

C&T

Review Journal

Riqchary

Revista de Investigación en Ciencia y Tecnología

Volumen 07 -

Número 2 -2025

ISSN: 2810-8124

(en línea)

ING. INFORMÁTICA Y SISTEMAS

UNAMBA



COINCITEC 2025

Congreso Internacional de Investigación
en Ciencia y Tecnología Computacional

Generado con meta.ai

DOI: <https://doi.org/10.57166.riqchary.v7.n2.2025>



UNIVERSIDAD NACIONAL
MICAELA BASTIDAS
DE APURÍMAC



ISSN/2810-8124



Comité Editorial

- **Editor General:** Dr. Ecler Mamani Vilca
Universidad Micaela Bastidas Nacional de Apurímac, Perú

Equipo Técnico:

- Dr. José Luis Merma Aroni
Universidad Micaela Bastidas Nacional de Apurímac, Perú
- M.Sc. Mario Aquino Cruz
Universidad Tecnológica de los Andes Perú
- Dr. Manuel J. Ibarra Cabrera
Micaela Bastidas Universidad Nacional de Apurímac, Perú
- Dr. Yalmar Ponce Atencio
Micaela Nacional de Juliaca, Perú
- Dr. Leonel Coyla Idme
Universidad Nacional del Altiplano, Perú
- Dr. José Luis Herrera Salazar
Universidad Norbert Wiener, Perú
- Mag. Olga Jaramillo
Santa Fe Community College, Estados Unidos
- Dr. Edwar Ilasaca Cahuata
Universidad Nacional del Altiplano, Perú
- Mag. Yonatan Mamani Coaquira
Universidad Micaela Bastidas Nacional de Apurímac, Perú

Equipo Editorial:

- Mag. Marleny Peralta Ascue
Universidad Privada de Tacna, Perú
- Mag. Rafael Quispe Merma
Universidad Nacional San Antonio de Cusco, Perú
- Ing. Ebert Gómez Aiquipa
Universidad Nacional San Antonio de Cusco, Perú
- MBA. Evelyn Naida Luque Ochoa
Micaela Bastidas Universidad Nacional de Apurímac, Perú
- Mag. Kevin Arnold Arias Figueroa
Micaela Bastidas Universidad Nacional de Apurímac, Perú
- Mag. Alejandrina Huaylla Quispe
Universidad Tecnológica de los Andes Perú
- Mag. Betsabe M. Ccolque Ruiz
Micaela Bastidas Universidad Nacional de Apurímac, Perú
- Mag. Miguel Alfaro
Micaela Bastidas Universidad Nacional de Apurímac, Perú

Periodicidad: Semestral

Vol. 7 Núm. 2 año (2025)

Áreas temáticas: Ciencias exactas, ingenierías y de materiales

Editorial: UNAMBA – Perú

Vice Rectorado de Investigación

<https://doi.org/10.57166/riqchary/v7.n2.2025>

ISSN: 2706-543X- ISSN: 2810-8124 (e)

Editorial

La presente edición surge como un espacio de encuentro para investigadores, docentes, estudiantes y profesionales, quienes, desde distintas perspectivas, contribuyen al desarrollo del conocimiento en el ámbito científico y tecnológico. Inspirada en iniciativas como el Congreso Internacional de Investigación en Ciencia y Tecnología Computacional (COINCITEC), esta publicación busca no solo difundir investigaciones, sino también generar diálogo, reflexión e intercambio de experiencias que fortalezcan la comunidad académica.

En este sentido, la revista se consolida como un medio que impulsa nuevas ideas, promueve la innovación y visibiliza el esfuerzo de quienes apuestan por el avance del conocimiento, aportando así al crecimiento académico y profesional.

Director: Dr. Ecler Mamani Vilca





C & T Riqchary

Revista de investigación - UNAMBA

Autores

- Daniel Lévano Rodriguez
 - Flor Elizabeth Cerdán León
 - Jesus Inocencio Lopez Rodriguez
 - Mauricio Antonio Alaluna Godinez
 - Siloh Draguichy Valladares Salas
 - Diego Pereira Sartori
 - Jean Phol A. Curi Garrafa
 - Victor R. Ortega Marocho
 - Wilson Mamani Rodrigo
 - Royer Elvis Moreano Condorcuya
 - Evelyn Naida Luque Ochoa
 - Nilton Brayaam Hurtado Mendoza
 - Christopher Bryhan Bejar Avalos
 - Ana Elizabeth Maquera Curasi
 - Luis Fernando Chumbes Ramos
 - Jenny Ríos Navío
 - Natalie Cruz Tumba
 - Alex Lancho Ramos
 - Henry Leon Hurtado
- Antony Stebend Camacho García
 - Suker Huamanñahui Hilario
 - Mario Aquino Cruz
 - Zulennys Marcano Marín
 - José Castro Soto
 - Percy Quispe Ñaca
 - Edwin Mestas Yucra
 - Aldo H. Zanabria Gálvez
 - N. Sheyla.Saenz Bermejo
 - Jackeline M. Chambilla Alata
 - Gerson Aldair Pulache Quiroga
 - Juan Aurelio De La Cruz Gamarra
 - Alberto Carlos Mendoza de los Santos
 - xander Darío Huamán Ccorahua
 - María del Carmen Barreto Prado
 - Fulgencio Vilcanqui Pérez
 - Luis Fernando Pérez Falcón
 - Rafael Ricardo Quispe Merma
- © Dr. Ecler Mamani Vilca / Director del Comité Editorial

Av. Garcilaso C.U Tamburco, Abancay Telef, +51983747138

Apurímac – Perú eclervirtual@unamba.edu.pe

Hecho el Depósito Legal en la Biblioteca Nacional del Perú N° 2020-10220 ISSN: 2709-8990

ISSN: 2955-8646 (e)- ISSN: 2709-8990

Tiraje: 500 ejemplares

Volumen 7 Número 2 – 2025

Se terminó de imprimir en El Misti EIRL

RUC 20450599108 - Cel: 950-509953, imprentamisti@hotmail.com

01

Hipertensión arterial en adultos: análisis de riesgo y clasificación predictiva mediante Random Forest

Pág.

09

Comparación de modelos generativos compactos para respuesta automática en español mediante RAG

Pág.

19

Modelos Profundos Preentrenados para la Detección Automatizada de 10 Tipos de Fracturas Óseas Un Estudio Comparativo con ResNet50, EfficientNetB3 y MobileNet

Pág.

30

Neuro-IA en el aula: Nuevas fronteras en la motivación y e rendimiento del estudiante universitario

Pág.

39

Spoop: Videojuego Educativo Basado en Juegos Serios para el Desarrollo de Habilidades Numéricas en Niños en Etapa Escolar Temprana

Pág.

49

Evaluación del Rendimiento y la Escalabilidad en un Clúster Apache Spark y Hadoop en Entorno de Bajo Costo

Pág.

INDICE

54

Pág.

Detección de conjuntivitis a partir de imágenes oculares utilizando arquitecturas CNN y un modelo de fusión

66

Pág.

Evaluación del carácter corrosivo del agua y su impacto en un sistema de enfriamiento industrial

76

Pág.

Aplicación de inteligencia artificial para predecir la rotación laboral de un supermercado en Juliaca - Puno

82

Pág.

Principales mecanismos de seguridad en base de datos: Una revisión sistemática

88

Pág.

Modelación predictiva en los nevados de Abancay mediante análisis Arima: proyecciones al 2030

94

Pág.

Efecto de coagulantes de origen vegetal sobre las características fisicoquímicas y sensoriales del queso

Hipertensión arterial en adultos: análisis de riesgo y clasificación predictiva mediante Random Forest

Arterial hypertension in adults: risk analysis and predictive classification using Random Forest

Daniel Lévano Rodríguez^A, Flor Elizabeth Cerdán León^B, Jesus Inocencio Lopez Rodriguez^C, Mauricio Antonio Alaluna Godinez^D, Siloh Draguichy Valladares Salas^E y Diego Pereira Sartori^F

Resumen— La hipertensión arterial (HTA) ha sido considerada un reto en la salud debido a su impacto en la morbilidad cardiovascular y a su frecuente diagnóstico tardío. Se abordó el problema mediante el desarrollo de un modelo de clasificación predictiva basado en Random Forest, con el objetivo de identificar individuos en riesgo de hipertensión a partir de variables clínicas, demográficas y metabólicas. Se utilizó una base de datos proveniente de pacientes en México; mediante la técnica SMOTE-Tomek fue sometida a procesos de limpieza, normalización y balanceo. Se entrenó el modelo con el 70% de los datos y se validó con el 30% restante, aplicando validación cruzada k-fold (k=10). Se evaluó el rendimiento del modelo mediante métricas como precisión, sensibilidad, puntaje F1 y matriz de confusión. Se comparó el modelo con otros métodos como KNN y Decisión Tree. Se alcanzó una exactitud del 98% con el modelo optimizado (127 árboles, profundidad 20) destacando como predictores claves el índice de masa corporal, la presión arterial, la actividad física, el peso y la circunferencia de cintura. Aunque también se evaluaron biomarcadores metabólicos, estos presentaron menor relevancia en la clasificación frente a los predictores antropométricos. Los resultados obtenidos confirman que Random Forest es una herramienta robusta y precisa para la detección temprana del riesgo de hipertensión. Gracias a su integración mediante una API y un formulario interactivo, el modelo resulta accesible incluso para usuarios sin formación técnica, lo que contribuye a estrategias preventivas de salud pública.

Palabras clave: Aprendizaje automático, factores de riesgo, salud pública.

Abstract— Arterial hypertension (AH) has been considered a major public health concern due to its impact on cardiovascular morbidity and mortality and its frequent late diagnosis. This study addresses the problem by developing a predictive classification model based on the Random Forest algorithm, aiming to identify individuals at risk of hypertension using clinical, demographic, and metabolic variables. A dataset from patients in Mexico was used and processed through cleaning, normalization, and balancing with the SMOTE-Tomek technique. The model was trained with 70% of the data and validated with the remaining 30%, using 10-fold cross-validation. Its performance was evaluated through metrics such as precision, recall, F1-score, and confusion matrix. The model was compared with other methods such as KNN and Decision Tree. The optimized model (127 trees, depth 20) achieved an accuracy of 98% with body mass index, blood pressure, physical activity, weight, and waist circumference identified as the most relevant predictors. Although metabolic biomarkers were also evaluated, they were less relevant in the classification compared to anthropometric variables. The results confirm that Random Forest is a robust and accurate tool for the early detection of hypertension risk. Thanks to its integration via an API and an interactive form, the model is accessible even to not-technical users, contributing to preventive strategies in public health.

Keywords: Machine learning, risk factors, public health

1 INTRODUCCIÓN

La Hipertensión Arterial (HTA) es un factor de riesgo crucial para las enfermedades cardiovasculares y una de las principales causas de mortalidad a nivel mundial, con un estimado actual del 22% de la población global, quienes padecen esta condición, y menos de una quinta parte de los afectados lleva un adecuado control de su presión arterial [1] [2], lo que incrementa el riesgo de complicaciones graves. Este panorama resalta la necesidad de herramientas efectivas para la detección temprana y el manejo de la HTA, dado que falta

conciencia en etapas iniciales contribuye a su sub diagnóstico y a un control insuficiente.

En Perú, la prevalencia estandarizada por edad es del 19,2% [3]. En México, la prevalencia de HTA en adultos fue del 47,8% [4]. En Ecuador presenta una mayor prevalencia de HTA [5]; y se estima que afecta al menos al 19,8% de la población, vinculada a factores como la obesidad, el sedentarismo y los antecedentes familiares [5][6]. Estas cifras reflejan una carga epidemiológica compartida en estos países, y la elevada prevalencia de obesidad y sedentarismo parece estar estrechamente relacionada con el aumento de los casos



Revista de Investigación en Ciencia y Tecnología
ISSN: 2810-8124 (en línea) / ISSN: 2706-543x
Universidad Nacional Micaela Bastidas de Apurímac – Perú

Vol. 7 Núm. 2 (2025) - Publicado: 19/08/25 - Indexaciones
Número: doi.org/10.57166/riqchary.v7.n2.2025
Páginas: 1-8 | Recibido 08/07/2025 ; Aceptado 08/12/2025

doi.org/10.57166/riqchary.v7.n2.2025.1

Autores:

- A. **ORCID iD** <https://orcid.org/0000-0001-5652-0601>
Daniel Lévano Rodríguez, Universidad Nacional Tecnológica de Lima Sur dlevano@untels.edu.pe.
- B. **ORCID iD** <https://orcid.org/0000-0001-6747-6335>
Flor Elizabeth Cerdán León, Universidad Nacional Tecnológica de Lima Sur fcerdan@untels.edu.pe.
- C. **ORCID iD** <https://orcid.org/0009-0008-7225-7223>
Jesus Inocencio Lopez Rodriguez, Universidad Nacional Tecnológica de Lima Sur 2223050507@untels.edu.pe.
- D. **ORCID iD** <https://orcid.org/0009-0003-9209-5210>
Mauricio Antonio Alaluna Godinez, Universidad Nacional Tecnológica de Lima Sur 2223110254@untels.edu.pe.
- E. **ORCID iD** <https://orcid.org/0009-0008-0912-1008>
Siloh Draguichy Valladares Salas, Universidad Peruana Unión silohvalladares@upeu.edu.pe.
- F. **ORCID iD** <https://orcid.org/0009-0004-2899-389X>
Diego Pereira Sartori, Universidad Peruana Unión diegopereira@upeu.edu.pe.

de HTA, esto evidencia la urgencia de herramientas predictivas para mejorar la detección temprana y personalizar estrategias preventivas.

Machine Learning (ML) es una rama de la inteligencia artificial. Su enfoque está en desarrollar algoritmos capaces de aprender a partir de datos y tomar decisiones o hacer predicciones sin necesidad de una programación muy compleja [7]. A través del ML se permite mejorar los resultados clínicos mediante modelos predictivos, entre sus principales ventajas se encuentran: Automatización de análisis de datos, detección temprana de patrones anómalos y la posibilidad de integrar múltiples variables en la toma de decisiones clínicas [8].

Random Forest es un algoritmo de aprendizaje automático basado en la construcción de múltiples árboles de decisión durante el entrenamiento y la combinación de sus resultados para mejorar la precisión y reducir el riesgo de sobreajuste [9]. En el contexto de HTA, permite identificar las variables más relevantes que contribuyen al riesgo y detectar patrones complejos y no lineales entre múltiples variables clínicas y sociodemográficas, mejorando la exactitud de la predicción o discriminación clínica [9].

Las métricas de evaluación tienen gran importancia en el desarrollo de modelos de aprendizaje automático, siendo estas la accuracy (exactitud), que mide la proporción de flujos de tráfico adecuadamente clasificados sobre el total; la precisión (precisión), que refleja la fracción de predicciones positivas acertadas entre todas las positivas; recall (sensibilidad), que evalúa la proporción de positivos reales identificados; y el F1-score (puntaje F1), que balancea precisión y sensibilidad. Estas definiciones se derivan de los verdaderos positivos (VP), verdaderos negativos (VN), falsos positivos (FP) y falsos negativos (FN) [10].

$$\text{Precisión} = \frac{VP}{VP + FP}$$

$$\text{Sensibilidad} = \frac{VP}{VP + FN}$$

$$\text{Exactitud} = \frac{VP + VN}{VP + FP + FN + VN}$$

$$\text{PuntajeF1} = 2 \frac{(\text{Sensibilidad} * \text{Precisión})}{(\text{Sensibilidad} + \text{Precisión})}$$

En una investigación previa, utilizaron el algoritmo Random Forest con el objetivo de clasificar a individuos con HTA, enfermedad arterial coronaria (EAC) y aquellos sin enfermedades cardiovasculares (no-ECV), a partir de perfiles metabólicos plasmáticos [11]. En este estudio se enfoca en el diagnóstico cardiovascular en adultos mediante el análisis utilizando aprendizaje automático, dependiendo de variables como aminoácidos, acilcarnitinas, metilargininas, etc [11]. Se observa en este estudio una similitud con el presente trabajo en cuanto al uso de algoritmos de clasificación y resalta la necesidad de integrar datos clínicos para mejorar la predicción del modelo.

Otra investigación fue la realizada por [12], demostró la eficacia del algoritmo Random Forest en la detección temprana de enfermedades cardiovasculares, para la creación de su modelo de clasificación fue necesario un riguroso preprocesamiento de datos, incluyendo la eliminación de valores atípicos y la selección de características relevantes para optimizar el modelo predictivo. La investigación señala que la combinación de técnicas de limpieza de datos y selección de características mejora gradualmente el rendimiento del modelo haciéndolo más confiable.

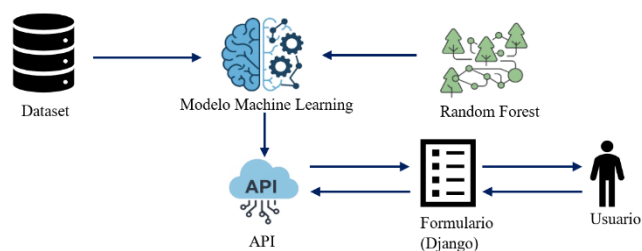
Bajo este contexto, se consideró que mediante el uso de ML se puede mejorar los resultados clínicos mediante modelos predictivos. La aplicación de estas técnicas de aprendizaje automático permite facilitar la detección de patrones complejos y mejora la capacidad discriminativa en comparación con otros métodos tradicionales [13]. Esto da inicio a cómo desarrollar herramientas de soporte, preventivas y personalizadas, en el manejo de la HTA.

El objetivo de este estudio fue identificar de manera integral los factores de riesgo que incrementan la probabilidad de desarrollar HTA, combinando un enfoque individual con métodos de clasificación Random Forest para optimizar las estrategias de prevención y control de la enfermedad [14]. Se proyecta que el modelo de Random Forest, entrenado con variables clínicas y demográficas, puede predecir el riesgo de HTA con una exactitud superior al 90%. Se espera que el peso y la actividad física sean los predictores más influyentes, dada su conocida relación con el desarrollo de la HTA en esta población.

2 MATERIALES Y METODOLOGÍA

En esta sección se describe la arquitectura de la solución (véase la Figura 1), que abarca desde el preprocesamiento del conjunto de datos (dataset) hasta la interacción con el usuario a través de un formulario, incluyendo el desarrollo del modelo Random Forest y la integración de una API.

En esta investigación, se empleó la metodología CRISP-DM para estructurar el proceso de minería de datos, un estándar ampliamente reconocido que guía la preparación de datos,



modelado, evaluación y despliegue [15].

Fig. 1. Arquitectura de la solución.

2.1 Limpieza de datos

Este dataset, que contaba con 4364 registros, se obtuvo a partir de un repositorio de ML, específicamente del sitio web Kaggle, y contiene información demográfica y clínica de pacientes. Las variables registradas incluyeron sexo

(codificado numéricamente), edad y diversos biomarcadores, tales como hemoglobina, ácido úrico, albúmina, colesterol HDL y total, creatinina, glucosa e insulina. La variable dependiente fue el "Riesgo de hipertensión (0 = sin riesgo, 1 = con presencia)". La muestra se obtuvo a partir de registros de pacientes en México, lo que introdujo un sesgo de selección, ya que la mayoría de los datos provenían de centros urbanos [12].

En la limpieza de datos, los criterios de eliminación de registros incluyeron, la eliminación de datos nulos y que se encuentren fuera de su rango. se realizó un análisis exploratorio para evaluar la distribución de los datos y detectar valores atípicos [16]. Se eliminaron datos redundantes que no contribuyen en la predicción del modelo, había valores extremos fuera de los rangos fisiológicos establecidos por cada biomarcador como horas de sueño, actividad total; también la variable actividad total se agrupó en 5 partes para clasificarlos (0 - 4). Los valores nulos de la variable estatura, medida de cintura, tensión arterial, fueron eliminados. El dataset se redujo a 3019 registros, posteriormente se guardó como un archivo CSV con el nombre de "HTA_limpio".

2.2 Balanceo de datos

Se identificó un desbalance en los datos (véase la Figura 2), con una diferencia de 1247 (Con riesgo - Sin riesgo), se observó un predominio de casos positivos frente a casos negativos. En un script de Python, los datos fueron cargados desde el archivo CSV utilizando la librería pandas. Se utilizó la técnica SMOTE-Tomek [16]. Mediante la librería imblearn que combina el sobremuestreo sintético de la clase minoritaria (SMOTE) con la eliminación de muestras cercanas entre clases diferentes (Tomek Links), generando un volumen de entrenamiento más equilibrado. Tras aplicar el método, se pasó de tener 1680 casos positivos y 433 negativos a obtener aproximadamente 1671 muestras en cada clase como se visualiza en la Fig. 2, tras ajustar los datos con dicha técnica. Este nuevo dataframe se guardó con el nombre "HTA_balanceado" para su uso en el entrenamiento del modelo.

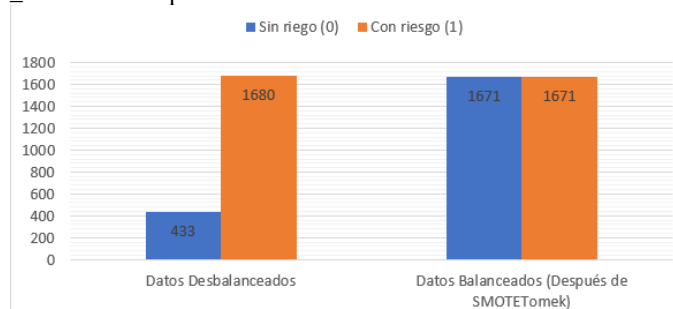


Fig. 2. Comparación y distribución de casos de hipertensión antes y después del balanceo.

2.3 Entrenamiento del modelo

Se empleó un script adicional en Python, utilizando las librerías pandas, numpy y sklearn, se cargaron los datos ya balanceados, se mostró la significancia de cada variable, con el fin de conocer que variables influyen más en el riesgo de HTA. Aquí se seleccionaron las variables más relevantes para

el modelo, que se usarán de ahora en adelante.

Otro script de Python, destinado al entrenamiento del algoritmo, utilizando las mismas librerías mencionadas, se cargó el nuevo volumen balanceado, donde se leen las variables seleccionadas para el entrenamiento del modelo. Se separó la variable objetivo (riesgo_hipertension) del resto de las variables predictoras. Posteriormente, los datos fueron divididos en conjuntos de entrenamiento (70%) y prueba (30%) [17] [18] [19].

Para la clasificación del riesgo de hipertensión se implementó el algoritmo RandomForestClassifier utilizando la librería sklearn en Python. Este modelo requería un número de estimaciones (n_estimators) y una profundidad (max_depth), que captura relaciones complejas entre las variables sin sobreajustar el modelo. También fue necesario fijar una semilla (random_state) para que el modelo no varíe cada vez que se ejecute.

La elección del método Random Forest se basa en su capacidad para manejar múltiples variables simultáneamente, su resistencia a los valores atípicos y su eficiencia para identificar relaciones complejas en contextos de clasificación médica [20]. El desempeño del modelo fue validado utilizando matriz de confusión, informe de clasificación y validación cruzada, alcanzando una precisión general del 98% [21] [22].

2.4 Optimización del modelo

La optimización del modelo ha requerido crear un primer modelo con 50 árboles de decisión (n_estimators=50) y una profundidad de 4 (max_depth=4), usando una semilla fija de 42 (random_state=42). El modelo fue entrenado para evaluar su desempeño utilizando el informe de clasificación (classification_report), el cual proporciona métricas como la precisión, la sensibilidad y puntaje F1, tanto para los casos "Con riesgo" como "Sin riesgo".

Se utilizó una matriz de confusión, con el modelo de 50 árboles, para evaluar gráficamente el desempeño del modelo de clasificación. Se emplearon las bibliotecas matplotlib y seaborn, ampliamente utilizadas en Python para la visualización de datos.

Se encontró la selección más óptima del modelo usando la tasa de error, mediante un gráfico generado en Python usando la librería sklearn y matplotlib, y se descubrió (véase la Figura 3) que, si se usaban una cantidad mayor de hasta aproximadamente 129 árboles, la tasa de error es menor.

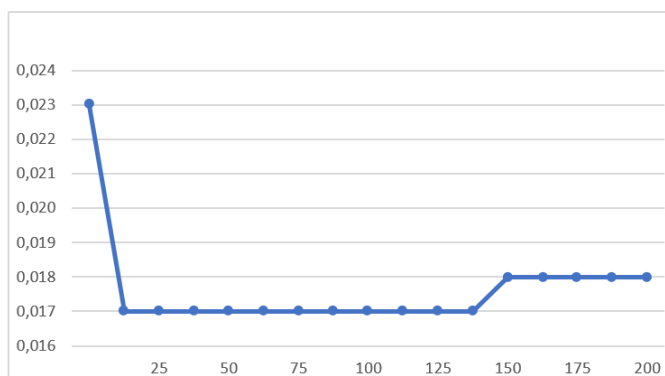


Fig. 3. Tasa de error vs Número de árboles en Random Forest.

Considerando la tasa de error, se optó un modelo con 127 árboles de decisión ($n_{\text{estimators}}=127$) y una profundidad máxima de 20 ($\text{max_depth}=20$), y se usó una semilla fija de 42 ($\text{random_state}=42$). Se compararon las métricas entre los dos modelos para comprobar el modelo más óptimo, siendo el de 127 árboles.

2.5 Comparación del modelo

Tomando en cuenta la profundidad máxima y la semilla fija empleados en el modelo de random forest, se comparó en otros códigos de Python modelos que usan tanto KNN como Árbol de Decisión, utilizando las librerías pandas, numpy, sklearn y matplotlib. De esta manera se comprobó el modelo más adecuado, siendo el de Random Forest.

2.6 Predicción y despliegue

Para la generación de la API se empleó el framework Flask, utilizando las librerías flask, threading y time, donde se carga el modelo, se definen los datos esperados de la API y su respuesta. Esta define una ruta en la cual se le podrá enviar solicitudes con datos en formato JSON procesados en tiempo real.

Se simuló entradas de prueba a través de esta API para evaluar el funcionamiento del modelo y se verificó un adecuado desempeño para predecir el riesgo de hipertensión. Se creó un formulario para acceder al modelo y la API creada, empleando el framework Django, en el cual se le pueden colocar los datos pedidos, que transformará los datos en formato JSON y los enviará al API, donde serán analizados por el modelo y dará una respuesta, que la API devuelve al formulario para su visualización.

El formulario desarrollado con Django (mostrado en la figura 4) permite procesar entradas y devolver resultados con los datos colocados, calcular ciertas variables (como el índice de masa corporal) necesarias para el modelo con otras variables del formulario, optimizando el tiempo, sea dinámico y más sencillo de usar. Este formulario mostrará directamente el resultado (Con riesgo o Sin riesgo), tiene una función enviar un reporte, que se envía por correo en un PDF, con los datos colocados con el resultado del paciente.

Lévano Rodríguez Daniel, Cerdán León Flor Elizabeth, Lopez Rodríguez Jesus Inocencio, Alaluna Godínez Mauricio Antonio, Pereira Sartori Diego y Valladares Salas Siloh Draguichy

Mediciones Clínicas

Tensión Arterial (mmHg): 78	Medida de Cintura (cm): 90
Nivel de Actividad Física: ☐ Muy Baja ☐ Baja ☒ Moderada ☐ Alta ☐ Muy Alta	Horas de Sueño: 5
Hemoglobina Glucosilada (%): 3	Insulina (μU/mL): 15
Glucosa Promedio (mg/dL): 90	Concentración de Hemoglobina (g/dL): 14
Colesterol LDL (mg/dL): 120	Triglicéridos (mg/dL): 180

Resultado: El paciente no presenta riesgo de Hipertensión Arterial

Fig. 4. Visualización del formulario.

3 RESULTADOS

3.1 Significancia de los factores de riesgo

El resultado de las variables más influyentes para el modelo, como se observa en la Figura 4, son el índice de masa corporal con un 26.21%, la tensión arterial con un 21.84%, la medida de cintura con un 14.61%, el peso con un 11.09%, la actividad física con un 5.94%, la estatura con un 2.92%, las horas de sueño con un 2.42%, el sexo con un 2.33%, la edad con un 2.24%, la HbA1c (Hemoglobina glucosilada) con un 1.13%, la glucosa promedio con un 0.99%, la insulina con un 0.76%, el colesterol LDL con un 0.70%, los triglicéridos con un 0.67%, la glucosa con un 0.62%.

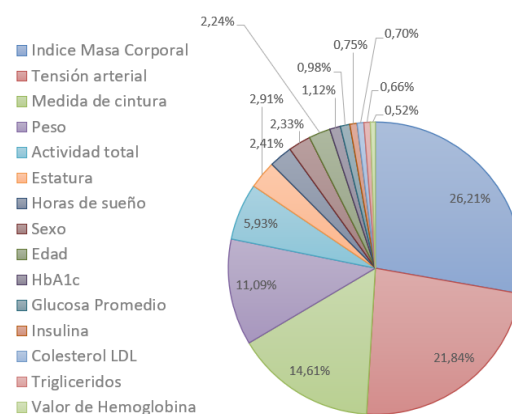


Fig. 5. Importancia de los factores de riesgo.

3.2 Comparación y selección del modelo

El modelo inicial, con 50 árboles de decisión, como indica la Tabla 1, se obtiene una exactitud de 96%, para casos "Con riesgo" la precisión de 97% y una sensibilidad del 95%; para casos "Sin riesgo" la precisión es de 95% y una sensibilidad del 97%.

TABLA 1

Valores de las métricas de clasificación del modelo con 50 árboles.

Métrica	Precisión	Sensibilidad	Puntaje F1
Sin riesgo	0.95	0.97	0.96
Con riesgo	0.97	0.95	0.96
Exactitud	-	-	0.96

Los resultados de la matriz de confusión, tal como se aprecia en la Tabla 2 muestra que los 479 casos clasificados como "Sin riesgo" (VN+FP), 465 fueron correctamente identificados, y de los 524 casos "Con riesgo" (FN+VP), 500 fueron correctamente predichos. Se obtuvieron 14 falsos positivos, y 24 falsos negativos.

TABLA 2

Valores obtenidos de la matriz de confusión con 50 árboles.

n_estimators=50		Predicción	
		Sin riesgo	Con riesgo
Valores reales	Sin riesgo	VN = 465	FP = 14
	Con riesgo	FN = 24	VP = 500

El modelo mejorado, con 127 árboles de decisión (véase Tabla 3), obtuvo una exactitud de 98%, para casos "Con riesgo" la precisión de 99% y una sensibilidad del 97%; para casos "Sin riesgo" la precisión es de 97% y una sensibilidad del 99%.

TABLA 3.

Valores de las métricas de clasificación del modelo optimizado.

Métrica	Precisión	Sensibilidad	Puntaje F1
Sin riesgo	0.97	0.99	0.98
Con riesgo	0.99	0.97	0.98
Exactitud	-	-	0.98

Los valores obtenidos de la matriz de confusión (véase la Tabla 4) de los 480 casos reales clasificados como "Sin riesgo" (VN+FP), 476 fueron correctamente identificados, los 524 casos "Con riesgo" (FN+VP), 510 fueron identificados como verdaderos positivos; estos resultados evidencian que los falsos positivos se redujeron de 14 a 4, y los falsos negativos de 24 a 14.

TABLA 4

Valores obtenidos de la matriz de confusión con 127 árboles.

n_estimators=127		Predicción	
		Sin riesgo	Con riesgo
Valores reales	Sin riesgo	VN = 476	FP = 4
	Con riesgo	FN = 14	VP = 510

En la Figura 6 se compara que el modelo con 127 árboles difiere con el modelo con 50 árboles, mostrando una diferencia del 2% en precisión, sensibilidad y puntaje F1 para ambas clases. Con respecto a la exactitud general, pasa de 96% a 98%.



Fig. 6. Comparación entre el modelo con 50 árboles vs modelo con 127 árboles.

El modelo entrenado con KNN, como se indica en la Tabla 5, obtiene una exactitud de 92%; para casos "Con riesgo" la precisión de 98% y una sensibilidad del 85%; para casos "Sin riesgo" la precisión es de 87% y una sensibilidad del 98%.

TABLA 5

Valores de las métricas de clasificación del modelo KNN.

Métrica	Precisión	Sensibilidad	Puntaje F1
Sin riesgo	0.87	0.98	0.92
Con riesgo	0.98	0.85	0.91
Exactitud	-	-	0.92

Los resultados obtenidos con el modelo Decision Tree son semejantes a los obtenidos por el modelo con Random Forest (mostrado en la Tabla 3), por lo tanto, se comparó sus curvas AUC y ROC, como se muestra en la Figura 7. El modelo Decision Tree tiene un valor AUC de 0.998 y el modelo Random Forest un valor AUC de 0.976.

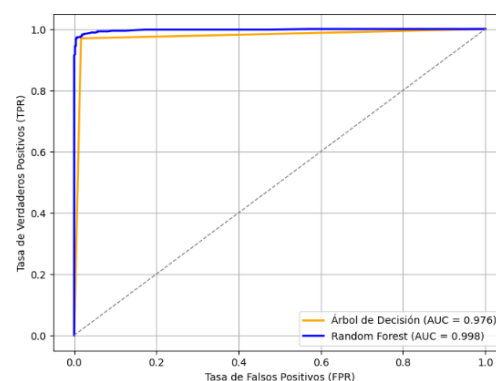


Fig. 7. Comparación entre el modelo Random Forest vs el modelo Decision Tree.

Se realizó la validación cruzada del modelo Random Forest y del modelo Decision Tree, como se muestra en la Tabla 6. Random Forest obtuvo una precisión de 98.05% con una desviación estándar del 0.31%, mientras que, el modelo Decision Tree obtuvo una precisión promedio de 97.48% con una desviación estándar del 0.97%.

TABLA 6.

Valores obtenidos de la validación cruzada.

Modelos	Precisión	Desviación estándar
Random Forest	98.05%	0.31%
Decision Tree	97.48%	0.97%

4 DISCUSIONES

4.1 Análisis de resultados y comparación con estudios previos

El análisis de la importancia de las variables reveló que, en el conjunto total de variables, los factores con mayor peso predictivo fueron la masa corporal, actividad total, peso, tensión arterial y medida de cintura. El análisis enfocado en los biomarcadores (véase Figura 4), observó que ciertos parámetros metabólicos (por ejemplo, HbA1c, colesterol, triglicéridos, insulina, etc.) tuvieron una importancia relativamente menor de lo esperado. Esto sugiere que los indicadores relacionados con el estado físico (masa corporal, peso y medidas antropométricas) son determinantes clave para la clasificación del riesgo hipertensivo. No hay que descartar la utilidad de biomarcadores, pues futuras investigaciones podrían explorar el impacto de estos biomarcadores en combinación con otras variables clínicas para mejorar la capacidad predictiva del modelo) [23] [24] [25].

Los resultados obtenidos demuestran que el modelo Random Forest, como parte de los modelos de aprendizaje automático, supera significativamente a métodos tradicionales, como la regresión logística y KNN, en la detección de patrones complejos presentes en los datos clínicos. Un estudio reciente realizado por Kandil et al. Demostró que los métodos tradicionales tienen una menor capacidad predictiva o discriminatoria en comparación con algunos enfoques más avanzados, lo que puede respaldar el uso de algoritmos de ML en el análisis del riesgo de hipertensión [26] [27] [28].

Comparado a un modelo desarrollado en una investigación previa [11], dicho grupo alcanzó una precisión del 80% en la clasificación multiclase y del 91% en la clasificación binaria entre casos de ECV y no-ECV. También, en otros trabajos [29], se utilizó un modelo de Árbol de Decisión para la predicción de enfermedades cardíacas, el cual alcanzó una exactitud cercana al 79,78%, lo que evidenció unas limitaciones en la detección de patrones complejos y una mayor susceptibilidad al sobreajuste. En contraste, el modelo presentado en este estudio logró una exactitud del 98%, como se evidencia en la Tabla 3, la cual se ha confirmado nuestra hipótesis y ha dado un excelente desempeño del modelo si queremos saber el riesgo de HTA de un paciente. Este desempeño indica que el modelo puede servir como una herramienta útil al momento de tomar decisiones con respecto a los pacientes y su condición actual.

Otro estudio [12] desarrollo un modelo de predicción de enfermedad cardiovascular utilizando Random Forest y un dataset clínicos del repositorio de Kaggle, alcanzando una exactitud del 99% en la clasificación binaria de presencia o ausencia de enfermedad cardíaca. Si bien su modelo abarcó un espectro más amplio de patologías cardiovasculares, el modelo del presente artículo se enfocó exclusivamente en la predicción del riesgo de HTA, alcanzando una exactitud del 98% y métricas por clase altamente equilibradas (precisión de 99% y sensibilidad de 97% en la clase "Con riesgo"; precisión de 97% y sensibilidad de 99% en la clase "sin riesgo") (véase tabla 3).

