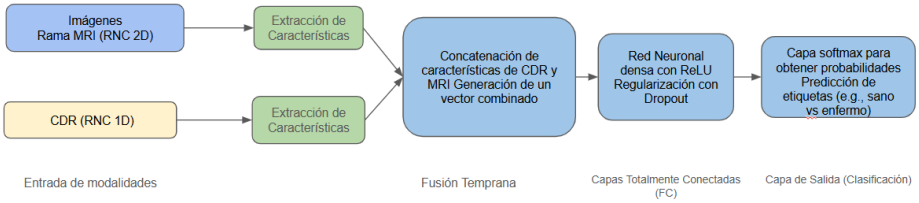

Fig. 2. Arquitectura de la fusión temprana.

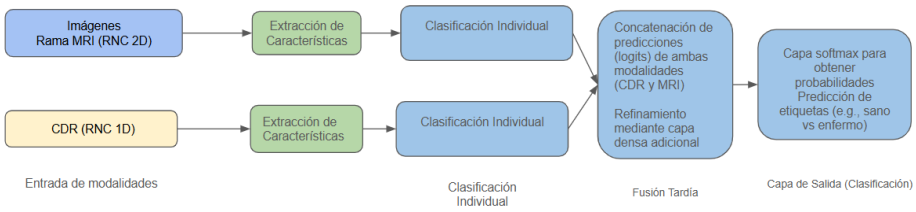

Fig. 3. Arquitectura de la fusión tardía.

paciente enfermo o sano; y, por otro lado, la rama de imágenes de resonancias magnéticas (MRI). La fusión se lleva a cabo en una etapa diferente dependiendo del tipo de fusión. Para la fusión temprana (Early Fusion), como se muestra en la Figura 2, se cuenta con una red neuronal 1D para procesar el CDR y una red neuronal convolucional 2D para las MRI. La fusión ocurre al concatenar las dos salidas y pasarlas por capas totalmente conectadas (fully connected). Cabe destacar que en este enfoque multimodal se utilizaron técnicas para mitigar el sobreajuste con dropout en las capas totalmente conectadas, se normaliza el batch tras la primera capa lineal que trabaja con los datos de CDR, y se emplea data augmentation, que genera rotaciones aleatorias en las imágenes para forzar al modelo a generalizar en vez de aprender de memoria.

La fusión tardía, como se muestra en la Figura 3, posee una arquitectura con una red neuronal convolucional de una capa que procesa el CDR, y otra red neuronal convolucional de dos capas que trabaja con MRI. A diferencia de la fusión temprana, que utiliza un vector de características para la concatenación, en la fusión tardía se obtienen los logits de cada red neuronal; luego, dichos logits se concatenan y se pasan por una capa final que genera la predicción.

Se crean las funciones que entrenan el modelo, permitiendo que procese la información del dataset, calcule posibles errores y ajuste los parámetros. Una segunda función se encarga de evaluar el modelo, registrando las predicciones en un log y calculando diferentes métricas como la exactitud, precisión, sensibilidad, especificidad y el área bajo la curva (AUC), para analizar su rendimiento.

5. Resultados

Se llevaron a cabo 30 experimentos independientes, cada uno con 100 épocas y un tamaño de lote (batch size) de 32, con el objetivo de evaluar el desempeño de

Tabla 2. Resultados del modelo con fusión temprana.

Pérdida entrenam.	Pérdida validación	Exactitud	Precisión	Sensibilid.	Especificidad	AUC
0.02	0.04	0.99	0.99	0.99	0.98	1
0.02	0.02	0.99	1	0.99	0.99	1
0.02	0.02	0.99	1	0.99	0.99	1
0.02	0.05	0.99	1	0.98	0.99	1
0.02	0.04	0.99	0.99	0.99	0.98	1
0.02	0.04	0.99	0.99	1	0.97	1
0.02	0.03	0.99	0.99	0.99	0.98	1
0.02	0.05	0.99	0.99	0.99	0.96	1
0.02	0.06	0.99	0.98	0.99	0.96	1
0.02	0.05	0.99	0.99	0.99	0.97	1
0.02	0.05	0.99	0.99	1	0.98	1
0.02	0.04	0.99	0.99	0.98	0.98	1
0.02	0.03	0.99	0.99	0.99	0.98	1
0.02	0.02	0.99	0.99	1	0.98	1
0.02	0.02	0.99	1	0.99	0.99	1
0.02	0.04	0.99	0.99	1	0.97	1
0.02	0.06	0.99	0.99	0.99	0.97	1
0.02	0.03	0.99	0.99	0.99	0.99	1
0.02	0.02	0.99	0.99	0.99	0.98	1
0.02	0.03	0.99	1	0.99	0.99	1
0.02	0.03	0.99	0.99	1	0.98	1
0.02	0.02	0.99	1	0.99	0.99	1
0.02	0.05	0.99	0.99	0.99	0.98	0.99
0.02	0.09	0.99	0.99	0.97	0.99	1
0.02	0.06	0.99	0.99	0.99	0.97	1
0.02	0.02	0.99	0.99	1	0.98	1
0.02	0.05	0.99	0.99	1	0.97	1
0.02	0.05	0.99	0.99	1	0.96	1
0.02	0.03	0.99	0.99	0.98	0.99	1
0.02	0.02	0.99	0.99	0.99	0.98	1

dos arquitecturas: fusión temprana y fusión tardía. El entrenamiento completo de la arquitectura de fusión temprana tomó aproximadamente 44 horas, mientras que la fusión tardía requirió alrededor de 39 horas El entrenamiento se llevo en un equipo con un procesador Intel Core i9-13900KF, con dos módulos de RAM combinados que suman 128 GB de memoria RAM, y una tarjeta de video NVIDIA GeForce RTX 4080 SUPER con 16 GB de memoria VRAM GDDR6X, con el framework PyTorch, el módulo Python OS para trabajar con los archivos del sistema, NumPy para operaciones matemáticas, Pandas para gestionar datos tabulares y PIL para trabajar con imágenes. Antes de seleccionar estas arquitecturas finales, se realizaron pruebas preliminares con distintas

Tabla 3. Resultados del modelo con fusión tardía.

Pérdida entrenam.	**Pérdida validación**	**Exactitud**	**Precisión**	**Sensibilid.**	**Especificidad**	**AUC**
0.05	0.03	0.99	0.99	0.99	0.97	1
0.05	0.05	0.98	0.98	0.99	0.95	0.99
0.04	0.03	0.99	0.99	0.99	0.97	1
0.05	0.03	0.99	0.99	0.99	0.98	1
0.05	0.02	0.99	0.99	0.99	0.98	1
0.05	0.02	0.99	0.99	0.99	0.98	1
0.04	0.07	0.97	0.97	0.99	0.93	0.99
0.04	0.05	0.98	0.98	0.99	0.96	1
0.04	0.03	0.99	0.99	1	0.97	1
0.05	0.03	0.99	0.99	1	0.98	1
0.05	0.03	0.99	0.99	1	0.97	1
0.04	0.04	0.98	0.98	0.99	0.96	1
0.05	0.02	0.99	1	0.99	0.99	1
0.04	0.03	0.99	0.99	1	0.98	1
0.04	0.03	0.99	0.99	1	0.97	1
0.04	0.02	0.99	0.99	1	0.98	1
0.04	0.03	0.99	0.99	0.99	0.97	1
0.04	0.02	0.99	0.99	0.99	0.98	1
0.04	0.03	0.99	0.99	1	0.97	1
0.05	0.02	0.99	0.99	0.99	0.98	1
0.06	0.03	0.99	0.99	1	0.96	1
0.04	0.03	0.99	0.99	1	0.96	1
0.04	0.05	0.98	0.99	0.99	0.96	1
0.04	0.05	0.98	0.99	0.99	0.97	1
0.05	0.02	0.99	0.99	0.99	0.98	1
0.04	0.03	0.99	0.99	0.99	0.98	1
0.05	0.04	0.98	0.98	1	0.96	1
0.05	0.03	0.99	0.99	0.99	0.97	1
0.04	0.04	0.98	0.98	0.99	0.95	1
0.05	0.03	0.99	0.99	1	0.97	1

configuraciones para identificar cuál ofrecía la mayor precisión, resultando seleccionadas las mencionadas por su superior desempeño. Para la evaluación de los experimentos, se consideraron las siguientes métricas: exactitud, precisión, sensibilidad, especificidad y área bajo la curva (AUC). Asimismo, se incluyeron la pérdida promedio durante el entrenamiento y la pérdida promedio en la fase de validación. Los resultados obtenidos con la arquitectura de fusión temprana se presentan en la Tabla 2, mientras que los correspondientes a la fusión tardía se muestran en la Tabla 3.

5.1. Análisis estadístico

En esta Sección se argumenta que la arquitectura de fusión temprana presenta una ventaja significativa en términos de rendimiento frente a la fusión tardía. Para evaluar si la diferencia en el rendimiento promedio es

Tabla 4. Resultados de precisión en 30 experimentos para fusión temprana y fusión tardía.