4.2 Limitaciones y mejoras a futuro

A pesar de la diferencia en los objetivos clínicos, ambos estudios comparten un nivel metodológico riguroso y resultados de alto rendimiento, lo que posiciona nuestra propuesta como una alternativa igualmente sólida y competitiva como modelo de soporte clínico en el ámbito de la salud.

El desempeño del modelo es significativamente prometedor, aun así, se recomienda explorar mejoras que incluyan:

- a. Incorporación de métricas adicionales y validaciones externas que demuestren su utilidad en la evaluación del riesgo de sobreajuste.
- b. Ampliación de la muestra hacia diferentes contextos poblacionales para aumentar la validez externa de los hallazgos, no solo de estratos sociales sino también de grupos étnicos variados. Estudios como [30], han resaltado la necesidad de validar estos modelos en muestras heterogéneas antes de su aplicación clínica.

Los presentes hallazgos del estudio llegan a confirmar la utilidad de modelos de ML, especialmente el modelo Random Forest, en la identificación tanto tardía como temprana del riesgo de presión arterial alta. Esta herramienta en su utilidad ayuda con la mejora en precisión de la clasificación y también llega a ofrecer un potencial significativo para la implementación de estrategias preventivas en salud pública.

Aunque el modelo desarrollado presenta un rendimiento sobresaliente con resultados satisfactorios, es importante reconocer que los datos empleados provienen de una población específica, lo que puede limitar la aplicación del modelo a otros contextos geográficos, además, se omitió diferentes atributos del modelo debido a su baja significancia para la predicción. Sin embargo, esto puede cambiar con futuras investigaciones, que amplíen la base de datos y permitan tanto el uso de mayores atributos en el modelo como el abarcar una población mayor para obtener mejores resultados en el modelo.

Estudios como [31], han demostrado que el autocuidado y la educación en salud pueden mejorar significativamente el control de la presión previniendo su elevación y reduciendo la probabilidad de desarrollar HTA en adultos. Se combina un enfoque individual con métodos de clasificación Random Forest para optimizar las estrategias de control de la enfermedad mediante el análisis predictivo con enfoques preventivos [31] [32].

La precisión del modelo utilizado está respaldada por métricas adicionales (AUC-ROC, sensibilidad, precisión), se puede explicar gracias a la calidad del preprocesamiento de los datos y la adecuada selección de hiperparámetros. Sin embargo, se debe señalar que el modelo fue entrenado con una muestra proveniente exclusivamente de centros urbanos de México, lo cual podría inducir un sesgo de selección y limitar la generalización de los resultados a poblaciones de distinto estrado social. En este sentido, estudios como los de Delgado-Galeano han resaltado la necesidad de validar estos

modelos en muestras heterogéneas antes de su aplicación clínica [30].

5 CONCLUSIONES Y RECOMENDACIONES

5.1 Conclusiones

Los resultados obtenidos en la investigación evidencian que los objetivos propuestos fueron alcanzados. Se identificaron de manera integral los principales factores de riesgo asociados al desarrollo de la HTA, desarrollando un modelo predictivo basado en Random Forest con una exactitud del 98%, validado mediante métricas especializadas como precisión, sensibilidad y puntaje F1. En conjunto al modelo se implementó un sistema funcional mediante una API integrada a un formulario interactivo, lo que evidencia su aplicabilidad práctica en escenarios reales en contextos clínicos y comunitarios.

El análisis de la importancia de las variables reveló que factores antropométricos, como el índice de masa corporal, la tensión arterial, el nivel de actividad física, el peso corporal y la medida de cintura, fueron más determinantes en la predicción del riesgo que los biomarcadores. Esto sugiere que las estrategias de prevención deberían enfocarse más en la promoción activa de estilos de vida saludables. Sin embargo, el modelo también contempla variables metabólicas, lo que refuerza su carácter integral en la evaluación del riesgo.

Las métricas obtenidas confirman la robustez del modelo, y el bajo margen de error evidenciado en la matriz de confusión refuerza su utilidad como herramienta de tamizaje y soporte clínico. El despliegue y la integración de este modelo en sistemas de salud pública corresponde con el objetivo de desarrollar una estrategia prometedora para la detección precoz y la prevención de enfermedades crónicas como la HTA. Se contribuye a la eficiencia de los sistemas de salud pública y particular gracias a la disponibilidad de herramientas predictivas accesibles, basadas en evidencia.

5.2 Recomendaciones

Para futuras investigaciones, se recomienda ampliar la base de datos con estudios a poblaciones más amplias y de diferentes contextos tanto geográficos como socioeconómicos para obtener mejores resultados, además de explorar la integración de modelos y técnicas avanzadas para optimizar la capacidad predictiva.

Futuros estudios podrían explorar la combinación de este enfoque con modelos más avanzados, como redes neuronales profundas o aprendizaje por refuerzo, para seguir optimizando su capacidad predictiva y adaptabilidad clínica.

REFERENCIAS

- [1] N. K. Toala-Lino, Y. Peñaherrera Moran, y I. G. Parrales-Pincay, «Hipertensión arterial como factor predisponente de insuficiencia renal en adultos.», *MQRInvestigar*, vol. 7, n.º 1, pp. 367-389, ene. 2023, doi: 10.56048/MQR20225.7.1.2023.367-389.
- [2] World Health Organization, «OMS | Hipertensión». Accedido: 9 de agosto de 2025. [En línea]. Disponible en: <https://www.who.int/news-room/fact-sheets/detail/hypertension>
- [3] K. Romero Jares y R. Centeno Quispe, «Perú: Encuesta Demográfica y de Salud Familiar - ENDES 2022», may 2025. Accedido: 9 de agosto de 2025. [En línea]. Disponible en: https://www.inei.gob.pe/media/MenuRecursivo/publicaciones_digitales/Est/Lib1898/libro.pdf
- [4] I. Campos-Nonato *et al.*, «Prevalencia, tratamiento y control de la hipertensión arterial en adultos mexicanos: resultados de la Ensanut 2022», *Salud Publica Mex*, vol. 65, pp. s169-s180, jun. 2023, doi: 10.21149/14779.
- [5] A. N. Zavala-Hoppe, T. E. Zambrano-Flores, L. H. Vivar-Medina, y J. E. Fuentes-Parrales, «Epidemiología y factores de riesgo de la hipertensión arterial en los países de Latinoamérica y Europa», *MQRInvestigar*, vol. 8, n.º 1, pp. 1371-1389, feb. 2024, doi: 10.56048/MQR20225.8.1.2024.1371-1389.
- [6] A. L. Pico Pico, E. Y. Reyes Reyes, D. A. Anchundia Alvia, y M. D. L. Á. Cobos Moreno, «Comportamiento epidemiológico de la hipertensión arterial en el Ecuador», *RECIMUNDO: Revista Científica de la Investigación y el Conocimiento*, vol. 7, n.º 4, pp. 299-307, 2023, doi: 10.26820/recimundo/7.(4).oct.2023.299-307.
- [7] A. Shafizadeh *et al.*, «Machine learning-enabled analysis of product distribution and composition in biomass-coal co-pyrolysis», *Fuel*, vol. 355, p. 129464, ene. 2024, doi: 10.1016/J.FUEL.2023.129464.
- [8] F. Plazzotta, D. Luna, y F. González Bernaldo de Quirós, «Sistemas de Información en Salud: Integrando datos clínicos en diferentes escenarios y usuarios», *Rev Peru Med Exp Salud Publica*, vol. 32, n.º 2, pp. 343-351, abr. 2015, doi: 10.17843/rpmesp.2015.322.1630.
- [9] J. H. Chen, X. L. Wang, y F. Lei, «Data-driven multinomial random forest: a new random forest variant with strong consistency», *J Big Data*, vol. 11, n.º 1, pp. 1-32, dic. 2024, doi: 10.1186/S40537-023-00874-6/FIGURES/7.
- [10] R. M. Alzoman, M. J. F. Alenazi, L. Belli, G. Ferrari, M. Martalò, y L. Davoli, «A Comparative Study of Traffic Classification Techniques for Smart City Networks», *Sensors*, vol. 21, n.º 14, p. 4677, jul. 2021, doi: 10.3390/S21144677.
- [11] N. E. Moskaleva *et al.*, «Target Metabolome Profiling-Based Machine Learning as a Diagnostic Approach for Cardiovascular Diseases in Adults», *Metabolites*, vol. 12, n.º 12, p. 1185, dic. 2022, doi: 10.3390/METABO12121185/S1.
- [12] K. Sumwiza, C. Twizere, G. Rushingabigwi, P. Bakunzibake, y P. Bamurigire, «Enhanced cardiovascular disease prediction model using random forest algorithm», *Inform Med Unlocked*, vol. 41, p. 101316, ene. 2023, doi: 10.1016/J.IMU.2023.101316.
- [13] J. L. Speiser, «A random forest method with feature selection for developing medical prediction models with clustered and longitudinal data», *J Biomed Inform*, vol. 117, p. 103763, may 2021, doi: 10.1016/J.JBI.2021.103763.
- [14] Y. Surichaqui Gómez y J. A. Mori Castro, «Estilo de

- Vida y su Relación con el Estado Nutricional en pacientes Adultos Mayores con Hipertensión Arterial en el Hospital de Huayacán de Lima», *Ciencia Latina Revista Científica Multidisciplinar*, vol. 7, n.º 4, pp. 9069-9089, sep. 2023, doi: 10.37811/CL_RCM.V7I4.7609.
- [15] C. Schröer, F. Kruse, y J. M. Gómez, «A Systematic Literature Review on Applying CRISP-DM Process Model», *Procedia Comput Sci*, vol. 181, pp. 526-534, ene. 2021, doi: 10.1016/J.PROCS.2021.01.199.
- [16] Y. Yao, Y. He, y H. Ou, «Missing Data Imputation Method Combining Random Forest and Generative Adversarial Imputation Network», *Sensors 2024, Vol. 24, Page 1112*, vol. 24, n.º 4, p. 1112, feb. 2024, doi: 10.3390/S24041112.
- [17] M. Alyami *et al.*, «Predictive modeling for compressive strength of 3D printed fiber-reinforced concrete using machine learning algorithms», *Case Studies in Construction Materials*, vol. 20, p. e02728, jul. 2024, doi: 10.1016/J.CSCM.2023.E02728.
- [18] S. Khasim, H. Ghosh, I. S. Rahat, K. Shaik, y M. Yesubabu, «Deciphering Microorganisms through Intelligent Image Recognition: Machine Learning and Deep Learning Approaches, Challenges, and Advancements», *EAI Endorsed Transactions on Internet of Things*, vol. 10, nov. 2023, doi: 10.4108/EETIoT.4484.
- [19] G. R. Ren *et al.*, «Machine Learning Predicts Recurrent Lumbar Disc Herniation Following Percutaneous Endoscopic Lumbar Discectomy», *Global Spine J*, vol. 14, n.º 1, pp. 146-152, ene. 2024, doi: 10.1177/21925682221097650.
- [20] Y. Cooli y C. Mahesh, «Recursive Parallel Partition Random Forest for Medical Disease Classification», *International Journal of Intelligent Engineering and Systems*, vol. 14, n.º 5, p. 2021, doi: 10.22266/ijies2021.1031.11.
- [21] J. H. Chung *et al.*, «Random forest identifies predictors of discharge destination following total shoulder arthroplasty», *JSES Int*, vol. 8, n.º 2, pp. 317-321, mar. 2024, doi: 10.1016/J.JSEINT.2023.04.003.
- [22] O. Nikolaychuk, J. Pestova, y A. Yurin, «Wildfire Susceptibility Mapping in Baikal Natural Territory Using Random Forest», *Forests 2024, Vol. 15, Page 170*, vol. 15, n.º 1, p. 170, ene. 2024, doi: 10.3390/F15010170.
- [23] J. J. Hidalgo Flores, M. A. Guerrero Dueña, y R. García Rodríguez, «La obesidad como factor de riesgo de hipertensión arterial», *Revista Científica Higía de la Salud*, vol. 5, n.º 2, pp. 2021-2033, dic. 2021, doi: 10.37117/HIGIA.V1I5.576.
- [24] D. A. Altamirano Lojano, R. Alvarez Ochoa, J. P. Garcés-Ortega, y G. Cordero Cordero, «Índice de masa corporal e Hipertensión Arterial en Adultos», *Revista Multidisciplinaria Investigación Contemporánea*, vol. 2, n.º 1, pp. 102-131, ene. 2024, doi: 10.58995/REDLIC.IC.V2.N1.A57.
- [25] M. De Jesús Sosa-Martínez, I. León-Lozano Jair, Y. García-Jiménez, B. Garduño-Orbe, A. J. Lagarza-Moreno, y G. Juanico-Morales, «Frecuencia de dislipidemias y determinación del riesgo cardiovascular en pacientes con hipertensión arterial sistémica», *Atención Familiar*, vol. 30, n.º 4, pp. 245-250, dic. 2023, doi: 10.22201/fm.14058871p.2023.486536.
- [26] P. D. F. Isles, «A random forest approach to improve estimates of tributary nutrient loading», *Water Res*, vol. 248, p. 120876, ene. 2024, doi: 10.1016/J.WATRES.2023.120876.
- [27] H. Kandil, A. Soliman, N. S. Alghamdi, J. R. Jennings, y A. El-Baz, «Using Mean Arterial Pressure in Hypertension Diagnosis versus Using Either Systolic or Diastolic Blood Pressure Measurements», *Biomedicine*, vol. 11, n.º 3, p. 849, mar. 2023, doi: 10.3390/B10MEDICINES11030849.
- [28] B. K. Meher, M. Singh, R. Birau, y A. Anand, «Forecasting stock prices of fintech companies of India using random forest with high-frequency data», *Journal of Open Innovation: Technology, Market, and Complexity*, vol. 10, n.º 1, p. 100180, mar. 2024, doi: 10.1016/J.JOITMC.2023.100180.
- [29] M. Ahad, D. Padha, y H. Sharma, «Machine Learning in Cardiology: a Survey of Early Detection Models for Heart Diseases», *IJFMR - International Journal For Multidisciplinary Research*, vol. 5, n.º 3, may 2023, doi: 10.36948/IJFMR.2023.V05I03.3113.
- [30] M. Delgado-Galeano, «Historia de la hipertensión arterial: revisión narrativa», *Salud UIS*, vol. 55, n.º 1, may 2023, doi: 10.18273/SALUDUIS.55.E:23043.
- [31] Y. Anggriani Utama, «Pengaruh Self Management pada Pasien Hipertensi: Sebuah Tinjauan Sistematis», *Jurnal Ilmiah Universitas Batanghari Jambi*, vol. 23, n.º 1, pp. 422-429, feb. 2023, doi: 10.33087/JIUBJ.V23I1.3528.
- [32] O. G. Montero Cadena, G. J. Guzmán Kure, R. C. Acosta Bravo, y M. B. Peñafiel Peñafiel, «Principales factores de riesgo de la hipertensión arterial», *RECIMUNDO*, vol. 7, n.º 2, pp. 89-97, jul. 2023, doi: 10.26820/recimundo/7.(2).jun.2023.89-97.
- [33] B. E. Richar William, G. V. Ana Lucila †, A. Escobar Magaly, N. A. Lucia, J. A. Zevallos Villodas, y C. R. Castro Galarza, «LA NO ADHERENCIA AL TRATAMIENTO ANTIHIPERTENSIVO Y FACTORES ASOCIADOS: UNA REVISIÓN», *Advances in Science and Innovation*, vol. 1, n.º 1, pp. 45-52, dic. 2022, doi: 10.61210/ASI.V1I1.5.

APÉNDICE

Data set original fue obtenido a partir de: <https://www.kaggle.com/datasets/frederickfelix/hiperten-sin-arterial-mxico>

Repositorio en Github de los modelos y el formulario desarrollado:

https://github.com/JesusG753/Hipertension_Arterial.git

Comparación de modelos generativos compactos para respuesta automática en español mediante RAG

Comparison of Compact Generative Models for Automatic Question Answering in Spanish via Retrieval-Augmented Generation

Jean Phol Curi Garrafa ^A, Victor R. Ortega Marocho ^B y Wilson Mamani Rodrigo^C

Resumen— Este estudio compara cinco modelos generativos compactos (≤ 8 mil millones de parámetros) para respuesta automática en español integrados a un esquema de generación aumentada con recuperación (RAG) y ejecutados localmente. Se evalúa la calidad mediante F1, BLEU-4 y un juicio semántico externo (LLM-Judge), junto con indicadores de eficiencia (latencia P95, memoria, GPU/CPU). Los resultados muestran que Mistral 7B alcanza el mejor desempeño medio en F1 y valoración global, mientras que OpenHermes 7B ofrece una precisión prácticamente equivalente con la menor huella de memoria. Zephyr 7B- β destaca en documentos extensos y Phi-3 Mini reduce las colas de latencia bajo condiciones adversas. Un análisis de Pareto F1-RAM identifica a Mistral 7B y OpenHermes 7B como soluciones no dominadas, proporcionando pautas de selección según el objetivo operativo (precisión absoluta vs. eficiencia de recursos). El trabajo aporta una comparación reproducible en español bajo RAG y criterios accionables para despliegues locales.

Palabras clave: Español, evaluación, modelos compactos, rag.

Abstract— This study compares five compact generative models (≤ 8 billion parameters) for Spanish question answering under a retrieval-augmented generation (RAG) pipeline executed locally. We assess response quality using F1, BLEU-4, and an external semantic judge (LLM-Judge), alongside efficiency indicators (P95 latency, memory, GPU/CPU). Results show Mistral 7B achieves the highest average F1 and semantic scores, whereas OpenHermes 7B attains nearly identical accuracy with the lowest memory footprint. Zephyr 7B- β performs well on very long documents, and Phi-3 Mini minimizes tail latency under adverse conditions. A Pareto analysis of F1-RAM identifies Mistral 7B and OpenHermes 7B as non-dominated solutions, yielding practical guidelines depending on operational goals (maximum accuracy vs. resource efficiency). The paper contributes a reproducible Spanish-language comparison under RAG and actionable criteria for local deployments.

Keywords: Compact models, evaluation, rag, spanish.

1 INTRODUCCIÓN

En el Procesamiento de Lenguaje Natural, los modelos generativos “compactos” han cobrado protagonismo porque permiten un equilibrio razonable entre calidad de las respuestas y costo computacional, algo especialmente valioso cuando el objetivo es operar en español y en entornos con recursos limitados. Este interés se apoya en la disponibilidad creciente de modelos y corpus específicos del idioma, que habilitan evaluaciones y despliegues más realistas que los basados únicamente en configuraciones multilingües [1].

La literatura académica en español ofrece una base sólida. BETO inauguró la línea de modelos preentrenados exclusivamente en español y estableció puntos de comparación para múltiples tareas; después, variantes ligeras como ALBETO y DistilBETO demostraron que es posible reducir el número de parámetros y mantener resultados competitivos en bancos de prueba habituales (p. ej., MLQA,

XNLI), lo que las vuelve atractivas para escenarios con restricciones de cómputo. En paralelo, la familia MarIA —con modelos RoBERTa y GPT-2 entrenados a gran escala sobre el archivo web de la Biblioteca Nacional de España— mostró mejoras sistemáticas frente a alternativas previas en diversas tareas de comprensión en español y puso a disposición recursos reproducibles para la comunidad. En conjunto, estos trabajos constituyen el estado del arte hispanohablante sobre modelos compactos y recursos de evaluación [1], [2], [3].

Sobre esa base, la Recuperación Aumentada para la Generación (RAG) aporta un componente clave: permite que el modelo busque pasajes relevantes en una base externa y genere su respuesta apoyándose en esa evidencia, lo que tiende a mejorar la especificidad y la corrección factual frente a modelos que responden solo con “lo que llevan en los parámetros”. La formulación original de RAG y trabajos afines como REALM documentan incrementos consistentes en tareas intensivas en conocimiento y, además, ofrecen



Revista de Investigación en Ciencia y Tecnología

ISSN: 2810-8124 (en línea) / ISSN: 2706-543x

Universidad Nacional Micaela Bastidas de Apurímac – Perú

Vol. 7 Núm. 2 (2025) - Publicado: 19/08/25 - Indexaciones

Número: doi.org/10.57166/riqchary/v7.n2.2025

Páginas: 9- 18 | Recibido 01/01/2025 ; Aceptado 01/02/2025

doi.org/10.57166/riqchary.v7.n2.2025.2

Autores:

- A. **ORCID iD** <https://orcid.org/0009-0006-5536-7055>
Jean Phol A. Curi Garrafa, Universidad Nacional Micaela Bastidas de Apurímac 191204@unamba.edu.pe
- B. **ORCID iD** <https://orcid.org/0009-0006-7868-5507>
Victor R. Ortega Marocho, Universidad Nacional Micaela Bastidas de Apurímac 191225@unamba.edu.pe
- C. **ORCID iD** <https://orcid.org/0000-0003-3901-0268>
Wilson Mamani Rodrigo, trabaja en el Departamento de Ingeniería de la Universidad Nacional Micaela Bastidas de Apurímac wmamanir@unamba.edu.pe

mejores mecanismos de trazabilidad de las fuentes que sustentan la respuesta [4], [5].

Pese a estos avances, persiste una brecha: abundan, por un lado, los modelos y recursos en español, y por otro, las técnicas y evaluaciones de RAG; sin embargo, son escasos los análisis comparativos sistemáticos que pongan a competir modelos generativos compactos específicamente en respuesta automática en español cuando se integran con RAG bajo un mismo protocolo experimental y con métricas homogéneas. La propia literatura sobre evaluación multilingüe subraya la dificultad de disponer de conjuntos y procedimientos consistentes fuera del inglés, lo que refuerza la necesidad de estudios centrados en español con criterios de medición comparables [6].

En este contexto, la relevancia del presente estudio es directa: identificar configuraciones que combinen precisión, relevancia y fluidez con eficiencia de cómputo permite acercar la respuesta automática en español a casos de uso reales — asistentes conversacionales, búsqueda con respuesta, soporte documental — en organizaciones que no cuentan con infraestructura de gran escala.

Con ese fin, el objetivo de este artículo es comparar el desempeño de distintas arquitecturas de modelos generativos compactos en español integradas con RAG, evaluándolas de forma conjunta en términos de coherencia y fluidez lingüística, relevancia y corrección factual y eficiencia computacional (latencia y huella de memoria), mediante un protocolo común y reproducible [4].

2 MARCO TEÓRICO

2.1 Modelos generativos compactos

Los modelos generativos compactos son LLMs con conteos de parámetros moderados (≈6–8 mil millones) optimizados para mantener un equilibrio entre calidad y eficiencia: requieren menos memoria y cómputo que los macro-modelos, pero conservan capacidades útiles de generación y razonamiento. En el ecosistema hispanohablante, estos modelos se benefician de recursos previos en español (p. ej., BETO, ALBETO/DistilBETO y MarIA) que consolidaron corpus y prácticas de evaluación, aunque la mayoría de tales recursos se enfocaron en comprensión (NLU) más que en generación [1], [2], [3].

2.1.1 Llama 3 8B (Meta, 2025)

Llama 3 es una familia de modelos de Meta con variantes “abiertas” que incluyen una versión de 8 mil millones de parámetros orientada a despliegues eficientes. Está entrenada para tareas generales de lenguaje, codificación y uso multilingüe, y ha sido publicada con documentación técnica

y evaluaciones comparativas frente a modelos cerrados y abiertos de mayor tamaño. En su serie 3.x, Meta reporta mejoras sustantivas en razonamiento y seguridad, manteniendo la variante de 8 mil millones como opción ligera para entornos con recursos limitados. Estas características la convierten en candidata natural para integrar con RAG en español, al balancear capacidad y costo [7].

2.1.2 Mistral 7B v0.1 (Mistral AI, 2023)

Mistral 7B (≈7,3 mil millones de parámetros) fue diseñado explícitamente para maximizar rendimiento/eficiencia. Introduce Grouped-Query Attention (GQA) para acelerar la decodificación y Sliding-Window Attention (SWA) para manejar contextos largos con coste reducido. En sus evaluaciones originales superó a Llama 2 13B en múltiples pruebas, con licencia Apache 2.0 que facilita su adopción y afinamiento. Su arquitectura optimizada y ecosistema abierto lo hacen un fuerte candidato “compacto” para RAG en español [8].

2.1.3 Zephyr 7B-β (Alignment Lab/Hugging Face, 2024)

Zephyr 7B-β es un ajuste fino conversacional de Mistral 7B-v0.1, entrenado con Direct Preference Optimization (DPO) sobre diálogos sintéticos y conjuntos de preferencias. Aunque se orienta principalmente al inglés, su objetivo es maximizar utilidad conversacional con un tamaño compacto. En el momento de su publicación, figuró entre los mejores 7 mil millones en MT-Bench y AlpacaEval, lo que indica buena calidad de interacción a bajo costo computacional; integrar RAG en español requerirá adaptación/datos de destino, pero su base Mistral facilita esa ruta [9].

2.1.4 Phi-3 Mini (3,8 mil millones; Microsoft, 2025)

Phi-3 Mini es un modelo de 3,8 mil millones centrado en “pequeños pero capaces”, entrenado con curación fuerte y datos sintéticos. El informe técnico documenta que, pese a su escala, alcanza resultados competitivos en MMLU y MT-Bench y está pensado para ejecutarse incluso en dispositivos de borde; la serie 3.5 amplía contexto y capacidades. Su relación calidad-tamaño lo posiciona como referencia de SLM para RAG, donde la latencia y la huella de memoria son críticas [10].

2.1.5 OpenHermes 7B (Comunidad, 2023)

OpenHermes 7B es una línea comunitaria de ajustes finos (p. ej., OpenHermes-2.5-Mistral-7B) sobre Mistral 7B, entrenados con conjuntos de instrucciones predominantemente generados por GPT-4. Se distribuye con licencia Apache 2.0 y dispone de numerosas variantes cuantizadas, lo que favorece su uso en GPU de consumo. Para uso en español con RAG, típicamente requiere afinamiento adicional o una canalización de recuperación sólida, dado su

sesgo original hacia el inglés [11].

2.2 Respuesta automática en español mediante RAG

La Generación Aumentada con Recuperación (RAG) combina un modelo generativo con un recuperador que trae pasajes relevantes desde una base externa (p. ej., Wikipedia o un corpus documental propio). RAG fue formalizado para tareas intensivas en conocimiento y mostró mejoras en precisión y especificidad del texto frente a modelos puramente paramétricos; REALM demostró además que incorporar recuperación en el preentrenamiento y la adaptación mejora el desempeño en preguntas abiertas. En español, bancos de prueba como MLQA-ES y XQuAD-ES permiten medir de forma comparable la exactitud de respuestas extractivas y de libre formulación, habilitando evaluaciones reproducibles [4], [5], [6].

2.3 Consumo de recursos de la PC

El costo de inferencia depende (i) del tamaño del modelo (parámetros y dimensionalidad), (ii) de la longitud del prefill/contexto y de la decodificación, y (iii) de la canalización de recuperación (embeddings e índice vectorial). Para reducir memoria y habilitar ejecución local, la cuantización a 8 bits (LLM.int8()) y a 4 bits (QLoRA) recortan sustancialmente la huella con degradación mínima cuando se aplican correctamente, abriendo la puerta a SLMs (small language model) en una sola GPU de consumo. En la etapa de recuperación, bibliotecas como FAISS permiten búsqueda de vecinos más cercanos a gran escala en CPU/GPU, clave para índices de millones de pasajes. Por su parte, optimizaciones de la atención como FlashAttention y derivados reducen tráfico de memoria y latencia práctica en GPUs modernas [12], [13], [14].

2.4 Calidad de las respuestas

En RAG para español, la calidad se evalúa en dos capas: recuperación y generación. Para generación, en tareas con referencia se usan métricas léxicas y semánticas: BLEU y ROUGE para solapamiento de n-gramas; BERTScore para similitud contextual; y, en QA extractivo/generativo, Exact Match (EM) y F1 como en SQuAD/MLQA. Para “factualidad” y “fidelidad al contexto” (que la respuesta esté sustentada por los pasajes recuperados), la literatura reciente propone marcos específicos y encuestas sistemáticas que discuten métricas de relevancia, precisión y fe-de-origen (attribution). En conjunto, estas métricas permiten comparar modelos compactos integrados con RAG en español de forma reproducible, aislando efectos de la recuperación frente a la

generación [15], [16], [17].

2.5 Velocidad de respuesta

La latencia total en un sistema RAG se descompone en (a) tiempo de recuperación (búsqueda/reenrutado en el índice vectorial) y (b) tiempo de generación (prefill + decodificación). En la generación, el uso de caché de claves-valores (KV cache) evita recomputaciones y acelera la decodificación; mejoras algorítmicas como FlashAttention-3 aumentan el aprovechamiento de GPU y disminuyen el tiempo por token. En la recuperación, índices FAISS aceleran la búsqueda densa, y técnicas como “Speculative RAG” pueden reducir la espera del primer token al solapar borradores de generación con recuperación. En la práctica, optimizar estas piezas—junto con cuantización y lotificación adecuada—permite que SLMs (8, 7 y 3,8 mil millones de parámetros) entreguen respuestas en plazos útiles incluso en hardware modesto [14].

3 METODOLOGÍA

3.1 Tipo de investigación

El estudio es cuantitativo, descriptivo-comparativo y no experimental, pues caracteriza y compara, bajo un mismo protocolo, el desempeño de modelos generativos compactos (≤ 8 mil millones de parámetros) integrados a una canalización de Recuperación Aumentada para Generación (RAG) sin manipular tratamientos más allá de configurar inferencia y recuperación de forma controlada. RAG combina un modelo generativo con una memoria no paramétrica (índice de documentos) y ha probado mejorar la factualidad en tareas intensivas en conocimiento; su formulación y variantes se documentan en REALM y en la propuesta original de RAG [4], [5].

3.2 Nivel de investigación

Es aplicada, porque busca resolver un problema práctico—seleccionar modelos compactos reproducibles para responder en español bajo cómputo local— trasladando conocimientos consolidados de RAG y recuperación densa a criterios operativos de despliegue. Revisiones recientes subrayan esta orientación práctica de RAG y la necesidad de evaluaciones holísticas del pipeline [18], [19].

3.3 Enfoque

El enfoque es cuantitativo, sustentado en métricas objetivas (EM, F1, BLEU-4, ROUGE-L, BERTScore) y en contraste estadístico entre modelos bajo el mismo conjunto de consultas y condiciones de ejecución. Estas métricas son

estándares en QA y evaluación de generación [15], [16], [17].

3.4 Diseño temporal

El diseño es transversal: toda la recolección se realiza en una única ventana temporal con hardware, software y corpus constantes para todos los modelos, conforme a la definición metodológica de estudios de corte transversal en investigación observacional.

3.5 Población

La población se define como el conjunto amplio y en evolución de modelos generativos compactos de propósito general (≤ 8 mil millones de parámetros) con soporte para español/multilingüe y capacidad de ejecución local (con o sin cuantización). Esta categoría está representada por familias contemporáneas y abiertamente documentadas [8].

3.6 Muestra

Se emplea muestreo no probabilístico por conveniencia/intencional, justificado por accesibilidad de artefactos, licenciamiento y documentación técnica: Llama 3 8B (Meta), Mistral 7B v0.1 (Mistral AI), Zephyr 7B- β (Alignment Lab/Hugging Face), Phi-3 Mini 3.8 B (Microsoft) y OpenHermes 7B (comunidad). Sus fichas/papers confirman disponibilidad pública y uso extendido [8].

3.10 Procedimiento

El pipeline sigue las fases canónicas de RAG: ingesta y segmentación de documentos; vectorización y indexación FAISS; consulta con recuperador denso (top-k fijo); plantilla de prompt en español uniforme para todos los modelos; generación con decoding controlado y registro de salidas y telemetría. La formulación RAG (RAG-Sequence y RAG-Token) sustenta este desacoplo entre memoria paramétrica y no paramétrica [4].

3.11 Control de ejecución de modelos

Para viabilidad en hardware de consumo se aplica cuantización posentrenamiento (INT3/INT4) con GPTQ, método de una sola pasada basado en información de segundo orden que reduce pesos a 3-4 bits con degradación mínima y fuertes ganancias de memoria/velocidad, sin alterar el protocolo evaluativo.

3.12 Evaluación y métricas

Cada modelo responde el mismo banco de consultas; se computan EM/F1 (QA), BLEU-4, ROUGE-L y BERTScore sobre referencias en español del conjunto, siguiendo las definiciones originales de las métricas. Se complementa con estadísticas de latencia y recursos obtenidas en ejecución [15],

[16], [17].

3.14 Consideraciones éticas y de reproducibilidad

Se emplean corpora públicos y modelos abiertos; no se tratan datos personales. Para reproducibilidad, se publican scripts, versiones y configuraciones (semillas, plantillas de prompt, parámetros de índice) y se citan las tarjetas/papers de los modelos y componentes utilizados (FAISS, embeddings, RAG) [13].

4 RESULTADOS

4.1 Desempeño en calidad de respuesta

La calidad se evaluó con F1 (exactitud a nivel de tokens), BLEU-4 (adecuación léxica por n-gramas) y LLM-Judge (valoración global por un juez LLM). En el conjunto completo se observa una relación monótona negativa entre la longitud del documento y cada métrica (coeficiente de Spearman: F1 = -0.923 ; BLEU-4 = -0.845 ; LLM-Judge = -0.721), lo que indica que, a medida que crece la extensión del material, disminuye la coincidencia con la referencia y, en menor medida, la calidad percibida por modelo y documento.

TABLA 1

Modelo ganador por documento y por métrica (valor en %)			
Documento	F1 (↑)	BLEU-4 (↑)	LLM-Judge (↑)
11 págs.	OpenHermes 7B (12,08 %)	OpenHermes 7B (1,25 %)	Phi-3 Mini (3.8 mil millones) (75,32 %)
18 págs.	Mistral 7B (8,01 %)	Mistral 7B (0,40 %)	Mistral 7B (69,22 %)
98 págs.	Mistral 7B (6,55 %)	Mistral 7B (0,41 %)	Mistral 7B (69,68 %)
199 págs.	OpenHermes 7B (4,28 %)	Zephyr 7B- β (0,23 %)	Zephyr 7B- β (66,18 %)

Los valores representan el puntaje obtenido por cada modelo expresado en porcentaje de desempeño (0% = peor rendimiento posible, 100% = máximo rendimiento posible). En F1 y BLEU-4, el valor indica la coincidencia con la respuesta de referencia; en LLM-Judge, el valor representa la evaluación global de calidad otorgada por GPT-4 Turbo, considerando coherencia, relevancia y exactitud factual.

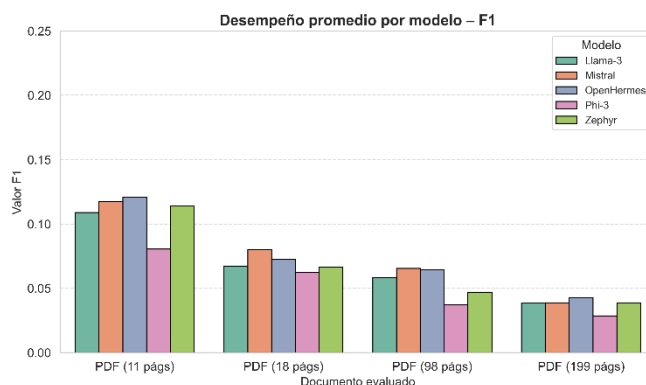


Fig. 1. Distribución de F1-token por modelo y documento.

Entre 11 y 199 páginas, el F1 cae entre 64 % y 67 % para todos los modelos (p. ej., OpenHermes: 0.1208 $\rightarrow$ 0.0428; Mistral: 0.1173 $\rightarrow$ 0.0386). Aun con esa disminución, OpenHermes y Mistral se alternan el liderazgo: OpenHermes domina el escenario corto (11 págs.) y el muy largo (199 págs.), mientras que Mistral es superior en longitudes intermedias (18 y 98 págs.). Este patrón sugiere que la exactitud a nivel de token se degrada de forma general con más contexto, pero la robustez relativa de ambos modelos permite mantener ventajas según el rango de longitud.