No. de experimento	Fusión Temprana (%)	Fusión Tardía (%)
1	0.99	0.99
2	0.99	0.98
3	0.99	0.99
4	0.99	0.99
5	0.99	0.99
6	0.99	0.99
7	0.99	0.97
8	0.99	0.98
9	0.99	0.99
10	0.99	0.99
11	0.99	0.99
12	0.99	0.98
13	0.99	0.99
14	0.99	0.99
15	0.99	0.99
16	0.99	0.99
17	0.99	0.99
18	0.99	0.99
19	0.99	0.99
20	0.99	0.99
21	0.99	0.99
22	0.99	0.99
23	0.99	0.98
24	0.99	0.98
25	0.99	0.99
26	0.99	0.99
27	0.99	0.98
28	0.99	0.99
29	0.99	0.98
30	0.99	0.99
Media (μ)	**0.990**	**0.987**
Desviación estándar (σ)	**0.000**	**0.0053**

estadísticamente significativa, se aplicó una prueba Z. La métrica utilizada en este análisis fue la precisión, cuyos resultados se detallan en la Tabla 4.

A partir de los resultados anteriores, se realizó una prueba estadística para evaluar si la diferencia entre ambas técnicas es significativa.

Hipótesis nula (H_0): No hay diferencia significativa entre los desempeños de la red de fusión temprana y la red de fusión tardía:

$$H_0 : \mu_1 = \mu_2.$$

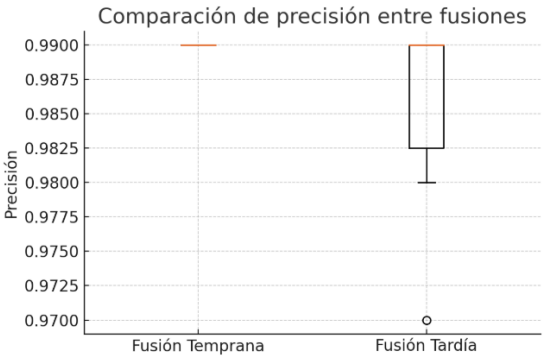

Fig. 4. Comparación de precisión entre fusión temprana y fusión tardía (Boxplot).

Hipótesis alternativa (H_a): Existe una diferencia significativa en el desempeño entre ambas redes:

$$H_a : \mu_1 \neq \mu_2.$$

Se utilizó una prueba Z para dos muestras independientes con los siguientes resultados:

- Estadístico Z: **3.071**,
- Valor p: **0.0021**.

Dado que el valor $p < 0{,}05$, se **rechaza la hipótesis nula**, indicando que la diferencia entre los modelos de fusión temprana y tardía es estadísticamente significativa.

En la Figura 4 se presenta una gráfica comparativa que ilustra la variabilidad y tendencia central de ambas técnicas.

Las redes neuronales multimodales logran obtener precisiones altas cuando se optimizan de manera correcta y se tienen datasets afinados y correctamente distribuidos. La fusión temprana nos da resultados consistentes; por otra parte, la fusión tardía permite mayor flexibilidad, aunque puede tener mayor variabilidad. A pesar de tener ventajas en rendimiento, también la complejidad técnica puede jugar en contra, ya que es necesario diseñar y distribuir los datos adecuadamente, puesto que no siempre estarán balanceados.

6. Conclusiones y trabajo futuro

Este estudio aborda el diagnóstico temprano de la enfermedad de Alzheimer con aprendizaje profundo multimodal, integrando dos tipos de datos, calificaciones clínicas de demencia (CDR) y resonancias magnéticas (MRI). Se evalúan dos tipos de arquitectura, la arquitectura de fusión temprana y la arquitectura de fusión tardía. La arquitectura de fusión temprana es más

lenta a la hora de entrenar, pero logró mejores resultados en la clasificación de la enfermedad de Alzheimer, ayudado por potente hardware y herramientas como PyTorch. En cuanto a la fusión tardía, se llevó menos tiempo de entrenamiento, pero su clasificación se vio reducida, lo cual nos deja con la fusión temprana como una opción viable para seguir trabajando y mejorando. La fusión de distintos tipos de datos cuidadosamente preprocesados logró resultados visiblemente buenos.

Esto es especialmente importante, ya que se estima que un gran número de personas mayores se verá afectado por enfermedades neurodegenerativas, y es crucial generar conciencia, ya que esto podría tener un impacto en muchas áreas de la sociedad. Las investigaciones futuras deberían integrar diversos tipos de datos, como imágenes médicas, información genética, registros clínicos e información de sensores, para mejorar la precisión de las predicciones y las capacidades diagnósticas en el ámbito de las enfermedades neurodegenerativas. Uno de los grandes desafíos del aprendizaje profundo es su naturaleza de caja negra.

El trabajo futuro debe centrarse en crear modelos más explicables que permitan a los profesionales médicos comprender y confiar mejor en los resultados. Además, fomentar una mayor colaboración entre científicos de datos, neurólogos y profesionales de la salud conducirá a modelos mejor alineados con las necesidades y prácticas clínicas. Finalmente, a medida que el aprendizaje profundo se vuelve más integral en la atención médica, abordar las preocupaciones éticas y garantizar la privacidad y seguridad de los datos será fundamental, especialmente al tratar con información médica sensible.

Agradecimientos. Agradecemos al TECNM/Instituto Tecnológico de Tijuana y a la SECIHTI por el apoyo financiero otorgado mediante el proyecto CF-2023-I-555.

Referencias

1. Anand, V. R., Priyan, S. T., Brahmam, M. G., Balusamy, B., Benedetto, F.: IMNMAGN: Integrative multimodal approach for enhanced detection of neurodegenerative diseases using fusion of multidomain analysis with graph networks. IEEE Access, (2024) doi: 10.1109/ACCESS.2024.3403860
2. Baltrušaitis, T., Ahuja, C., Morency, L.-P.: Multimodal machine learning: A survey and taxonomy. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 41, no. 2, pp. 423–443 (2019) doi: 10.1109/TPAMI.2018.2798607
3. Candelise, N., Baiardi, S., Franceschini, A., Rossi, M., Parchi, P.: Towards an improved early diagnosis of neurodegenerative diseases: the emerging role of in vitro conversion assays for protein amyloids. Acta Neuropathologica Communications, vol. 8, no. 1, pp. 1–16 (2020) doi: 10.1186/s40478-020-00940-w
4. Chattopadhyay, T., Ozarkar, S. S., Buwa, K., Thomopoulos, S. I., Thompson, P. M.: Predicting brain amyloid positivity from T1 weighted brain MRI and MRI-derived gray matter, white matter and CSF maps using transfer learning on 3D CNNs. bioRxiv, (2023) doi: 10.1101/2023.02.15.528705
5. Diogo, V. S., Ferreira, H. A., Prata, D.: Early diagnosis of alzheimer's disease using machine learning: a multi-diagnostic, generalizable approach. Alzheimer's Research & Therapy, vol. 14, no. 107, pp. 1–15 (2022) doi: 10.1186/s13195-022-01047-y

6. Franciotti, R., Nardini, D., Russo, M., Onofrj, M., Sensi, S. L., Alzheimer's Disease Neuroimaging Initiative, Alzheimer's Disease Metabolomics Consortium: Comparison of machine learning-based approaches to predict the conversion to Alzheimer's disease from mild cognitive impairment. Neuroscience, vol. 514, pp. 143–152 (2023)
7. González Berrelleza, C. I.: Método de detección de bordes por medio de lógica difusa tipo-2 generalizada. Ph.D. thesis, Universidad Autónoma de Baja California (2016), tesis de doctorado, Facultad de Ciencias Químicas e Ingeniería, Tijuana, Baja California
8. Hajdu Macelaru, M., Chiuzbaian, R., Pop, P.: Machine learning approaches in the detection of amyotrophic lateral sclerosis disease using orofacial gestures (2024), manuscript
9. Huang, W.: Multimodal contrastive learning and tabular attention for automated Alzheimer's disease prediction. http://arxiv.org/abs/2308.15469v1 (2023)
10. LaMontagne, P. J., Benzinger, T. L. S., Morris, J. C., Keefe, S., Hornbeck, R., Xiong, C., Grant, E., Hassenstab, J., Moulder, K., Vlassenko, A., Raichle, M. E., Cruchaga, C., Marcus, D.: OASIS-3: Longitudinal neuroimaging, clinical, and cognitive dataset for normal aging and Alzheimer disease. medRxiv, (2019) doi: 10.1101/2019.12.13.19014902
11. Li, J., Xu, H., Yu, H., Jiang, Z., Zhu, L.: Multi-modal feature selection with anchor graph for Alzheimer's disease. Frontiers in Neuroscience, vol. 16, pp. 1036244 (2022) doi: 10.3389/fnins.2022.1036244
12. Martí-Juan, G., Lorenzi, M., Piella, G.: Mc-rvae: Multi-channel recurrent variational autoencoder for multimodal Alzheimer's disease progression modelling. NeuroImage, vol. 268, pp. 119892 (2023) doi: 10.1016/j.neuroimage.2023.119892
13. McFall, G. P., Bohn, L., Gee, M., Drouin, S. M., Fah, H., Han, W., Dixon, R. A.: Identifying key multi-modal predictors of incipient dementia in Parkinson's disease: a machine learning analysis and tree SHAP interpretation. Frontiers in Aging Neuroscience, vol. 15, pp. 1124232 (2023) doi: 10.3389/fnagi.2023.1124232
14. Pan American Health Organization: Dementia in Latin America and the Caribbean: prevalence, incidence, impact, and trends over time. PAHO, Washington, DC (2023), https://doi.org/10.37774/9789275326657
15. Pao, P., Patnaik, D., Watson, L., Gao, F., Pan, L., Wang, J., Tsai, L.: HDAC1 modulates OGG1-initiated oxidative DNA damage repair in the aging brain and Alzheimer's disease. Nature Communications, vol. 11, no. 1, pp. 2484 (2020) doi: 10.1038/s41467-020-16298-6
16. Qiao, J., Wang, T., Shao, Z., Zhu, Y., Zhang, M., Huang, S., Zeng, P.: Genetic correlation and gene-based pleiotropy analysis for four major neurodegenerative diseases with summary statistics. Neurobiology of Aging, vol. 124, pp. 117–128 (2023) doi: 10.1016/j.neurobiolaging.2023.03.017
17. Romo-Galindo, D. A., Padilla-Moya, E.: Usefulness of brief cognitive tests for detecting dementia in the Mexican population. Archives of Neurosciences, vol. 23, no. 4, pp. 26–34 (2018)
18. Sarraf, S., DeSouza, D. D., Anderson, J., Tofighi, G.: Alzheimer's disease neuroimaging initiative. bioRxiv, (2017) doi: 10.1101/070441
19. Vong, W. K., Lake, B. M.: Cross-situational word learning with multimodal neural networks. Cognitive Science, vol. 46, no. 4, pp. e13122 (2021) doi: 10.1111/cogs.13122
20. Wadekar, S. N., Chaurasia, A., Chadha, A., Culurciello, E.: The evolution of multimodal model architectures. https://arxiv.org/abs/2405.17927 (2024)

ISSN 1870-4069

Aplicación de modelos de lenguaje de gran escala en capacitación personalizada para entrevistas técnicas

Itzel Cabrera, Diego Flores, Bella Martinez-Seis, Obdulia Pichardo-Lagunas

Instituto Politécnico Nacional,
Unidad Profesional Interdisciplinaria en Ingeniería y Tecnologías Avanzadas,
México

{bcmartinez,opichardola}@ipn.mx

Resumen. Prepararse para una entrevista técnica es un desafío que requiere no solo conocimientos teóricos, sino también práctica efectiva y personalizada. El presente trabajo propone integrar Modelos de Lenguaje de Gran Escala (LLMs, por sus siglas en inglés) en el proceso de capacitación para entrevistas técnicas. El prototipo propuesto proporciona experiencias de aprendizaje personalizadas basadas en los conocimientos seleccionados del candidato. El sistema web ofrece una interfaz flexible, interactiva y atractiva. Analizamos el diseño y la implementación del sistema en un entorno de entrevista técnica en el área de Desarrollo Web. Especialistas del área evaluaron el rendimiento del sistema en tres aspectos: preguntas generadas, respuestas generadas y evaluación de las respuestas.

Palabras clave: Modelos de lenguaje de gran escala, entrevistas técnicas, aplicaciones AI.

Leveraging Large Language Models for Personalized Technical Interview Preparation

Abstract. Preparing for a technical interview is a challenging task that requires not only theoretical knowledge but also effective and personalized practice. This study proposes the integration of Large Language Models (LLMs) into the training process for technical interviews. The proposed prototype provides personalized learning experiences based on the candidate's selected knowledge areas. The web-based system offers a flexible, interactive, and engaging interface. We analyze the system's design and implementation in the context of technical interviews in the field of Web Development. Experts in the field evaluated the system's performance across three key aspects: the quality of generated questions, the quality of generated answers, and the evaluation of those answers.

Keywords: Large language models, technical interviews, AI application.

1. Introducción

El proceso de reclutamiento técnico representa un desafío tanto para las empresas como para los candidatos, especialmente en la evaluación de conocimientos específicos y habilidades prácticas. Con la evolución de la Inteligencia Artificial, los Modelos de Lenguaje de Gran Escala (LLMs, por sus siglas en inglés) han emergido como una herramienta que podría mejorar la eficiencia y precisión de estas evaluaciones.

Las LLMs ofrecen una solución más versátil y efectiva para la capacitación en entrevistas técnicas en comparación con otras tecnologías de Inteligencia Artificial (IA). Por ejemplo, los chatbots convencionales pueden responder preguntas pero con respuestas predefinidas. Los sistemas de evaluación automatizada basados en redes neuronales, requieren conjuntos de datos especializados y entrenamiento costoso; mientras otros sistemas expertos requieren conjuntos de reglas predefinidas que limitan su capacidad de adaptación a distintos perfiles de candidatos. En contraste, los LLMs pueden ajustar la dificultad de las preguntas, generar explicaciones según el nivel del usuario y ofrecer escenarios específicos según el área de especialización.

Por otro lado, en el caso particular de México, durante el 2020 hubo una aceleración en la digitalización de las empresas y es por ello que, las vacantes en el sector tecnológico aumentaron un 57 %; colocando a los desarrolladores de front-end y back-end como dos de las profesiones más solicitadas [6]. Con relación a lo expuesto, se observa que el área de Desarrollo de Web es una de las áreas tecnológicas más importantes en el panorama socioeconómico.

En este artículo, exploramos cómo los LLMs pueden aplicarse en entrevistas técnicas personalizadas en el área de Desarrollo Web, ajustando preguntas y respuestas según el nivel de conocimiento del practicante. La presente propuesta abarca desde la generación de preguntas hasta la evaluación de ellas.

Para potenciar la LLM se hace uso del proceso conocido como prompt engineering para refinar y mejorar la entrada (prompt) a la Inteligencia Artificial Generativa (GenAI). En este desarrollo, se hace uso de los LLMs en tres procesos: (1) generación de preguntas correspondientes al área y nivel del candidato en torno al Desarrollo Web, (2) generación de respuestas esperadas para las preguntas y (3) evaluación de las respuestas del candidato a través del asistente de voz.

El desarrollo final corresponde al asistente inteligente de voz para entrevistas técnicas que abarca la extracción de datos en el currículum vitae del usuario, la generación de preguntas personalizadas a su perfil técnico, la transformación de las respuestas orales del usuario a texto y la evaluación de las preguntas.

El resto del artículo se compone de: Sección 2 explora el estado del arte de los LLMs enfocado principalmente a reclutamiento de personal, Sección 3 describe la propuesta del sistema, en esta se describirán los módulos que la componen así como la arquitectura usada en su implementación, la Sección 4 muestra la evaluación y resultados del LLM y la Sección 5 presenta las conclusiones.

2. LLMs para el reclutamiento de personal

En esta sección, presentaremos brevemente los antecedentes de los LLMs junto con una descripción general de los estudios relacionados centrados en el uso de LLMs enfocados al proceso de reclutamiento de personal como la generación de currículums, emparejamiento de vacantes y entrevistas entre reclutador y candidatos.

Los Modelos de Lenguaje Extendidos o Modelos de Lenguaje de Gran Escala (LLM) son una categoría de modelos funcionales entrenados sobre enormes cantidades de datos que los hacen capaces de generar lenguaje natural, entre otros tipos de contenidos, para realizar una amplia gama de tareas [4]. Los LLMs estiman la probabilidad de la generación de un símbolo o secuencia de símbolos dentro de una secuencia más extensa de símbolos, un LLM toma mayor potencial cuando se introduce el concepto de transformador (Transformer en inglés) que comprende de un codificador y decodificador.

Los LLMs se encuentran inmersos en varios procesos enfocados a la educación, capacitación y aprendizaje [11]. El uso de ellos para el aprendizaje de otros idiomas es muy concurrido, por ejemplo [15] integra GPT API para personalizar cursos de inglés. También se está usando para integrarla en metaversos [1] sobre todo involucrados en educación. Es evidente que cualquier proceso que involucre el intercambio de preguntas y respuestas podría ser apoyado por los LLMs sin embargo, aún hay aspectos por mejorar [5] como la identificación de respuestas o comunicación persuasiva en entrevistas periodísticas [7].

En el área de reclutamiento laboral existen trabajos que hacen uso de los LLMs como parte del proceso principal. Aguinalde et al. [2] las usaron para obtener las responsabilidades y tareas requeridas en ofertas para pasantes de ingeniería aeroespacial y posteriormente, consolidando la entrada para entrenar algoritmos de clasificación. DeBaer et al. [3] utilizan LLMs para generación de ejemplos de conversaciones para entrevistas de trabajo.

MockLLM [13] realiza generación de entrevistas simuladas y evaluación bilateral, con el objeto de encontrar las mejores relaciones persona-trabajo. Enfocado hacia la generación de preguntas para reclutadores de TI, Pasaribu et al. [10] dividen las entrevistas en conductual y técnico. Para el primero, usó fine-tuning en el modelo Longformer y para el segundo few-shot learning en el LLM GPT-4. Sin embargo, dichos trabajos se enfocan en las reclutadoras y no en los candidatos, como se propone en este trabajo.