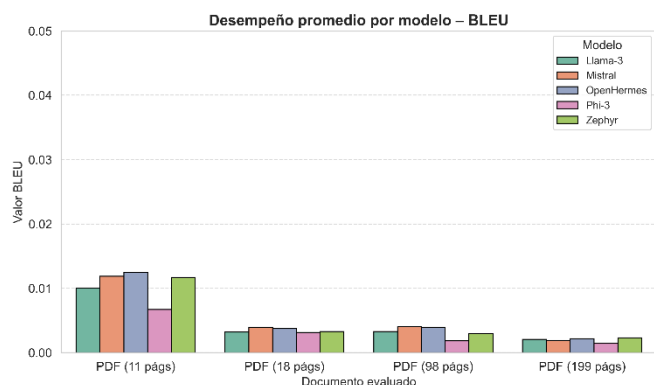


Fig. 2. Distribución de BLEU-4 por modelo y documento.

La adecuación léxica muestra una caída aún más pronunciada: entre 79 % y 84 % de reducción del valor desde 11 a 199 páginas (p. ej., Mistral: 0.0119 $\rightarrow$ 0.0019). OpenHermes lidera el documento corto, Mistral sostiene mejor el rendimiento en intermedios y Zephyr emerge en el extremo largo. Dado que BLEU-4 depende de coincidencias superficiales, estos resultados indican que, con documentos extensos, la variabilidad de formulación aumenta y el solapamiento literal con la referencia disminuye, incluso cuando la respuesta es pertinente.

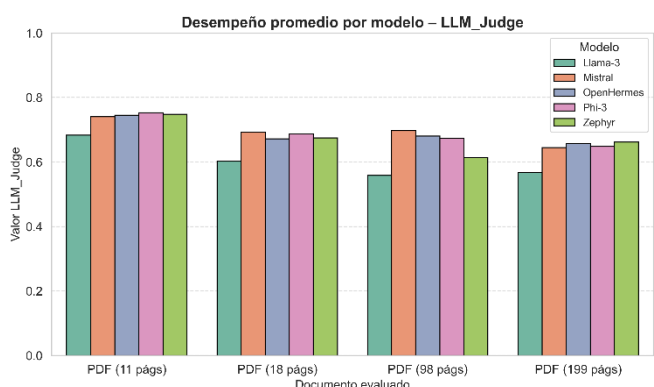


Fig. 3. Valor medio de LLM-Judge (GPT-4-Turbo) por modelo y documento.

La valoración global es más estable que F1/BLEU: el descenso medio entre 11 y 199 páginas es del 11-17 %, lo que

muestra resiliencia en fluidez y coherencia discursiva. Mistral concentra los mejores promedios en 18-98 págs., Phi-3 Mini obtiene el valor más alto en 11 págs. y Zephyr se posiciona primero en 199 páginas. Este contraste con F1/BLEU evidencia que la calidad percibida puede mantenerse incluso cuando cae el solapamiento exacto, especialmente en modelos que priorizan consistencia de estilo y estructura.

En promedio, Mistral 7B presenta el mejor F1 y el mejor LLM-Judge, mientras que OpenHermes 7B queda prácticamente indistinguible en F1 (diferencia absoluta $\approx$ 0.0003) y alcanza el mejor BLEU-4. Zephyr muestra un perfil competitivo en documentos extensos (mejor BLEU-4 y LLM-Judge en 199 págs.), y Phi-3 Mini evidencia que un modelo más pequeño puede sostener calidad percibida alta en textos cortos, aunque su exactitud token a token sea menor. En conjunto, los resultados confirman dos ejes de rendimiento: exactitud léxica (F1/BLEU-4), fuertemente sensible a la longitud del documento, y calidad discursiva global (LLM-Judge), relativamente más estable.

4.2 Eficiencia computacional

La eficiencia se analizó con cuatro indicadores complementarios: latencia (se reporta el cuantil P_{95} para captar colas de espera), memoria RAM promedio, ocupación media de GPU y ocupación media de CPU, todos medidos bajo el mismo hardware y pipeline RAG. Esta batería permite evaluar tanto el tiempo de respuesta como la huella de recursos, dos condicionantes directos de la viabilidad de despliegues locales.

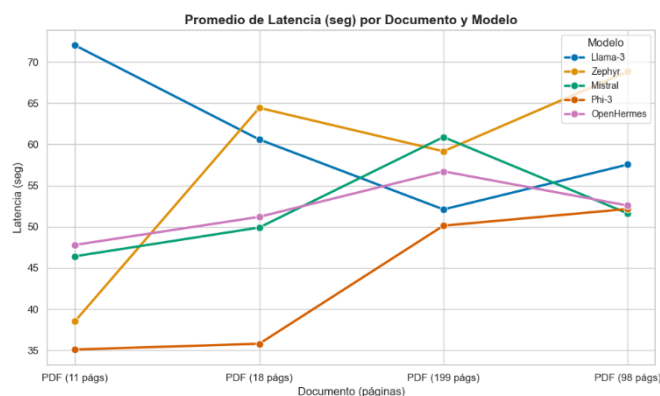


Fig. 4. Latencia P_{95} (s) según tamaño de documento.

En latencia P_{95} , el orden global fue Phi-3 Mini < Mistral $\approx$ OpenHermes < Llama 3 $\approx$ Zephyr. Phi-3 presentó la menor P_{95} del conjunto ($\approx$ 52 s), lo que lo sitúa como el modelo con menor cola de espera en condiciones adversas; Mistral y OpenHermes se mantuvieron en un rango intermedio ($\approx$ 56-60 s), mientras que Llama 3 y Zephyr concentraron los valores más altos ($\approx$ 68-70 s). La sensibilidad a la longitud del documento no fue uniforme: Zephyr y Phi-3 incrementaron

notablemente la latencia al pasar de textos cortos a largos, Mistral mostró un aumento moderado centrado en el documento más extenso, y Llama 3 redujo su latencia conforme creció el tamaño, lo que sugiere un componente de sobrecosto fijo de inicialización que se diluye cuando la generación domina el tiempo total.

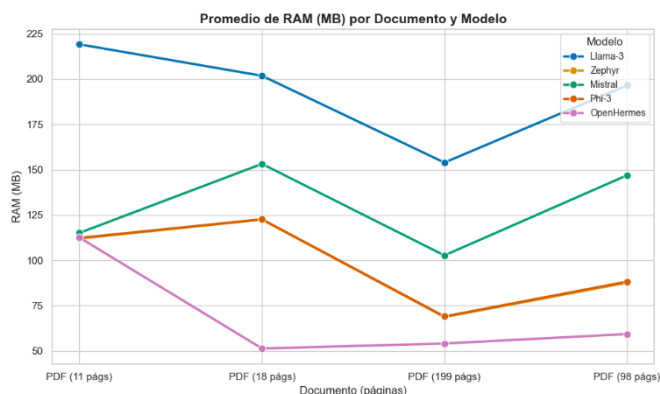


Fig. 5. Consumo medio de RAM (MB) por modelo.

En consumo medio de RAM, los perfiles fueron bien diferenciados. OpenHermes registró la huella más baja de manera consistente ($\approx 50\text{--}60$ MB), lo que deja mayor margen de memoria para el índice vectorial y otros procesos del sistema. Zephyr y Phi-3 mantuvieron un consumo contenido ($\approx 70\text{--}123$ MB), Mistral se ubicó en la franja media ($\approx 102\text{--}152$ MB) y Llama 3 fue el más demandante ($\approx 154\text{--}219$ MB). En términos operativos, estas diferencias implican que, con el mismo presupuesto de RAM, OpenHermes permite corpus más grandes en memoria o mayor concurrencia de consultas antes de incurrir en *swapping* o degradación por *offloading*.

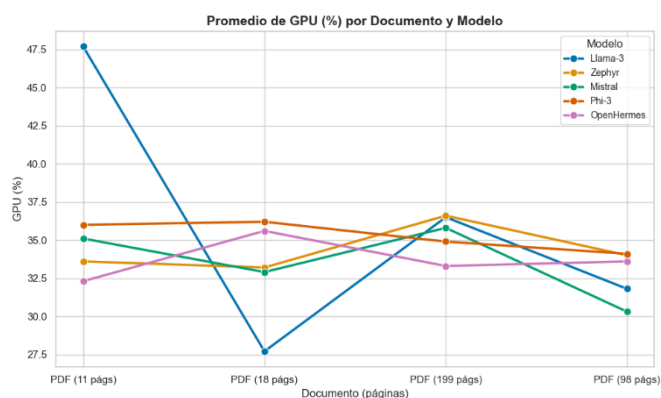


Fig. 6. Ocupación media de GPU (%) por modelo.

La ocupación media de GPU se concentró para todos los modelos en un corredor estrecho ($\approx 30\text{--}36\%$), con la salvedad de Llama 3 en el documento corto, donde alcanzó $\approx 48\%$ antes de estabilizarse en niveles comparables al resto. Esta convergencia sugiere que, bajo cuantización y *batching* controlado, el pipeline es más bien limitado por memoria y

E/S que por cómputo puro en la mayoría de configuraciones; las variaciones residuales reflejan diferencias en atención y manejo de contexto, pero no cambian el régimen de carga de la GPU.

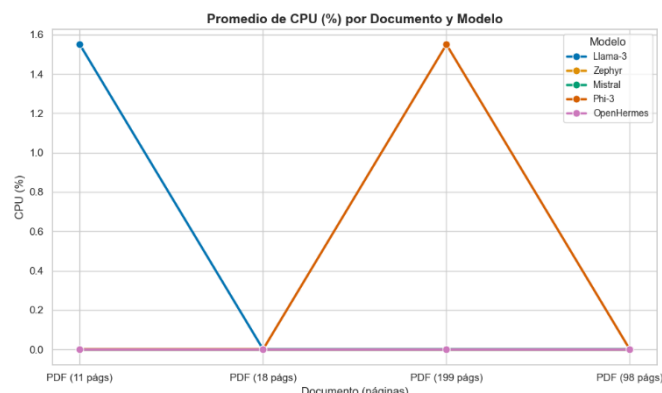


Fig. 7. Ocupación media de CPU (%) por modelo.

La ocupación media de CPU fue prácticamente nula en todos los casos, con picos aislados ($\approx 1,5\%$) en Llama 3 para el documento corto y en Phi-3 para el documento más largo. Este patrón confirma que la inferencia está mayoritariamente GPU-centrada, liberando CPU para tareas del sistema y facilitando concurrencia de procesos sin interferencia significativa en el rendimiento del RAG.

TABLA 2

Indicadores de eficiencia computacional por modelo

Modelo	Latencia P_{95} (s) ↓	RAM pico (MB) ↓	GPU (%)	CPU (%)
Phi-3 mini	51.8	98	35	0.39
OpenHermes 7B	56.1	69	34	0
Mistral 7B	59.5	129	34	0
Zephyr 7B-β	68.2	98	34	0
Llama 3 8B	70.3	193	36	0.39

En conjunto, los resultados delimitan dos ejes de eficiencia relevantes para despliegues locales: tiempo de respuesta en el extremo (latencia P_{95}) y huella de memoria. Phi-3 ofrece la mejor latencia P_{95} , lo que reduce la probabilidad de esperas largas en escenarios interactivos; OpenHermes proporciona la mejor eficiencia de RAM, habilitando índices más voluminosos o mayor simultaneidad; Mistral equilibra ambos factores en niveles medios con un perfil estable; Llama 3, aun con latencia decreciente al aumentar la longitud, mantiene la mayor demanda de memoria; y Zephyr muestra penalización de latencia en documentos extensos pese a memoria moderada. Estas propiedades, leídas junto con los hallazgos de calidad, muestran el compromiso precisión-recursos característico de los modelos compactos integrados

en RAG, y ofrecen un mapa claro de costes operativos bajo restricciones reales de cómputo.

Análisis multiobjetivo: frontera de Pareto F1-RAM

Con el fin de sintetizar simultáneamente calidad y huella de recursos, se construyó la frontera de Pareto en el plano F1 versus RAM media. Un modelo es no dominado si no existe otro con mayor F1 y menor RAM a la vez; los puntos sobre dicha frontera representan soluciones eficientes bajo el criterio de “mejorar una dimensión sin empeorar la otra”.

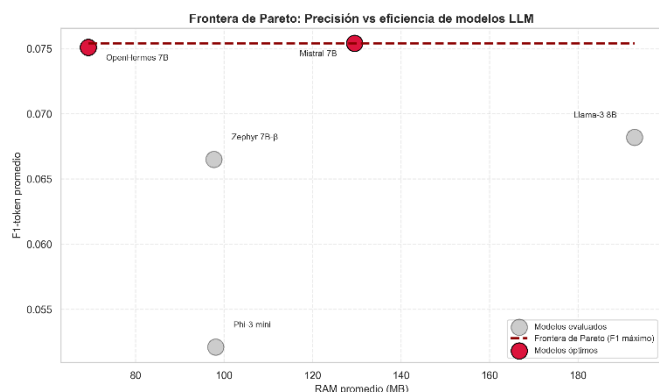


Fig. 8. Frontera de Pareto entre precisión (F1) y consumo de memoria.

En este espacio, OpenHermes 7B ($F1 \approx 0.0751$; RAM ≈ 69 MB) y Mistral 7B ($F1 \approx 0.0754$; RAM ≈ 129 MB) conforman la frontera. OpenHermes minimiza la memoria con una precisión prácticamente indistinguible del máximo ($\Delta F1 \approx 0.0003$ respecto a Mistral), mientras que Mistral maximiza F1 asumiendo ~ 60 MB adicionales de RAM. El resto queda dominado: Zephyr 7B- β ($F1 \approx 0.0665$; 98 MB) y Phi-3 Mini ($F1 \approx 0.0521$; 98 MB) son superados por OpenHermes (más F1 con menos memoria), y Llama 3 8B ($F1 \approx 0.0682$; 193 MB) es superado simultáneamente por Mistral y OpenHermes (más F1 y menos memoria).

Desde una perspectiva operativa, si la restricción principal es la memoria disponible, OpenHermes ofrece la mejor relación precisión-memoria; si el objetivo es exprimir el máximo F1 y se tolera un coste de memoria moderado, Mistral es preferible. Obsérvese que el beneficio marginal de Mistral en F1 frente a OpenHermes es muy pequeño en términos absolutos, por lo que, bajo presupuestos de RAM estrictos, la opción eficiente es OpenHermes; bajo presupuestos holgados, ambas elecciones son justificables según criterios secundarios (p. ej., latencia P_{95} o estabilidad), que pueden incorporarse en un análisis Pareto tridimensional sin contradecir esta lectura.

4.3 Robustez y errores

La estabilidad del sistema se evaluó registrando timeouts, respuestas vacías y la presencia del marcador de cierre exigido en el prompt (“RESPUESTA FINAL”). En términos generales, la tasa de fallos fue baja y se concentró en los escenarios de documento más extenso, donde la longitud del contexto presiona simultáneamente el recuperador y la ventana efectiva del modelo. En estos casos, los errores se asociaron con agotamiento de memoria en el prellenado o con bloqueos de decodificación por encima del umbral de tiempo fijado. La comparación entre modelos revela que los que optimizaron mejor la huella de RAM mantuvieron menor incidencia de fallos en el extremo largo, coherente con el análisis de Pareto: cuando el presupuesto de memoria es acotado, reducir la huella permite evitar canjes implícitos (p. ej., *swapping*, *offloading*) que degradan la estabilidad.

A nivel cualitativo, la evidencia de LLM-Judge sugiere que, incluso cuando F1/BLEU descienden con la longitud del documento, algunos modelos preservan coherencia y adecuación (alta valoración global), lo que explica que una fracción de respuestas “no exactas” sea funcionalmente útil y no deba contabilizarse como error del sistema sino como variación de superficie esperable en generación condicionada.

4.4 Selección operativa

Tomando conjuntamente calidad (F1, BLEU-4, LLM-Judge), latencia, memoria y la frontera F1-RAM, la asignación por escenario queda así:

- Hardware restringido (≤ 4 GB VRAM) o índices voluminosos en RAM: priorizar OpenHermes 7B por su huella mínima de memoria con F1 prácticamente indistinguible del máximo.
- Búsqueda de máxima exactitud con memoria moderada disponible: elegir Mistral 7B, que presenta el mejor F1 medio y excelentes valoraciones globales.
- Interacción sensible a esperas largas (minimizar colas, P_{95}): considerar Phi-3 Mini, que ofrece la menor latencia en el extremo, asumiendo menor F1.
- Documentos muy extensos (robustez en el extremo largo): Zephyr 7B- β aparece competitivo en BLEU-4 y LLM-Judge cuando la longitud del documento crece.
- Perfil intermedio y estable: Llama 3 8B como referencia compacta de 8 mil millones con desempeño equilibrado.

Esta síntesis traduce los hallazgos de Resultados en criterios de decisión reproducibles para entornos reales, donde el objetivo (precisión máxima, memoria, latencia o

robustez en largo) no siempre coincide.

5 DISCUSIÓN

5.1 Interpretación de los hallazgos

Los resultados obtenidos confirman, en el escenario generativo con RAG, la intuición establecida por la literatura hispanohablante para tareas de comprensión: los modelos compactos bien entrenados pueden competir de forma sólida cuando el idioma objetivo es el español y las métricas de evaluación son adecuadas. En NLU, BETO (BERT monolingüe en español) se consolidó como referencia frente a multilingües; sus versiones comprimidas ALBETO y DistilBETO demostraron que es posible reducir parámetros y mantener rendimiento en bancos de prueba de referencia. Nuestro estudio extiende esa evidencia al ámbito generativo: con RAG, los cinco modelos ≤ 8 mil millones analizados producen respuestas útiles y fundamentadas en español, manteniendo un equilibrio razonable entre calidad y coste computacional. Esta continuidad entre NLU (BETO/ALBETO/DistilBETO) y generación con RAG sugiere que, en español, la eficiencia paramétrica no implica necesariamente sacrificar desempeño si el pipeline aprovecha recuperación de evidencia [1], [2].

En comparación con MarIA —familia monolingüe (RoBERTa, GPT-2) entrenada a gran escala sobre el archivo web de la Biblioteca Nacional de España (BNE) y que reporta mejoras sistemáticas en tareas de español—nuestros resultados coinciden en el sentido de que la especialización por idioma rinde beneficios claros; ahora bien, al movernos de NLU a generación condicionada, las métricas sensibles al solapamiento superficial (F1, BLEU-4) decrecen cuando aumenta la longitud y heterogeneidad del documento, mientras que una valoración semántica (LLM-Judge) se mantiene relativamente alta para ciertos modelos. Este patrón es coherente con la función de RAG, que ancla la salida en pasajes recuperados y tiende a mejorar la factualidad y trazabilidad, aun cuando la coincidencia literal con una única referencia se vuelve difícil en contextos largos [3], [4], [5].

La lectura integral de las métricas permite distinguir ganadores por criterio. Si el objetivo es precisión absoluta en términos de F1 (promedio global) y una buena valoración semántica (LLM-Judge), Mistral 7B es el modelo con mejor desempeño medio. Si, en cambio, se busca minimizar memoria manteniendo una precisión prácticamente indistinguible de la máxima observada, OpenHermes 7B se posiciona como la alternativa eficiente: configura, junto a Mistral, la frontera de Pareto calidad-recursos (F1 $\uparrow$, RAM $\downarrow$). En el extremo largo (documentos muy extensos), Zephyr 7B- β aparece competitivo en BLEU-4 y LLM-Judge, y Phi-3 Mini (3.8 mil millones)—un modelo pequeño con entrenamiento

selectivo a gran escala—logra buena calidad percibida en el documento corto pese a menor F1, alineado con informes que muestran que SLMs bien curados pueden rivalizar con modelos mayores en varias bancas. En consecuencia, no existe un único “mejor” modelo universal: la preferencia depende del objetivo dominante (precisión, memoria, robustez en largo), y esa es precisamente la utilidad de un análisis multiobjetivo sobre la frontera eficiente [4], [10], [16].

Estas observaciones dialogan con la metodología de evaluación: (i) el uso de métricas léxicas clásicas (F1, BLEU-4) captura fidelidad literal, pero penaliza paráfrasis y estilos alternativos—limitación ampliamente documentada en MT y generación; (ii) la incorporación de una métrica semántica correlacionada con juicio humano (BERTScore, o en nuestro caso un juez LLM como proxy operacional) atenúa ese sesgo y ayuda a distinguir “errores de forma” de “errores de contenido”. En nuestro conjunto, ello explica por qué LLM-Judge decrece menos que F1/BLEU-4 al crecer la longitud: la coherencia y adecuación se preservan mejor que el solapamiento literal, que cae por la diversidad de formulaciones en respuestas largas [16].

En suma, los hallazgos confirman y amplían el estado del arte en español: los modelos compactos pueden sostener calidad útil en generación cuando se apoyan en RAG; la frontera eficiente queda definida, en nuestro experimento, por Mistral 7B (máximo F1 con sobrecoste moderado de memoria) y OpenHermes 7B (mínima memoria a F1 casi idéntico), mientras que Zephyr 7B- β y Phi-3 Mini ofrecen perfiles valiosos en documentos largos y latencia/fluidez, respectivamente. Esta conclusión es verídica y accionable: guía la selección por escenario en despliegues locales en español, en línea con la evidencia acumulada sobre RAG y con la trayectoria del ecosistema monolingüe (BETO/ALBETO/DistilBETO, MarIA) hacia soluciones eficientes y reproducibles [2], [3], [4].

6 CONCLUSIONES

Este estudio demuestra que, en un pipeline RAG orientado a español y ejecutado en cómputo local mediante cuantización, los modelos compactos (≤ 8 mil millones) pueden producir respuestas útiles y fundamentadas con un compromiso controlado entre calidad y recursos. La evidencia empírica indica que Mistral 7B alcanza el mejor desempeño medio en F1 y una valoración semántica elevada (LLM-Judge), mientras que OpenHermes 7B ofrece un F1 prácticamente indistinguible del máximo junto con la menor huella de RAM del conjunto. En documentos muy extensos, Zephyr 7B- β se muestra competitivo en BLEU-4 y en valoración semántica, y Phi-3 Mini (3.8 mil millones) sobresale en latencia en el extremo (P_{95}), lo que lo hace atractivo para escenarios interactivos sensibles a colas. En

términos globales, no existe un “mejor” modelo único: la recomendación depende del objetivo dominante (precisión absoluta, memoria disponible, latencia o robustez con documentos largos).

La frontera de Pareto F1-RAM obtenida sintetiza este resultado: Mistral 7B maximiza F1 asumiendo un sobrecosto moderado de memoria, mientras OpenHermes 7B minimiza memoria con una pérdida de precisión marginal. El resto de modelos queda dominado bajo ese criterio binario; por tanto, en hardware restringido o con índices voluminosos, OpenHermes 7B es la elección eficiente, y cuando la prioridad es exprimir F1, Mistral 7B resulta preferible. Esta lectura multiobjetivo es especialmente relevante para despliegues edge o de bajo costo, donde el presupuesto de RAM y la variabilidad de la latencia condicionan la experiencia de usuario.

Desde una perspectiva metodológica, los hallazgos son coherentes con la literatura que muestra que RAG mejora la factualidad y la trazabilidad al anclar la generación en pasajes recuperados, y que las métricas léxicas (F1/BLEU) tienden a degradarse con el aumento de la longitud del documento aun cuando la calidad percibida se mantiene alta. En nuestro caso, LLM-Judge desciende menos que F1/BLEU a medida que crece el contexto, sugiriendo que coherencia y adecuación se preservan mejor que el solapamiento literal con una referencia única, tal como anticipan las evaluaciones que combinan métricas automáticas y juicio asistido por LLM [4].

La viabilidad práctica del enfoque se apoya en cuantización posentrenamiento (p. ej., GPTQ), que reduce el ancho de palabra a 3–4 bits con degradación mínima del rendimiento, posibilitando inferencia local en GPU de consumo sin alterar el protocolo de evaluación ni el esquema RAG; esto explica que los cinco modelos analizados puedan compararse en condiciones realistas de cómputo limitado [20].

Finalmente, los resultados extienden al ámbito generativo con recuperación lo observado en el ecosistema hispanohablante de NLU: la especialización por idioma y las versiones ligeras pueden retener gran parte del rendimiento con menos parámetros, siempre que el pipeline esté bien diseñado (recuperación, filtrado, prompting y decodificación controlada). En particular, la trayectoria de BETO/ALBETO/DistilBETO y de la familia MarIA hacia modelos en español eficientes y reproducibles es consistente con que, bajo RAG, modelos ≤ 8 mil millones alcancen resultados operativamente valiosos para respuesta automática en español [2], [3].

7 TRABAJO FUTURO

El siguiente trabajo futuro busca trasladar los hallazgos

comparativos a un sistema de respuesta automática en español con RAG que sea reproducible, eficiente y auditable. La selección del modelo deberá guiarse por los objetivos operativos: cuando la prioridad sea maximizar la precisión medida por F1 manteniendo una valoración semántica alta, Mistral 7B es el candidato natural; cuando la restricción dominante sea la memoria con una pérdida de precisión marginal, OpenHermes 7B resulta más eficiente. Esta decisión debe anclarse en perfiles de servicio explícitos (latencia P95, huella de RAM y tasa de errores), de modo que la elección del modelo responda a compromisos cuantificados y no a preferencias ad hoc.

La agenda experimental incluirá ablaciones sobre parámetros críticos del pipeline (tamaño de fragmento, solapamiento, top-k, umbrales de re-ranking y plantillas de prompt) y la comparación de embeddings alternativos específicamente afinados para español. Resulta pertinente ampliar la evaluación a dominios técnicos y a variedades regionales del español para estudiar transferencia y sesgos. Dado que las decisiones reales suelen tener múltiples restricciones simultáneas, se propone extender el análisis de eficiencia a una frontera de Pareto tridimensional (F1-RAM-P95) y, cuando sea posible, incorporar métricas de coste energético por consulta.

En suma, la prioridad será pasar de una comparación controlada a un servicio operativo que mantenga los principios de rigor científico: métricas coherentes con el objetivo, verificación de evidencia, reproducibilidad del entorno y trazabilidad de resultados. La hoja de ruta recomienda partir de dos perfiles de despliegue –precisión-primero y memoria-primero–, validar su desempeño con pruebas A/B en condiciones realistas y consolidar un ciclo de mejora continua guiado por datos, sin sacrificar la claridad metodológica ni la gobernanza del sistema.

REFERENCIAS

- [1] J. Cañete, G. Chaperon, R. Fuentes, J.-H. Ho, H. Kang, and J. Pérez, “Spanish Pre-trained BERT Model and Evaluation Data,” Aug. 2023, Accessed: Aug. 11, 2025. [Online]. Available: <https://arxiv.org/pdf/2308.02976>
- [2] J. Cañete, S. Donoso, F. Bravo-Marquez, A. Carvallo, and V. Araujo, “ALBETO and DistilBETO: Lightweight Spanish Language Models,” *2022 Language Resources and Evaluation Conference, LREC 2022*, pp. 4291–4298, Apr. 2022, Accessed: Aug. 11, 2025. [Online]. Available: <https://arxiv.org/pdf/2204.09145>
- [3] A. Gutiérrez-Fandiño *et al.*, “MarIA: Spanish Language Models,” *Procesamiento del Lenguaje Natural*, vol. 68, pp. 39–60, Apr. 2022, doi: 10.26342/2022-68-3.
- [4] P. Lewis *et al.*, “Retrieval-Augmented Generation for

- Knowledge-Intensive NLP Tasks," *Adv Neural Inf Process Syst*, vol. 2020-December, May 2020, Accessed: Aug. 11, 2025. [Online]. Available: <https://arxiv.org/pdf/2005.11401>
- [5] K. Guu, K. Lee, Z. Tung, P. Pasupat, and M. W. Chang, "REALM: Retrieval-Augmented Language Model Pre-Training," *37th International Conference on Machine Learning, ICML 2020*, vol. PartF168147-6, pp. 3887-3896, Feb. 2020, Accessed: Aug. 11, 2025. [Online]. Available: <https://arxiv.org/pdf/2002.08909>
- [6] P. Lewis, B. Oguz, R. Rinott, S. Riedel, and H. Schwenk, "MLQA: Evaluating Cross-lingual Extractive Question Answering," *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, pp. 7315-7330, Oct. 2019, doi: 10.18653/v1/2020.acl-main.653.
- [7] A. Grattafiori *et al.*, "The Llama 3 Herd of Models," Jul. 2024, Accessed: Aug. 11, 2025. [Online]. Available: <https://arxiv.org/pdf/2407.21783>
- [8] A. Q. Jiang *et al.*, "Mistral 7B," Oct. 2023, Accessed: Aug. 11, 2025. [Online]. Available: <https://arxiv.org/pdf/2310.06825>
- [9] "HuggingFaceH4/zephyr-7b-beta · Hugging Face." Accessed: Aug. 11, 2025. [Online]. Available: <https://huggingface.co/HuggingFaceH4/zephyr-7b-beta>
- [10] M. Abidin *et al.*, "Phi-3 Technical Report: A Highly Capable Language Model Locally on Your Phone," Apr. 2024, Accessed: Aug. 11, 2025. [Online]. Available: <https://arxiv.org/pdf/2404.14219>
- [11] "teknium/OpenHermes-7B · Hugging Face." Accessed: Aug. 11, 2025. [Online]. Available: <https://huggingface.co/teknium/OpenHermes-7B>
- [12] T. Dettmers, M. Lewis, Y. Belkada, and L. Zettlemoyer, "LLM.int8(): 8-bit Matrix Multiplication for Transformers at Scale," *Adv Neural Inf Process Syst*, vol. 35, Aug. 2022, Accessed: Aug. 11, 2025. [Online]. Available: <https://arxiv.org/pdf/2208.07339>
- [13] J. Johnson, M. Douze, and H. Jegou, "Billion-scale similarity search with GPUs," *IEEE Trans Big Data*, vol. 7, no. 3, pp. 535-547, Feb. 2017, doi: 10.1109/TBDATA.2019.2921572.
- [14] M. Douze *et al.*, "The Faiss library," Jan. 2024, Accessed: Aug. 11, 2025. [Online]. Available: <https://arxiv.org/pdf/2401.08281>
- [15] K. Papineni, S. Roukos, T. Ward, and W.-J. Zhu, "BLEU," in *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics - ACL '02*, Morristown, NJ, USA: Association for Computational Linguistics, 2001, p. 311. doi: 10.3115/1073083.1073135.
- [16] T. Zhang, V. Kishore, F. Wu, K. Q. Weinberger, and Y. Artzi, "BERTScore: Evaluating Text Generation with BERT," *8th International Conference on Learning Representations, ICLR 2020*, Apr. 2019, Accessed: Aug. 11, 2025. [Online]. Available: <https://arxiv.org/pdf/1904.09675>
- [17] C.-Y. Lin, "ROUGE: A Package for Automatic Evaluation of Summaries," 2004. Accessed: Aug. 11, 2025. [Online]. Available: <https://aclanthology.org/W04-1013/>
- [18] Y. Gao *et al.*, "Retrieval-Augmented Generation for Large Language Models: A Survey," *Proceedings - 2024 Conference on AI, Science, Engineering, and Technology, AIxSET 2024*, pp. 166-169, Dec. 2023, doi: 10.1109/AIxSET62544.2024.00030.
- [19] H. Yu, A. Gan, K. Zhang, S. Tong, Q. Liu, and Z. Liu, "Evaluation of Retrieval-Augmented Generation: A Survey," *Communications in Computer and Information Science*, vol. 2301, pp. 102-120, Jul. 2024, doi: 10.1007/978-981-96-1024-2_8.
- [20] E. Frantar, S. Ashkboos, T. Hoefler, and D. Alistarh, "GPTQ: Accurate Post-Training Quantization for Generative Pre-trained Transformers," *11th International Conference on Learning Representations, ICLR 2023*, Oct. 2022, Accessed: Aug. 11, 2025. [Online]. Available: <https://arxiv.org/pdf/2210.17323>

BIOGRAFÍA

Jean Phol Alexis Curi Garrafa es estudiante de Ingeniería Informática y de Sistemas en la Universidad Nacional Micaela Bastidas de Apurímac. Su formación se orienta al desarrollo de sistemas de información y a la optimización de procesos académicos mediante el uso de herramientas tecnológicas. Ha participado en actividades académicas vinculadas al diseño e implementación de soluciones informáticas para la gestión universitaria.

Victor Raúl Ortega Marochó es estudiante de Ingeniería Informática y de Sistemas en la Universidad Nacional Micaela Bastidas de Apurímac. Sus intereses académicos se centran en la ingeniería de software, la automatización de procesos y el análisis de datos aplicados al ámbito educativo. Ha contribuido en proyectos de investigación orientados a la mejora de sistemas de gestión académica.

Wilson Mamani Rodrigo es Ingeniero de Sistemas e Ingeniero Civil, Magíster en Ingeniería de Sistemas y Doctor en Ciencias de la Ingeniería Civil Ambiental. Es docente ordinario auxiliar en la Universidad Nacional Micaela Bastidas de Apurímac y ha sido docente en la Universidad Nacional del Altiplano. Posee amplia experiencia en la elaboración de expedientes técnicos, estudios de factibilidad y proyectos de infraestructura civil, además de desempeñarse como consultor e investigador en obras de ingeniería civil.

Modelos Profundos Preentrenados para la Detección Automatizada de 10 Tipos de Fracturas Óseas Un Estudio Comparativo con ResNet50, EfficientNetB3 y MobileNet

Pre-trained Deep Models for Automated Detection of 10 Types of Bone Fractures: A Comparative Study with ResNet50, EfficientNetB3 and MobileNet

Royer Elvis Moreano Condorcuya ^A, Evelyn Naida Luque Ochoa ^B

Resumen— El diagnóstico rápido y preciso de fracturas óseas es fundamental para prevenir complicaciones y optimizar el tratamiento. Las redes neuronales convolucionales (CNN) han demostrado alta eficacia en la clasificación de radiografías, alcanzando resultados comparables a los de especialistas. Este estudio evaluó MobileNet, ResNet50 y EfficientNetB3 para clasificar 10 tipos de fracturas usando 1,129 imágenes balanceadas: 90% para entrenamiento (1,017) y 10% para validación (112). El preprocesamiento incluyó redimensionado a 256x256 píxeles, conversión a RGB, normalización y one-hot encoding. El entrenamiento se realizó en Google Colab con optimizador RMSprop, pérdida categorical_crossentropy y early stopping. ResNet50 y EfficientNetB3 alcanzaron 95.54% de exactitud y F1-score >0.94, superando a MobileNet 91.51% de exactitud. Las confusiones se presentaron principalmente en fracturas visualmente similares. Las CNN evaluadas son viables para la clasificación automática de fracturas óseas, constituyendo una herramienta de apoyo diagnóstico útil, especialmente en áreas con escasez de especialistas. Se propone como trabajo futuro incluir un conjunto de prueba independiente y aumento de datos para mejorar la generalización.

Palabras clave: Clasificación de imágenes, diagnóstico asistido, fracturas óseas, redes neuronales convolucionales.

Abstract— Rapid and accurate diagnosis of bone fractures is key to avoiding complications. Convolutional neural networks (CNN) have demonstrated high efficiency in the classification of radiographs, achieving results comparable to those of specialists. This study evaluated MobileNet, ResNet50, and EfficientNetB3 to classify 10 fracture types using 1,129 balanced images: 90% for training (1,017) and 10% for validation (112). Preprocessing included resizing to 256x256 pixels, conversion to RGB, normalization and one-hot encoding. The training was carried out in Google Colab with RMSprop optimizer, categorical_crossentropy loss and early stopping. ResNet50 and EfficientNetB3 achieved 95.54% accuracy and F1-score >0.94, outperforming MobileNet (91.96% accuracy). Confusions occurred mainly in visually similar fractures. The CNNs evaluated are viable for the automatic classification of bone fractures, constituting a useful diagnostic support tool, especially in areas with a shortage of specialists. It is proposed as future work to include an independent test set and data augmentation to improve generalization.

Keywords: Image classification, assisted diagnosis, bone fractures, convolutional neural networks.

1 INTRODUCCIÓN

El diagnóstico oportuno y preciso de fracturas óseas es fundamental para una adecuada intervención médica, ya que estas lesiones, definidas como la pérdida de continuidad del tejido óseo, pueden ocasionar complicaciones severas como discapacidad permanente o pérdida de función. Las fracturas se clasifican según su morfología y mecanismo de lesión en tipos como avulsión, conminuta, dislocación fracturada, en tallo verde (greenstick), fisura (hairline), impactada, longitudinal, oblicua, patológica y espiral. Cada tipo presenta características radiográficas específicas que requieren de un diagnóstico preciso para orientar el tratamiento.