Enfocado a los candidatos existen propuestas como la creación de currículums [14]. La Universidad Politécnica de Bucharest [12] desarrolló un simulador para entrevistas de trabajo combinando realidad virtual, reconocimiento de emociones y chatbots, este último a través del framework Pandarobots. ITEM [8] es una propuesta de entrenamiento para entrevistas con Realidad Virtual usando GenAI, donde el LLM de GPT generar preguntas para evaluar las habilidades de comunicación y liderazgo del estudiante. La inmersión en el metaverso genera un enfoque transformador, sin embargo, deja a un lado las entrevistas técnicas que son retomadas en esta propuesta.

3. Entrenamiento para entrevistas técnicas

Las entrevistas laborales técnicas se pueden clasificar en tres: estructuradas, no estructuradas y semiestructuradas. El trabajo fue desarrollado para entrevistas estructuradas cuyas preguntas se enfocan en las aptitudes del candidato evitando opiniones sesgadas, además de ser de corta duración ya que las preguntas se concretan antes de comenzar. La estructura de la aplicación está compuesta por cuatro módulos:

1. **Tzin**[1]. Captura de campos clave de CV.
2. **Waach**[2]. Generación de preguntas técnico-conceptuales personalizados.
3. **Marli**[3]. Procesamiento de voz a texto.
4. **Chari**[4]. Evaluación de las respuestas.

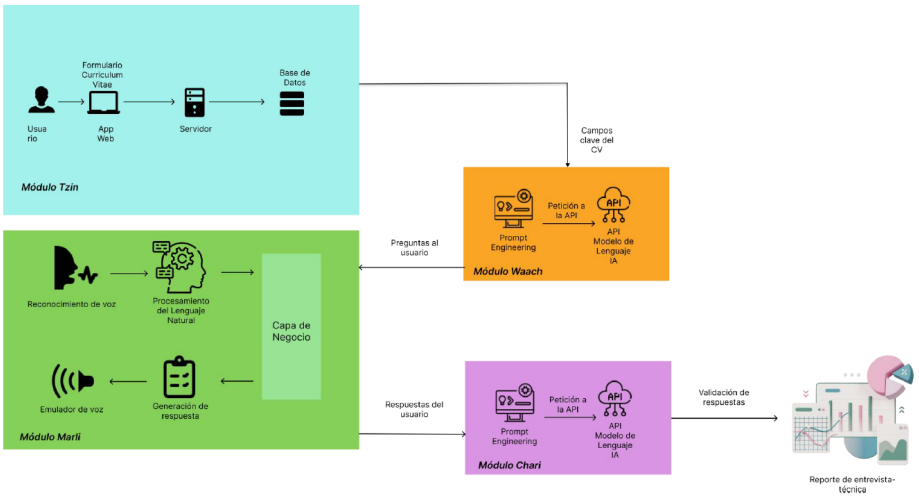

Fig. 1. Diagrama a bloques del prototipo de entrenamiento para entrevistas técnicas.

A continuación, se detalla cada módulo y se describe cómo se comunican. El uso de LLMs se presenta en los módulos Waach (Generación de preguntas) y Chari (Evaluación de respuestas). Posteriormente se describe la arquitectura usada para la implementación.

[1] Del Náhuatl clásico, significa *amado, respetado, querido.*
[2] Del maya yucateco, significa *soltar.*
[3] Unión de dos nombres.
[4] Diminutivo de Charango, instrumento musical.

3.1. Módulo Tzin

El módulo Tzin permite obtener los campos claves de un CV por competencias y nivel, el cual permite el registro de conocimiento y aptitudes a partir de categorías (con respecto a lo profesional) y a partir de cada una se pueden ingresar campos con su respectivo nivel de conocimiento.

3.2. Módulo Waach

El Módulo Waach obtiene las preguntas técnico-conceptuales a partir de los campos clave proporcionados por el módulo Tzin. Se evaluaron cualitativamente Gemini y GPT4 como se describe en la Sección 4.1. Para ello, se enlaza a la LLM cuyo diseño y evaluación de *promt* se detalla en la Sección 4.2. La LLM utilizada es GPT4-o mini.

3.3. Módulo Marli

El módulo Marli es el encargado de procesamiento de voz. Marli requiere como entrada las preguntas generadas por Waach. El usuario responde verbalmente a las preguntas y se realiza la transformación a texto con la API SpeechRecognition de NPM debido a que es multilingüistico y no tiene costo ni límite.

3.4. Módulo Chari

El módulo Chari tiene como objetivo mostrar la evaluación de las respuestas (correcta o incorrecta); con el fin de determinar el desempeño del usuario en dicha entrevista. Para ello se hace uso nuevamente de la LLM GPT4-o mini. Finalmente, esta evaluación se guarda y se muestra en la aplicación web; de esta forma, se busca que el usuario visualice su progreso dadas las entrevistas que haya realizado.

3.5. Aquitectura e implementación

Para el proyecto se cuenta con una base de datos relacional donde el usuario tiene el histórico de sus currículums vitae, cada uno con una o varias categorías con su respectivo nivel. Además, se tiene un registro de las entrevistas donde se almacenan las preguntas y evaluación. El manejador de base de datos usado es SQL Server 2019, debido a su compatibilidad con Spring JPA y el entorno distribuido.

Para la arquitectura, se consideró la de microservicios (ver Figura 2) debido a que la propuesta involucra diferentes módulos que deben integrarse a una aplicación web. De esta forma, la aplicación cuenta con los siguientes principios: resiliencia, alta escalabilidad y disponibilidad.

La plataforma distribuida para la transmisión de datos utilizada es Kafka debido a su capacidad para manejar flujos de datos en tiempo real y de alta

velocidad entre módulos como la generación de preguntas, el procesamiento de respuestas y la evaluación de desempeño. Kafka permite almacenar mensajes durante un tiempo configurable, lo que facilita reprocesar datos históricos y escalar de manera eficiente a medida que aumenta el volumen de usuarios y entrevistas. Además, su modelo de consumo independiente (donde los consumidores rastrean su propio progreso) asegura flexibilidad para manejar tareas asincrónicas complejas, como la interacción con modelos de lenguaje y el procesamiento de entrevistas, características críticas para la aplicación.

La comunicación se lleva a cabo de la siguiente forma según la Figura 2:

1. El usuario realiza la conexión hacia el servidor de aplicaciones donde se encuentre el aplicativo bajo el protocolo HTTPS.
2. El aplicativo front-end realiza la solicitud a un recurso que contenga algún microservicio dentro del back-end, este será interceptado por el Gateway.
3. El Gateway es el filtro, para autenticar todas las peticiones que quieran acceder a las diferentes instancias de los microservicios y denegar aquellas que no estén autenticadas.
4. El Gateway conoce los microservicios que se hayan descubierto/registrado.
5. 6. 7. El Gateway realiza la petición y el balanceo de carga hacia la instancia, junto con el endpoint del recurso al cual quiere acceder, así mismo con el body request o query params que llegara a ocupar la API.
8. Publica en el bus de datos los campos del CV.
9. Se avalúa la entrevista con ayuda del LLM.

4. Evaluación y resultados

En este capítulo, se presenta el análisis cualitativo entre LLMs y los resultados obtenidos a partir de la evaluación del modelo de lenguaje de gran escala seleccionado. La evaluación se divide en varias secciones dentro del contexto de entrevistas: en primer lugar, se explora el diseño y evaluación de prompts, elementos fundamentales que guían la interacción con el modelo; en segundo lugar, se evalúa por expertos en el área, la capacidad del modelo para generar preguntas y respuestas relevantes en el contexto de entrevistas técnicas; y finalmente, se proporciona una visión general sobre el rendimiento del modelo.

4.1. Análisis cualitativo de LLMs

Los LLM son una categoría de modelos funcionales que permiten comprender y generar lenguaje natural. Se analizaron las LLM más competitivas en el año 2024: GPT, Gemini y BERT [9]. Para los dos primeros se hizo un análisis cualitativo en tres rubros: (1) generación de preguntas, (2) alcance de contenidos y (3) calificar respuestas. Siendo relevantes los dos primeros para el módulo Waach y el último para el módulo Marli. El análisis cualitativo comprende la prueba de prompts relacionados con el área y la evaluación de la respuesta por parte de expertos con respecto a la completitud, correctitud y objetividad. En

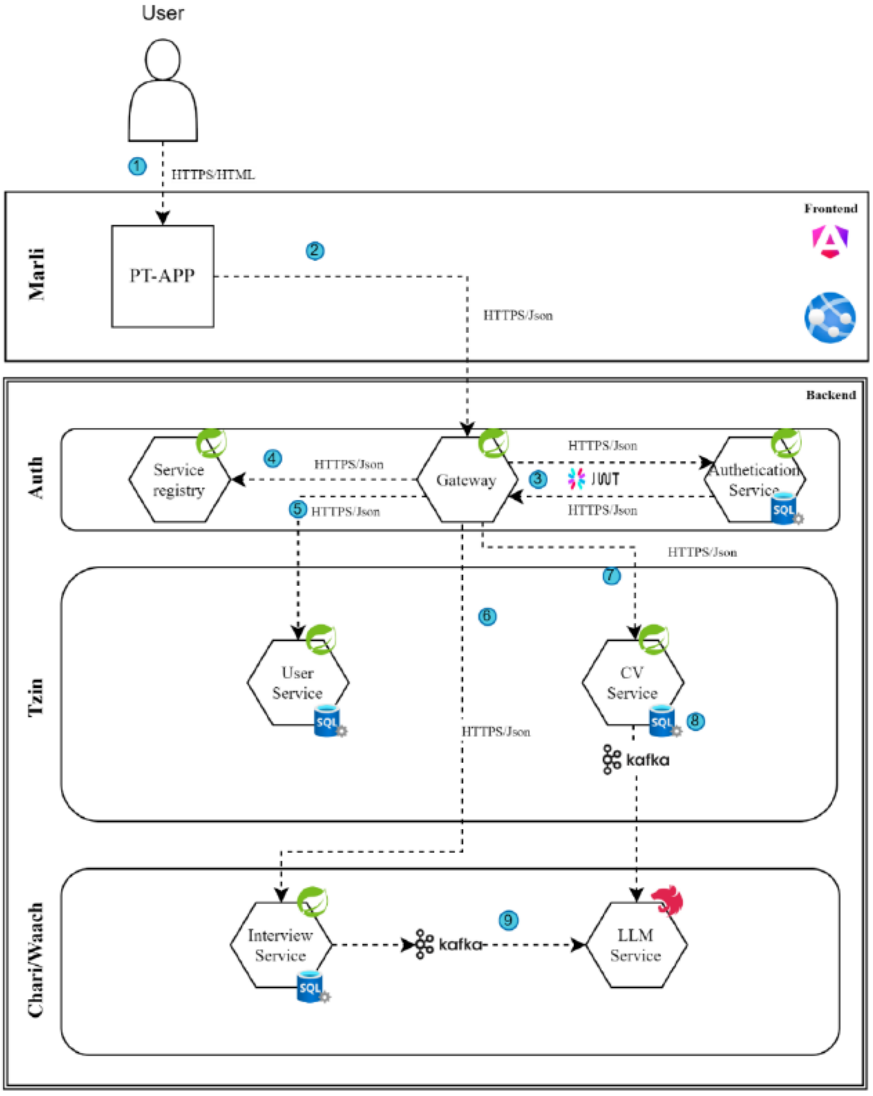

Fig. 2. Diagrama de arquitectura del sistema de entrenamiento para entrevistas técnicas.

la evaluación de **generación de preguntas**, GPT tuvo un mejor rendimiento que Gemini, ya que siguió directrices específicas y generó preguntas conceptuales en el formato indicado, mientras que Gemini se enfocó en preguntas prácticas fuera del alcance requerido. En la **evaluación de alcance de contenidos**, ambos modelos tuvieron un rendimiento similar, pero Gemini ofreció respuestas más contextuales y con enlaces de respaldo, lo que puede reducir alucinaciones

y validar resultados. En la **calificacación de respuestas**, ambos modelos son comparables, pero Gemini proporciona contexto adicional en sus evaluaciones, mientras que GPT es más estricto con errores tipográficos, ortográficos y gramaticales. En conclusión, tanto GPT como Gemini son opciones competitivas para generar preguntas y evaluar respuestas. Sin embargo, debido a que se requieren respuestas concretas por parte del LLM, se optó por GPT.

4.2. Diseño y evaluación de *promt*s

De acuerdo con la metodología GPEI: metodología para un Prompt Engineering eficiente, en el diseño del prompt de la aplicación se siguieron los siguientes pasos:

1. Definición de metas:
   - Generar preguntas-respuestas de acuerdo a los campos clave del CV.
   - Evaluar respuestas del usuario, teniendo como referencia las respuestas correctas que el LLM haya generado
2. Elementos a incluidos en el prompt:
   - Incluir detalles.
   - Pedir al modelo que tome un rol.
   - Proveer ejemplos.
   - Especificar la extensión de la respuesta.
   - Pedir al modelo que responda usando algún texto de referencia.
3. Etapa de iteración consiste en probar una versión del prompt y si la respuesta no cumple con los criterios de evaluación, entonces el prompt se modifica de forma inmediata.

La evolución del *promt* para generar las **preguntas con sus respuestas** fue la siguiente: El prompt inicial contenía los elementos enumerados con anterioridad considerando los campos del CV del usuario. En la segunda iteración se solicitó expresamente el número de preguntas y se dieron ejemplos. En la tercera, se aumentó el contexto y se agregó el enfoque de las preguntas y se solicitó la respuesta correcta. En la cuarta se limitó la extensión de la respuesta. La quinta se clarificó que se requiere en idioma español.

El *promt* inicial para **calificar respuestas** indicaba qué rol tenía, qué realizarías, descripción de qué se le dará, en qué formato, cuál es el objetivo, cuál es el formato de respuesta y se intercalan definiciones para evitar ambiegüedades. En la segunda iteración se aclaró que la evaluación la hará desde su perspectiva considerando correcta si la respuesta tiene noción del tema. En la tercera, se agregó la respuesta correcta para tenerlo como referencia en la evaluación. La cuarta, se aclara que el enfoque sea el contenido. Finalmente, en la quinta, se especifica el idioma.

4.3. Evaluación de Generación de Preguntas-Respuestas por la LLM

En la evaluación participaron 3 expertos que se describen a continuación:

1. Primer experto.12 años de trayectoria, cargos como desarrollador Java, líder técnico y arquitecto de soluciones, desempeñando tareas de planeación y gestión de personal, así mismo validando el conocimiento y capacidades técnicas del equipo que tuviera a su cargo.
2. Segundo experto. 6 años de trayectoria, desempeñándose como desarrollador Java y líder de proyectos, participe de procesos de reclutamiento para labores dentro del Desarrollo Web.
3. Tercer experto. 26 años de experiencia, gerente de gestión de proyectos y diferentes gerencias, pasante de maestría en ciberseguridad, participante en subáreas del desarrollo web. Ha diseñado y validado de habilidades técnicas en procesos de reclutamiento.

Se realiza el análisis de los resultados obtenidos con el modelo GPT-4-o mini. Se evaluaron los resultados proporcionados por la LLM en 3 aspectos:

1. Evaluación de preguntas generadas. Se evaluúa la correspondencia con la categoría y nivel. Se espera que la pregunta generada corresponda al tópico y nivel ingresado, los niveles considerados fueron: Trainee, Junior y Senior.
2. Evaluación de respuestas generadas. Se validala respuesta generada por la LLM es correcta con respecto a la pregunta generada.
3. Evaluación de calificación de respuestas. Se valida que la calificación asignada por la LLM de las respuestas del usuario sea correcta.

Se proporcionó a los expertos las respuestas generadas por el LLM. Ellos etiquetaron como correcto o incorrecto según el aspecto que estén evaluando. La participación de tres expertos permite discernir si lo arrojado por el LLM es correcto, ya que podrían los tres concidir en que se correcto o no, pero podría darse el caso de tener diferencia de opinión, en ese caso si dos de tres coinciden es la evaluación que se compara con el LLM.

Evaluación de preguntas generadas. Para el nivel Trainee se tuvo una coincidencia del 88.8 % de la evaluación manual de los expertos con respecto al nivel. Para el nivel de Junior la coincidencia fue también de 88.8 % y para el nivel de Senior fue del 100 %. Se puede concluir que las preguntas que GPT 4-o mini proporciona son las adecuadas al nivel correspondiente del usuario. Esto sugiere que el Prompt para la generación de preguntas-respuestas permite al modelo generar cuestionamientos coherentes y relevantes para el contexto técnico.

Evaluación de respuestas generadas. Los expertos evaluaron las respuestas generadas en los tres niveles, de tal forma que el 92.59 % de las respuestas era considerada correcta por los expertos. Se puede concluir que las respuestas que GPT4-o mini proporciona son las adecuadas tanto en contenido como en extensión. Este resultado confirma que el modelo no solo interpreta correctamente las preguntas, sino que también genera respuestas claras, completas y bien estructuradas. La capacidad del modelo para mantener consistencia en la calidad de las respuestas respalda la idea de que el diseño del Prompt para la generación de preguntas-respuestas es robusto y eficiente.

Evaluación de calificaciones generadas. El 66 % de las calificaciones son correctas según el promedio de los expertos. Aunque este porcentaje es significativo, también evidencia un área de mejora en el diseño del Prompt final para la evaluación de respuestas y en la configuración del modelo para evaluar respuestas de manera más precisa. Este resultado sugiere la necesidad de realizar ajustes al Prompt final para la evaluación de respuestas para mejorar la precisión de las calificaciones, así como considerar estrategias complementarias, como la incorporación de ejemplos adicionales o criterios más específicos de evaluación.