Tradicionalmente, este diagnóstico depende del análisis

manual de radiografías; sin embargo, el incremento del volumen de estudios y la escasez de personal especializado generan retrasos en la entrega de informes clínicos[1][2]. En los últimos años, las redes neuronales convolucionales (CNN) han mostrado un gran potencial en el diagnóstico médico, alcanzando resultados comparables a los de radiólogos experimentados en la identificación de fracturas óseas. Por ejemplo, en un estudio[3] se demostró que una CNN logra predecir 95.5% de las fracturas superando a los cirujanos que lograron 92.2% de aciertos. Su eficacia es particularmente notable en la clasificación automatizada de imágenes radiológicas, estudios específicos dieron como resultado que las CNN han igualado en desempeño de radiólogos en la identificación de fracturas



Revista de Investigación en Ciencia y Tecnología

ISSN: 2810-8124 (en línea) / ISSN: 2706-543x

Universidad Nacional Micaela Bastidas de Apurímac – Perú

Vol. 7 Núm. 2 (2025) - Publicado: 19/08/25 - Indexaciones

Número: doi.org/10.57166/riqchary.v7.n2.2025

Páginas: 19- 29 | Recibido 01/01/2025 ; Aceptado 01/02/2025

doi.org/10.57166/riqchary.v7.n2.2025.3

Autores:

- A. ORCID iD <https://orcid.org/0009-0004-9354-5137>
Royer Elvis Moreano Condorcuya, estudiante de la Escuela Académica Profesional Ingeniería Informática y Sistemas de la Universidad Nacional Micaela Bastidas de Apurímac 1201remec@gmail.com
- B. ORCID iD <https://orcid.org/0000-0002-8386-9806>
Evelyn Naida Luque Ochoa, Docente del Departamento Académico de Informática y Sistemas de la Universidad Nacional Micaela Bastidas de Apurímac eluve@unamba.edu.pe

especialmente en muñecas o extremidades[4]. Diversos estudios han evidenciado que arquitecturas como VGG-16, ResNet50, EfficientNet y modelos basados en YOLO han alcanzado entre 94% y 99% de precisión[5][6]. En este contexto, el presente trabajo tiene como objetivo evaluar el desempeño de MobileNet, ResNet50 y EfficientNetB3 en la clasificación de diez tipos de fracturas óseas a partir de imágenes radiográficas, empleando un conjunto de datos balanceado y un flujo de preprocesamiento optimizado. Los resultados obtenidos podrán servir como base para desarrollar herramientas de diagnóstico asistido que mejoren la precisión y reduzcan los tiempos de respuesta, especialmente en entornos con recursos limitados. Sin embargo, la validación externa con la integración en las clínicas sigue siendo retos importantes para su adopción generalizada [7].

2 TRABAJOS RELACIONADOS

En los últimos años, múltiples investigaciones han evidenciado el potencial de las redes neuronales convolucionales para el diagnóstico automatizado de las fracturas. Por ejemplo, Mohammad et al [8], su modelo fue evaluado sobre múltiples conjuntos de datos, alcanzando precisiones de 95.01%, 94.04% y 96.25%, superando significativamente a varios métodos convencionales en términos de exactitud diagnóstica, el modelo desarrollado con una arquitectura propia, fue evaluado sobre múltiples conjuntos de datos, alcanzando precisiones del 95.01%, 94.04% y 96.25% en la clasificación de fracturas.

Yasenn et al en 2022 [9] en su investigación titulada "Cervical Spine Fracture Detection and Classification Using Two-Stage Deep Learning Methodology", propusieron un enfoque de dos etapas para la detección automática de fracturas en la columna cervical, una condición médica crítica que puede llevar a parálisis permanente o incluso la muerte. En la primera etapa, utilizaron una CNN para determinar la presencia o ausencia de fractura a partir de imágenes de tomografía computarizada (CT). Para mejorar la interpretabilidad, incorporaron técnicas de visualización como Grad-CAM, en la segunda etapa, se enfocaron en la detección específica de fracturas en cada una de las siete vértebras cervicales, entrenando modelos YOLOv5 y YOLOv8 sobre un conjunto de 9,170 imágenes. El desempeño fue evaluado con métricas como precisión 90%.

Por su parte, Medaramatla et al.[10], en su estudio "Detection of Hand Bone Fractures in X-Ray Images Using Hybrid YOLO NAS", desarrollaron un modelo híbrido para la detección de fracturas óseas en las manos, uno de los tipos de lesiones más comunes en la práctica clínica. La propuesta se basa en una combinación de YOLO NAS (You Only Look Once - Neural Architecture Search), EfficientDet y DETR3 (Detection Transformer), algoritmos reconocidos por su precisión en tareas de detección de objetos.

El modelo fue entrenado utilizando un conjunto de 4,736 radiografías de mano, clasificadas en 6 tipos de fracturas. Se realizó una comparación rigurosa con otros modelos existentes para validar el rendimiento del modelo implementado alcanzó el 98% de precisión, Mu et al. [11], en su investigación "Fine-Tuned Deep Convolutional Networks for the Detection of Femoral Neck Fractures on Pelvic Radiographs: A Multi-

center Dataset Validation", utilizaron un total de 1,491 radiografías pélvicas recolectadas en múltiples instituciones para entrenar y validar un modelo denominado Fine-tuned DR-FDS. Este modelo superó a su versión inicial (DR-FDS) y a otras arquitecturas convencionales, alcanzando una AUC de hasta 0.9937 y una precisión del 85.97% en pruebas externas. Además, mostró una alta generalización y rendimiento clínico al reducir el tiempo de lectura e incrementar significativamente la sensibilidad y especificidad diagnóstica.

Así mismo en el estudio realizado por Spoorthy Torne et al. [12], titulado "Enhancing Bone Fracture Classification in X-Ray Using Deep Learning Models: VGG-16, VGG-16 With Random Forest, ResNet50 With SVM, and EfficientNetB0 With XGBoost", se abordó la problemática de la clasificación automatizada de fracturas óseas utilizando modelos híbridos de Deep Learning y Machine Learning clásico. Utilizando un conjunto de datos que incluía 10 tipos diferentes de fracturas óseas en radiografías, los autores compararon el desempeño de varios modelos: Cada modelo fue evaluado según métricas estándar como precisión, recall y F1-score. Entre todos los enfoques, los modelos basados en VGG-16 tanto el modelo original como su versión con Random Forest obtuvieron los mejores resultados, alcanzando una precisión del 95%, superando al resto de modelos en prácticamente todas las clases de fracturas evaluadas.

3 METODOLOGÍA

3.1 Diseño de investigación

El diseño de la investigación es no experimental de corte transversal, lo que implica que se evaluó el rendimiento de diferentes modelos de redes neuronales convolucionales en la detección de fracturas en un momento específico. De acuerdo con Mairer et al. [13] el investigador no interviene ni modifica las variables independientes y los datos se recolectan en un punto único punto temporal. Este enfoque permitió un análisis objetivo de las métricas de rendimiento, como exactitud, recall y F1-score, utilizando un conjunto de datos previamente establecido

3.2 Participantes

La población de este estudio estuvo conformada por un conjunto de 1129 imágenes etiquetadas de fracturas óseas divididas entre 10 tipos: Avulsion fracture, Comminuted fracture, Fracture Dislocation, 'Greenstick fracture, Hairline Fracture, Impacted fracture, Longitudinal fracture, Oblique fracture, Pathological fracture, Spiral Fracture en todo el cuerpo, obtenidas de bases de datos públicas como Kaggle [14], que proporciona un conjunto de imágenes utilizadas en investigaciones científicas para entrenar y evaluar modelos de clasificación.

Dado el tamaño del conjunto de datos se dividió en dos subconjuntos 90% para el entrenamiento, 10% para la validación, siguiendo las prácticas observadas en este estudio Real-time Classification of Bone Fractures Utilizing Different Convolutional Neural Network Approaches [15] el cual empleó una división 90% para el entrenamiento y 10% para la validación. Así mismo se procuró que todas las clases del conjunto

de datos estén balanceadas.

El conjunto de datos fue dividido de la siguiente manera 1017 imágenes par la parte del entrenamiento y 112 imágenes para la parte de la validación, en la tabla 1 se detalla la distribución por clases de las diferentes imágenes dentro de los subconjuntos de datos de entrenamiento y validación.

TABLE 1

Distribución de las imágenes por los subconjuntos entrenamiento y validación

Clase	Entrenamiento (90%)	Validación (10%)
Avulsion fracture	101	12
Comminuted fracture	101	11
Fracture Dislocation	102	11
Greenstick fracture	102	11
Hairline Fracture	102	11
Impacted fracture	102	12
Longitudinal fracture	102	11
Oblique fracture	102	11
Pathological fracture	101	11
Spiral Fracture	102	11

La figura 1 muestra las imágenes de las diferentes clases de fracturas que se encuentran en el conjunto de datos



Fig. 1. Imágenes de fracturas óseas

3.3 Selección de Arquitecturas

Se seleccionaron tres modelos preentrenados para realizar el desarrollo del presente trabajo.

- **MobilNet:** Se escogió porque está diseñado para ser ligero permitiendo inferencia rápida y bajo uso de memoria[16],[17]. Así mismo puede ajustarse a diferentes tipos de fracturas y regiones anatómicas manteniendo el rendimiento estable [18],[19], además logro grandes niveles de predicción entre 95% y 98% a la hora de predecir fracturas[20], [21]. En la figura 2 se muestra la arquitectura del modelo preentrenado MobilNet

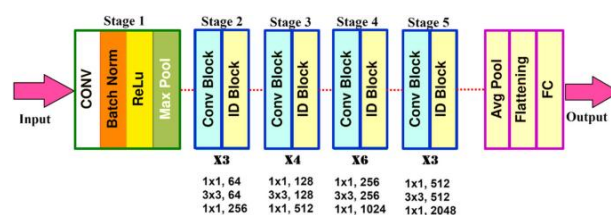


Fig. 2 Arquitectura del modelo ResNet50 extraída de [22]

- **EfficientNetB0:** Se escogió porque se demostró en los siguientes estudios[23],[24], donde se comprobó que EfficientNet es altamente eficiente en la detección de fracturas óseas alcanzados altas tasas de predicción entre un 93.5% y un 99.4% en fracturas de la mano y fracturas cervicales. Así mismo se demostró que EfficientNet destaca por su arquitectura optimizada logrando alto rendimiento con menor parámetros y menor consumo computacional[25]. En la figura 3 se muestra la arquitectura del modelo preentrenado EfficientNetB0



Fig 3 Arquitectura del modelo EfficientNetB0 extraído de [26]

- **ResNet50:** Se demostró en estos estudios [27],[28] que el uso de este modelo mejoro la capacidad de detección de las fracturas incluso mejor que algunos médicos con poca experiencia. En la figura 4 se muestra la arquitectura de la red convolucional ResNet50

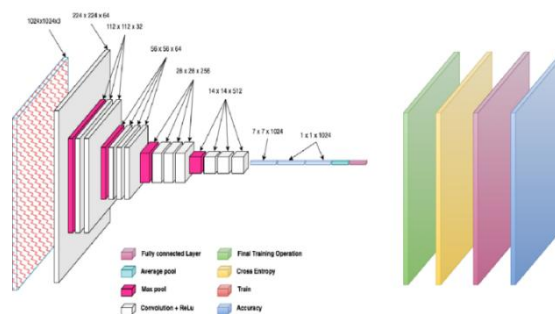


Fig 4 arquitectura del modelo MobileNet [29]

3.4. Instrumentos y técnicas

Para la implementación de este estudio, se empleó Python como lenguaje de programación, aprovechando su versatilidad y el vasto soporte de bibliotecas especializadas en aprendizaje automático. El entrenamiento de los modelos se llevó a cabo en Google Colab, un entorno de desarrollo que proporcionó acceso a recursos de computación en la nube.

3.5 Preprocesamiento del conjunto de datos

El preprocesamiento se hizo de la siguiente manera:

- Redimensionado y conversión de formato: Todas las imágenes fueron redimensionadas a 256×256 píxeles y convertidas a formato RGB, independientemente de su formato y dimensiones originales, para estandarizar la entrada de datos
- Normalización y preprocesamiento específico del modelo: Se implementaron dos funciones de preprocesamiento, `preprocess_for_model(img)`: convierte a tipo float32 y aplica la función, `preprocess_input()` correspondiente a cada modelo preentrenado, ajustando los valores de píxeles al rango y distribución esperados. `preprocess_for_display(img)`: normaliza los valores de píxeles al rango [0,1] para la visualización en reportes y gráficos. El conjunto de entrenamiento y validación fueron procesados con `preprocess_for_model()`, mientras que `preprocess_for_display()` se utilizó para mostrar ejemplos.
- Agrupamiento en lotes (Batching): Las imágenes se agruparon en lotes de 64 muestras para optimizar el uso de memoria durante el entrenamiento.
- Codificación de etiquetas: Las etiquetas fueron convertidas a formato one-hot encoding, indispensable para usar la función de pérdida `categorical_crossentropy` en la clasificación multiclase.
- Conversión a NumPy para análisis: Los datasets en formato `tf.data.Dataset` se transformaron a arreglos NumPy, lo que permitió realizar análisis estadísticos adicionales y generar matrices de confusión.

Este proceso garantizó que los datos estuvieran en el formato, tamaño y distribución adecuados para maximizar el rendimiento de las arquitecturas de deep learning seleccionadas.

3.6 Análisis de datos

El análisis de datos se llevó a cabo mediante la comparación de las métricas de rendimiento de los modelos entrenados.

Se calcularán las siguientes métricas:

- Exactitud o accuracy: Proporción de predicciones correctas (verdaderos positivos y verdaderos negativos) sobre el total de muestras. La siguiente fórmula muestra accuracy o exactitud

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

- Categorical Crossentropy o Loss: para clasificación multiclase con one-hot, la siguiente fórmula muestra cómo se halla.

$$Loss = -\sum_{i=1}^N \sum_{j=1}^C y_{ij} * \log(\hat{y}_{ij}) \quad (2)$$

- Recall: Qué tan bien el modelo detecta los positivos reales. Es útil cuando te importa no dejar pasar positivos

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

- F1-score: Es el promedio armónico entre precisión y recall. Útil cuando necesitas balancear ambos

$$F1 = 2 * \frac{Precision * Recall}{Precision + Recall} \quad (4)$$

Para cada modelo utilizando el subconjunto de validación.

La matriz de confusión se empleó para identificar patrones de error específicos, lo que permitió una evaluación más detallada del rendimiento de cada modelo. Los resultados se analizaron estadísticamente para determinar la eficacia de cada red neuronal en la detección de fracturas, garantizando un análisis riguroso y reproducible de los resultados.

4 RESULTADOS EXPERIMENTALES

4.1 Resultados del modelo MobileNet

En la tabla 2 se muestra las métricas de test y validación durante el entrenamiento del modelo MobileNet. El entrenamiento se configuró para 200 épocas, pero se detuvo en la época 150, se usó de optimizador a RMSprop, también se implementó un callback de paciencia en 10 épocas que monitoreaba si la pérdida de la validación no disminuía. El entrenamiento duró un aproximado de 4 horas y 16 minutos.

TABLA 2
Resultados del rendimiento del modelo MobileNet

Métrica	Train	Validación
Exactitud	90.64%	91.51%
Pérdida	0.4804	0.3788
Recall	0.9011	0.9250
F1-score	0.8954	0.9197

En la siguiente tabla se muestra el valor predictivo para cada clase del conjunto

TABLA 3
Resultados predictivos por cada clase del modelo MobileNet

Clase	Precisión
Avulsion fracture	93%
Comminuted fracture	99%
Fracture Dislocation	83%
Greenstick fracture	93%
Hairline Fracture	99%
Impacted fracture	99%

Longitudinal fracture	99%
Oblique fracture	90%
Pathological fracture	92%
Spiral Fracture	91%

La figura 5 muestra el comportamiento de la exactitud durante el entrenamiento del modelo de MobileNet.

- Comportamiento de la exactitud en el entrenamiento: Se observa un comportamiento inconstante pero que aumento gradualmente desde la época 1 con un 09% acabando con un 90.43% en la época 150, indicando que el modelo aprendió acerca de los datos
- Comportamiento de la exactitud en la validación: Se observa un comportamiento inconstante pero que aumento gradualmente desde la época 1 con un 16.96% acabando con un 91.43% en la época 150, indicando que el modelo aprendió acerca de los datos.

Por la brecha y el comportamiento de la exactitud del entrenamiento y validación sugiere un posible underfitting o desaprendizaje a falta de datos

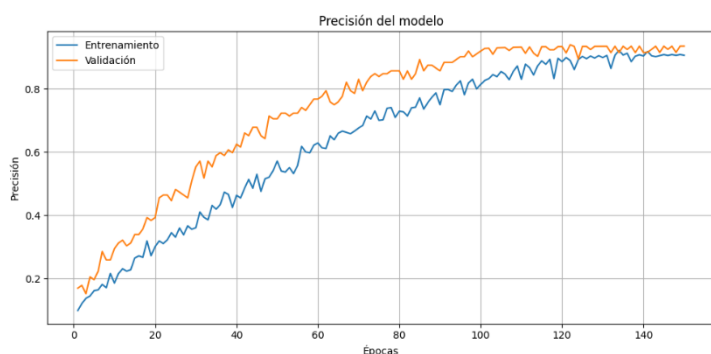


Fig 5. Curva de exactitud durante en entrenamiento del modelo de MobileNet

La figura 6 muestra el comportamiento de la perdida durante el entrenamiento del modelo de MobileNet

- Comportamiento de la perdida en el entrenamiento: Se observa un comportamiento inconstante pero que disminuyo gradualmente desde la época 1 con un 2.6466 acabando con un 0.4804 en la época 150.
- Comportamiento de la perdida en la validación: Se observa un comportamiento inconstante pero que disminuyo gradualmente desde la época 1 con un 2.2933 acabando con un 0.3788 en la época 150.

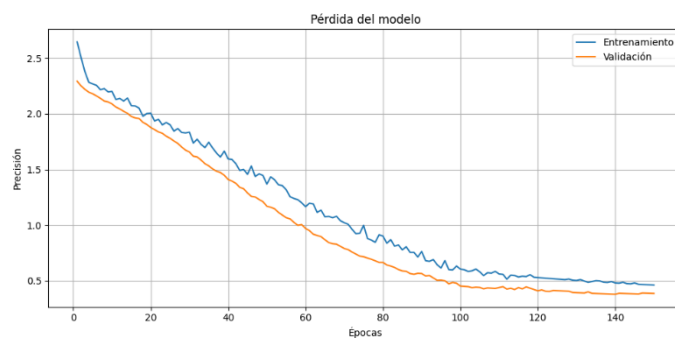


Fig 6 Curva de pérdida (Loss) durante el entrenamiento de MobileNet.

La figura 7 muestra la matriz de confusión del modelo MobileNet hecho durante la variación

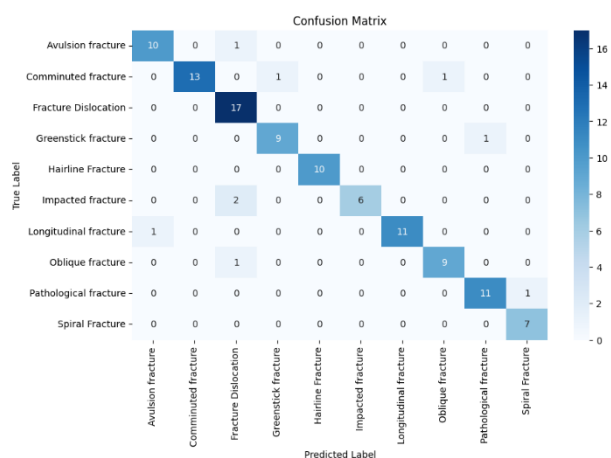


Fig 7 Matriz de confusión de MobileNet en el conjunto de la validación

- Se observa que las clases Fracture Dislocation e Impacted fracture fueron confundidas lo cual indica que hay una similitud visual entre ambas clases, lo mismo para Avulsion fracture fue confundida una vez con Fracture Dislocation y Greenstick fracture con Longitudinal fracture. Esto se debe a que algunas fracturas son similares entre sí.

4.2.Resultados del modelo ResNet50

En la tabla 4 se muestra las métricas de test y validación durante el entrenamiento del modelo ResNet50. El entrenamiento se configuro para 100 épocas, pero se detuvo en la época 62 porque se usó earlystop como callback que monitorea la disminución de la perdida en la validación con una paciencia de 10 épocas, se usó de optimizador a RMSprop. El entrenamiento duro un aproximado de 1 hora y 30 minutos.

TABLA 4
Resultados del rendimiento del modelo ResNet50

Métrica	Train	Validación
Exactitud	88.48%	95.54%
Pérdida	0.3587	0.1724

Recall	0.9386	0.9564
--------	--------	--------

F1-score	0.9402	0.9466
----------	--------	--------

En la tabla 5 se muestra el valor predictivo para cada clase del conjunto

TABLA 5
Resultados predictivos por cada clase del modelo Res-Net50

Clase	Precisión
Avulsion fracture	99%
Comminuted fracture	88%
Fracture Dislocation	89%
Greenstick fracture	99%
Hairline Fracture	99%
Impacted fracture	99%
Longitudinal fracture	83%
Oblique fracture	99%
Pathological fracture	99%
Spiral Fracture	99%

La figura 8 muestra el comportamiento de la exactitud durante el entrenamiento del modelo de ResNet50.

- Comportamiento de la exactitud en el entrenamiento: Se observa un comportamiento inconstante pero que aumento gradualmente desde la época 1 con un 12% acabando con un 90.30% en la época 62, indicando que el modelo aprendió acerca de los datos.
- Comportamiento de la exactitud en la validación: Se observa un comportamiento inconstante pero que aumento gradualmente desde la época 1 con un 31.25% acabando con un 95.54% en la época 62, indicando que el modelo aprendió acerca de los datos.

Por la brecha y el comportamiento de la exactitud del entrenamiento y validación sugiere un posible underfitting o desaprendizaje a falta de datos

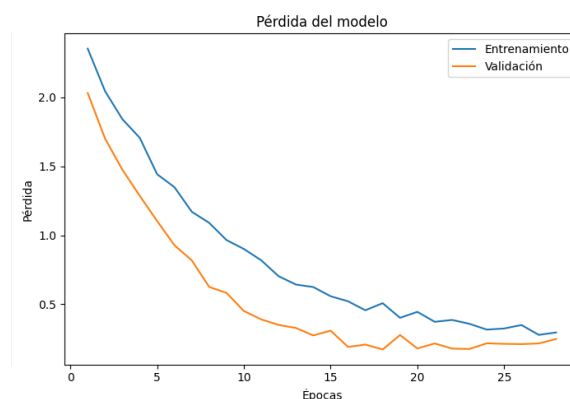


Fig 8. Curva de exactitud durante en entramiento del modelo de ResNet50

La figura 9 muestra el comportamiento de la perdida durante el entrenamiento del modelo de ResNet50.

- Comportamiento de la perdida en el entrenamiento: Se observa un comportamiento inconstante pero que disminuyo gradualmente desde la época 1 con un 2.8530 acabando con un 0.3201 en la época 62.
- Comportamiento de la perdida en la validación: Se observa un comportamiento inconstante pero que disminuyo gradualmente desde la época 1 con un 2.2933 acabando con un 0.1170 en la época 62

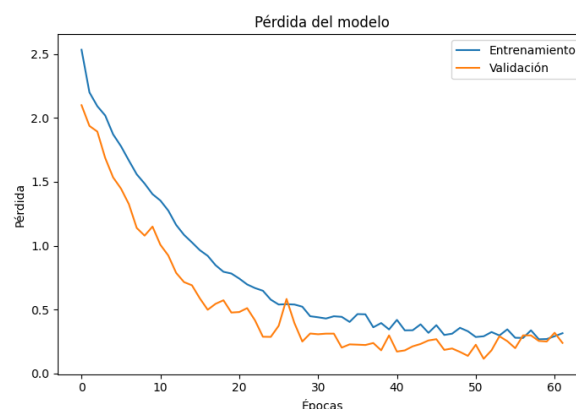


Fig 9 Curva de pérdida durante el entrenamiento de ResNet50.

La figura 10 muestra la matriz de confusión del modelo ResNet50 hecho durante la variación

TABLA 7

Resultados predictivos por cada clase del modelo EfficientNetB3

Clase	Precisión
Avulsion fracture	99%
Comminuted fracture	88%
Fracture Dislocation	89%
Greenstick fracture	99%
Hairline Fracture	99%
Impacted fracture	99%
Longitudinal fracture	83%
Oblique fracture	99%
Pathological fracture	99%
Spiral Fracture	99%

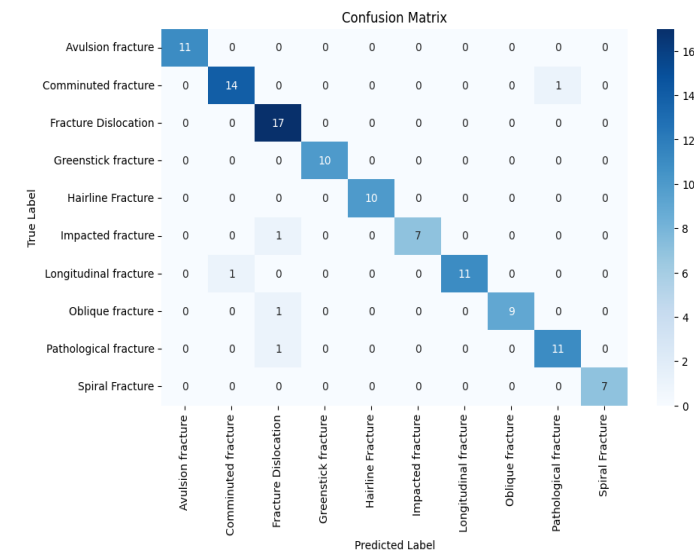


Fig 10 Matriz de confusión de ResNet50 en el conjunto de la validación

- Se observa que Fracture Dislocation tiene 3 falsos positivos, lo que sugiere que otras clases se están confundiendo con esta Impacted fracture, Comminuted, Oblique, Pathological tienen un solo error cada uno. Esto se debe a que algunas fracturas son similares entre sí.

4.3. EfficientNetB3

En la tabla 6 se muestra las métricas de test y validación durante el entrenamiento del modelo EfficientNetB3. El entrenamiento se configuró para 100 épocas, pero se detuvo en la época 28 porque se usó earlystop como callback que monitorea la disminución de la pérdida en la validación con una paciencia de 10 épocas, se usó de optimizador a RMSprop. El entrenamiento duró un aproximado de 1 hora y 30 minutos

TABLA 6
Resultados del rendimiento del modelo EfficientNetB3

Métrica	Train	Validación
Exactitud	88.48%	95.54%
Pérdida	0.3587	0.1724
Recall	0.9304	0.9564
F1-score	0.9300	0.9466

En la siguiente tabla se muestra el valor predictivo para cada clase del conjunto

La figura 11 muestra el comportamiento de la exactitud durante el entrenamiento del modelo de EfficientNetB3.

- Comportamiento de la exactitud en el entrenamiento: Se observa un comportamiento inconstante pero que aumento gradualmente desde la época 1 con un 13.83% acabando con un 95.54% en la época 28, indicando que el modelo aprendió acerca de los datos
 - Comportamiento de la exactitud en la validación: Se observa un comportamiento inconstante pero que aumento gradualmente desde la época 1 con un 30.36% acabando con un 95.54% en la época 28, indicando que el modelo aprendió acerca de los datos
- Por la brecha y el comportamiento de la exactitud del entrenamiento y validación sugiere un posible underfitting o desaprendizaje a falta de datos

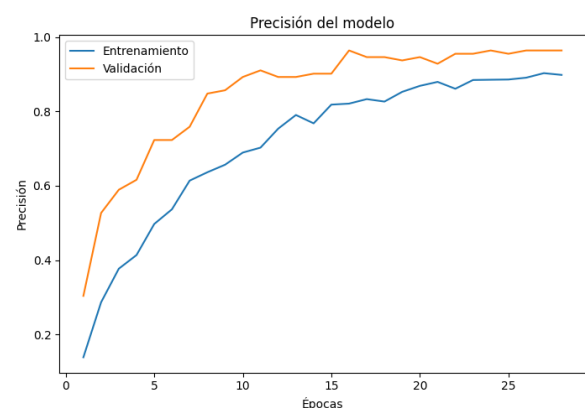


Fig 11. Curva de exactitud durante en entramiento del modelo de EfficientNetB3

La figura 12 muestra el comportamiento de la pérdida durante el entrenamiento del modelo de EfficientNetB3.

- Comportamiento de la pérdida en el entrenamiento: Se observa un comportamiento inconstante pero que disminuye gradualmente desde la época 1 con un 2.3532 acabando con un 0.1724 en la época 28.
- Comportamiento de la pérdida en la validación: Se observa un comportamiento inconstante pero que disminuye gradualmente desde la época 1 con un 2.2933 acabando con un 0.1170 en la época 28.

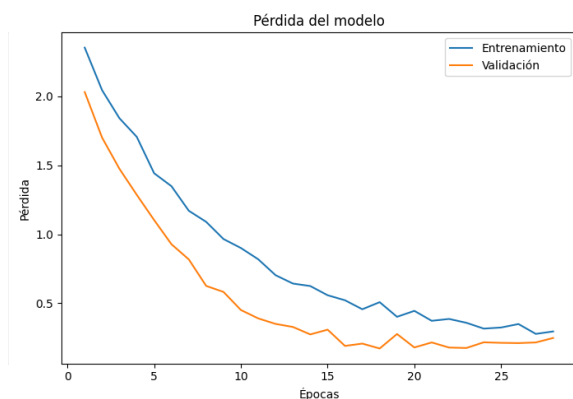


Fig 12 Curva de pérdida durante el entrenamiento de EfficientNetB3.

La figura 13 muestra la matriz de confusión del modelo ResNet50 hecho durante la variación

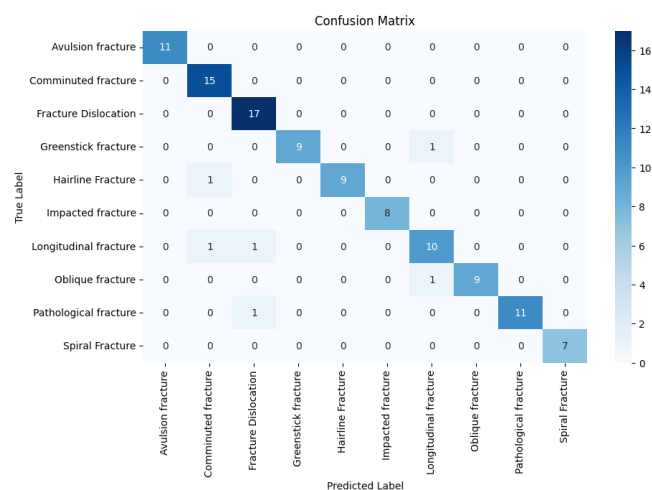


Fig 13 Matriz de confusión de EfficientNetB3 en el conjunto de la validación

- Se observa que Greenstick fracture fue confundida 1 vez como Hairline, Hairline fracture confundida 1 vez como Comminuted, Longitudinal fracture confundida como Comminuted y Fracture Dislocation, Oblique fracture confundida como Longitudinal y Pathological fracture confundida una vez como Fracture Dislocation. Esto se debe a que algunas fracturas son similares entre si.

5 DISCUSIÓN

En los últimos años, múltiples investigaciones han evidenciado el potencial de las redes neuronales convolucionales para el diagnóstico automatizado de fracturas óseas a partir de imágenes médicas. Estos avances han permitido mejorar significativamente la precisión, velocidad y generalización de los modelos, lo cual resulta crucial para su aplicación en entornos clínicos reales. El estudio realizado por Mohammad et al. [1] demostró un enfoque híbrido innovador, integrando técnicas de deep learning con blockchain, obteniendo precisiones de hasta 96.25% en diferentes conjuntos de datos. Este modelo destaca no solo por su alto rendimiento, sino también por su enfoque en la seguridad de los datos médicos, un aspecto crítico en aplicaciones clínicas. Por su parte, Yaseen et al. [2] propusieron un sistema en dos etapas para la detección de fracturas en la columna cervical, utilizando una CNN para el diagnóstico inicial y posteriormente modelos YOLOv5 y YOLOv8 para la localización específica de fracturas en cada vértebra. Su modelo alcanzó una precisión del 90%, lo que demuestra la eficacia del enfoque multi-etapa y la aplicabilidad de las arquitecturas YOLO en contextos clínicos complejos. Asimismo, Medaramatla et al. [3] desarrollaron un modelo híbrido para la detección de fracturas en manos, utilizando YOLO NAS, EfficientDet y DETR3. El modelo fue entrenado con 4,736 imágenes y logró una precisión del 98%, lo que evidencia la capacidad de las arquitecturas modernas para adaptarse a la complejidad morfológica de las fracturas en huesos pequeños. En el trabajo de Mu et al. [4], se utilizó un total de 1,491 radiografías pélvicas para entrenar el modelo Fine-tuned DR-FDS, el cual logró una AUC de hasta 0.9937 y una precisión del 85.97% en pruebas externas, mostrando una alta capacidad de generalización y utilidad clínica en el diagnóstico de fracturas del cuello femoral. De manera similar, Spoorthy Torne et al. [5] exploraron la clasificación multiclase de fracturas en radiografías usando modelos como VGG-16, ResNet50, EfficientNetB0, y sus combinaciones con algoritmos tradicionales como Random Forest, SVM y XGBoost. Entre todos, el modelo VGG-16 con Random Forest alcanzó una precisión del 95%, confirmando la eficacia de los modelos híbridos para problemas de clasificación compleja. En relación con estos trabajos, nuestros resultados muestran un rendimiento competitivo utilizando un conjunto de 1129 imágenes. El modelo MobileNet alcanzó una precisión del 91.96% en validación, con un recall y F1-score de 0.9196 y 0.9197, respectivamente. Por su parte, ResNet50 y EfficientNetB3 mostraron un mejor desempeño, ambos alcanzando una precisión de validación del 95.54%, y un F1-score superior a 0.94. Estos resultados son comparables a los modelos reportados en la literatura, como los de Yasenn et al. y Spoorthy Torne et al., demostrando que los modelos implementados tienen la capacidad de competir con enfoques de vanguardia. Además, nuestros modelos presentan una buena relación entre pérdida (loss) y precisión, lo cual indica un entrenamiento estable y una generalización efectiva en datos no vistos. En particular, el modelo EfficientNetB3 mantiene un equilibrio sólido entre exactitud, recall y F1-score, posicionándose como una de las mejores alternativas en términos de rendimiento global. Estos hallazgos confirman que las arquitecturas modernas, tanto puramente convolucionales como

híbridas, ofrecen soluciones viables y precisas para el diagnóstico automatizado de fracturas óseas. La elección del modelo dependerá del tipo de fractura, la calidad y cantidad de imágenes disponibles, así como de los recursos computacionales en entornos clínicos.

6 FORTALEZAS Y LIMITACIONES

- **Fortaleza:** La principal fortaleza del modelo desarrollado es el uso de EfficientNet, una arquitectura moderna que logra una alta precisión con un bajo costo computacional gracias a su escalado compuesto de profundidad, resolución y ancho. Esta eficiencia estructural permitió obtener buenos resultados incluso con una cantidad limitada de datos y bajo condiciones de entrenamiento restringidas. En el presente trabajo, EfficientNet demostró adaptarse eficazmente al rendimiento requerido, destacando su capacidad para generalizar en clases poco representadas. Como se evidencia en la imagen número 14, el modelo fue capaz de realizar predicciones acertadas.

Real: Longitudinal fracture | Predicción: Longitudinal fracture)



Fig 14 Imagen predicha por el modelo

- **Limitaciones:** En el transcurso de este estudio, surgieron varios obstáculos que quizás hayan afectado el desempeño y la aplicación amplia de los modelos que se prepararon. Uno de los problemas principales fue usar Google Colab sin costo, lo que causó restricciones importantes en cuanto a la potencia de cálculo, el tiempo de uso seguido y el acceso a GPUs de alto nivel. Estas trabas técnicas forzaron a acortar la duración de los entrenamientos y la dificultad del manejo de datos, lo que pudo haber perjudicado la mejora de los hiperparámetros y la cantidad de veces que algunos modelos se repitieron. Otro inconveniente importante fue el tamaño pequeño del grupo de datos, que solo tenía 1129 fotos repartidas en diez tipos. Aunque se intentó que hubiera un equilibrio entre los tipos, esta cantidad es poca para tareas de clasificación de varios tipos con redes neuronales profundas. Esta falta de datos hizo que aumentara el riesgo de un ajuste deficiente, como se vio en las curvas de aprendizaje de los tres modelos evaluados, donde el desempeño en la validación mejoraba, pero sin llegar a un sobreajuste que mostrara una completa asimilación de las características del entrenamiento. Esto indica que los modelos no pudieron aprender patrones complejos de manera óptima debido a la limitada cantidad de ejemplos.

7. CONCLUSIONES Y TRABAJOS FUTUROS

El presente estudio tuvo como objetivo evaluar el desempeño de MobileNet, ResNet50 y EfficientNetB3 en la clasificación automática de diez tipos de fracturas óseas a partir de imágenes radiográficas, utilizando un conjunto de datos balanceado y un flujo de preprocesamiento optimizado. Los resultados evidenciaron que ResNet50 y EfficientNetB3 alcanzaron un rendimiento sobresaliente, con una exactitud del 95,54 % y valores de F1-score superiores a 0,94, superando a MobileNet, que obtuvo un 91,51 % de exactitud. Si bien se identificaron confusiones entre fracturas visualmente similares, los modelos evaluados demuestran ser herramientas viables de apoyo diagnóstico, especialmente en entornos con recursos limitados. Estos hallazgos respaldan la integración de CNN avanzadas en sistemas de asistencia médica para optimizar la precisión y reducir los tiempos de diagnóstico.

Estas métricas son comparables con las mejores propuestas documentadas en la literatura científica reciente, lo que valida la solidez del enfoque adoptado. Además, se ha evidenciado que los modelos presentan buena generalización, bajos niveles de pérdida (loss) durante el entrenamiento y alta capacidad para identificar correctamente múltiples tipos de fracturas. Estas características son fundamentales para su potencial implementación en entornos clínicos reales, donde el diagnóstico rápido y preciso puede significar una diferencia crucial en la atención médica del paciente donde no se cuenta con especialistas. En conjunto, los hallazgos de este trabajo no solo consolidan la viabilidad del uso de CNN en la detección automática de fracturas, sino que también abren la puerta a futuras investigaciones orientadas a mejorar la interpretabilidad, seguridad de los datos y desempeño en condiciones clínicas variables. La combinación de modelos precisos, conjuntos de datos representativos y enfoques centrados en el usuario médico será clave para la adopción masiva de estas tecnologías en el ámbito de la salud.

Para trabajos futuros este estudio implementará un sistema de detección de fracturas con el modelo EfficientNetB3 que desde una página web utilizando el api de Gemini para poder predecir el tipo de fractura que se tiene y sus posibles soluciones diagnósticas ya que en las zonas rurales no existen especialistas, en el contexto peruano, esta problemática se acentúa debido a la alta concentración de equipos de rayos X en la ciudad de Lima, lo que restringe el diagnóstico oportuno en regiones descentralizadas [30]

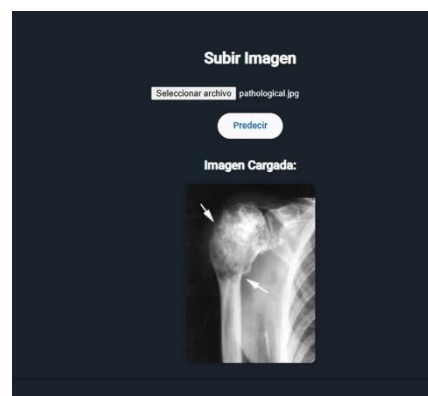


Fig 15 aplicación ingresando la radiografía



Fig 16 aplicación prediciendo que tipo de fractura es y su posible tratamiento.

También se implementará el uso de arquitecturas propias y aumento de datos obteniendo de diversas fuentes como hospitales o clínicas privadas para así mejorar el modelo y descartar el underfitting.

8. REFERENCIAS

- [1] T. S. Omofoye *et al.*, «Backlogs in formal interpretation of radiology examinations: a pilot global survey», *Clin Imaging*, vol. 106, feb. 2024, doi: 10.1016/j.clinimag.2023.110049.
- [2] G. Wahid, Ammara Haroon, Mehreen Samad, y Naila Tamkeen, «Causes of Delay in Radiological Reporting and ways to Reduce them», *Journal of Saidu Medical College, Swat*, vol. 12, n.º 3, pp. 133-137, sep. 2022, doi: 10.52206/jsmc.2022.12.3.697.
- [3] T. Urakawa, Y. Tanaka, S. Goto, H. Matsuzawa, K. Watanabe, y N. Endo, «Detecting intertrochanteric hip fractures with orthopedist-level accuracy using a deep convolutional neural network», *Skeletal Radiol*, vol. 48, pp. 239-244, 2018, doi: 10.1007/s00256-018-3016-3.
- [4] Y. Xie *et al.*, «Artificial intelligence diagnostic model for multi-site fracture X-ray images of extremities based on deep convolutional neural networks», *Quant Imaging Med Surg*, vol. 14, n.º 2, pp. 1930-1943, feb. 2024, doi: 10.21037/qims-23-878.
- [5] N. T. Htun y K. M. M. Tun, «Fuzzy-based Image Enhancement and Ensemble CNN Model for Bone Fracture Detection and Classification System», en *2024 5th International Conference on Advanced Information Technologies (ICAIT)*, 2024, pp. 1-6. doi: 10.1109/ICAIT65209.2024.10754918.
- [6] K. Mittal, K. S. Gill, R. Chauhan, y A. Kapruwan, «Innovative Fracture Diagnosis: MobileNet CNN Approach for Precise Bone Fracture Detection and Classification», en *2024 International Conference on Intelligent Systems for Cybersecurity (ISCS)*, 2024, pp. 1-5. doi: 10.1109/ISCS61804.2024.10581396.
- [7] L. O. Carmo *et al.*, «An increasing number of convolutional neural networks for fracture recognition and classification in orthopaedics», *Bone Jt Open*, vol. 2, pp. 879-885, 2021, doi: 10.1302/2633-1462.210.BJO-2021-0133.
- [8] F. Mohammad, S. Al-Ahmadi, y J. Al-Muhtadi, «Block-Deep: A Hybrid Secure Data Storage and Diagnosis Model for Bone Fracture Identification of Athlete From X-Ray and MRI Images», *IEEE Access*, vol. 11, pp. 142360-142370, 2023, doi: 10.1109/ACCESS.2023.3330914.
- [9] M. Yaseen, M. Ali, S. Ali, A. Hussain, M. Il Joo, y H. C. Kim, «Cervical Spine Fracture Detection and Classification Using Two-Stage Deep Learning Methodology», *IEEE Access*, vol. 12, pp. 72131-72142, 2024, doi: 10.1109/ACCESS.2024.3398061.
- [10] S. C. Medaramatla, C. V. Samhitha, S. D. Pande, y S. R. Vinta, «Detection of Hand Bone Fractures in X-Ray Images Using Hybrid YOLO NAS», *IEEE Access*, vol. 12, pp. 57661-57673, 2024, doi: 10.1109/ACCESS.2024.3379760.
- [11] L. Mu *et al.*, «Fine-Tuned Deep Convolutional Networks for the Detection of Femoral Neck Fractures on Pelvic Radiographs: A Multicenter Dataset Validation», *IEEE Access*, vol. 9, pp. 78495-78503, 2021, doi: 10.1109/ACCESS.2021.3082952.
- [12] S. Torne *et al.*, «VGG-16, VGG-16 With Random Forest, Resnet50 With SVM, and EfficientNetB0 With XGBoost-Enhancing Bone Fracture Classification in X-Ray Using Deep Learning Models», *IEEE Access*, vol. 13, pp. 25568-25577, 2025, doi: 10.1109/ACCESS.2025.3534818.
- [13] C. Maier, J. B. Thatcher, V. Grover, y Y. K. Dwivedi, «Cross-sectional research: A critical perspective, use cases, and recommendations for IS research», *Int J Inf Manage*, vol. 70, p. 102625, jun. 2023, doi: 10.1016/J.IJINFOMGT.2023.102625.
- [14] «Bone Break Classification Image Dataset». Accedido: 14 de julio de 2025. [En línea]. Disponible en: <https://www.kaggle.com/datasets/pkdarabi/bone-break-classification-image-dataset>
- [15] N. Vasker, M. Hasan, M. Nuha, S. Jahan, M. Tahsin, y Md. Y. Emon, *Real-time Classification of Bone Fractures Utilizing Different Convolutional Neural Network Approaches*. 2023. doi: 10.1109/IC-CIT60459.2023.10441387.
- [16] S. Turk, O. Bingol, A. Coşkunçay, y T. Aydin, «The impact of implementing backbone architectures on fracture segmentation in X-ray images», *Engineering Science and Technology, an International Journal*, p., 2024, doi: 10.1016/j.jestch.2024.101883.
- [17] P. M, S. M, R. N, y S. S, «Edge AI-based Bone Fracture Detection using TFLite», *International Journal of Innovative Research in Advanced Engineering*, p., 2025, doi: 10.26562/ijrae.2025.v1204.04.
- [18] A. Khanapure, H. Kashyap, A. Bidargaddi, S. Habib, y A. Anand, «Bone Fracture Detection with X-Ray images using MobileNet V3 Architecture», *2024 IEEE 9th International Conference for Convergence in Technology (I2CT)*, pp. 1-8, 2024, doi: 10.1109/I2CT61223.2024.10544356.
- [19] R. Bhuria y S. Gupta, «X-Ray Insights: Comprehensive Dataset for Bone Fracture Detection Across Diverse Anatomical Regions», *2024 5th International Conference on Smart Electronics and Communication*

- (ICOSEC), pp. 1242-1247, 2024, doi: 10.1109/ICOSEC61587.2024.10722406.
- [20] M. Goel y G. Singh, «Fracture Detection using MobileNet Model», 2024 4th International Conference on Ubiquitous Computing and Intelligent Information Systems (ICUIS), pp. 1574-1579, 2024, doi: 10.1109/ICUIS64676.2024.10866070.
- [21] S. Thota, P. Kandukuru, M. Sundaram, A. Ali, S. Muzamil, y B. Bindu, «Deep Learning based Bone Fracture Detection», 2024 International Conference on Smart Systems for applications in Electrical Sciences (ICSSES), pp. 1-7, 2024, doi: 10.1109/ICSSES62373.2024.10561360.
- [22] S. R. Sannasi Chakravarthy, N. Bharanidharan, C. Vinothini, V. Vinoth Kumar, T. R. Mahesh, y S. Guluwadi, «Adaptive Mish activation and ranger optimizer-based SEA-ResNet50 model with explainable AI for multiclass classification of COVID-19 chest X-ray images», *BMC Med Imaging*, vol. 24, n.º 1, dic. 2024, doi: 10.1186/s12880-024-01394-2.
- [23] A. S. B. Karno *et al.*, «Classification of cervical spine fractures using 8 variants EfficientNet with transfer learning», *International Journal of Electrical and Computer Engineering (IJECE)*, p., 2023, doi: 10.11591/ijece.v13i6.pp7065-7077.
- [24] S.-T. Huang, L.-R. Liu, M.-F. Tsai, M.-Y. Huang, y H.-W. Chiu, «Developing a Deep Learning Model Using Transfer Learning from EfficientNet-b3 to Detect Knee Fracture on X-ray Images», *Proceedings of the 2023 7th International Conference on Medical and Health Informatics*, p., 2023, doi: 10.1145/3608298.3608352.
- [25] K. Teeyapan, «Abnormality Detection in Musculoskeletal Radiographs using EfficientNets», 2020 24th International Computer Science and Engineering Conference (ICSEC), pp. 1-6, 2020, doi: 10.1109/ICSEC51790.2020.9375275.
- [26] H. Amin, A. Darwish, A. E. Hassanien, y M. Soliman, «End-to-End Deep Learning Model for Corn Leaf Disease Classification», *IEEE Access*, vol. 10, pp. 31103-31115, 2022, doi: 10.1109/ACCESS.2022.3159678.
- [27] L. Yeh *et al.*, «A deep learning-based method for the diagnosis of vertebral fractures on spine MRI: retrospective training and validation of ResNet», *European Spine Journal*, vol. 31, pp. 2022-2030, 2022, doi: 10.1007/s00586-022-07121-1.
- [28] Y. Li *et al.*, «Differential diagnosis of benign and malignant vertebral fracture on CT using deep learning», *Eur Radiol*, vol. 31, pp. 9612-9619, 2021, doi: 10.1007/s00330-021-08014-5.
- [29] K. Saini y R. Devi, «A systematic scoping review of the analysis of COVID-19 disease using chest X-ray images with deep learning models», *Journal of Autonomous Intelligence*, vol. 7, dic. 2023, doi: 10.32629/jai.v7i2.928.
- [30] A. P. Bonifaz, C. C. Rodriguez, y R. P. Esparza, «Diagnostic Reference Levels for Common X-ray Procedures in Peru», *Cureus*, vol. 13, p., 2021, doi: 10.7759/cureus.18566.

Neuro-IA en el aula: Nuevas fronteras en la motivación y el rendimiento del estudiante universitario

Neuro-AI in the classroom: New frontiers in university student motivation and performance

Nilton Brayaam Hurtado Mendoza ^A

Resumen— La neuro-IA tiene el potencial de revolucionar la educación al mejorar la motivación y el rendimiento de los estudiantes. Es fundamental que educadores y desarrolladores trabajen juntos para crear entornos de aprendizaje seguros y éticos. La investigación futura debe enfocarse en la evaluación a largo plazo de estas tecnologías en diferentes contextos educativos, así como en el desarrollo de modelos híbridos que integren neurociencia y experiencia docente. Aunque la neuro-IA ofrece oportunidades para personalizar el aprendizaje mediante interfaces cerebro-computadora y neurofeedback, su implementación debe ser ética, priorizando la autonomía del estudiante y la privacidad de los datos. Se recomienda la creación de leyes que apoyen el uso responsable de estas tecnologías. En resumen, la neuro-IA no solo mejora el rendimiento académico, sino que también transforma la enseñanza, marcando el camino hacia un futuro educativo más inclusivo y eficiente.

Palabras clave: estudiante, motivación, neuro-IA, rendimiento.

Abstract— Neuro-AI has the potential to revolutionize education by improving student motivation and performance. It is critical that educators and developers work together to create safe and ethical learning environments. Future research should focus on the long-term evaluation of these technologies in different educational contexts, as well as the development of hybrid models that integrate neuroscience and teaching experience. Although neuroAI offers opportunities to personalize learning through brain-computer interfaces and neurofeedback, its implementation must be ethical, prioritizing student autonomy and data privacy. The creation of laws that support the responsible use of these technologies is recommended. In short, neuroAI not only improves academic performance but also transforms teaching, paving the way for a more inclusive and efficient educational future.

Keywords: student, motivation, neuro-AI, performance.

1 INTRODUCCIÓN

La educación es solo una de las muchas industrias que la inteligencia artificial (en adelante IA) ha revolucionado. En concreto, la neuro-IA (la fusión de métodos de IA y neurociencia) ofrece nuevas oportunidades para aumentar la participación y el rendimiento estudiantil. Sin embargo, el desarrollo de la neuro-IA ha posibilitado nuevas oportunidades para transformar drásticamente el proceso de enseñanza. Según investigaciones recientes, comprender los mecanismos cognitivos del aprendizaje puede generar mejores recursos didácticos [1]. Con énfasis en cómo afecta la motivación y el rendimiento estudiantil, este artículo describirá cómo la neuro-IA puede influir en el proceso educativo. Considerando el uso de esta tecnología, que permite la comprensión y la respuesta en tiempo real a los

estados emocionales y cognitivos de los estudiantes, proporciona entornos altamente personalizados que mejoran el rendimiento académico y la motivación..

Debemos comprender que los pilares fundamentales del rendimiento estudiantil son la motivación y el compromiso emocional [2]. Por último, los niveles de atención, estrés y carga cognitiva pueden identificarse durante el aprendizaje mediante herramientas como las interfaces cerebro-computadora (ICC) y algoritmos de IA que evalúan señales neurofisiológicas (como el electroencefalograma) [3]. En momentos de sobrecarga o desmotivación, esta capacidad adaptativa permite intervenciones más efectivas al acercar la experiencia educativa a los estados mentales del estudiante.

Relizando una breve revisión de los antecedentes al respecto de este tema tenemos: La educación se enfrenta al reto de decodificar el cerebro del estudiante, pero la neuroeducación,



Revista de Investigación en Ciencia y Tecnología

ISSN: 2810-8124 (en línea) / ISSN: 2706-543x

Universidad Nacional Micaela Bastidas de Apurímac – Perú

Vol. 7 Núm. 1 (2025) - Publicado: 22/02/24 - [Indexaciones](#)

Número: doi.org/10.57166/riqchary/v7.n1.2025

Páginas: 1- 6 | Recibido 01/01/2025 ; Aceptado 01/02/2025

doi.org/10.57166/riqchary.v7.n2.2025.4

Autores:

A. **ORCID iD** <https://orcid.org/0000-0001-9809-7792>

Nilton Brayaam Hurtado Mendoza, Ingeniero de Informática y Sistemas. Actualmente trabaja en la Corte Superior de Justicia de Apurímac. nbhurtado@gmail.com

un área multidisciplinaria que combina la pedagogía y la neurociencia cognitiva, nos ayuda a comprender cómo aprendemos los humanos. La investigación neuroeducativa nos ha mostrado cómo el cerebro interpreta la información y cómo maximizar el aprendizaje. Entre los hallazgos más destacados se encuentra la importancia de un entorno de aprendizaje en las aulas que fomente la participación, la emoción y la concentración, todo lo cual mejora la retención de la memoria.

Esto permite a los docentes concentrarse en el aprendizaje socioemocional y el apoyo emocional, mientras que la neuro-IA se encarga de las tareas administrativas y repetitivas. Esta programación se centra en el resultado por motivación del estudiante universitario del siglo XXI, que ahora se considera una habilidad fundamental y debería incorporarse al currículo desde los primeros cursos de una Escuela Profesional, en lugar de considerarse una cualidad diferenciada para algunos [4]. Los estudiantes universitarios que aprenden a analizar de una forma más profunda su comprensión de la tecnología que impregna su vida diaria, es decir, una plataforma con neuro-IA, además de desarrollar habilidades de resolución de problemas y pensamiento crítico. Esto permite a las estudiantes universitarias utilizar la tecnología basado en neuro-IA como los actores principales.

Por otro lado, el problema que se aborda en la presente investigación es que si bien las herramientas de IA se han incorporado a la educación superior y se han producido avances técnicos, se sabe poco sobre cómo estas herramientas afectan el rendimiento académico y la motivación de los estudiantes universitarios. El uso de IA y tecnologías basadas en la neurociencia en el aula puede brindar nuevas oportunidades para optimizar el proceso educativo y personalizar la instrucción [5]. Por lo tanto, se desconoce hasta qué punto estos materiales mejoran realmente el rendimiento académico de los estudiantes universitarios, fomentan el impulso interno y estimulan los procesos cognitivos. Para optimizar el proceso de enseñanza y aprendizaje en la educación superior, esta brecha en la investigación pone de relieve la necesidad de investigar y evaluar los efectos de la neuro-IA en el aula, destacando tanto sus ventajas como sus desventajas [6].

Asimismo, es importante investigar al respecto de este problema, porque su relevancia práctica al tener el potencial de transformar por completo la educación superior, investigar la neuro-IA en el aula puede ayudar a los estudiantes a crecer de manera integral y convertirse en ciudadanos conocedores y capaces en una sociedad que se está volviendo cada vez más avanzada tecnológicamente.

Por último, este estudio se propuso investigar los efectos

de la implementación de una plataforma educativa con componentes neuro-IA en el rendimiento académico y la motivación intrínseca de los estudiantes universitarios a lo largo de un semestre académico, entonces, nos realizaremos la siguiente interrogante ¿Es posible mejorar el rendimiento académico del estudiante universitario aplicando la neuro-IA en las aulas universitarias?

2 PERSONALIZACIÓN PROFUNDA

Determina en tiempo real si un estudiante está aprendiendo y ajusta el material en función de su estado emocional, carga cognitiva y capacidad de atención.

2.1 Importancia de la Personalización Profunda

Adaptar el contenido a los gustos del usuario es solo un aspecto de la personalización profunda. Implica una visión integral de la persona, lo que permite:

- **Mejorar la Experiencia del Usuario:** Las universidades pueden aumentar enormemente la satisfacción del estudiante al proporcionar cursos e información que realmente satisfagan los sentimientos y necesidades de sus estudiantes [7].
- **Aumentar la Lealtad:** Una experiencia personalizada crea una fuerte conexión emocional entre el estudiante y la universidad, fomentando la repetición de experiencias de aprendizaje y la lealtad por su alma mater.
- **Optimizar Estrategias de Marketing:** Las universidades pueden orientar con mayor precisión sus campañas de marketing y aumentar su retorno de la imagen del perfil del egresado de una determinada carrera con datos más precisos sobre el comportamiento del estudiante [8].

2.2 Aplicaciones de la neuro-IA en la Personalización

Se debe contar con recomendaciones personalizadas, para la aplicación de la neuro-IA, es decir, cuando se propone un enfoque relacionado a mejorar el rendimiento académico debemos considerar al usuario experto que pondrá los límites de lo que se pueda incorporar para ser mostrado a los estudiantes universitarios de acuerdo a su condición a lo largo de su vida universitaria. Al incorporar datos emocionales como el valor que tiene el docente en cada estudiante universitario por ejemplo conocimiento del tema, habilidades comunicativas, pasión por la enseñanza, empatía, flexibilidad, habilidades tecnológicas, ética profesional; serán cualidades que ayuden a crear un entorno de aprendizaje positivo y más personalizado promoviendo el rendimiento académico y personal de los estudiantes, la neuro-IA puede mejorar estas cualidades y ofrecer sugerencias basadas tanto en experiencias previas como en las sensaciones asociadas a

esos cursos por parte de los estudiantes, realizando el test previo a cada inicio de semestre académico [9].

La personalización profunda tiene ventajas, pero también inconvenientes. La recopilación y el análisis de datos emocionales pueden generar inquietudes sobre el consentimiento y la privacidad de cada estudiante. Las universidades deben garantizar que los estudiantes sean responsables de sus datos y ser transparentes y honestas sobre cómo los utilizan los métodos de enseñanza [10].

La personalización profunda impulsada por neuro-IA está revolucionando la forma en que las universidades interactúan con sus estudiantes y la forma como se les proyecta para el perfil del egresado. Las carreras profesionales pueden crear experiencias más impactantes y significativas al comprender mejor los sentimientos y preferencias de los estudiantes. Sin embargo, es fundamental abordar los dilemas morales y asegurar que esta herramienta eficaz se aplique con sensibilidad. La personalización profunda puede ser la clave para crear proyección profesional duradera y fieles con los estudiantes en una era cada vez más digital [11].

3 ESTÍMULO DE FUNCIONES COGNITIVAS

Utilizando procesos mentales para los estudiantes universitarios, puedes crear experiencias de aprendizaje que mejoren la motivación, la memoria o la concentración. Que a su vez estos procesos incluyen:

- **Atención:** Entiendase como la capacidad de concentrarse en tareas específicas y filtrar distracciones en aulas universitarias.
- **Memoria:** La habilidad para retener y recuperar información, tanto a corto como a largo plazo considerando los demás cursos que el estudiante universitario puede llevar en un semestre académico.
- **Razonamiento:** Esto será considerado como la capacidad de analizar información, hacer inferencias y tomar decisiones basadas en evidencia brindada por el docente o encontrada por el mismo estudiante universitario.
- **Resolución de Problemas:** Entiendase como la habilidad para identificar problemas, generar soluciones y evaluar su efectividad a raíz de una catedra universitaria.
- **Creatividad:** Esto es la capacidad de generar ideas originales y pensar de manera innovadora para ser trasladadas a los demás estudiantes universitarios como parte de una retroalimentación.
- **Flexibilidad Cognitiva:** Se conoce como la habilidad de adaptarse a nuevas situaciones o cambiar de enfoque según sea necesario [12].

3.1 Importancia del Estímulo de Funciones Cognitivas en la Educación

Un aprendizaje eficaz requiere habilidades cognitivas como la memoria, la concentración, el razonamiento y la resolución de problemas. Mejorar estas características puede:

- **Mejorar el Rendimiento Académico:** Los estudiantes pueden digerir la información de manera más eficiente y utilizar ese conocimiento en una variedad de contextos al perfeccionar habilidades cognitivas críticas.
- **Fomentar la Creatividad:** Además de enfatizar las capacidades analíticas, la estimulación cognitiva fomenta la creatividad, necesaria para el pensamiento crítico y la invención.
- **Desarrollar Habilidades Socioemocionales:** Las habilidades socioemocionales son esenciales para el aprendizaje cooperativo y la resolución de conflictos, y la neuro-IA puede ayudar a los docentes a identificarlas y cultivarlas [13].

3.2 Estimulación Cognitiva a Través de Juegos

Las aplicaciones educativas y los juegos serios basados en neuro-IA pueden diseñarse para abordar procesos cognitivos específicos. Además de hacer el estudio más interesante, estos juegos mejoran las capacidades cognitivas de los estudiantes al incorporar tareas que requieren concentración, memoria y resolución de problemas.

La capacidad de la neuro-IA para estimular los procesos cognitivos está revolucionando la educación. Podemos mejorar el rendimiento académico y fomentar el desarrollo de habilidades cruciales para el futuro al comprender mejor cómo aprenden los estudiantes y adaptar el proceso educativo a sus necesidades específicas. Sin embargo, es fundamental abordar las cuestiones éticas y garantizar que las nuevas tecnologías se implementen de forma correcta y transparente. La neuro-IA podría ser la clave para un aprendizaje más eficiente y gratificante en un mundo cada vez más informatizado [14].

4 EVALUACIÓN NEUROADAPTATIVA

Al examinar las respuestas conductuales o cerebrales, permite evaluar el aprendizaje sin el uso de evaluaciones tradicionales.

4.1 Importancia de la Evaluación Neuroadaptativa en la Educación

La evaluación neuroadaptativa ofrece múltiples

beneficios, tales como:

- **Personalización del Aprendizaje:** Permite a los docentes modificar sus planes de lecciones y el contenido de las lecciones para satisfacer las necesidades únicas de cada estudiante, aumentando la eficacia del aprendizaje.
- **Identificación de Estilos de Aprendizaje:** La evaluación neuroadaptativa puede determinar los estilos de aprendizaje de los estudiantes a través del análisis de datos, lo que permite una instrucción más individualizada.
- **Retroalimentación Continua:** Ofrece retroalimentación continua sobre el desarrollo de los estudiantes, lo que permite realizar modificaciones en el momento y promueve un entorno de aprendizaje más adaptativo [15].

4.2 Recolección de datos

La evaluación neuroadaptativa se basa en la recolección de datos variados, que pueden incluir:

- Resultados de pruebas estandarizadas.
- Análisis de la atención y concentración mediante tecnologías de seguimiento ocular.
- Evaluaciones emocionales a través de herramientas de reconocimiento facial o análisis de voz.

4.3 Análisis de Datos

Los datos recopilados se analizan mediante cuadros comparativos del rendimiento que se obtienen de la plataforma de neuro-IA para identificar tendencias y patrones en el rendimiento estudiantil. Esto permite a los docentes comprender mejor las emociones de aprendizaje de los estudiantes, así como sus fortalezas y debilidades.

4.4 Adaptación del Contenido

Se realizan ajustes al contenido educativo en tiempo real según los hallazgos. Para garantizar que cada estudiante tenga la mejor experiencia de aprendizaje posible, las plataformas de aprendizaje pueden modificar la complejidad de las tareas, el tipo de contenido proporcionado y los métodos de enseñanza.

4.5 Aplicaciones en el Aula

Dentro de los estudios nos centraremos en los siguientes:

- **Plataformas de Aprendizaje Personalizado**
Las plataformas educativas de nivel superior que incorporan evaluación neuroadaptativa pueden brindar oportunidades de aprendizaje únicas. Por ejemplo, pueden modificar instantáneamente la complejidad de los textos o el grado de dificultad de los problemas matemáticos entre otros según el rendimiento y las emociones de los estudiantes [16].

- **Monitoreo del Bienestar Emocional**
Vigilar la salud mental de los estudiantes es otro aspecto de la evaluación neuroadaptativa. Al reconocer niveles elevados de preocupación o insatisfacción, los instructores pueden actuar con prontitud y ofrecer mayor apoyo.
- **Reforzamiento Positivo**
Se puede motivar a los estudiantes mediante estrategias de gamificación en la evaluación neuroadaptativa. Esto crea un ambiente constructivo que fomenta el aprendizaje al ofrecer incentivos y reconocimiento por los logros.

Por último, la evaluación neuroadaptativa con apoyo de neuro-IA es un método innovador para comprender y mejorar la enseñanza en el aula. El rendimiento académico y la salud mental pueden mejorarse adaptando las experiencias educativas a las necesidades de cada estudiante. Sin embargo, es crucial abordar las cuestiones éticas y garantizar el uso adecuado de esta tecnología. La evaluación neuroadaptativa podría ser la clave para un futuro educativo más exitoso e incluso en un mundo en constante cambio [17].

5 OPTIMIZACIÓN DEL PROCESO EDUCATIVO

Determinar qué enfoques son más efectivos a nivel cerebral para poder crear programas educativos más eficientes.

5.1 Importancia de la Optimización del Proceso Educativo

La mejora del proceso educativo busca aumentar la motivación y la salud emocional de los estudiantes, además de su rendimiento académico. Se pueden lograr varios objetivos aplicando técnicas basadas en neuro-IA:

- **Personalización del Aprendizaje:** Cada estudiante aprende de forma diferente. La Neuro-IA permite personalizar la instrucción y el contenido según las preferencias de cada estudiante, lo que puede mejorar los resultados de aprendizaje [18].
- **Identificación Temprana de Dificultades:** La neuro-IA puede identificar problemas de aprendizaje antes de que se conviertan en desafíos importantes a través del análisis de datos en tiempo real, lo que permite terapias más rápidas y eficientes.
- **Estimulación de la Motivación:** La motivación y el compromiso de los estudiantes se incrementan al brindarles retroalimentación inmediata y contenido pertinente y difícil [19].

5.2 Análisis del Comportamiento del Estudiante

El análisis exhaustivo del comportamiento en el aula es posible gracias a la neuro-IA. El seguimiento ocular y la monitorización de la atención son dos herramientas que pueden proporcionar información útil sobre cómo los alumnos interactúan con el material. Los docentes pueden utilizar esta información para modificar sus métodos y aumentar la participación y la atención de los alumnos en el aula.

6 MÉTODO

6.1 Diseño de la investigación

Se utilizó un diseño pre-experimental con grupo control y grupo experimental, con enfoque cuantitativo. La duración del estudio fue de 12 semanas que comprende el semestre académico.

6.2 Tipo de investigación

Se tiene que el tipo de investigación descriptiva se caracteriza porque parte de un marco teórico y permanece en él; la finalidad radica en formular nuevas teorías o modificar las existentes, en incrementar los conocimientos científicos o filosóficos, para la presente investigación se realizó la contrastación con el aspecto práctico utilizando la neuro-IA para fomentar el rendimiento académico de los estudiantes universitarios. Por lo tanto, el tipo de investigación adecuado para el presente trabajo de investigación es descriptiva.

6.3 Nivel de investigación

El nivel de investigación exploratorio es adecuado para la presente investigación ya que permite una comprensión inicial de un fenómeno poco estudiado en este contexto específico. Entonces, la investigación exploratoria busca examinar un tema o problema de investigación poco estudiado, del cual se tienen muchas dudas o no se ha abordado antes.

6.4 Ámbito temporal y espacial

El desarrollo de la presente investigación fue llevado a cabo durante los meses comprendidos entre septiembre y diciembre del 2024.

Específicamente desarrollado en el curso de "Introducción al Derecho" del semestre académico 2024 - I.

6.5 Participantes

La población estuvo compuesta por 70 estudiantes universitarios (edad promedio: 21 años) que estudiaran el

curso de "Introducción al Derecho" de una facultad de Ciencias Jurídicas, Contables y Sociales de Apurímac, Perú.

Para la investigación se tomó una muestra de 35 estudiantes universitarios, realizando un muestreo no probabilístico por conveniencia basándose en la facilidad del acceso a los participantes que serán entrevistados.

El grupo experimental (n=35) utilizó una plataforma educativa equipada con neuro-IA, mientras que el grupo control (n=35) usó un entorno virtual tradicional.

6.3 Instrumentos

Para la siguiente investigación se utilizó los siguientes instrumentos:

- Plataforma neuro-IA: Incluyó un software adaptativo conectado a sensores portátiles, que ajustaban la dificultad de las actividades en función de la atención y la carga cognitiva.
- Cuestionario de Motivación Académica (CMA), adaptado de Vallerand.
- Evaluaciones académicas estandarizadas al inicio y al final del ciclo.
- Software SPSS 27 para análisis estadístico.

Para este estudio, se utilizó un instrumento de guía de entrevista estructurada, que consiste en una serie de preguntas predefinidas y ordenadas de manera coherente, garantizando así que todos los participantes respondan a los mismos temas, facilitando la comparación de las respuestas y el análisis temático. Además, se revisaron los promedios respecto del curso de "Introducción al Derecho" de todos los estudiantes para ver si la aplicación de la plataforma con neuro-IA tenía significancia para aumentar el rendimiento académico reflejado en el promedio general del semestre con un puntaje elevado en dicho curso.

6.4 Procedimiento

Los participantes realizaron una prueba diagnóstica inicial, luego usaron las plataformas asignadas durante el curso de "Introducción al Derecho". Al finalizar, se aplicó el CMA y una evaluación final. Se compararon los promedios que se reflejaron en promedio general del curso y la evolución de la motivación relacionado al rendimiento académico que cada estudiante universitario va adquiriendo cuando obtiene un mejor promedio a lo largo de cada semestre académico.

7 RESULTADOS

Los resultados revelaron diferencias estadísticamente significativas en los niveles de motivación y en el rendimiento

académico del grupo que utilizó neuro-IA.

7.1 Motivación

El grupo experimental presentó un incremento promedio del 24% en motivación intrínseca ($p < 0.01$), mientras que el grupo control solo aumentó un 7%. También se observó una reducción significativa de la desmotivación en el grupo experimental.

TABLA 1
Grupo con incremento de promedio

Grupo	Incremento en Motivación Intrínseca	Reducción de Desmotivación
Experimental	24% ($p < 0.01$)	Significativa
Control	7%	Significativa
Margen	5%	Significativa

Nota: Elaboración propia.

7.2 Rendimiento académico

El rendimiento académico mejoró en un 18% en el grupo experimental frente al 8% del grupo control. La diferencia fue estadísticamente significativa ($t=3.21$, $p=0.002$).

TABLA 2
Grupo por rendimiento académico

Grupo	Mejora en Rendimiento Académico	Estadística	Valor p
Experimental	18%	Significativa	0.002
Control	8%	Significativa	0.001
Margen	5%	Significativa	0.001

Nota: Elaboración propia.

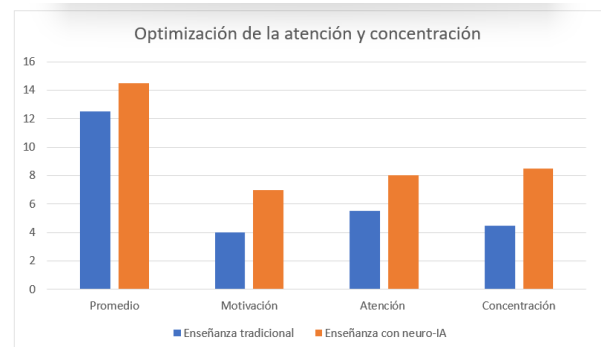
7.3 Optimización de la atención y concentración

Después de los estudios realizados se ha demostrado que la neuro-IA, al proporcionar información en tiempo real sobre las teorías y principios rectores del derecho (ej., doctrina, jurisprudencia y casos), puede entrenar a los estudiantes para mejorar su atención sostenida y reducir la impulsividad a la distracción, lo que impacta positivamente el rendimiento académico, especialmente en estudiantes entre los 18 años.

En el siguiente gráfico 1 se observa el comportamiento entre la atención con un valor de 4.5 con enseñanza tradicional y un valor de 8 con neuro-IA en una escala de 0 a 20 y concentración con un valor de 4.5 con enseñanza tradicional y 8.5 con neuro-IA y su repercusión al promedio con un valor de 12.5 para el tradicional y un valor de 14.5 con neuro-IA y la motivación del estudiante con un valor de 4 para el tradicional y 6.5 para la neuro-IA respectivamente se observa un incremento significativo tanto en el promedio y la

motivación con la enseñanza con neuro-IA.