4.4. Análisis de evaluación.

Los tres expertos consultados coinciden en que el modelo realiza las tareas asignadas de manera coherente y eficiente en la mayoría de los casos. Sin embargo, mientras que los resultados de los dos primeros rubros reflejan un alto nivel de desempeño, el tercer rubro indica que existe margen para optimizar el modelo en la tarea de calificación. Esto resalta la importancia de iterar en el diseño del prompt para evaluar las respuestas y de explorar ajustes adicionales para maximizar la efectividad del modelo en todos los aspectos evaluados. Este análisis subraya el potencial del modelo para aplicaciones prácticas, mientras que también enfatiza la necesidad de una mejora continua para alcanzar una mayor precisión y consistencia en sus resultados.

5. Conclusión

La propuesta implemntó una arquitectura de microservicios que integra Kafka para la comunicación en tiempo real y el uso de LLMs para dar solución al entrenamiento personlizado en entrevistas técnicas en el área de Desarrollo Web.

Las tareas principales del proyecto son la generación de preguntas y evaluación de respuestas. La generación de preguntas se realiza a partir de los campos del CV proporcionados por los usuarios, lo que asegura que las entrevistas sean personalizadas y se adapten a las competencias de cada candidato. Este enfoque hace que las entrevistas sean más relevantes y efectivas en la evaluación de habilidades técnicas. Al mismo tiempo, las respuestas correctas generadas por el modelo sirven como referencia para la evaluación posterior de las respuestas del usuario, asegurando que la evaluación sea coherente y alineada con el conocimiento técnico esperado.

Durante el análisis, un desafío importante fue la valoración y comparación de los diferentes LLM existentes en el mercado; esto debido a la constante actualización de modelos. Tanto GPT como Gemini fueron opciones competitivas para generar preguntas y evaluar respuestas. Se presenta una evaluación manual dada por expertos en el área según sus conocimientos, lo cual genera una validación que depende de los etiquetadores. Sin embargo, no se requirió un banco de datos para el desarrollo del proyecto.

Referencias

1. Abdullakutty, F., Qayyum, A., Qadir, J.: Trustworthy AI for educational metaverses. Authorea Preprints, (2024)
2. Aguinalde, P., Shin, J., Carroll, B. F., Crippen, K. J.: Leveraging large language models to automatically investigate core tasks within undergraduate engineering work-integrated learning experiences. In: 2024 IEEE Frontiers in Education Conference (FIE). pp. 1–8. IEEE (2024)
3. De Baer, J., Doğruöz, A. S., Demeester, T., Develder, C.: Single-vs. dual-prompt dialogue generation with LLMs for job interviews in human resources. arXiv preprint arXiv:2502.18650, (2025)
4. IBM: Modelos de Lenguaje Grande (LLM) (2025), `https://www.ibm.com/mx-es/topics/large-language-models`, consultado el 25 de marzo de 2025
5. Kamerlin, S. C. L., Ratcliff, W. C.: Using AI to prepare for academic interviews–don't trade authenticity for polish (2025)
6. LinkedIn Business: Jobs on the Rise - Tendencias en Contratación (2025), `https://business.linkedin.com/es-mx/talent-solutions/resources/talent-acquisition/jobs-on-the-rise-cont-fact`, consultado el 25 de marzo de 2025
7. Lu, M., Cho, H. J., Shi, W., May, J., Spangher, A.: Newsinterview: a dataset and a playground to evaluate LLMs' ground gap via informational interviews. arXiv preprint arXiv:2411.13779, (2024)
8. Nofal, A. B., Ali, H., Hadi, M., Ahmad, A., Qayyum, A., Johri, A., Al-Fuqaha, A., Qadir, J.: Ai-enhanced interview simulation in the metaverse: Transforming professional skills training through VR and generative conversational AI. Computers and Education: Artificial Intelligence, vol. 8, pp. 100347 (2025)
9. Ozdemir, S.: Quick start guide to large language models: strategies and best practices for using ChatGPT and other LLMs. Addison-Wesley Professional (2023)
10. Pasaribu, D. K. H., Dewandaru, A., Saptawati, G. A. P.: Development of LLM-based system for IT talent interview. In: 2024 IEEE International Conference on Data and Software Engineering (ICoDSE). pp. 108–113. IEEE (2024)
11. Sallam, M.: ChatGPT utility in healthcare education, research, and practice: systematic review on the promising perspectives and valid concerns. In: Healthcare. vol. 11, pp. 887. MDPI (2023)
12. Stanica, I., Dascalu, M.-I., Bodea, C. N., Moldoveanu, A. D. B.: VR job interview simulator: where virtual reality meets artificial intelligence for education. In: 2018 Zooming innovation in consumer technologies conference (ZINC). pp. 9–12. IEEE (2018)
13. Sun, H., Lin, H., Yan, H., Zhu, C., Song, Y., Gao, X., Shang, S., Yan, R.: Facilitating multi-role and multi-behavior collaboration of large language models for online job seeking and recruiting. arXiv preprint arXiv:2405.18113, (2024)
14. Sunico, R. J., Pachchigar, S., Kumar, V., Shah, I., Wang, J., Song, I.: Resume building application based on LLM (large language model). In: 2023 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS). pp. 486–492. IEEE (2023)
15. Yan, H.: Let's chat: Integrating large language models into blended learning of English for specific purposes. In: 2023 International Symposium on Educational Technology (ISET). pp. 187–192. IEEE (2023)

ISSN 1870-4069

Implementación de algoritmos de ecualización para imágenes industriales

Jonathan Villanueva Tavira[1], Damian Macedo García[1], Antonio Martínez Santos[2], Manuela Calixto Rodríguez[2], Marilú Chávez Castillo[2], Claudia Ayala Vázquez[3]

[1] Tecnológico Nacional de México, Centro Nacional de Investigación y Desarrollo Tecnológico, México

[2] Universidad Tecnológica Emiliano Zapata del Estado de Morelos, México

[3] Centro Morelense de Innovación y Transferencia Tecnológica (CemiTT), México

jonathan.vt@cenidet.tecnm.mx

Resumen. En este artículo se presentan los resultados obtenidos de las imágenes resultantes obtenidas después de la aplicación de diferentes algoritmos de ecualización para imágenes de escala de grises e imágenes a color. La finalidad de estos algoritmos es obtener mejores imágenes para su procesamiento en etapas posteriores en un sistema de visión por computadora. Para la obtención de las imágenes, se implementa cada uno de los algoritmos en el lenguaje de programación MATLAB empleando interfaces GUIDE con la finalidad de proporcionar una vista al usuario para elegir la ecualización que más le convenga al usuario.

Palabras clave: Procesamiento digital de imágenes, visión por computadora.

Implementation of Equalization Algorithms for Industrial Images

Abstract. This article presents the results obtained from the resulting images obtained after applying different equalization algorithms to grayscale and color images. The purpose of these algorithms is to obtain better images for subsequent processing in a computer vision system. To obtain the images, each algorithm is implemented in the MATLAB programming language using GUIDE interfaces to provide a user-friendly interface for choosing the most suitable equalization.

Keywords: Digital image processing, computer vision.

1. Introducción

Las imágenes digitales han sido por años, uno de los elementos más importantes para ser utilizadas en conjunto con alguna técnica de visión por computadora o actualmente

con el aprendizaje profundo. El contar con imágenes de excelentes características, contribuye de manera importante en los procesos donde se utilizan, facilitando así, la etapa de reconocimiento o clasificación en un sistema de visión por computadora. En este artículo, se abordan los diferentes métodos de ecualización del histograma, su implementación, y al final se muestran cada uno de los resultados obtenidos.

El histograma, es una herramienta que permite visualizar la distribución de cada uno de los niveles de gris o color en una imagen y en consecuencia su contenido. Algunas de las técnicas de realzado, se orientan hacia la mejora de la calidad de la imagen. Para ello, estas técnicas procuran eliminar efectos no deseados tales como: sombras y reflejos, a la vez que aumentan el contraste. El histograma de una imagen es la función discreta que representa el número de pixeles en la imagen en función de los niveles de intensidad, g se define como [1, 2]:

$$P(g) = \frac{N(g)}{M}, \tag{1}$$

donde M es el número de pixeles en la imagen y N(g) es el número de pixeles en el nivel de intensidad g. Como con cualquier distribución de probabilidad todos los valores de P(g) son menores o iguales que 1 y la suma de todos los valores de P(g) es 1 [1, 2].

2. Marco teórico

Las siguientes propiedades estadísticas informan sobre la distribución de los niveles de gris en la imagen basándonos en el histograma [1].

Media. En una imagen de escala de grises, es el valor promedio de los niveles de gris de intensidad de la imagen. Finalmente, este dato nos proporciona información sobre el brillo general, está definida por [1]:

$$\bar{g} = \sum_{g=0}^{L-1} gP(g) = \sum_{i} \sum_{j} \frac{I(i,j)}{M}. \tag{2}$$

L es el número total de niveles de gris, así para una imagen con valores de gris entre 0 y 255; L sería 256. Una imagen brillante tendrá una media alta y viceversa [1].