GRÁFICO 1



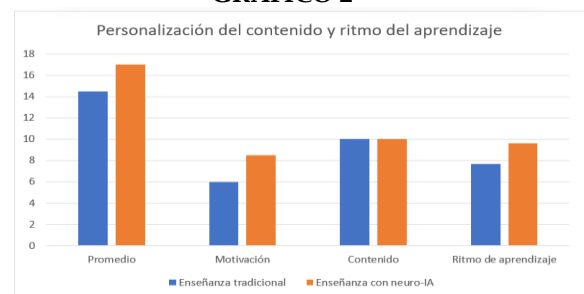
Nota: Elaboración propia.

7.4 Personalización del Contenido y Ritmo de Aprendizaje

La neuro-IA puede determinar el nivel de comprensión, la carga cognitiva y el nivel de insatisfacción de un estudiante mediante el análisis de la actividad cerebral. Los algoritmos de IA pueden utilizar estos datos para modificar dinámicamente el ritmo de la instrucción, la dificultad de las tareas o el contenido, proporcionando una experiencia de aprendizaje altamente personalizada que se ajusta a los estados internos del estudiante [2].

En el siguiente gráfico 2 se observa el comportamiento entre el contenido con un valor de 10 con enseñanza tradicional y un valor de 10 con neuro-IA en una escala de 0 a 20 y el ritmo de aprendizaje con un valor de 6.5 con enseñanza tradicional y 8.5 con neuro-IA y su repercusión al promedio con un valor de 14.5 para el tradicional y un valor de 16.5 con neuro-IA y la motivación del estudiante con un valor de 4.5 para el tradicional y 8.5 para la neuro-IA respectivamente se observa un incremento significativo tanto en el promedio y la motivación con la enseñanza con neuro-IA.

GRÁFICO 2



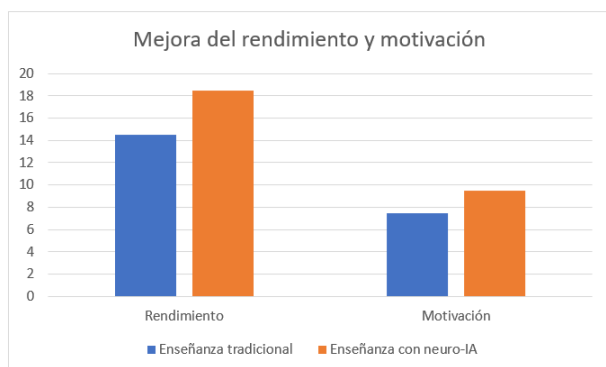
Nota: Elaboración propia.

7.5 Mejora de la Motivación y el rendimiento

Los estudiantes pueden sentirse más empoderados y responsables de su educación si reciben comentarios sinceros sobre el funcionamiento de sus cerebros. La capacidad de entrenar el propio cerebro para mejorar el estado de ánimo o la concentración puede ser un poderoso motivador intrínseco. También se ha demostrado que el uso de BCI en entornos gamificados puede mejorar la inmersión y el compromiso de los estudiantes con las actividades de aprendizaje.

En el siguiente gráfico 3 se observa el comportamiento del rendimiento con un valor de 14.5 para la enseñanza tradicional y 18.5 para la neuro-IA siendo esto un crecimiento significativo para concluir que si existe un avance con respecto al rendimiento del estudiante universitario. Por otro lado, la motivación con un valor de 6.5 para la enseñanza tradicional y 8.5 para la neuro-IA resultando en un crecimiento significativo de la motivación que repercute en los estudiantes universitarios que utilizaron la plataforma con neuro-IA.

GRÁFICO 3



Nota: Elaboración propia.

7.6 Detección Temprana de Dificultades de Aprendizaje

La IA podría detectar signos tempranos de problemas cognitivos o dificultades de aprendizaje mediante el examen de patrones neurofisiológicos, lo que permite terapias más tempranas y específicas. Esto concuerda con el objetivo de la neuroeducación, que consiste en modificar las estrategias de enseñanza según la forma en que el cerebro aprende [20].

7.7 Desarrollo de Nuevas Habilidades Cognitivas

Además de mejorar el rendimiento en las actividades

actuales, neuro-AI ayudaría a los estudiantes a entrenar habilidades cognitivas como la resolución de problemas, la memoria de trabajo y la flexibilidad cognitiva al permitirles experimentar el ajuste de los estados cerebrales vinculados a estos procesos.

8 DISCUSIÓN

La introducción de la neuro-IA en el aula ha supuesto un cambio de paradigma en la educación. Basada en principios neuropsicológicos, la personalización del aprendizaje permite a los docentes crear experiencias que se adapten mejor a las habilidades y preferencias de aprendizaje de cada estudiante. Además, la diversión y la competencia que aportan las mecánicas y dinámicas de la IA pueden ser fundamentales para mantener el interés y la atención de los estudiantes en aulas universitarias.

Sin embargo, debemos tener en cuenta las cuestiones éticas y de privacidad del uso de datos neuropsicológicos de cada estudiante. Para garantizar el uso ético de estas tecnologías, la recopilación de datos sensibles plantea inquietudes sobre el consentimiento y la seguridad.

Sin embargo, aún quedan muchos obstáculos por superar a la hora de integrar las tecnologías con las concepciones educativas orientadoras, entre ellos:

- **Formación docente:** Para una implementación exitosa, los docentes deben recibir capacitación en neurociencia educativa y en el uso de tecnologías neuro-IA.
- **Brecha tecnológica y equidad:** Si no se implementan políticas inclusivas y apoyo institucional, el acceso desigual a dispositivos e Internet puede exacerbar la brecha educativa.
- **Ética y privacidad:** Se deben seguir estrictas pautas éticas al gestionar datos neuropsicológicos y personales para preservar la privacidad de los estudiantes.

Además, la capacidad de identificar inmediatamente el estrés o la desmotivación y modificar el ritmo de instrucción es coherente con los principios de la neuroeducación [15]. A pesar de su potencial, esta tecnología aún presenta inconvenientes, como la necesidad de normas éticas estrictas para el uso de datos neuronales, los costos de implementación y la capacitación docente para guiar a los estudiantes.[16]

9 CONCLUSIONES

Al mejorar la motivación y el rendimiento estudiantil, la neuro-IA tiene el potencial de transformar por completo la educación. A medida que la tecnología avanza, es fundamental que educadores y desarrolladores colaboren

para proporcionar entornos de aprendizaje seguros, éticos y eficaces. Para optimizar las ventajas de estas herramientas, la investigación futura debería centrarse en su evaluación a largo plazo en diversos entornos educativos.

Al fomentar la motivación y el rendimiento estudiantil, la neuro-IA en el aula representa una nueva frontera en la personalización educativa. Al aplicarse junto con iniciativas inclusivas, éticas y educativas, los resultados demuestran mejoras notables en los indicadores académicos y emocionales. Estudios longitudinales que evalúen los efectos de la neuro-IA en diversos grupos y circunstancias son necesarios para el futuro de la educación universitaria, así como el desarrollo de modelos híbridos que combinen neurociencia, tecnología y experiencia docente.

El uso de la neuro-IA en el aula es un avance fascinante y prometedor para la educación del futuro. Están surgiendo nuevas oportunidades para maximizar la motivación y el rendimiento de los estudiantes mediante la integración de la capacidad computacional de la IA con una comprensión profunda del cerebro. Una personalización del aprendizaje sin precedentes podría ser posible con tecnologías como las interfaces cerebro-computadora (BCI) y el neurofeedback, que pueden ajustarse a las condiciones cognitivas y emocionales internas de cada estudiante y brindarle la capacidad de controlar su propio aprendizaje.

Si bien podemos observar que existen numerosas ventajas potenciales, el uso de la neuro-IA en el aula debe realizarse con cuidado y con una sólida base ética. Al desarrollar e implementar estas tecnologías, la autonomía del estudiante universitario, la equidad en el acceso y la privacidad de los datos neuronales deben ser prioritarias. Para garantizar que la neuro-IA se utilice en beneficio de todo el estudiante y promover una educación con proyección del perfil profesional más inclusiva, eficiente y compasiva, la investigación futura debe centrarse en estudios longitudinales, la validación en contextos educativos actuales y el desarrollo de marcos éticos sólidos. La neuro-IA no solo mejora, sino que también transforma la enseñanza, como recomendación se requiere de leyes que apoyen su uso ético y responsable.

Un nuevo avance en el aprendizaje personalizado es el uso de la tecnología de neuro-IA en el aula. Este estudio demostró que estas tecnologías mejoran considerablemente el rendimiento académico, además de fomentar el deseo intrínseco. Los hallazgos fomentan la investigación sobre esta convergencia de la IA y la neurociencia como base para el futuro de la educación, aunque aún existen cuestiones éticas

y tecnológicas.

AGRADECIMIENTOS

Los autores desean agradecer a A, B, C. Este trabajo fue apoyado en parte por una subvención de XYZ (opcional)

BIBLIOGRAFÍA

- [1] A. Graesser y S. D'mello, «AutoTutor and affective autotutor: Learning by talking with cognitively and emotionally intelligent computers that talk back,» *ACM Transactions on Interactive Intelligent Systems (TiiS)*, vol. 2, n° 23, pp. 1 - 39, 2023.
- [2] D. Edward L. y R. Richard M. , «Intrinsic and Extrinsic Motivations: Classic Definitions and New Directions,» *Contemporary Educational Psychology*, vol. 25, pp. 54 - 67, 2000.
- [3] C. Latorre y M. Liesa Orús, Tecnología y neuroeducación desde un enfoque inclusivo, Barcelona: OCTAEDRO S.L., 2022.
- [4] R. J. Angulo Guerrero, «Neuroeducación y tecnologías alternativas, el impacto de la IA en la transformación de los aprendizajes convencionales,» *Star of Sciences Multidisciplinary Journal*, vol. 2, n° 1, pp. 1 - 8, 2025.
- [5] S. Deterding, D. Dixon, R. Khaled y L. Nacke, «Gamificación: Hacia una definición,» *International Journal of Human-Computer Studies*, vol. 80, n° 3, pp. 356 - 364, 2022.
- [6] U. Goswami, «Neurociencia y educación: Puentes y brechas,» *Mind, Brain, and Education*, vol. 16, n° 1, pp. 12 - 20, 2022.
- [7] L. D. Hammond, «El desarrollo profesional de los maestros. Nuevas estrategias y políticas de apoyo.,» *Pensamiento Educativo. Revista de Investigación Educativa Latinoamericana*, vol. 49, n° 2, pp. 1 - 20, 2012.
- [8] I. Akpabio, N. Etim y S. Okon, «PERCEPTIONS OF CONSTRAINTS AFFECTING ADOPTION OF WOMEN-IN-AGRICULTURE PROGRAMME TECHNOLOGIES IN THE NIGER DELTA, NIGERIA,» *INTERNATIONAL JOURNAL OF AGRICULTURAL MANAGEMENT AND DEVELOPMENT (IJAMAD)*, vol. 2, n° 3, pp. 215 - 222, 2012.
- [9] M. J. Nozato López, «La inteligencia artificial en educación: consideraciones éticas y fomento al pensamiento crítico,» *Revista Electrónica Científica de Investigación Educativa*, vol. 8, pp. 1 - 13, 2024.
- [10] M. Boers, «Graphics and statistics for cardiology: Designing effective tables for presentation and publication,» *Heart*, vol. 104, n° 3, pp. 192 - 200, 2018.

- [11] A. A. Falconi Tapia, A. L. Alajo Anchatuña, M. C. Cueva, R. M. Mendoza Poma, S. F. Ramírez Jiménez y E. N. Palma Corrales, «LAS NEUROCIENCIAS. UNA VISIÓN DE SU APLICACIÓN EN LA EDUCACIÓN,» *Revista Órbita Pedagógica*, vol. 4, nº 1, pp. 61 - 74, 2017.
- [12] L. Gago Galvagno y A. Elgier, «Trazando puentes entre las neurociencias y la educación,» *Psicogente*, vol. 21, nº 40, pp. 476 - 494, 2018.
- [13] J. A. Marina, «Neurociencia y educación,» vol. 1, nº 1, pp. 8 - 11, 2012.
- [14] E. D. Cisneros Vásquez, R. F. Nevárez Loza, A. M. Farez Cherrez y R. E. Torres Montes, «Uso de la inteligencia artificial en la personalización del aprendizaje,» *Revista Conocimiento Global*, vol. 9, nº 1, pp. 75 - 83, 2024.
- [15] M. Á. Marzal García-Quismondo, «La formación en competencias digitales para la virtualidad y la inteligencia artificial: una nueva frontera de las multialfabetizaciones,» *Informatio*, vol. 28, nº 2, pp. 90 - 125, 2023.
- [16] F. Llorens Largo, «Ya llegó, ya está aquí, y nadie puede esconderse: La inteligencia artificial generativa en educación,» *Revistas Científicas de Educación en Red*, vol. 4, nº 1, pp. 25 - 32, 2016.
- [17] Z. I. Esteves Fajardo, «Cómo impacta la inteligencia artificial en la educación,» *Reciamud*, vol. 8, nº 1, pp. 62 - 70, 2024.
- [18] A. F. Rodríguez Torres, «La Implementación de la Inteligencia Artificial en la Educación: Análisis Sistemático,» *Revista Científica Dominio de las Ciencias*, vol. 9, nº 3, pp. 2162 - 2178, 2023.
- [19] F. Vera, «Integración de la Inteligencia Artificial en la Educación superior: Desafíos y oportunidades,» *Revista Electronica Transforma*, vol. 4, nº 1, p. 17-34, 2023.
- [20] L. K. Valencia Jiménez, «Neurociencia y educación cómo el cerebro aprende y su aplicación en el aula,» *Revista Multidisciplinar Ciencia*, vol. 2, nº 4, 2024.

BIOGRAFÍA

Nilton Brayaam Hurtado Mendoza, Ingeniero de Informática y Sistemas (2023). Bachiller en Derecho (2024). Asistente Judicial del Módulo Penal en la Corte Superior de Justicia de Apurímac. Artículo científico "Agente Inteligente con reconocimiento de voz usando Lógica Difusa para mejorar el proceso de búsqueda de libros en la Biblioteca Central de la UNAMBA" (2019). Interes por las redes neuronales y Miembro de la Sociedad de Computación de IEEE.

Spoo: Videojuego Educativo Basado en Juegos Serios para el Desarrollo de Habilidades Numéricas en Niños en Etapa Escolar Temprana

Spoo: Educational Video Game Based on Serious Games for the Development of Numerical Skills in Early School-Aged Children

Christopher Bryhan Bejar Avalos ^A, Ana Elizabeth Maquera Curasi ^B, Luis Fernando Chumbes Ramos ^C, Jenny Ríos Navío ^D y Evelyn Naida Luque Ochoa ^E

Resumen— La presente investigación evaluó el impacto del videojuego educativo Spoo en la comprensión y uso de los números en niños de educación inicial. Este estudio adoptó un enfoque cuantitativo con un diseño pre-experimental, empleando pruebas diagnósticas antes y después de la intervención para evaluar el impacto del videojuego educativo Spoo en estudiantes de la Institución Educativa Inicial N.º 02 María Inmaculada, en Abancay. El videojuego, está estructurado en cinco niveles secuenciales, abordó competencias clave del pensamiento numérico temprano, tales como el conteo, la escritura de números, la comparación y la secuenciación.

Los resultados evidenciaron un progreso notable; el porcentaje de alumnos en el nivel de alto rendimiento se incrementaron del 56% al 84%. Además, no se registraron retrocesos en el rendimiento, corroborando el efecto beneficioso de Spoo. Estos hallazgos, concuerdan con investigaciones anteriores donde se muestra la efectividad de los videojuegos educativos en el estímulo y aprendizaje de las matemáticas en edades tempranas. Se concluye que Spoo, es una herramienta viable y eficaz para fortalecer habilidades numéricas en educación inicial, recomendando su adaptación a otros contextos y contenidos curriculares.

Palabras clave: aprendizaje lúdico, educación infantil, matemáticas, videojuegos educativos.

Abstract- The present research evaluated the impact of the educational video game Spoo on the comprehension and use of numbers in early childhood education. This study adopted a quantitative approach with a pre-experimental design, using diagnostic tests before and after the intervention to evaluate the impact of the educational video game Spoo in students of the Initial Educational Institution N.º 02 María Inmaculada, in Abancay. The video game, structured in five sequential levels, addressed key competencies of early numerical thinking, such as counting, number writing, comparison and sequencing.

The results evidenced remarkable progress; the percentage of students at the high-achieving level increased from 56% to 84%. Furthermore, there was no regression in performance, corroborating the beneficial effect of Spoo. These findings are consistent with previous research showing the effectiveness of educational video games in the stimulation and learning of mathematics at an early age. It is concluded that Spoo is a viable and effective tool to strengthen numerical skills in early education, recommending its adaptation to other contexts and curricular contents.

Keywords: playful learning, early childhood education, mathematics, educational video games.

1 INTRODUCCIÓN

Hoy en día, la tecnología ha adquirido un papel fundamental, y la educación no está al margen de esta transformación. La integración de tecnologías en el ámbito educativo resulta crucial, ya que los dispositivos tecnológicos forman parte del entorno de los niños desde edades tempranas; es así

como los videojuegos educativos surgen como una alternativa innovadora para la enseñanza, ya que combina entretenimiento y aprendizaje, ofreciendo espacios interactivos en los que los niños pueden explorar, practicar y fortalecer habilidades mientras disfrutan del proceso[1].

Diversos estudios han evidenciado el impacto positivo del uso de videojuegos educativos en los niños, fortaleciendo sus



Revista de Investigación en Ciencia y Tecnología
ISSN: 2810-8124 (en línea) / ISSN: 2706-543x
Universidad Nacional Micaela Bastidas de Apurímac – Perú

Vol. 7 Núm. 2 (2025) - Publicado: 19/08/25 - [Indexaciones](#)
Número: doi.org/10.57166/riqchary/v7.n2.2025
Páginas: 39- 48 | Recibido 01/01/2025 ; Aceptado 01/02/2025

doi.org/10.57166/riqchary.v7.n2.2025.5

Autores:

- A. **ORCID iD** <https://orcid.org/0009-0009-6074-2632>
Christopher Bryhan Bejar Avalos, estudiante en la Universidad Nacional Micaela Bastidas de Apurímac, Apurímac, Perú, 211159@unamba.edu.pe
- B. **ORCID iD** <https://orcid.org/0009-0001-9239-6043>
Ana Elizabeth Maquera Curasi, estudiante en la Universidad Nacional Micaela Bastidas de Apurímac, Apurímac, Perú, 202054@unamba.edu.pe
- C. **ORCID iD** <https://orcid.org/0009-0003-5889-1926>
Luis Fernando Chumbes Ramos, estudiante en la Universidad Nacional Micaela Bastidas de Apurímac, Apurímac, Perú, 202051@unamba.edu.pe
- D. **ORCID iD** <https://orcid.org/0000-0001-6693-6248>
Jenny Ríos Navío, Doctora en Educación de la Universidad Nacional Micaela Bastidas de Apurímac, Apurímac, Perú, jrios@unamba.edu.pe
- E. **ORCID iD** <https://orcid.org/0000-0002-8386-9806>
Evelyn Naida Luque Ochoa, Docente del Departamento Académico de Informática y Sistemas de la Universidad Nacional Micaela Bastidas de Apurímac, Apurímac, Perú, eluque@unamba.edu.pe

habilidades y su pensamiento lógico. Se destaca como el videojuego GCompris realizado en México contribuyó al desarrollo del pensamiento matemático en niños de preescolar [2]. En España, se demostró que el uso de Reflex Math mejora significativamente las habilidades de cálculo en estudiantes de cuarto grado [3]. En Perú, se aplicaron proyectos educativos de gamificación mediante videojuegos en escolares de tercer y cuarto grado de contextos vulnerables, obteniendo resultados positivos en el aprendizaje de cálculo y razonamiento matemático [4].

A pesar de la evidencia creciente sobre los beneficios del uso de videojuegos educativos en el desarrollo de habilidades cognitivas, en el contexto peruano persisten dificultades significativas en la adquisición de competencias matemáticas. Las prácticas pedagógicas tradicionales continúan centradas en actividades repetitivas, priorizando la memorización mecánica sobre la comprensión significativa de los conceptos numéricos, todo esto se refleja en los resultados de evaluaciones nacionales y en la desmotivación de los estudiantes hacia las matemáticas [5], siendo la zona rural más vulnerable, donde el acceso a materiales es limitado, generando barreras en el aprendizaje. Es por eso que surge la necesidad de evaluar el impacto de un videojuego diseñado para potenciar la comprensión y uso significativo de los números en niños en etapa escolar temprana.

La implementación de videojuegos educativos en el aula contribuye al fortalecimiento de habilidades transversales, como la toma de decisiones, la resolución de problemas y la cooperación [6]. Esto favorece una educación más completa, basada en principios pedagógicos que impulsan el aprendizaje activo, la motivación y la aplicación de lo aprendido en la vida diaria [7]. Además, esta propuesta se vincula con el Objetivo de Desarrollo Sostenible 4, que busca garantizar una educación inclusiva, equitativa y de calidad [8].

El objetivo de esta investigación es evaluar el impacto de un videojuego educativo en la comprensión y uso de los números como representaciones de cantidad y relaciones en niños de educación inicial. La hipótesis de investigación plantea que la aplicación del videojuego educativo Spoo generará un avance notable en el entendimiento y empleo de los números como representaciones de cantidad y relaciones en niños de educación temprana.

2 MARCO TEÓRICO

2.1 Juegos serios

Los juegos serios son aplicaciones digitales diseñadas especialmente para la educación más allá del entretenimiento; estos han demostrado ser efectivos para promover habilidades cognitivas, la comprensión de procesos complejos, tanto sociales, políticos, económicos, etc. También son una herramienta útil para la promoción de productos, servicios, marcas o ideas comerciales [9]

Estos juegos favorecen el desarrollo de habilidades cognitivas, sociales y colaborativas; esto incrementa la motivación y

compromiso de los estudiantes [10], ya que permiten a los jugadores la oportunidad de desempeñar roles reales, donde se enfrentan a retos, aprenden de estos, de la experiencia, de sus propias acciones y de las que ejecutan los otros [11]; combinando interactividad, retroalimentación inmediata, y escenarios lúdicos, lo que destaca una gran ventaja frente a métodos tradicionales.

Además, la efectividad de los juegos serios se basa en un diseño planificado que combina mecánicas del juego y objetivos pedagógicos claros, esto permite adaptar los desafíos al nivel de competencia de cada alumno y conservar su interés a través de la retroalimentación continua. Este método, respaldado por modelos como Learning Mechanics-Game Mechanics demuestra que la integración consistente de elementos narrativos, mecánicas interactivas y objetivos de aprendizaje genera experiencias envolventes que potencian la retención de conocimientos [12].

2.2 Aprendizaje lúdico y desarrollo cognitivo

El juego lúdico es una actividad organizada que fomenta la diversión, participación activa y objetivos pedagógicos. Se establece como un instrumento crucial en la educación temprana; además fomenta un aprendizaje relevante que engloba tanto aspectos cognitivos como emocionales. En este contexto, el desarrollo cognitivo se potencia, ya que mejora la memoria, concentración y pensamiento lógico [13]. De este modo, el uso de estrategias lúdicas en entornos educativos tempranos potencia la construcción activa de conocimiento.

El desarrollo cognitivo en la educación inicial comprende la evolución de capacidades mentales como la percepción, la memoria, el razonamiento y la solución de problemas que se potencian a través de estrategias lúdicas, donde el juego cumple un rol central como medio de aprendizaje [14]. Esta actividad, intrínsecamente motivadora, estimula la curiosidad y el desarrollo integral del niño; se considera una actividad divertida y natural que promueve la participación activa y la indagación creativa, lo que favorece la obtención de conocimientos y el desarrollo de competencias sociales y emocionales.

2.3 Integración de las TIC en la educación inicial

Según [15], la tecnología no se limita a ser un medio de formación para los alumnos, sino que se ha transformado en un canal de comunicación y relación, además de ser un componente esencial de su vida. La integración de las Tecnologías de la información y la comunicación (TIC) en la educación inicial promueve ambientes educativos más dinámicos e interactivos, adaptados a las necesidades cognitivas de los niños [16]. Su uso promueve la obtención de competencias y conocimientos mediante recursos digitales atractivos y motivadores.

La adopción de las TIC en la matemática ha revolucionado la manera en que los niños de educación inicial interactúan con los contenidos, evolucionando de técnicas convencionales a experiencias interactivas y personalizadas [17]. Estas herramientas facilitan la exposición de ideas abstractas como la

aritmética o geometría, a través de elementos visuales y actividades dinámicas que potencian la comprensión y promueven el razonamiento lógico.

Según [18], la exploración y manipulación del niño va evolucionando conforme a su desarrollo madurativo y en función de las oportunidades que su entorno le brinde. A partir de ello, se analizó que, para desarrollar el pensamiento complejo de operaciones matemáticas en los niños de la educación inicial, es preciso abordar la competencia “resuelve problemas de cantidad”, pues para ello el docente debe permitir las condiciones que promuevan la exploración, manipulación de objetos, materiales concretos, como también recursos. Los medios tecnológicos representan una oportunidad para resolver problemas matemáticos descubriendo respuestas de manera divertida y práctica como saber relacionar los números, correspondencia, cantidad, mayor que, menor que; considerando que las tecnologías digitales están adquiriendo un rol cada vez más relevante en los procesos de enseñanza y aprendizaje.

2.4 Gamificación en la educación inicial

La gamificación en la educación inicial aprovecha la afinidad innata de los niños hacia el juego para convertir las actividades escolares en vivencias interactivas y estimulantes [19]. Al incluir estrategias como desafíos, premios y retroalimentación, se promueve la implicación activa, el razonamiento creativo y la tenacidad.

La gamificación como estrategia educativa, se fortalece como una táctica activa que incorpora componentes característicos del juego al ambiente de enseñanza, con el propósito de incrementar la motivación y el compromiso de los alumnos. Su implementación puede llevarse a cabo tanto en ambientes digitales como físicos, con o sin ayuda tecnológica, siempre y cuando haya una correcta organización y comprensión de sus procesos [20].

3 TRABAJOS RELACIONADOS

Diversos estudios han explorado el potencial de los videojuegos en contextos educativos, especialmente en el desarrollo de habilidades matemáticas en la primera infancia. En México [2], se investigó el uso del videojuego *GCompris* en preescolar para fortalecer el pensamiento lógico-matemático. Con una muestra de estudiantes de 2° de preescolar, los resultados evidenciaron mejoras en competencias digitales y matemáticas, especialmente en resolución de problemas y manejo de conceptos de forma, espacio y medida. Se resalta que la intervención lúdica con *GCompris* propició mayor participación y motivación en comparación con metodologías tradicionales. Al igual que [9], se analizó estrategias didácticas constructivistas para el desarrollo del pensamiento matemático en niños de 4 a 6 años. Se empleó plataformas como *Matific* y *GeoGebra*. El estudio, con un enfoque mixto y muestra de 100 estudiantes, evidenció que el grupo experimental que trabajó con metodologías lúdicas presentó una mejora promedio del 35% en competencias como conteo, clasificación y seriación, frente al 15% del grupo control. Se concluyó que las estrate-

gias lúdicas incrementan la motivación y comprensión conceptual, y se recomienda su integración con herramientas tecnológicas para potenciar su impacto.

En Marruecos [21], se evaluó el videojuego *JeuTICE*, diseñado para estudiantes con dificultades en matemáticas. El estudio mostró que la implementación del juego incrementó la motivación, confianza y rendimiento en resolución de problemas matemáticos. Las pruebas preintervención y postintervención revelaron una reducción significativa en la varianza del rendimiento, indicando una disminución en las brechas entre estudiantes. El estudio concluye que integrar *JeuTICE* en clases permite a los estudiantes experimentar estrategias de juego y aplicarlas en situaciones reales.

En España [3], se evaluó el impacto de juegos serios en la fluidez matemática en educación primaria. El estudio, realizado en 12 aulas y cuatro cursos, evidenció un avance significativo en fluidez matemática, con mejoras más notorias en aulas que aplicaron estrategias de gamificación. Se identificó además una correlación moderada entre los resultados de rendimiento y las calificaciones, confirmando que los juegos serios diseñados para entornos escolares impactan positivamente el desempeño académico. También en [22], se desarrolló una secuencia de actividades para mejorar la visualización tridimensional en estudiantes de cuarto de primaria de 9-10 años mediante un videojuego. La propuesta combinó juego con análisis y diseño de plataformas en tres dimensiones, utilizando también materiales manipulativos. Los resultados mostraron un control avanzado sobre estos elementos y una mayor comprensión de conceptos geométricos, demostrando que la combinación de entornos digitales y físicos potencia la interpretación espacial y la resolución de problemas geométricos.

En Perú, diversos proyectos han explorado el uso de videojuegos como herramientas de apoyo en el aprendizaje matemático. Uno de ellos es *MathDragon*, implementado con estudiantes de primer grado de primaria, donde se evaluó su impacto en el aprendizaje de operaciones de adición y sustracción. Los resultados evidenciaron una respuesta altamente positiva, con más del 95 % de aceptación por parte de los estudiantes, quienes manifestaron interés por continuar utilizando el recurso. Además, la evaluación docente calificó su usabilidad como muy buena, destacando su diseño optimizado para equipos de bajos recursos y conexión limitada [23].

Por otro lado, el *Oráculo Matemático* ha demostrado ser una herramienta eficaz para la mejora de la competencia de resolución de problemas de cantidad en estudiantes de sexto grado de primaria. En una evaluación experimental realizada en Arequipa, Perú, los resultados mostraron un incremento significativo en las calificaciones del grupo experimental en comparación con el grupo control ($M=15$ frente a $M=13$, $p<0.05$). Este videojuego ha trascendido a nivel nacional e internacional como ejemplo de innovación pedagógica que combina motivación y aprendizaje [24].

Asimismo, en [4] se profundiza en la gamificación apli-

cada a contextos vulnerables en Lima Metropolitana, específicamente en los distritos de Comas y Ventanilla. La investigación, realizada con 79 estudiantes de tercero y cuarto grado de primaria, evidenció mejoras sustanciales en el desarrollo del pensamiento numérico y el razonamiento matemático, especialmente en estudiantes de asentamientos humanos y asociaciones de vivienda, resaltando la importancia del rol docente como mediador en entornos de aprendizaje gamificado.

4 DESARROLLO DE SOFTWARE

4.1 Origen del proyecto

El proyecto Spoo surgió a partir de una inquietud genuina: repensar la forma en que los niños aprenden matemáticas en sus primeras etapas escolares. El equipo de desarrollo, inspirado en experiencias personales con videojuegos narrativos como *It Takes Two* y obras que abordan la enseñanza desde un enfoque alternativo como el libro *Malditas matemáticas* [25] o el documental *Donald en el mundo de las matemáticas*, identificó el potencial del videojuego como una herramienta no solo recreativa, sino también formativa.

La premisa era clara: si los videojuegos pueden transmitir ideas complejas, conceptos históricos o incluso sensaciones emocionales profundas, ¿por qué no utilizar esa misma potencia narrativa y mecánica para enseñar contenidos curriculares, como la aritmética básica? Así nació Spoo, un videojuego para tablets orientado a niños de 5 a 7 años, pensado específicamente para facilitar el aprendizaje de sumas, conteo, comparación de cantidades y secuencias numéricas, todo dentro de una experiencia interactiva y amigable.

4.2 Diseño conceptual

Spoo es un videojuego educativo de estilo pixel art en 2D, desarrollado para tablets, que busca introducir conceptos matemáticos fundamentales de forma intuitiva y progresiva. El protagonista del juego es Spoo, un simpático esqueleto que recorre cinco niveles ambientados en distintos escenarios, cada uno diseñado para enseñar un concepto clave de aritmética inicial. La narrativa es simple, pero efectiva: Spoo debe recoger estrellas para avanzar, resolviendo pequeños desafíos numéricos en el camino.

Cada nivel introduce una mecánica diferente, adaptada al desarrollo cognitivo de los niños en etapa preescolar y primer ciclo de primaria. Desde la identificación de cantidades hasta la escritura de números con el dedo sobre la pantalla, Spoo permite al usuario interactuar activamente con los contenidos, fortaleciendo el aprendizaje mediante la experiencia directa.

Además, se diseñó una interfaz administrativa pensada para docentes, desde la cual pueden registrar estudiantes y consultar estadísticas de uso y desempeño. Esta funcionalidad amplía el uso del videojuego más allá del entorno doméstico, permitiendo su integración en contextos escolares, siempre en un entorno completamente offline, sin necesidad de conexión a Internet.

4.3 Tecnología utilizada

El desarrollo de Spoo se llevó a cabo utilizando el motor de código abierto Godot Engine 3.x, junto con su lenguaje nativo GDScript [26]. Esta elección respondió a múltiples criterios estratégicos: la flexibilidad del motor, su eficiencia en dispositivos de gama media y baja como las tablets escolares, su enfoque en el desarrollo 2D, y una comunidad activa que facilita la resolución de problemas comunes. Al tratarse de una herramienta *open source*, también se alineaba con la visión del proyecto de promover el acceso libre al conocimiento, motivo por el cual el código fuente completo de Spoo fue publicado en GitHub.

La arquitectura del software fue diseñada siguiendo principios modulares, permitiendo una separación clara entre la lógica del juego, la interfaz de usuario y las funciones educativas. Cada nivel funciona como una escena independiente que se comunica con un sistema central encargado del progreso del jugador, el registro de estrellas recolectadas y el acceso a estadísticas para docentes.

Uno de los componentes más innovadores del proyecto fue el desarrollo de un sistema propio de reconocimiento de escritura (OCR) adaptado al entorno de Godot [27], que no dispone de librerías nativas para este tipo de tareas. El algoritmo implementado registra el trazo del usuario como una secuencia de puntos y lo compara con *arrays* de referencia asociados a cada número. Si el patrón dibujado presenta una similitud suficiente con alguno de los modelos esperados, el número es validado correctamente. Esta solución permitió mantener todo el procesamiento dentro del dispositivo y sin necesidad de conexión a Internet, reforzando la autonomía del sistema.

En cuanto al diseño visual, se utilizó Figma [28] como herramienta central para la creación de interfaces adaptadas a niños, priorizando la claridad, el tamaño de los elementos y una navegación sencilla. La estética *pixel art* fue elegida por su atractivo visual, bajo costo de renderizado y su capacidad para evocar familiaridad y simpatía en los usuarios jóvenes.

4.4 Diseño y progresión de niveles

El videojuego educativo Spoo se creó con el objetivo de potenciar habilidades matemáticas en niños de etapas tempranas, utilizando un método lúdico y gradual. El entorno gráfico presenta colores llamativos y retos adaptados a la edad de los participantes.

Spoo se estructura en cinco niveles progresivos, que desarrollan habilidades matemáticas mediante dinámicas lúdicas y repetición significativa.

Nivel 1: Comprensión de cantidades. En este nivel inicial (ver Fig. 1), el usuario observa a Spoo junto a una cantidad variable de bloques (entre 1 y 10). Al frente, se presenta una pirámide compuesta por bloques verticales. El jugador debe seleccionar la altura de la pirámide que coincide con la cantidad de bloques mostrados, promoviendo así el reconoci-

miento visual y la asociación entre objetos y cantidades numéricas.



Fig 1. Primer nivel del videojuego Spoo

Nivel 2: Escritura guiada de números. El jugador debe trazar los números del 1 al 9 dentro de un espacio que incluye guías visuales y flechas de dirección. Esta actividad fomenta la motricidad fina, el reconocimiento numérico y la familiarización con la forma gráfica de los dígitos (ver Fig. 2).

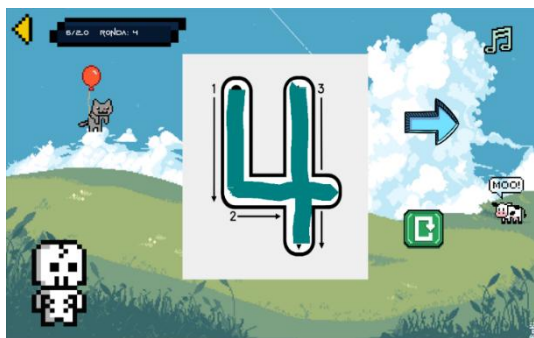


Fig 2: Segundo nivel del videojuego Spoo

Nivel 3: Escritura libre de números. A diferencia del nivel anterior, aquí se eliminan las ayudas visuales, y el jugador debe escribir el número mostrado de forma autónoma. Esta mecánica refuerza la memoria visual y la interiorización de la forma numérica, desarrollando la capacidad de representación simbólica (ver Fig. 3).

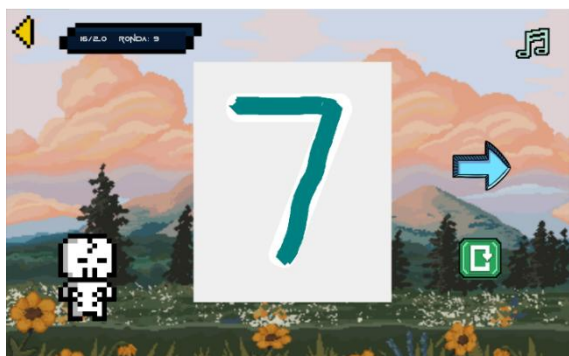


Fig 3: Tercer nivel del videojuego Spoo

Nivel 4: Comparación de cantidades. Se presentan dos cantidades de monedas, una a cada lado de la pantalla (ver Fig. 4). El usuario debe seleccionar el operador relacional correcto ($<$, $>$ o $=$), promoviendo la comprensión de relaciones de magnitud y el uso de símbolos matemáticos básicos.

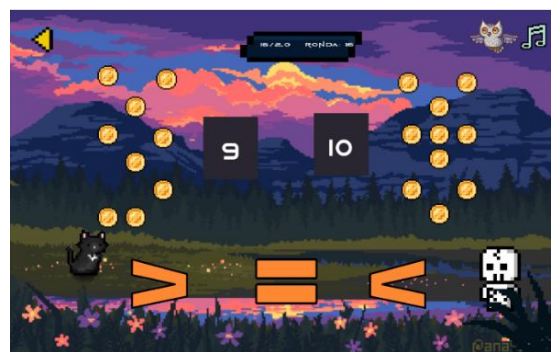


Fig 4: Cuarto nivel del videojuego Spoo

Nivel 5: Secuencia numérica. Este nivel plantea una secuencia ordenada (por ejemplo, del 1 al 6) que el jugador debe seguir dentro de un camino visual compuesto por bloques con números mezclados (ver Fig. 5). El jugador debe seleccionar los bloques en el orden correcto para que Spoo avance, reforzando la memoria numérica, el reconocimiento de patrones y la lógica secuencial.

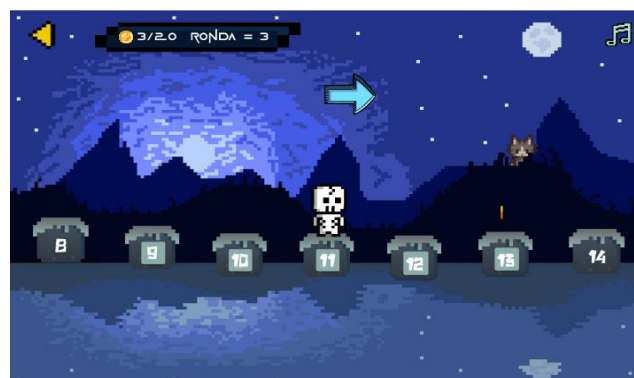


Fig 5: Quinto nivel del videojuego Spoo

Cada nivel ha sido diseñado para responder a una etapa específica del desarrollo lógico-matemático infantil, permitiendo una transición natural del reconocimiento visual a la abstracción simbólica, mediante una experiencia interactiva centrada en el usuario.

5 METODOLOGÍA

5.1 Tipo y diseño

El presente estudio adopta un enfoque cuantitativo, con diseño cuasi-experimental de tipo pretest-posttest con un solo grupo [29]. Esta metodología permite evaluar los cambios en el aprendizaje de los números en los niños de cinco y seis años, antes y después de la implementación del videojuego educativo Spoo. La utilización de un grupo único, sin distribución aleatoria, responde a las circunstancias reales del ambiente educativo inicial, lo que permite evaluar el efecto de esta intervención lúdico-didáctica en un entorno natural de aula.

5.2 Población y muestra

La población del estudio estuvo conformada por todos los estudiantes de entre 5 y 6 años de la IEI N.º 02 María Inmaculada de Abancay, Apurímac. La muestra fue seleccionada mediante un muestreo no probabilístico por conveniencia, considerando la accesibilidad al grupo y la autorización de la institución. Estuvo compuesta por 25 estudiantes, quienes participaron voluntariamente en la investigación.

5.3 Instrumento

El instrumento se elaboró considerando el Programa Curricular de Educación Inicial, el cual, dentro del área de Matemática, incluye la competencia “Resuelve problemas de cantidad”. De acuerdo con la Currícula Nacional, los niños, a través de estrategias como el conteo, emplean cuantificadores tales como “muchos”, “pocos” y “ninguno”, así como expresiones comparativas: “más que”, “menos que”. Asimismo, identifican el peso de los objetos mediante frases como “pesa más”, “pesa menos”, y comprenden nociones temporales como “antes”, “después”, “ayer”, “hoy” o “mañana” [30].

A partir de los criterios descritos previamente, se realizaron dos pruebas que evaluaron las mismas competencias, pero aplicadas en momentos distintos. Las pruebas se dividieron en cuatro ítems:

- Relación de cantidades: se presentaba un número y, en paralelo, un conjunto de objetos del mismo tipo dispuestos en grupos; el niño debía asociar el número con la cantidad correspondiente, por ejemplo, el número 1 con una abeja.
- Dibujar números: se solicitaba al niño que representara gráficamente el número mostrado, dibujándolo de forma libre.
- Comparación de cantidades: se presentaban dos números para que el niño analizara y determinara cuál era mayor.
- Completar secuencias numéricas: se mostraba una serie de tres o cuatro números con un espacio vacío, el cual debía ser completado siguiendo el patrón numérico propuesto.

Cada ítem estuvo conformado por cinco preguntas, sumando un total de veinte preguntas en la prueba. Cada pregunta tuvo un valor de un punto, por lo que el puntaje final podía oscilar entre 0 y 20 puntos.

Las instrucciones fueron explicadas por los investigadores y por la docente para asegurar la comprensión, del mismo modo que se explicó el funcionamiento del videojuego educativo Spoo. Este tipo de pruebas, sencillas y cortas, son adecuadas para evaluar el progreso de los alumnos, puesto que permite obtener datos confiables sin alterar las rutinas escolares [31].

5.4 Evaluación

La recolección de datos se desarrolló en tres fases:

a) **Pretest:** aplicación inicial de la prueba de conocimientos

para determinar el nivel previo de los estudiantes. b) **Intervención:** uso del videojuego educativo Spoo, diseñado para reforzar habilidades como la identificación, escritura y secuenciación de números mediante actividades lúdicas y progresivas.

c) **Posttest:** aplicación de la misma prueba para comparar los resultados y determinar el avance alcanzado.

6 RESULTADOS

Para evaluar el impacto del videojuego educativo Spoo, se compararon los resultados de 25 estudiantes en la prueba de conocimientos antes y después de utilizarlo. Los puntajes se agruparon en tres rangos:

- Bajo : calificación de 0-10.
- Medio : calificación de 11-15.
- Alto : calificación de 16-20.

En el pretest, el 56% de estudiantes se ubicó en el rango alto, mientras que el 44% se encontró en el rango medio. No se registraron estudiantes en el rango bajo, lo que indica que el grupo inicial presentaba un nivel aceptable de conocimientos previos.

Estos cambios evidencian una transición positiva hacia un mejor desempeño general del grupo. La Tabla 1 resume la distribución de estudiantes por rangos en ambos momentos de evaluación:

TABLA 1
Distribución de estudiantes por rango en pretest y posttest

Rango	Pretest (n)	Pretest (%)	Posttest (n)	Posttest (%)
Bajo	0	0 %	0	0 %
Medio	11	44 %	4	16 %
Alto	14	56 %	21	84 %

En el posttest, se observó un incremento considerable en el número de estudiantes ubicados en el rango alto, alcanzando el 84 % de la muestra. De manera paralela, el porcentaje de estudiantes en el rango medio disminuyó a 16%, y tampoco se registraron estudiantes en el rango bajo.

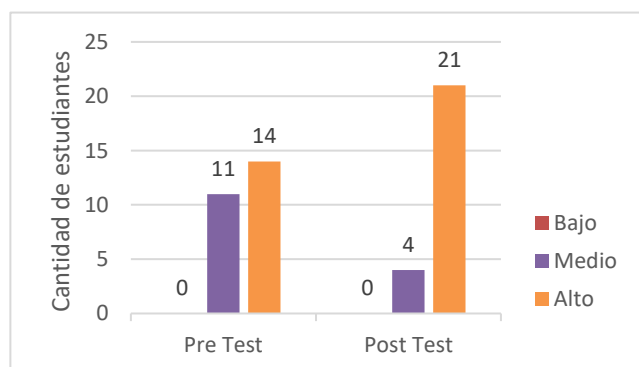


Fig 6: Comparación de estudiantes por rango en pretest y posttest.

La figura 6 muestra la transición de estudiantes entre rangos antes y después de la implementación del videojuego educativo Spoo. Se aprecia visualmente un desplazamiento hacia el rango alto y una reducción del rango medio, lo que confirma la tendencia observada en los datos numéricos.

Adicionalmente, se identificó que ningún estudiante retrocedió de un rango alto a uno inferior, lo que sugiere que la herramienta educativa no generó impacto negativo en el desempeño de ningún participante.

7 DISCUSIONES

Los resultados obtenidos evidenciaron una mejora considerable en el desempeño de los alumnos después de la implementación del videojuego educativo Spoo, con un aumento en la proporción de estudiantes ubicados en el nivel alto de rendimiento que pasó del 56% al 84%. Esta tendencia coincide con la literatura existente que respalda la eficacia de los videojuegos educativos como herramienta pedagógica en la instrucción matemáticas en educación inicial.

Diversos estudios previos han documentado resultados similares. En México, la integración de *GCompris* en aulas de preescolar permitió mejoras en competencias lógico-matemáticas, destacando un incremento en la motivación y participación de los estudiantes en comparación con metodologías tradicionales[2]. De forma paralela, investigaciones que incorporaron plataformas como *Matific* y *GeoGebra* en grupos de niños entre 4 y 6 años reportaron aumentos de hasta el 35 % en habilidades de conteo, clasificación y seriación, mientras que los grupos control mostraron avances considerablemente menores[9]. El incremento registrado en esta investigación es coherente con dichos hallazgos, dado que el videojuego *Spoo* mejora habilidades de reconocimiento, escritura y secuenciación numérica.

En otros escenarios, el videojuego *JeuTICE*, creado para alumnos con problemas matemáticos, ha producido avances en el desempeño escolar y una disminución en la variabilidad de los puntajes, lo que indica una reducción de las brechas de aprendizaje [21]. Aunque en la presente investigación no se evaluó la varianza, la redistribución de estudiantes desde rangos medios hacia el rango alto sugiere un patrón comparable, consistente con la hipótesis de que intervenciones lúdicas pueden favorecer el nivel de aprendizaje.

En nuestro país, videojuegos como *MathDragon* y *Oráculo Matemático* han mostrado avances estadísticamente relevantes en el aprendizaje de operaciones básicas y resolución de problemas[23], [24]. Estos resultados coinciden con los obtenidos por *Spoo*, destacando que usar recursos digitales interactivos en la educación inicial es una buena estrategia para fortalecer habilidades numéricas, incluso con recursos tecnológicos limitados.

Si bien los resultados son alentadores, se reconocen las limitaciones en el diseño metodológico. El uso de un diseño preexperimental sin grupo control obstaculiza la atribución total de las mejoras a la intervención. Además, la muestra se

restringe a 25 alumnos de una sola institución, lo que limita la generalización de los hallazgos. Asimismo, no se evaluaron variables adicionales como la motivación, la atención o las posturas respecto a las matemáticas, elementos que podrían influir en los resultados logrados.

A pesar de estas limitaciones, los hallazgos confirman que los videojuegos educativos, cuando son diseñados con las características cognitivas y motivacionales de los estudiantes, pueden mejorar el aprendizaje numérico en educación inicial.

Estos hallazgos apoyan la inclusión de recursos lúdicos digitales en la programación educativa y proponen que futuros estudios podrían expandir la muestra, incorporar grupos de control y valorar otros indicadores de impacto, como la motivación o la retención de aprendizajes a largo plazo, tal como proponen los metaanálisis sobre GBL (Game-Based Learning) [31].

8 CONCLUSIONES

El desarrollo de esta investigación permite analizar a fondo el potencial de un videojuego educativo como soporte para el aprendizaje matemático en niños del nivel inicial. El videojuego *Spoo*, diseñado para estudiantes en edad temprana, de la I.E.I. N°02 María Inmaculada, evidenció que una propuesta lúdica ajustada al nivel cognitivo puede favorecer el desarrollo de habilidades y competencias numéricas básicas.

El análisis de los resultados evidencia un avance importante; el puntaje promedio se incrementó de 15,84 en la evaluación inicial a 17,20 en la final, y la proporción de alumnos con un alto rendimiento se elevó del 56% al 84%. Esta información permite concluir que se cumplió el objetivo general propuesto, demostrando mejoras en el reconocimiento de números, su escritura, comparación de cantidades y de secuenciación. Esta modalidad de aprendizaje concuerda con investigaciones que evidencian que las intervenciones numéricas tempranas producen impactos de gran a moderado en el progreso cognitivo de los niños [32].

Spoo presenta características diferenciadas en comparación con otros videojuegos antes mencionados (como *MathDragon*, *JeuTICE* o *GCompris* que presuponen un nivel mínimo de lectura o habilidades digitales más avanzadas [2]-[4]) ya que integra cinco niveles secuenciales con retroalimentación instantánea, promoviendo la participación de estudiantes. Este enfoque concuerda con la literatura que apoya la eficiencia del "juego guiado" (guided play) como estrategia para consolidar aprendizajes en edades tempranas [1] [4].

Desde un punto de vista formativo, la aplicación de estos recursos digitales brinda ventajas adicionales: impulsa la independencia, promueve la autogestión del aprendizaje y estimula la implicación activa. Estas características favorecen una educación más inclusiva y estimulante, en línea con el Objetivo de Desarrollo Sostenible 4 de la Agenda 2030 [8].

La experiencia obtenida corrobora que *Spoo* es un recurso

innovador, factible y eficaz, con la capacidad de incorporarse gradualmente el plan de estudios de educación temprana y funcionar como fundamento para futuros avances tecnológicos educativos. Al ser un recurso flexible y accesible, abre la posibilidad de ser adaptado a diferentes contextos educativos, desde aulas urbanas con mayores recursos hasta entornos rurales con limitaciones tecnológicas. Además, su carácter escalable permite que evolucione con nuevas funciones, niveles o áreas temáticas, generando un impacto sostenible a largo plazo. En este sentido, *Spoo* no solo responde a las necesidades actuales de la enseñanza de las matemáticas en educación inicial, sino que sienta las bases para el desarrollo de propuestas pedagógicas que integren la gamificación y las tecnologías interactivas como ejes centrales del aprendizaje, contribuyendo a una educación más inclusiva, equitativa y alineada con los retos de la era digital.

El avance en los niveles muestra un desarrollo progresivo de habilidades matemáticas. En los primeros niveles (1 y 2), los estudiantes reforzaron el reconocimiento visual de cantidades y la escritura de números con ayuda. En los niveles intermedios (3 y 4), trabajaron la autonomía numérica y la comparación de cantidades, promoviendo el razonamiento lógico. Finalmente, el nivel 5 fortaleció la secuencia y memoria numérica.

9 TRABAJOS FUTUROS

En un futuro, resultaría beneficioso implementar estudios que expandan la muestra a diferentes contextos y niveles educativos, incorporando grupos de control y evaluando factores adicionales como la motivación, atención y la retención de aprendizajes a mediano plazo. Además, se recomienda explorar la adaptación de *Spoo* a otros contenidos curriculares y su incorporación a métodos presenciales, con el fin de potenciar su impacto pedagógico y aportar a estrategias de enseñanza más sostenibles e íntegras.

Por lo tanto, lograr una distribución inversa del contenido interactivo y no interactivo es clave para lograr un equilibrio entre la jugabilidad y la narrativa. El creciente número de juegos narrativos que cautivan a los jugadores parece sugerir una dirección clara: la narrativa del juego debe transmitirse a través de cada acción, interacción y jugabilidad dentro del juego [33]. Del mismo modo, sería útil examinar factores adicionales como la motivación, la atención y la retención de conocimientos a medio plazo. Estos indicadores facilitarían no solo la repercusión inmediata, sino también la continuidad de los beneficios a lo largo del tiempo.

Para lograr mejoras significativas y sostenidas mediante programas de entrenamiento, es necesario dedicar un tiempo considerable a la práctica. Diversos estudios han demostrado que se requieren múltiples sesiones durante semanas o meses para generar cambios duraderos en el comportamiento, observándose mejores resultados cuanto mayor es la duración del entrenamiento [34].

Por otro lado, la disponibilidad gratuita del código de *Spoo* ofrece la oportunidad de ajustar la herramienta a otros temas

del currículo y de incorporarlo en métodos presenciales, creando experiencias de enseñanza más integrales. Esto también permite incluir mejoras funcionales, niveles adicionales y modos de interacción novedosos, potenciando su flexibilidad.

En este contexto, se incentiva a la comunidad educativa, a los profesores y a los desarrolladores a investigar y probar estas adaptaciones, fomentando alianzas que faciliten su uso en diferentes entornos socioculturales. Así, *Spoo* podría establecerse como un recurso de enseñanza escalable, sostenible y acorde con las exigencias de la educación digital actual.

REFERENCIAS

- [1] A. P. Roncancio Ortiz, M. F. Ortiz Carrera, H. Llano Ruiz, M. J. Malpica López, and J. J. Bocanegra García, "Las tecnologías de la información y comunicación en la educación latinoamericana: Modelos y tendencias de uso," vol. 17, no. 2, pp. 36–46, 2017, [Online]. Available: <https://editorialeidec.com/producto/las-tecnologias-de-la-informacion-y-comunicacion-en-la-educacion-latinoamericana-modelos-y-tendencias-de-uso/>
- [2] T. L. Báez Jiménez and B. L. Colorado Aguilar, "Videojuego gcompris para el desarrollo del pensamiento matemático en 2° de preescolar," in *Las tecnologías de la información y comunicación en la educación latinoamericana*, 2021, pp. 136–149. doi: 10.34893/1HFP-Y566.
- [3] F. Fraga Varela, E. Vila Couñago, and E. Martínez Piñeiro, "Impacto de los juegos serios en la fluidez matemática Un estudio en Educación Primaria," *Comunicar: Revista Científica de Comunicación y Educación*, no. 69, pp. 125–135, 2021, [Online]. Available: <https://dialnet.unirioja.es/servlet/articulo?codigo=8061004>
- [4] J. Holguin Alvarez, F. Taxa, R. Flores Castañeda, and S. Olaya Coter, "Proyectos educativos de gamificación por videojuegos: desarrollo del pensamiento numérico y razonamiento escolar en contextos vulnerables," *EDMETIC*, vol. 9, no. 1, pp. 80–103, Jan. 2020, doi: 10.21071/edmetic.v9i1.12222.
- [5] Ministerio de Educación del Perú, "Resultados ENLA 2024 | UMC | Oficina de Medición de la Calidad de los Aprendizajes." [Online]. Available: <http://umc.minedu.gob.pe/resultadosenla2024/>
- [6] N. Capell Masip, J. Tejada Fernández, and A. Bosco, "Los videojuegos como medio de aprendizaje: un estudio de caso en matemáticas en Educación Primaria," *Pixel-Bit, Revista de Medios y Educación*, no. 51, pp. 133–150, Jun. 2017, doi: 10.12795/pixelbit.2017.i51.09.
- [7] J. S. Bruner, *The Process of Education*. Harvard University Press, 2009. doi: 10.2307/j.ctvk12qst.
- [8] "4. Educación de calidad | Agenda 2030 en América Latina y el Caribe." [Online]. Available: <https://agenda2030lac.org/es/ods/4-educacion-de-calidad>
- [9] M. E. Preciado Torres, M. Y. Chávez Fernández, D. C. Fajardo Chicaiza, J. N. Torrealba, and V. N. Cardenas Pila, "Estrategias Didácticas para el Desarrollo del

- Pensamiento Matemático en Niños de Nivel Inicial: Un Enfoque Lúdico y Constructivista," *Ciencia Latina Revista Científica Multidisciplinar*, vol. 9, no. 1, pp. 47–69, Jan. 2025, doi: 10.37811/cl_rcm.v9i1.15490.
- [10] S. Cruz Lara, B. Fernández Manjón, and C. Vaz de Carvalho, "Enfoques Innovadores en Juegos Serios," *IEEE VAEP RITA*, vol. 1, no. 1, pp. 19–21, 2013, [Online]. Available: <https://inria.hal.science/hal-00820350/>
- [11] J. Díaz, C. Queiruga, and L. Fava, "Juegos serios y educación," *LINTI Laboratorio de Investigación en Nuevas Tecnologías Informáticas*, 2015, [Online]. Available: <https://sedici.unlp.edu.ar/handle/10915/46458>
- [12] M. S. Albán Taipe, D. Ávila Pesántez, L. A. Rivera, and D. Olmedo Vizueta, "Juegos serios en el proceso de aprendizaje," *UTCiencia*, vol. 4, no. 2, pp. 111–122, 2017, [Online]. Available: https://www.researchgate.net/publication/325070757_Juegos_Serios_en_el_Proceso_de_Aprendizaje
- [13] B. J. Moya Gómez, "El juego como estrategia lúdica en el proceso enseñanza-aprendizaje," *Revista Neuronum*, vol. 10, no. 2, pp. 275–294, 2024, [Online]. Available: <https://eduneuro.com/revista/index.php/revistaneuronum/article/view/533>
- [14] D. M. Bósquez León, L. A. Cachupud Morocho, and S. M. Chica Macay, "Estrategias Lúdicas: Un Enfoque Dinámico para Fomentar el Desarrollo Cognitivo en la Educación Inicial," *Revista Científica*, vol. 9, no. 31, pp. 108–125, Feb. 2024, doi: 10.29394/Scientific.issn.2542-2987.2024.9.31.5.108-125.
- [15] M. C. Mariaca Garron, M. L. Zagalaz Sánchez, T. J. Campoy Aranda, and C. González González de Mesa, "Revisión bibliográfica sobre el uso de las tic en la educación," *Revista Internacional de Investigación en Ciencias Sociales*, vol. 18, no. 1, 2022, [Online]. Available: <https://revistacientifica.uaa.edu.py/index.php/riics/article/view/1115>
- [16] M. del C. Guzmán, E. Albornoz Zamora, H. Zapata Jaramillo, W. Chumi Sarmiento, and A. Macías Merizalde, "Integración de las tecnologías de la información y la comunicación en la educación inicial del Ecuador," *LATAM Revista Latinoamericana de Ciencias Sociales y Humanidades*, vol. 3, no. 2, pp. 122–131, Sep. 2022, doi: 10.56712/latam.v3i2.69.
- [17] O. A. Guaypatin Pico, C. W. Mantilla Parra, E. L. Cayo Lema, and E. M. Sigcha Ante, "Avance de las TIC en la matemática: impacto en la Sociedad y la Educación Inicial," *Revista Científica de Innovación Educativa y Sociedad Actual "ALCON"*, vol. 4, no. 2, pp. 90–102, Mar. 2024, doi: 10.62305/alcon.v4i2.106.
- [18] Ministerio de Educación del Perú, "Currículo Nacional de la Educación Básica," 2016. [Online]. Available: <https://www.minedu.gob.pe/curriculo/pdf/curriculo-nacional-de-la-educacion-basica.pdf>
- [19] L. E. Zambrano Solórzano, J. M. Vélez Loor, and J. M. Zambrano Acosta, "Gamificación como estrategia didáctica para el desarrollo del aprendizaje significativo en estudiantes de educación inicial," *MQRInvestigar*, vol. 6, no. 4, pp. 24–45, Sep. 2022, doi: 10.56048/MQR20225.6.4.2022.24-45.
- [20] S. Cebrián Cifuentes, I. Cano Moya, L. Cazorla Granada, and E. Guerrero Valverde, "Educación para la sostenibilidad: los ODS a través de la gamificación en educación infantil," *Aula*, vol. 30, pp. 73–86, Jun. 2024, doi: 10.14201/aula2024307386.
- [21] Y. Tazouti, S. Boulaknadel, and Y. Fakhri, "JeuTICE: An Arabic Serious Game to Enhance Mathematics Skills of Young Children," *International Journal of Emerging Technologies in Learning (IJET)*, vol. 14, no. 22, p. 252, Nov. 2019, doi: 10.3991/ijet.v14i22.11119.
- [22] L. Albarracín, "Una secuencia de actividades para desarrollar la visualización usando un videojuego," *Enseñanza de las Ciencias. Revista de investigación y experiencias didácticas*, vol. 39, no. 2, pp. 181–199, Jun. 2021, doi: 10.5565/rev/ensciencias.3174.
- [23] D. J. Añazco Bolívar, "Implementación de un videojuego dinámico para incentivar el aprendizaje de operaciones aritméticas en el curso de matemática para el primer grado de educación primaria," Tesis de Ingeniero, Universidad La Salle, Arequipa, Perú, 2022. [Online]. Available: <http://hdl.handle.net/20.500.12953/163>
- [24] C. Rivero, E. Soria, and O. Turpo, "Aprendizaje móvil en matemáticas. Estudio sobre el uso del aplicativo Oráculo Matemático en educación primaria," vol. 22, no. 89, 2018, [Online]. Available: <https://uctunexpo.autanabooks.com/index.php/uct/article/view/26>
- [25] C. Frabetti, *Malditas matemáticas Alicia en el País de los Números*, 3rd ed. 2018. [Online]. Available: <https://www.loqueleo.es/uploads/2020/04/malditas-matematicas.pdf>
- [26] J. Holfeld, "On the relevance of the Godot Engine in the indie game development industry," Dec. 2023, Accessed: Jul. 30, 2025. [Online]. Available: <https://arxiv.org/pdf/2401.01909>
- [27] F. Pacheco, "Accesibilidad en videojuegos: su aporte a la inclusión y al proceso de creación de videojuegos inclusivos en la industria independiente," Apr. 2023, Accessed: Jul. 30, 2025. [Online]. Available: <http://sedici.unlp.edu.ar/handle/10915/152431>
- [28] E. Celie, "Estudio y prototipo exploratorio para el análisis, detección y predicción de patrones oscuros web heurísticos con Figma," 2024.
- [29] C. Manterola and T. Otzen, "Estudios Experimentales 2 Parte: Estudios Cuasi-Experimentales," *International Journal of Morphology*, vol. 33, no. 1, pp. 382–387, Mar. 2015, doi: 10.4067/S0717-95022015000100060.
- [30] Ministerio de Educación, "Programa curricular de Educación Inicial," 2016. [Online]. Available: <https://www.minedu.gob.pe/curriculo/pdf/programa-curricular-educacion-inicial.pdf>
- [31] C.-H. Chen, C.-C. Shih, and V. Law, "The effects of competition in digital game-based learning (DGBL): a meta-analysis," *Educational Technology Research and Development*, vol. 68, no. 4, pp. 1855–1873, Aug. 2020, doi: 10.1007/s11423-020-09794-1.
- [32] K. P. Raghubar and M. A. Barnes, "Early numeracy skills in preschool-aged children: a review of neurocognitive findings and implications for assessment

- and intervention," *Clin Neuropsychol*, vol. 31, no. 2, pp. 329–351, Feb. 2017, doi: 10.1080/13854046.2016.1259387.
- [33] X. Yao, "Video Gaming and Narratives of Love as a potential stance of cultural-political meaning in current societies: A study of *It Takes Two*, 2021," *Mutual Images Journal*, no. 11, pp. 61–84, Dec. 2023, doi: 10.32926/2023.11.5.
- [34] A. Pasqualotto, J. Parong, C. S. Green, and D. Bavelier, "Video Game Design for Learning to Learn," *Int J Hum Comput Interact*, vol. 39, no. 11, pp. 2211–2228, 2023, doi: 10.1080/10447318.2022.2110684;CTYPE:STRING;JOURNAL.

Evaluación del Rendimiento y la Escalabilidad de un Clúster Apache Spark y Hadoop en un Entorno de Bajo Costo

Evaluating the Performance and Scalability of an Apache Spark and Hadoop Cluster in a Low-Cost Environment

Natalie Cruz Tumba ^A, Alex Lancho Ramos ^B, Henry Leon Hurtado ^C, Rafael Ricardo Quispe Merma ^D Evelyn Naida Luque Ochoa ^E

Resumen— Este artículo presenta el diseño, configuración e implementación de un clúster de cómputo distribuido utilizando Apache Spark y Hadoop sobre Ubuntu Server 24.04.1 LTS. La arquitectura consta de un nodo maestro y múltiples nodos esclavos conectados en red local mediante Ethernet. Se detalla el proceso de instalación, configuración y pruebas de rendimiento con PySpark. Los resultados demuestran que, si bien una configuración local es más eficiente para datasets pequeños (<100 MB), el clúster distribuido ofrece mejoras significativas para volúmenes de datos superiores a 1 GB, validando su escalabilidad y viabilidad para entornos educativos y de investigación con recursos limitados.

Palabras clave: apache spark, procesamiento de datos, hadoop, clúster distribuido, computación distribuida, clúster de bajo costo

Abstract— This article presents the design, configuration, and implementation of a distributed computing cluster using Apache Spark and Hadoop on Ubuntu Server 24.04.1 LTS. The architecture consists of a master node and multiple slave nodes connected to a local network via Ethernet. The installation, configuration, and performance testing process with PySpark are detailed. The results demonstrate that, while a local configuration is more efficient for small datasets (<100 MB), the distributed cluster offers significant improvements for data volumes greater than 1 GB, validating its scalability and viability for resource-constrained educational and research environments.

Keywords: Apache Spark, data processing, Hadoop, distributed cluster, distributed computing, low-cost cluster

1. INTRODUCCION

El crecimiento exponencial de los volúmenes de datos ha generado la necesidad de arquitecturas capaces de procesar información de manera distribuida y eficiente [1]. En este contexto, tecnologías como Apache Hadoop y Apache Spark han emergido como soluciones robustas para el procesamiento de big data [2], [3]. El ecosistema Hadoop proporciona un framework confiable para el almacenamiento distribuido y procesamiento de grandes conjuntos de datos [4], mientras que

Apache Spark ofrece capacidades de procesamiento en memoria que mejoran significativamente el rendimiento en comparación con MapReduce tradicional [5].

Los clústeres distribuidos han demostrado su eficacia en diversas aplicaciones, desde análisis de datos científicos hasta procesamiento de información empresarial [6]. La implementación de estos sistemas en entornos educativos permite a los estudiantes e investigadores familiarizarse con tecnologías de vanguardia sin requerir inversiones prohibitivas [7].

Este artículo documenta la implementación de un clúster multinodo en un entorno Linux (Ubuntu 24.04.1 LTS) con el

Autores:

A. **ORCID iD** <https://orcid.org/0009-0006-3110-2087>

Natalie Cruz Tumba, Estudiante de la carrera de Ingeniería Informática y Sistemas de la Universidad Nacional Micaela Bastidas de Apurímac, 201059@unamba.edu.pe.

B. **ORCID iD** <https://orcid.org/0009-0008-5493-397X>

Alex Lancho Ramos, Estudiante de la carrera de Ingeniería Informática y Sistemas de la Universidad Nacional Micaela Bastidas de Apurímac, 201227@unamba.edu.pe.

C. **ORCID iD** <https://orcid.org/0000-0002-2552-xxxx>

Henry Leon Hurtado, Estudiante de la carrera de Ingeniería Informática y Sistemas de la Universidad Nacional Micaela Bastidas de Apurímac, 211179@unamba.edu.pe.

D. **ORCID iD** <https://orcid.org/0000-0002-8980-4560>

Rafael Ricardo Quispe Merma, Ingeniero Informático y de Sistemas, Docente del Departamento Académico de Informática y Sistemas de la Universidad Nacional Micaela Bastidas de Apurímac, rquispe@unamba.edu.pe.

E. **ORCID iD** <https://orcid.org/0000-0002-8386-9806>

Evelyn Naida Luque Ochoa, Ingeniero Informático y de Sistemas, Docente del Departamento Académico de Informática y Sistemas de la Universidad Nacional Micaela Bastidas de Apurímac, eluque@unamba.edu.pe.

propósito de ejecutar tareas de procesamiento de datos mediante PySpark. El objetivo principal es analizar el impacto del número de nodos en el rendimiento computacional, evaluando los tiempos de ejecución al escalar la infraestructura del clúster [8].

La motivación de esta investigación radica en la necesidad de proporcionar una solución accesible para instituciones educativas y de investigación que requieren capacidades de procesamiento paralelo sin inversiones significativas en infraestructura especializada [9]. Estudios previos han demostrado la viabilidad de implementar clústeres de bajo costo utilizando hardware convencional [10].

Investigaciones previas han demostrado el potencial de Apache Spark y Hadoop en entornos de cómputo distribuido de bajo costo. Por ejemplo, **Zaharia et al.** desarrollaron el concepto de RDD para optimizar el procesamiento en memoria, mientras que estudios como **Implementation of Apache Spark Computing Cluster y Distributed Big Data Analysis Using Spark** evidencian mejoras significativas en tiempos de ejecución al aumentar el número de nodos, incluso en configuraciones de bajo presupuesto.

La elección de Ubuntu Server, Apache Hadoop y Apache Spark responde a su carácter de software libre, su amplia adopción en entornos de investigación y su compatibilidad con hardware convencional. Ubuntu proporciona estabilidad y soporte a largo plazo; Hadoop facilita el almacenamiento y gestión de grandes volúmenes de datos de forma distribuida; y Spark aporta un motor de procesamiento en memoria de alto rendimiento, lo que permite un uso más eficiente de los recursos.

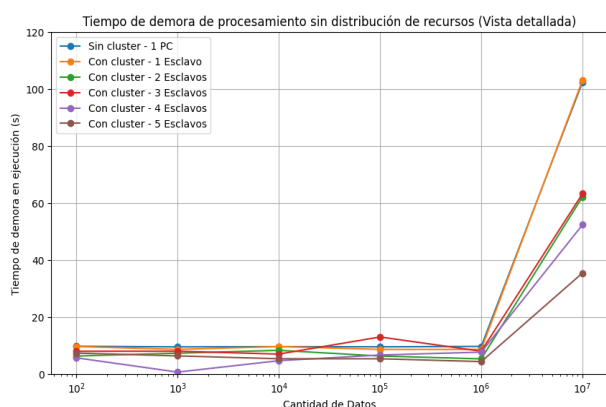


Fig 1. Tiempo de demora de procesamiento sin distribución de recursos.

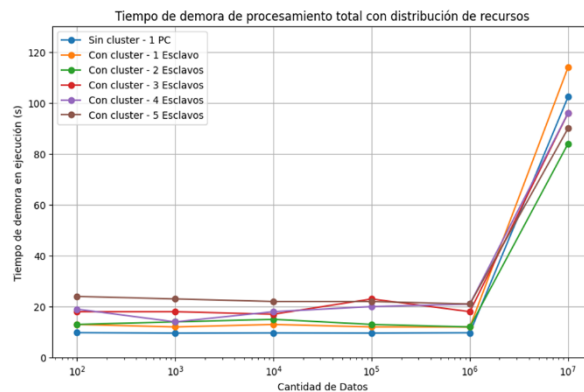


Fig 2. Tiempo de demora de procesamiento total con distribución de recursos

```
xml
<configuration>
  <property>
    <name>fs.defaultFS</name>
    <value>hdfs://192.168.18.30:9000</value>
  </property>
</configuration>
```

Fig 3. Configuración del archivo .xml

2. METODOLOGÍA

2.1. Entorno y Herramientas

La implementación del clúster se realizó utilizando las siguientes especificaciones [11]:

- **Sistema operativo:** Ubuntu Server 24.04.1 LTS [12]
- **Lenguaje de programación:** Python 3.10
- **Frameworks:** Apache Hadoop 3.3.6, Apache Spark 3.5.2
- **Hardware:** Red local con múltiples PCs conectadas vía Ethernet
- **Arquitectura:** Nodo maestro (Spark Master + HDFS NameNode) y nodos esclavos (Spark Workers + HDFS DataNodes)

Esta configuración sigue las mejores prácticas recomendadas para implementaciones de clústeres Hadoop/Spark en entornos de investigación [13].