Varianza. Mide que tanto varían los valores de los pixeles, respecto a la media de la imagen. Esta característica mide la dispersión de los alrededores de la media. Una varianza alta corresponde a una imagen con contraste alto y al contrario. Está definida por la siguiente expresión [1]:

$$\sigma^2 = \sum_{g=0}^{L-1} (g - \bar{g})^2 P(g). \tag{3}$$

Asimetría. También llamada sesgo de una imagen, es la media en la distribución de los niveles de gris. Un valor absoluto alto indica una gran asimetría [1]:

$$a = \sum_{g=0}^{L-1} (g - \bar{g})^3 P(g). \tag{4}$$

Energía. Es una medida que refleja qué tan uniforme o estructurada es la imagen. Esta nos informa sobre la distribución de los niveles de gris. La energía tiene un valor máximo 1 para una imagen con un único nivel de gris y disminuye a medida que aumenta el número de niveles de grises [1]:

$$E = \sum_{g=0}^{L-1} (P(g))^2. \tag{5}$$

Entropía. Informa sobre la distribución de los niveles de gris, cuanto mayor es el número de niveles de gris en la imagen mayor es la entropía. Esta medida tiende a variar inversamente con la energía [1]:

$$e = \sum_{g=0}^{L-1} P(g)\, log_2[P(g)]. \tag{6}$$

Una de las técnicas más utilizadas para la mejora del contraste de la imagen original es la de igualación de histogramas [2]. Se trata de una técnica que realza la imagen original mediante una determinada transformación o modificación del histograma denominada igualación o ecualización. Este procedimiento busca encontrar una función F(g) que realce el contraste general en la imagen original expandiendo la distribución de los niveles de gris. La expansión debe de ser lo más suave posible en el sentido que idealmente debería haber el mismo número de pixeles por niveles de gris. En consecuencia, el objetivo es distribuir los niveles de gris de una manera uniforme a lo largo de todo el rango de valores de niveles de gris. Se puede deducir la función de F(g) mediante la simple inspección del histograma original, pero es deseable una función analítica [1]. Por otra parte, a partir del histograma podemos definir la función de densidad de probabilidad de la siguiente forma [3]:

En el supuesto que la imagen de 256 niveles de gris en el rango 0 a 255, se cumple:

$$\sum_{g=0}^{g=255} N(g) = NxM. \tag{7}$$

La probabilidad por cada nivel de gris g viene dada por:

$$p(g) = \frac{N(g)}{NxM}; g = 0,1, \dots, 255. \tag{8}$$

La función de densidad de probabilidad resulta ser:

Tabla 1. Formas de ecualización del Histograma

Distribuciones	**Expresiones**
Uniforme	$F(g)=[gmax-gmin]P(g)+gmin$
Exponencial	$F(g) = gmin - \frac{1}{\propto} ln[1 - P_g(g)]$
Rayleigh	$F(g) = gmin + \left[2 \propto^2 ln\left\{\frac{1}{1 - P_g(g)}\right\}\right]^{1/2}$
Hipercúbica	$F(g) = \left(\left[\sqrt[3]{gmax} - \sqrt[3]{gmin}\right]P_g(g) + \sqrt[3]{gmin}\right)^3$
Logaritmo Hiperbólica	$(g) = gmin - \frac{1}{\propto} ln[1 - P_g(g)]$

$$P_x(x) \cong \sum_{g=0}^{x} p(g). \tag{9}$$

Se trata de realizar una transformación entre funciones de densidad de probabilidad *Px(x)* y *Py(x)*, si se impone la condición de que la función de transformación sea monótona creciente, para cada valor de x e y se cumple. A modo de ejemplo, se analiza los tipos de ecualización derivados de [3, 4, 5, 6].

a) Ecualización uniforme. Se realiza mediante la modificación de los pixeles de la imagen resultante de forma que estos se repartan de forma equitativa en todo el rango de valores establecidos, es decir 0 a 255:

$$F(g)=[gmax-gmin]P(g)+gmin. \tag{10}$$

b) Ecualización exponencial. Esta se trata de distribuir los pixeles según una función de tipo exponencial, por lo que el valor de α se ajusta con el fin de acumular los niveles de gris con más o menos densidad cerca del origen:

$$F(g) = gmin - \frac{1}{\propto} ln[1 - P_g(g)]. \tag{11}$$

c) Ecualización Rayleight. Para este tipo de ecualización cuanto mayor sea el valor de α, mayor es el valor de los pixeles más frecuentes en la imagen resultante:

$$F(g) = gmin + \left[2 \propto^2 ln\left\{\frac{1}{1 - P_g(g)}\right\}\right]^{1/2}. \tag{12}$$

d) Ecualización Hipercúbica. Utiliza un enfoque basado en diferentes factores y características de la imagen en más dimensiones, tratando diferentes regiones de la imagen como entidades independientes en un espacio de muchas dimensiones:

$$F(g) = \left(\left[\sqrt[3]{gmax} - \sqrt[3]{gmin}\right]P_g(g) + \sqrt[3]{gmin}\right)^3. \tag{13}$$

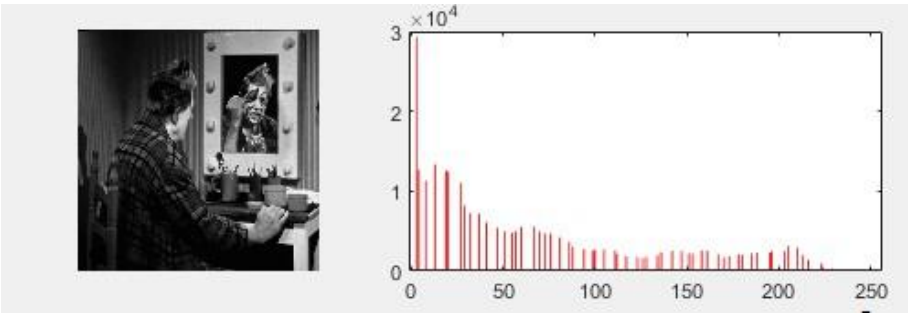

Fig. 1. Interfaz gráfica realizada en MATLAB para imágenes a escala de grises.

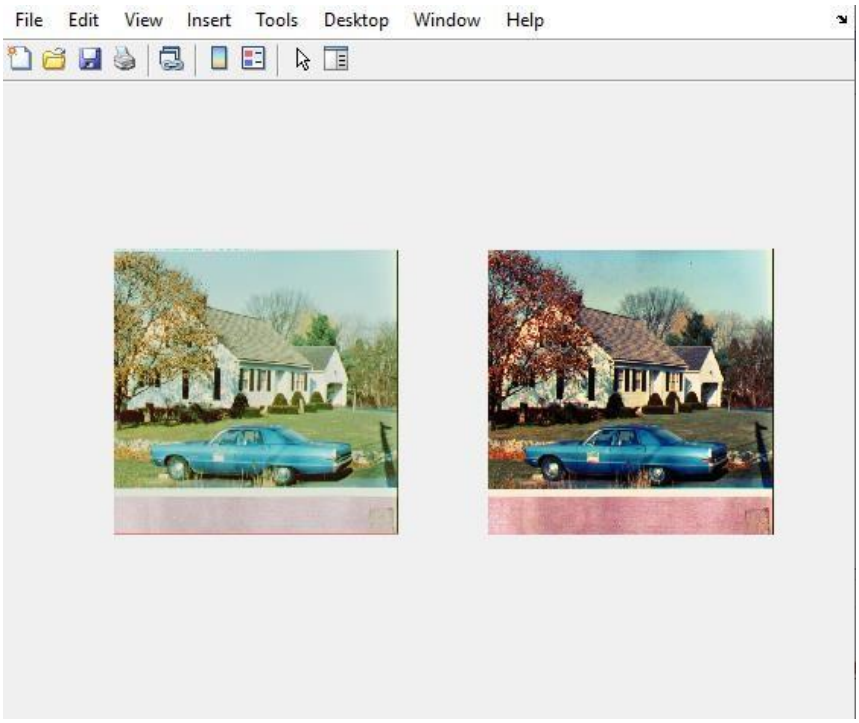

Fig. 2. Interfaz realizada en MATLAB para la ecualización de imágenes a color.

e) Ecualización Logaritmo Hiperbólica. Es una técnica utilizada para aumentar el contraste de imágenes, especialmente en aquellas que tienen detalles oscuros. El término "hiperbólica" se refiere a una forma específica de la función matemática utilizada en el proceso:

$$F(g) = gmin - \frac{1}{\propto} ln\left[1 - P_g(g)\right]. \tag{13}$$

Fig. 3. Imágenes resultantes de aplicar cada ecualización. a) Uniforme, b) Exponencial, c) Rayleigh, d) Hipercúbica, e) Logaritmo Hiperbólica.

Finalmente, en la Tabla 1 se enlistan todos los tipos de ecualización anteriormente mencionados.

3. Desarrollo

Para la implementación y ejecución de los experimentos, se utiliza el software MATLAB, aprovechando su capacidad para el procesamiento de imágenes y la

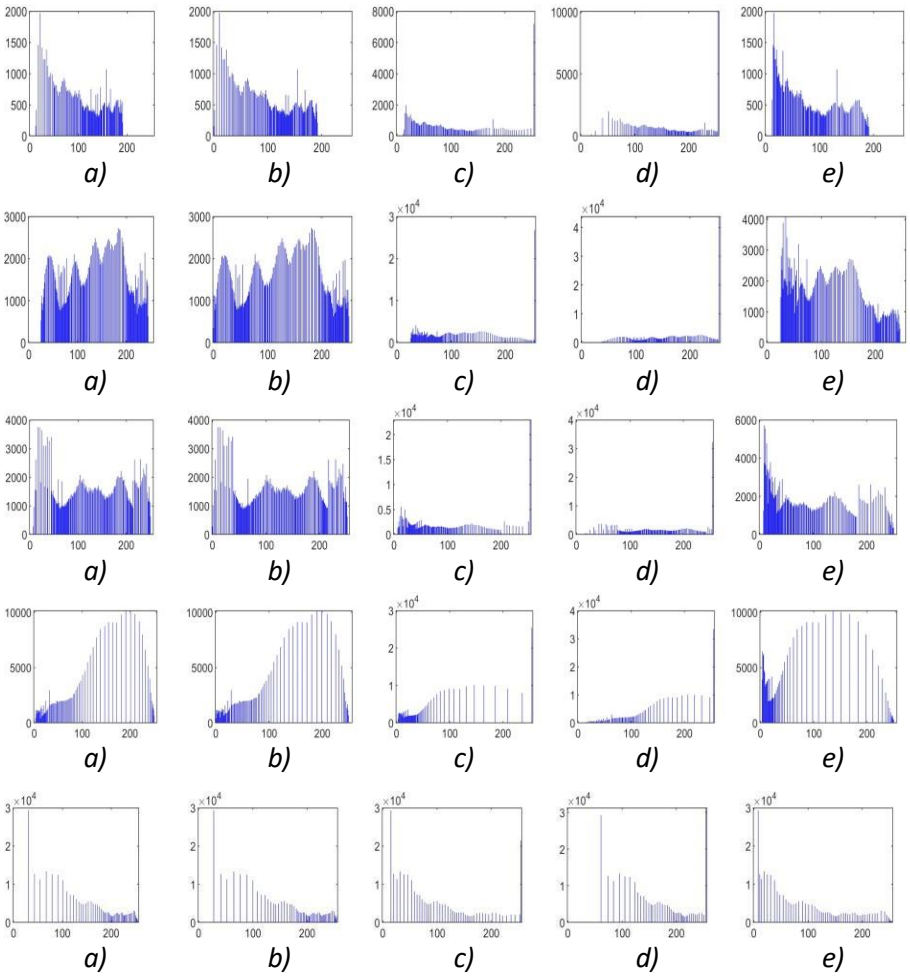

Fig. 4. Histogramas de imágenes resultantes al aplicar cada ecualización. a) Uniforme, b) Exponencial, c) Rayleigh, d) Hipercúbica, e) Logaritmo Hiperbólica.

implementación de los algoritmos de ecualización. Además, se diseña una interfaz gráfica de usuario (GUI) mediante GUIDE (Graphical User Interface Development Environment), lo que permite una interacción más intuitiva con los parámetros de los algoritmos y además facilita la visualización de los resultados obtenidos (ver Figura 1 y 2).

4. Experimentación y resultados

En esta sección se presentan los resultados obtenidos tras la aplicación de los algoritmos de ecualización. Estos permiten visualizar el impacto de las técnicas

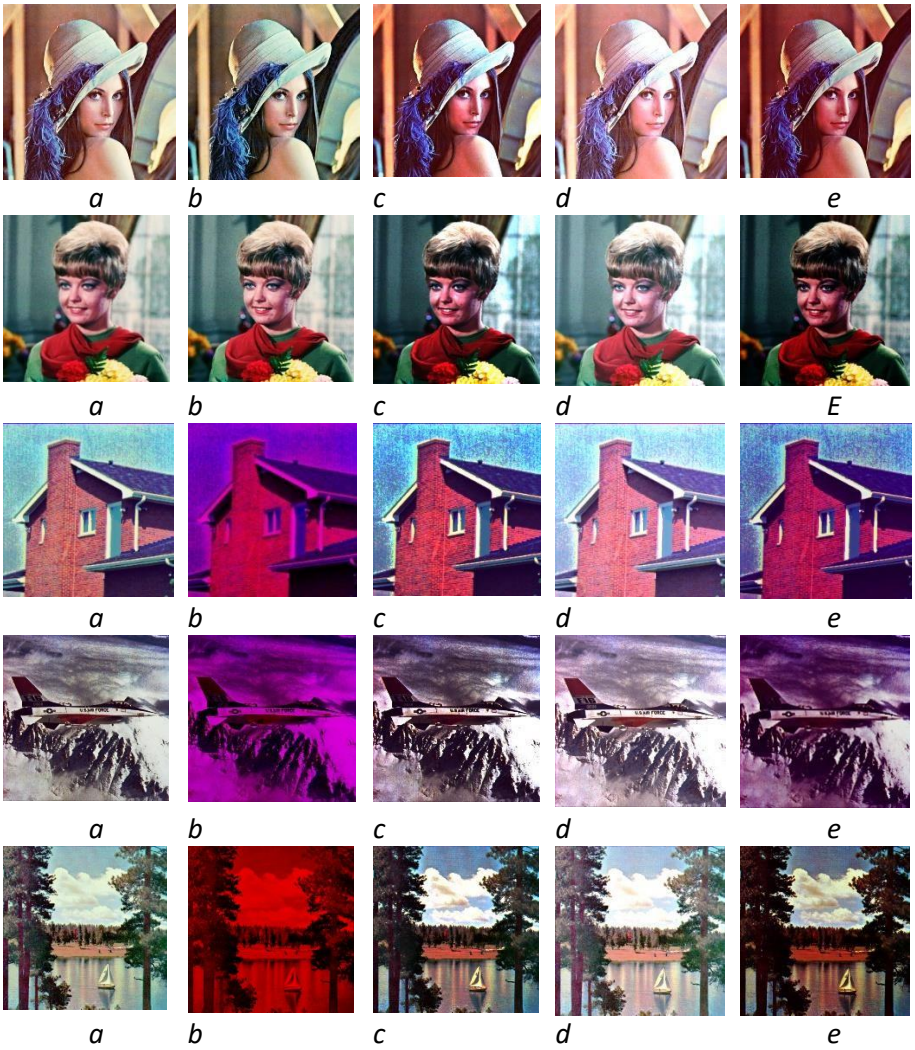

Fig. 5. Imágenes resultantes al aplicar cada tipo de ecualización a imágenes de color. a) Uniforme, b) Exponencial, c) Rayleigh, d) Hipercúbica, e) Logaritmo Hiperbólica.

empleadas en la mejora de la distribución de los valores de intensidad, evidenciando los cambios en la calidad y contraste de las imágenes procesadas (ver Figura 3).

Adicionalmente, en la figura 4 se muestran cada uno de los histogramas obtenidos al aplicar el proceso de ecualización para cada una de las imágenes anteriormente mostradas. La finalidad de desplegar los histogramas radica en analizar de forma gráfica la distribución de los pixeles para cada una de las imágenes resultantes.

A continuación, se presentan las imágenes resultantes tras aplicar los algoritmos de ecualización a las imágenes de color. Se puede observar a simple vista el impacto de las técnicas en cada uno de los canales de color, observando cómo se se modifica la

distribución de los niveles de intensidad y cómo esto influye en la mejora del contraste y la percepción visual de las imágenes procesada (ver Figura 5).

5. Conclusiones

Los resultados obtenidos tras la aplicación de los algoritmos de ecualización en imágenes de escala de grises y a color demuestran su efectividad en la mejora del contraste y la distribución de los niveles de intensidad. En imágenes en escala de grises, se observa una mejora significativa. En el caso de las imágenes a color, la ecualización aplicada de manera independiente a cada canal de color genera cambios perceptibles en la tonalidad y saturación, lo que puede ser beneficioso en algunas aplicaciones, pero también puede introducir distorsiones no deseadas. Esto resalta la importancia de seleccionar el método de ecualización más adecuado según el tipo de imagen y la aplicación específica.

Finalmente, los resultados obtenidos confirman que la ecualización es una herramienta útil para mejorar la calidad visual de las imágenes, aunque su impacto puede variar dependiendo de las características de la imagen original y del algoritmo utilizado. Como trabajo futuro, se podría explorar el uso de técnicas avanzadas como la ecualización adaptativa.

Referencias

1. Pajares Martinsanz, G., de la Crúz García, J. M.: Visión por Computador: Imágenes Digitales y Aplicaciones. Alfaomega-Ra-Ma, México D.F (2002)
2. Visión Artificial. Servicio de Publicaciones de la Universidad de Alcalá de Henares, Alcalá de Henares, Madrid (1996)
3. Image Contrast Enhancement by Constrained Local Histogram Equalization. Computer Vision Image Understanding, 73(2), 281-290 (1999)
4. Pajares Martinsanz, G., de la Crúz García, J. M.: Ejercicios resueltos de visión por computador. RA-MA (2007)
5. Gonzalez, R. C., Woods, R. E.: Tratamiento digital de imágenes. (1996)
6. Pratt, W.K.: Digital Image Processing. John Willey and Sons, New York (1991)

Electronic edition
Available online: http://www.rcs.cic.ipn.mx

http://rcs.cic.ipn.mx

Centro de Investigación
en Computación