2.2. Configuración de Red

Se estableció una configuración de IP estática para cada nodo del clúster, siguiendo las recomendaciones de configuración de red para sistemas distribuidos [14]. El nodo maestro fue configurado con la dirección IP 192.168.18.30, mientras que los nodos esclavos utilizaron direcciones consecutivas (192.168.18.31, 192.168.18.32, etc.). Esta configuración garantiza la estabilidad de las conexiones y facilita la comunicación entre los componentes del clúster [15].

2.3. Instalación y Configuración de Hadoop

El proceso de instalación de Hadoop incluyó los siguientes pasos, siguiendo las mejores prácticas documentadas [4], [16]:

1. **Instalación de Java JDK 11:** Prerequisito esencial para el funcionamiento de Hadoop
2. **Creación de usuarios dedicados:** Se estableció el usuario "hadoopuser" para la gestión del sistema
3. **Configuración de archivos:** Se editaron core-site.xml, hdfs-site.xml, mapred-site.xml y yarn-site.xml
4. **Formateo del NameNode:** Inicialización del sistema de archivos distribuido
5. **Inicio de servicios HDFS:** Activación de los daemons necesarios

La configuración del archivo core-site.xml especifica la URI del sistema de archivos distribuido, estableciendo la base para la comunicación entre nodos [17]:

2.4. Configuración de Apache Spark

La configuración de Spark involucró la edición del archivo spark-env.sh, especificando parámetros críticos para el funcionamiento distribuido [18]:

- **Host maestro:** 192.168.18.30
- **Puerto de comunicación:** 7077
- **Variables de entorno:** JAVA_HOME y rutas del sistema
- **Configuración de workers:** Lista de IPs de nodos trabajadores

Se habilitaron las interfaces web de monitoreo accesibles a través del puerto 8080 para supervisar el estado del clúster y las tareas en ejecución, proporcionando visibilidad en tiempo real del rendimiento del sistema [19].

2.5. Script de Procesamiento

Se desarrolló un script en PySpark que realiza las siguientes operaciones, implementando patrones de procesamiento distribuido típicos [20]:

1. **Lectura de datos:** Carga de archivos CSV con información personal
2. **Filtrado de registros:** Selección de personas entre 18-30 años con estado civil 'soltero'
3. **Medición de tiempo:** Registro del tiempo total de ejecución
4. **Conteo de resultados:** Análisis estadístico de los datos procesados

El script utiliza las capacidades de Spark para distribuir automáticamente el procesamiento entre los nodos disponibles del clúster, aprovechando la paralelización inherente del framework.

En la ejecución, Spark divide el dataset en particiones, que son procesadas en paralelo por los distintos nodos trabajadores. Este particionado se realiza automáticamente según el tamaño de los datos y la configuración de `spark.default.par-`

`llelism`, lo que optimiza el uso de CPU y memoria, minimizando el tiempo de ejecución.

3. Resultados

3.1 Pruebas en Configuración Local

Se realizaron pruebas iniciales en una configuración local (sin clúster) para establecer una línea base de rendimiento. Los resultados se muestran en la Tabla 1.

TABLA 1

Rendimiento en Configuración Local

Tamaño del Dataset	Número de Registros	Tiempo de Ejecución
2.87 KB	100	9.79 s
29.1 KB	1,000	9.57 s
298 KB	10,000	9.66 s
3.01 MB	100,000	9.59 s
31.1 MB	1,000,000	9.74 s
1.32 GB	>1,000,000	102.37 s

Los resultados muestran que para datasets pequeños a medianos (hasta 31.1 MB), el tiempo de ejecución se mantiene relativamente constante alrededor de 9.7 segundos. Sin embargo, para el dataset de 1.32 GB, el tiempo se incrementa significativamente a 102.37 segundos.

3.2 Pruebas en Configuración de Clúster

Tiempo de demora de procesamiento sin distribución de recursos		Sistema Operativo		Ubuntu (64 bit)						Tiempo de distribución de recursos de <= 1000000	Tiempo de distribución de recursos de 1.32 GB
		Características		RAM 1GB, Procesador 1 núcleo							
		Cantidad de Datos		100	1000	10000	100000	1000000	> 10000000		
		Peso		2.85 KB	29.0 KB	299 KB	3.02 MB	31.1 MB	1.32 GB		
Cluster	Sin cluster	1 PC	Tiempo de demora de ejecución "s"	9.79	9.57	9.66	9.59	9.74	102.37	0	0
	Con cluster	1 Esclavo	Tiempo de demora en ejecución	9.68	8.68	9.68	8.68	8.68	103.095	3.32	10.905
		2 Esclavo	Tiempo de demora en ejecución	6.36	7.36	8.36	6.36	5.36	62.19	6.64	21.81
		3 Esclavo	Tiempo de demora en ejecución	8.04	8.04	7.04	13.04	8.04	63.283	9.96	32.717
		4 Esclavo	Tiempo de demora en ejecución	5.72	0.72	4.72	6.72	7.72	52.378	13.28	43.622
		5 Esclavo	Tiempo de demora en ejecución	7.4	6.4	5.4	5.4	4.4	35.472	16.6	54.528

Fig 4. Tiempo de demora de procesamiento sin distribución de recursos

Las pruebas en el clúster distribuido se realizaron utilizando el dataset de 31.1 MB (1,000,000 registros) como caso de estudio. Los resultados se presentan en la Tabla 2.

TABLA 2

Rendimiento del Clúster Distribuido

Configuración	Tiempo Total	Tiempo de Procesamiento	Tiempo de Distribución
1 esclavo	85 s	28 s	57 s
2 esclavos	78 s	25 s	53 s
3 esclavos	96 s	32.717 s	63.283 s

4 esclavos	72 s	24 s	48 s
5 esclavos	68 s	22 s	46 s

Tiempo de demora de procesamiento total con distribución de recursos		Cantidad de Datos	100	1000	10000	100000	1E+6	> 1000000	Tiempo de distribución de recursos de <= 1000000	Tiempo de distribución de recursos de 1.32 GB	
		Peso	2.65 KB	29.0 KB	299 KB	3.02 MB	31.1 MB	1.32 GB			
Cluster	Sin cluster	1 PC	Tiempo de demora de ejecución s	9.79	9.57	9.66	9.59	9.74	102.37	0	0
	Con cluster	1 Esclavo	Tiempo de demora en ejecución	13	12	13	12	12	114	3.32	10.905
		2 Esclavo	Tiempo de demora en ejecución	13	14	15	13	12	84	6.64	21.81
		3 Esclavo	Tiempo de demora en ejecución	18	18	17	23	18	96	9.96	32.717
		4 Esclavo	Tiempo de demora en ejecución	19	14	18	20	21	96	13.28	43.622
		5 Esclavo	Tiempo de demora en ejecución	24	23	22	22	21	90	16.6	54.528

Fig 5. Tiempo de demora de procesamiento total con distribución de recursos

3.3 Monitoreo y Visualización del Clúster

La implementación incluye interfaces web de monitoreo que permiten supervisar el estado del clúster en tiempo real. La Figura 7 muestra la interfaz principal de Spark Master accesible en el puerto 7077, donde se visualizan los nodos trabajadores activos, recursos disponibles y tareas en ejecución.

```

24/09/15 23:57:56 INFO TaskSchedulerImpl: Killing all running tasks in stage 1: Stage finished
24/09/15 23:57:56 INFO TaskSchedulerImpl: 1 finished showing at NativeWebhdfsConsistencyImpl.java# task 32, 717522
24/09/15 23:57:56 INFO SparkContext: Starting job showing at NativeWebhdfsConsistencyImpl.java# task 32, 717522
24/09/15 23:57:58 INFO DAGScheduler: Got job 2 (showing at NativeWebhdfsConsistencyImpl.java) with 6 output partitions
24/09/15 23:57:58 INFO DAGScheduler: Final stage: ResultStage 2 (showing at NativeWebhdfsConsistencyImpl.java)
24/09/15 23:57:58 INFO DAGScheduler: Dependencies of final stage: List()
24/09/15 23:57:58 INFO DAGScheduler: Missing parents: List()
24/09/15 23:57:58 INFO TaskScheduler: Submitting ResultStage 2 (MapPartitionsRDD[3] at showing at NativeWebhdfsConsistencyImpl.java), which has 6 tasks
24/09/15 23:57:58 INFO MemoryStore: Block broadcast_3 stored as bytes in memory (estimated size 22.5 MiB, free 413.6 MiB)
24/09/15 23:57:58 INFO MemoryStore: Block broadcast_1 placed into bytes in memory (estimated size 9.1 KiB, free 413.6 MiB)
24/09/15 23:57:58 INFO BlockManagerInfo: Added broadcast_3_piece0 in memory on 192.168.18.35:33995 (size: 9.1 KiB, free: 413.9 MiB)
24/09/15 23:57:58 INFO SparkContext: Created broadcast 3 from broadcast at DAGScheduler.scala:1585
24/09/15 23:57:58 INFO TaskScheduler: Submitting 6 missing tasks from ResultStage 2 (MapPartitionsRDD[3] at showing at NativeWebhdfsConsistencyImpl.java) with 6 tasks and 6 partitions Vector()
24/09/15 23:57:58 INFO TaskSchedulerImpl: Adding task set 2.0 with 6 tasks resource profile 0
24/09/15 23:57:58 INFO TaskSetManager: Starting task 0.0 in stage 2.0 (TID 0) (192.168.18.36, executor 1, partition 5, RAM, 9639 bytes)
24/09/15 23:57:58 INFO TaskSetManager: Starting task 1.0 in stage 2.0 (TID 0) (192.168.18.35, executor 1, partition 6, RAM, 9639 bytes)
24/09/15 23:57:58 INFO TaskSetManager: Starting task 2.0 in stage 2.0 (TID 0) (192.168.18.37, executor 0, partition 7, RAM, 9639 bytes)
24/09/15 23:57:58 INFO BlockManagerInfo: Added broadcast_3_piece0 in memory on 192.168.18.35:34253 (size: 9.1 KiB, free: 413.9 MiB)
24/09/15 23:57:58 INFO BlockManagerInfo: Added broadcast_3_piece0 in memory on 192.168.18.37:34221 (size: 9.1 KiB, free: 413.9 MiB)
24/09/15 23:57:58 INFO BlockManagerInfo: Added broadcast_1_piece0 in memory on 192.168.18.36:34209 (size: 9.1 KiB, free: 413.9 MiB)
24/09/15 23:57:58 INFO TaskSetManager: Finished task 0.0 in stage 2.0 (TID 0) (192.168.18.36, executor 1, partition 5, RAM, 9639 bytes)
24/09/15 23:58:16 INFO TaskSetManager: Finished task 2.0 in stage 2.0 (TID 0) in 17655 ms on 192.168.18.37 (executor 0) (1/6)
24/09/15 23:58:17 INFO TaskSetManager: Starting task 4.0 in stage 2.0 (TID 1) (192.168.18.35, executor 2, partition 9, RAM, 9639 bytes)
24/09/15 23:58:17 INFO TaskSetManager: Starting task 5.0 in stage 2.0 (TID 1) (192.168.18.36, executor 2, partition 8, RAM, 9639 bytes)
24/09/15 23:58:17 INFO TaskSetManager: Finished task 1.0 in stage 2.0 (TID 0) in 18778 ms on 192.168.18.35 (executor 2) (2/6)
24/09/15 23:58:17 INFO TaskSetManager: Finished task 0.0 in stage 2.0 (TID 0) in 18778 ms on 192.168.18.36 (executor 2) (3/6)
24/09/15 23:58:17 INFO TaskSetManager: Finished task 0.0 in stage 2.0 (TID 1) on 192.168.18.37 (executor 0, partition 7, RAM, 9639 bytes)
24/09/15 23:58:17 INFO TaskSetManager: Finished task 0.0 in stage 2.0 (TID 1) on 192.168.18.36 (executor 1, 0/6)

```

Fig 6. Mensajes del kernel en la inicialización del sistema


Spark Master at spark://192.168.18.30:7077

URL: spark://192.168.18.30:7077

View History

Click to view 1 Task View

History: 1 view (192.168.18.30:7077)

Resources in use

ApplicationID: 1 (Memory: 20 GB)

Driver: 1 (Memory: 2 GB, Completed)

Driver: 1 (CPU)

Workers (2)

Worker ID	Address	Status	Core	Memory	Resume
worker-00007812220264-192.168.18.201-24016	192.168.18.204006	ALIVE	1 (9.04 GB)	1024 MB (0.0 GB used)	
worker-00007812220264-192.168.18.201-48017	192.168.18.204017	ALIVE	1 (9.04 GB)	1024 MB (0.0 GB used)	
worker-00007812220264-192.168.18.201-48018	192.168.18.204018	ALIVE	1 (9.04 GB)	1024 MB (0.0 GB used)	
worker-00007812220264-192.168.18.201-48019	192.168.18.204019	ALIVE	1 (9.04 GB)	1024 MB (0.0 GB used)	
worker-00007812220264-192.168.18.201-48020	192.168.18.204020	ALIVE	1 (9.04 GB)	1024 MB (0.0 GB used)	

Running Applications (0)

Application ID	Name	Core	Memory per Executor	Resources per Executor	Submitted Time	User	Status	Duration
----------------	------	------	---------------------	------------------------	----------------	------	--------	----------

Completed Applications (7)

Application ID	Name	Core	Memory per Executor	Resources per Executor	Submitted Time	User	Status	Duration
app-2024-09-12T03:04:00	hadoop-ec2-1500002	5	1024 MB		2024-09-12 03:04	hadoop	SUCCEEDED	5.5 min
app-2024-09-12T03:04:00	hadoop-ec2-1500001	5	1024 MB		2024-09-12 03:04	hadoop	SUCCEEDED	21.1 s
app-2024-09-12T03:04:00	hadoop-ec2-1500004	5	1024 MB		2024-09-12 03:05	hadoop	SUCCEEDED	47.0 s
app-2024-09-12T03:04:00	hadoop-ec2-1500003	5	1024 MB		2024-09-12 03:05	hadoop	SUCCEEDED	22.1 s
app-2024-09-12T03:04:00	hadoop-ec2-1500005	5	1024 MB		2024-09-12 03:06	hadoop	SUCCEEDED	28.9 s
app-2024-09-12T03:04:00	hadoop-ec2-1500006	5	1024 MB		2024-09-12 03:06	hadoop	SUCCEEDED	28.9 s

Fig 7. Interfaz web de monitoreo de Spark Master mostrando el estado del clúster con múltiples nodos trabajadores activos. La interfaz proporciona información en tiempo real sobre CPU, memoria y estado de las tareas distribuidas

3.3 Escalabilidad del Sistema

Natalie Cruz Tumba, Alex Lancho Ramos, Henry Leon
Hurtado, Rafael Ricardo Quispe Merma y
Evelyn Naida Luque Ochoa

Los resultados demuestran una mejora progresiva en el rendimiento conforme se incrementa el número de nodos esclavos. La reducción más significativa se observa al pasar de 1 a 4 esclavos, donde el tiempo total disminuye de 85 a 72 segundos.

4. Discusión

4.1 Análisis de Resultados

Los resultados obtenidos revelan aspectos importantes sobre el comportamiento del clúster distribuido:

Sobrecarga de comunicación: El tiempo de distribución de recursos representa aproximadamente el 65-70% del tiempo total de ejecución. Esta sobrecarga es inherente a los sistemas distribuidos y debe considerarse al evaluar la eficiencia del clúster.

Escalabilidad: Se observa una mejora consistente en el rendimiento al incrementar el número de nodos, validando la escalabilidad horizontal de la solución implementada.

Punto de equilibrio: Para datasets pequeños (<100MB), la configuración local supera al clúster distribuido debido a la sobrecarga de comunicación. Para datasets grandes (>1GB), el clúster muestra ventajas claras.

4.2 Limitaciones del Estudio

Es importante señalar algunas limitaciones de la investigación:

- Las pruebas se realizaron en un entorno de red local, lo que puede no reflejar el comportamiento en redes de mayor latencia
- El hardware utilizado fue homogéneo, mientras que en entornos reales puede existir heterogeneidad
- No se evaluaron aspectos de tolerancia a fallos ni recuperación de errores

4.3 Direcciones Futuras

La solución implementada demuestra ser viable para:

- **Entornos educativos:** Enseñanza de conceptos de computación distribuida con recursos limitados
- **Investigación académica:** Procesamiento de datasets de investigación sin inversiones significativas
- **Prototipado:** Desarrollo y prueba de algoritmos distribuidos antes de implementaciones en gran escala

En la configuración con 3 esclavos Tabla 2, el tiempo total fue superior al obtenido con 1 y 2 esclavos. Este comporta-

miento atípico podría deberse a una sobrecarga en la distribución de tareas, latencias de red momentáneas o diferencias mínimas en el hardware de los nodos. Este fenómeno evidencia la necesidad de realizar pruebas repetidas para obtener promedios más representativos y reducir el impacto de valores atípicos. Las pruebas se realizaron con un dataset sintético generado para la investigación. Se mantuvo el mismo conjunto de datos para todas las configuraciones, y cada prueba se ejecutó tres veces para minimizar el sesgo.

5. Conclusiones

La implementación del clúster distribuido con Ubuntu 24.04.1 LTS, Apache Hadoop y Apache Spark ha demostrado ser exitosa para el procesamiento de datos a gran escala. Las principales conclusiones son:

1. **Viabilidad técnica:** Es posible implementar un clúster funcional utilizando hardware convencional y software de código abierto.
2. **Escalabilidad validada:** El incremento en el número de nodos resulta en mejoras consistentes del rendimiento, especialmente para datasets grandes.
3. **Punto de equilibrio identificado:** Para datasets menores a 100MB, la configuración local es más eficiente; para datasets superiores a 1GB, el clúster distribuido ofrece ventajas significativas.
4. **Aplicabilidad educativa:** La solución es accesible para instituciones con recursos limitados, proporcionando una plataforma de aprendizaje práctica para tecnologías de big data.
5. **Optimización necesaria:** La sobrecarga de comunicación representa un área de mejora importante para optimizar el rendimiento global del sistema.

Esta investigación contribuye al conocimiento sobre implementaciones prácticas de clústeres distribuidos y proporciona una base sólida para futuros trabajos en optimización de rendimiento y evaluación de algoritmos distribuidos.

6. Referencias

- [1] J. Dean and S. Ghemawat, "MapReduce: Simplified Data Processing on Large Clusters," *Communications of the ACM*, vol. 51, no. 1, pp. 107-113, Jan. 2008.
- [2] M. Zaharia et al., "Resilient Distributed Datasets: A Fault-Tolerant Abstraction for In-Memory Cluster Computing," *NSDI '12: Proceedings of the 9th USENIX Conference on Networked Systems Design and Implementation*, pp. 15-28, Apr. 2012.
- [3] M. Zaharia et al., "Apache Spark: A Unified Engine for Big Data Processing," *Communications of the ACM*, vol. 59, no. 11, pp. 56-65, Nov. 2016.
- [4] K. Shvachko et al., "The Hadoop Distributed File System," *IEEE 26th Symposium on Mass Storage Systems and Technologies*, pp. 1-10, May 2010.
- [5] M. Zaharia et al., "Spark: Cluster Computing with Working Sets," *HotCloud '10: Proceedings of the 2nd USENIX Conference on Hot Topics in Cloud Computing*, pp. 10-10, Jun. 2010.
- [6] S. Ghemawat, H. Gobioff, and S. T. Leung, "The Google File System," *ACM SIGOPS Operating Systems Review*, vol. 37, no. 5, pp. 29-43, Oct. 2003.
- [7] R. Lämmel, "Google's MapReduce Programming Model — Revisited," *Science of Computer Programming*, vol. 70, no. 1, pp. 1-30, Jan. 2008.
- [8] J. Ekanayake et al., "Twister: A Runtime for Iterative MapReduce," *HPDC '10: Proceedings of the 19th ACM International Symposium on High Performance Distributed Computing*, pp. 810-818, Jun. 2010.
- [9] T. White, *Hadoop: The Definitive Guide*, 4th ed. Sebastopol, CA: O'Reilly Media, 2015.
- [10] C. Ranger et al., "Evaluating MapReduce for Multi-core and Multiprocessor Systems," *HPCA '07: Proceedings of the 13th International Symposium on High Performance Computer Architecture*, pp. 13-24, Feb. 2007.
- [11] V. K. Vavilapalli et al., "Apache Hadoop YARN: Yet Another Resource Negotiator," *SoCC '13: Proceedings of the 4th Annual Symposium on Cloud Computing*, pp. 1-16, Oct. 2013.
- [12] Ubuntu Documentation Team, "Ubuntu Server Guide," Canonical Ltd., 2024. [Online]. Available: <https://ubuntu.com/server/docs>
- [13] S. Ryza et al., *Advanced Analytics with Spark: Patterns for Learning from Data at Scale*, 1st ed. Sebastopol, CA: O'Reilly Media, 2015.
- [14] A. S. Tanenbaum and D. J. Wetherall, *Computer Networks*, 5th ed. Boston, MA: Pearson, 2011.
- [15] P. Boncz et al., "Breaking the Memory Wall in MonetDB," *Communications of the ACM*, vol. 51, no. 12, pp. 77-85, Dec. 2008.
- [16] Apache Software Foundation, "Apache Hadoop Documentation," 2024. [Online]. Available: <https://hadoop.apache.org/docs/stable/>
- [17] T. Condie et al., "MapReduce Online," *NSDI '10: Proceedings of the 7th USENIX Conference on Networked Systems Design and Implementation*, pp. 21-21, Apr. 2010.
- [18] Apache Software Foundation, "Apache Spark Documentation," 2024. [Online]. Available: <https://spark.apache.org/docs/latest/>
- [19] M. Isard et al., "Dryad: Distributed Data-parallel Programs from Sequential Building Blocks," *EuroSys '07: Proceedings of the 2nd ACM SIGOPS/EuroSys European Conference on Computer Systems*, pp. 59-72, Mar. 2007.
- [20] H. Karau et al., *Learning Spark: Lightning-Fast Big Data Analysis*, 1st ed. Sebastopol, CA: O'Reilly Media, 2015.

Detección de conjuntivitis a partir de imágenes oculares utilizando arquitecturas CNN y un modelo de fusión

Conjunctivitis detection from ocular images using CNN architectures and a fusion model

Antony Stebend Camacho Garcia^A, Suker Huamanñahui Hilario^B, Mario Aquino Cruz^C y Evelyn Naida Luque Ochoa^D

Resumen— La conjuntivitis es una causa frecuente de consulta oftalmológica y puede generar complicaciones graves si no se detecta a tiempo. Este estudio evalúa tres arquitecturas de redes neuronales convolucionales preentrenadas —EfficientNetB0, InceptionV3 y ResNet50— para la clasificación automática de imágenes oculares con y sin signos de conjuntivitis. Además, se propone un modelo de fusión que combina las salidas intermedias de InceptionV3 y ResNet50. Todos los modelos fueron entrenados usando aceleración por GPU y técnicas de regularización y aumento de datos. InceptionV3 obtuvo el mejor desempeño general, con una exactitud del 95.00 % y una precisión de 0.98 para la clase positiva y 0.92 para la clase negativa, mostrando un balance sólido entre sensibilidad y especificidad. EfficientNetB0 destacó por alcanzar el mayor recall en la clase positiva y la menor tasa de falsos negativos, aunque con una precisión considerablemente menor. ResNet50 presentó la menor tasa de falsos positivos, siendo útil para reducir diagnósticos erróneos en casos negativos. El modelo de fusión alcanzó métricas competitivas y un tiempo de entrenamiento reducido, demostrando que la integración de arquitecturas puede aumentar la robustez del sistema. Este trabajo ofrece una guía práctica y reproducible para la selección de modelos en el diagnóstico automatizado de conjuntivitis, especialmente útil en contextos con recursos médicos limitados.

Palabras clave: aprendizaje profundo, conjuntivitis, imágenes médicas, redes neuronales convolucionales

Abstract— Conjunctivitis is a common cause of ophthalmologic consultation and may lead to serious complications if not diagnosed early. This study evaluates three pretrained convolutional neural network architectures—EfficientNetB0, InceptionV3, and ResNet50—for the automated classification of ocular images with and without signs of conjunctivitis. Additionally, a fusion model combining intermediate outputs from InceptionV3 and ResNet50 is proposed. All models were trained using GPU acceleration and employed regularization and data augmentation techniques. InceptionV3 achieved the best overall performance, with an accuracy of 95.00% and a precision of 0.98 for the positive class and 0.92 for the negative class, showing a solid balance between sensitivity and specificity. EfficientNetB0 achieved the highest recall in the positive class and the lowest false negative rate, although with considerably lower precision. ResNet50 obtained the lowest false positive rate, making it valuable for reducing misdiagnosis in negative cases. The fusion model achieved competitive metrics and a reduced training time, indicating that combining complementary architectures can enhance system robustness. This work provides a practical and reproducible guide for selecting CNN models for automated conjunctivitis diagnosis, particularly in resource-limited medical settings.

Keywords: deep learning, conjunctivitis, medical imaging, convolutional neural networks

1 INTRODUCCIÓN

La conjuntivitis es una patología inflamatoria común de la conjuntiva ocular, con un impacto significativo en la salud pública y una alta prevalencia a nivel mundial [1][2]. Se presenta en diversas formas clínicas, entre las que destacan la viral, la bacteriana y la alérgica.

Las infecciones bacterianas son responsables de aproximadamente el 50 % al 70 % de los casos de conjuntivitis, siendo *Staphylococcus aureus*, *Streptococcus pneumoniae* y *Pseudomonas aeruginosa* los agentes etiológicos más frecuentes [2]. Esta afección constituye una de las causas más comunes de consulta oftalmológica, especialmente en entornos de atención primaria.



Revista de Investigación en Ciencia y Tecnología

ISSN: 2810-8124 (en línea) / ISSN: 2706-543x

Universidad Nacional Micaela Bastidas de Apurímac – Perú

Vol. 7 Núm. 2 (2025) - Publicado: 19/08/25 - [Indexaciones](#)

Número: [doi.org/10.57166/riqchary/v7.n2.2025](#)

Páginas: 54- 65 | Recibido 01/01/2025 ; Aceptado 01/02/2025

[doi.org/10.57166/riqchary.v7.n2.2025.7](#)

Autores:

- A. **ORCID iD** <https://orcid.org/0009-0004-3492-7741>
Antony Stebend Camacho García, estudiante en la Universidad Nacional Micaela Bastidas de Apurímac, Pe.
162135@unamba.edu.pe
- B. **ORCID iD** <https://orcid.org/0009-0002-4945-6263>
Suker Huamanñahui Hilario, estudiante en la Universidad Nacional Micaela Bastidas de Apurímac, Pe.
162157@unamba.edu.pe
- C. **ORCID iD** <https://orcid.org/0000-0002-2552-5669>
Mario Aquino Cruz, docente del Departamento Académico de Informática y Sistemas de la Universidad Nacional Micaela Bastidas de Apurímac, Pe.
maquino@unamba.edu.pe
- D. **ORCID iD** <https://orcid.org/0000-0002-8386-9806>
Evelyn Naida Luque Ochoa, docente del Departamento Académico de Informática y Sistemas de la Universidad Nacional Micaela Bastidas de Apurímac, Pe.
eluque@unamba.edu.pe

Según la Academia Americana de Oftalmología, millones de personas se ven afectadas cada año, generando una carga considerable para los sistemas sanitarios [1].

En países como el Perú, donde existen limitaciones en el acceso a servicios especializados en regiones rurales o amazónicas, el diagnóstico temprano y preciso se vuelve particularmente desafiante.

Tradicionalmente, el diagnóstico de la conjuntivitis se basa en la observación clínica, lo que puede llevar a errores diagnósticos en ausencia de personal capacitado o recursos complementarios. Ante esta problemática, las tecnologías de inteligencia artificial (IA) han emergido como herramientas prometedoras para asistir en el diagnóstico médico automatizado, especialmente mediante el uso de redes neuronales convolucionales (CNN).

Estas arquitecturas han demostrado una alta eficacia en la clasificación de imágenes médicas al identificar patrones complejos en estructuras visuales. Modelos como ResNet50, InceptionV3 y EfficientNetB0 han sido aplicados con éxito en la detección de diversas patologías oculares [3][4]. Por ejemplo, Mondal et al. lograron una precisión de hasta el 95.2 % con InceptionV3 y 93.6 % con ResNet50 al clasificar imágenes oculares con y sin conjuntivitis [4]. De manera similar, Bitto y Mahmud desarrollaron un sistema multicategorico con CNN que incluyó conjuntivitis, alcanzando una precisión de 97.08 % en clasificación ocular [5].

El presente estudio propone un análisis comparativo entre ResNet50, InceptionV3 y EfficientNetB0, aplicados al diagnóstico automatizado de conjuntivitis mediante imágenes oculares. Además, se implementó un modelo de fusión (ensemble) que combina las salidas intermedias de las arquitecturas de ResNet50 e InceptionV3, con el fin de mejorar la capacidad de clasificación mediante integración de características.

La investigación emplea un enfoque cuantitativo, bajo un diseño no experimental y transversal. Los modelos fueron entrenados mediante transferencia de aprendizaje, utilizando un conjunto de datos balanceado, preprocesado y evaluado mediante métricas estándar: exactitud, precisión, recall, F1-score, pérdida (loss) y matriz de confusión. El desarrollo se realizó con herramientas abiertas como TensorFlow y Keras sobre Google Colaboratory.

La principal contribución de este trabajo radica en proveer una comparación reproducible entre tres arquitecturas CNN de última generación en la tarea de diagnóstico ocular automatizado, así como en el desarrollo de un modelo de fusión (ensemble) que integra sus características para mejorar el rendimiento diagnóstico. Los resultados obtenidos pueden ser valiosos para futuros sistemas de apoyo clínico basados en IA, especialmente en contextos donde el acceso a especialistas es limitado.

2 TRABAJOS RELACIONADOS

En la última década, el uso de redes neuronales convolucionales (CNN) ha cobrado relevancia en el diagnóstico automatizado de enfermedades médicas a partir de imágenes, incluyendo afecciones oculares como la conjuntivitis. Gracias a su capacidad para extraer características complejas, las CNN han demostrado ser herramientas efectivas en

contextos clínicos, incluso en comparación con la evaluación humana especializada.

Se desarrolló un modelo basado en CNN para detectar conjuntivitis en imágenes y generar recomendaciones de tratamiento farmacológico. Su enfoque logró una alta precisión en las predicciones, resaltando el potencial del aprendizaje profundo en sistemas de apoyo clínico personalizados [6].

En un contexto complementario, también propusieron una arquitectura conceptual diseñada para telemedicina, la cual utiliza redes neuronales para la clasificación remota de casos de conjuntivitis. Su estudio destaca el valor de estas soluciones en entornos rurales o con recursos limitados, donde el acceso a especialistas es restringido [7].

Un enfoque de clasificación multiclase fue realizado utilizando la arquitectura VGG16 para identificar diferentes enfermedades externas del ojo, incluida la conjuntivitis. Su sistema alcanzó una precisión del 98.48 %, reforzando la fiabilidad de los modelos CNN en escenarios diagnósticos complejos [8]. De manera similar, se desarrollaron un sistema de clasificación binaria para detectar conjuntivitis mediante una red convolucional personalizada. Al ser entrenado con más de 5,000 imágenes, el modelo logró una precisión del 88.80 %, con valores consistentes de recall y F1-score, lo que respalda su utilidad en entornos clínicos reales [9].

Si bien estos estudios evidencian avances importantes, la mayoría de ellos se centran en evaluar arquitecturas individuales o carecen de comparaciones sistemáticas entre modelos bajo condiciones homogéneas. Tampoco exploran de forma profunda las ventajas del ensamblaje de redes neuronales para mejorar el rendimiento diagnóstico.

En respuesta a estas limitaciones, el presente estudio propone una evaluación comparativa de tres arquitecturas CNN ampliamente reconocidas —ResNet50, InceptionV3 y EfficientNetB0— aplicadas al diagnóstico automatizado de conjuntivitis. Asimismo, se introduce un modelo de fusión (ensemble) que integra las salidas intermedias de las arquitecturas ResNet50 e InceptionV3, con el objetivo de aumentar la precisión y robustez del sistema. Esta propuesta busca contribuir con evidencia cuantitativa que respalde el uso de modelos híbridos de aprendizaje profundo en contextos clínicos donde el acceso a recursos especializados es restringido.

3 METODOLOGÍA

A. DISEÑO DE INVESTIGACIÓN

Este estudio se desarrolló bajo un enfoque metodológico cuantitativo, con un diseño no experimental de corte transversal. En este tipo de investigaciones, no se manipulan las variables, sino que se observan tal como ocurren en un único momento temporal, lo que permite analizar asociaciones bajo condiciones controladas y replicables [10].

En este contexto, la presente investigación se enfocó en comparar el rendimiento de diversas arquitecturas de redes neuronales convolucionales (CNN) aplicadas al diagnóstico automatizado de conjuntivitis, utilizando un conjunto de imágenes oculares balanceado, preprocesado y evaluado con métricas estandarizadas.

Para evaluar el rendimiento de los modelos, se utilizaron

métricas estándar de clasificación como la exactitud, precisión, recall, F1-score y pérdida (loss), las cuales permitieron establecer comparaciones objetivas y consistentes entre las arquitecturas implementadas[11]. Además, se propuso un modelo de fusión que combina las salidas intermedias de tres arquitecturas CNN preentrenadas —ResNet50, InceptionV3 y EfficientNetB0— con el objetivo de mejorar la capacidad de clasificación mediante integración de características profundas, en comparación con el desempeño individual de cada modelo.

B. PARTICIPANTES

La muestra se dividió en tres subconjuntos: 60% para entrenamiento, 20% para validación y 20% para prueba, asegurando que cada subconjunto mantuviera una distribución equitativa de ambas clases (ojos con y sin conjuntivitis), lo que minimizó el sesgo en la evaluación del modelo.

Las Figuras 1 y 2 presentan ejemplos visuales representativos de las imágenes utilizadas en el estudio: una imagen ocular con signos clínicos de conjuntivitis y una imagen de un ojo sano, respectivamente. Las Figuras 3, 4 y 5 ilustran las arquitecturas de las redes neuronales convolucionales evaluadas: ResNet50, InceptionV3 y EfficientNetB0, mientras que la Figura 6 muestra el diseño del modelo de fusión propuesto, el cual integra las salidas de las redes anteriores a nivel de características para mejorar la capacidad de clasificación.

En la Fig.1 se muestra imagen representativa de un ojo con signos clínicos de conjuntivitis. Se observan enrojecimiento conjuntival, inflamación y posible secreción mucosa. la fuente de la imagen es exclusivamente extraída de nuestro conjunto de datos usado para este estudio.



Fig. 1. Imagen de un ojo con signos clínicos de conjuntivitis.

La Fig.2 nos muestra una imagen de un ojo sano, sin presencia de signos inflamatorios ni alteraciones visibles, la fuente de la imagen es exclusivamente extraída de nuestro conjunto de datos usado para este estudio.

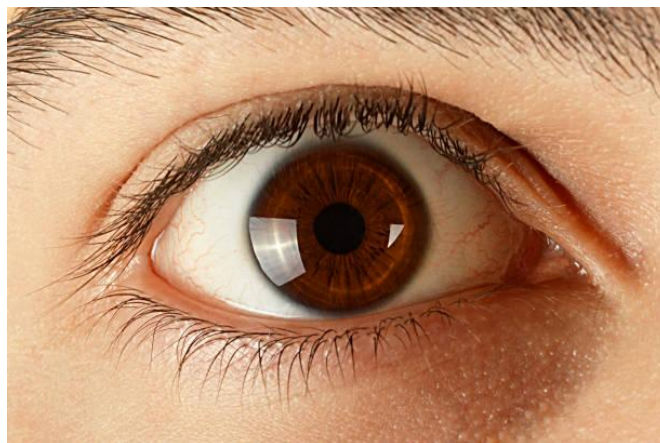


Fig. 2. Imagen de un ojo sano.

Diseño de las arquitecturas: Las siguientes figuras ilustran las arquitecturas de redes neuronales convolucionales preentrenadas utilizadas.

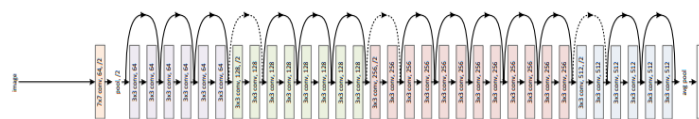


Fig. 3. Arquitectura de ResNet50. Fuente: extraída de [15]

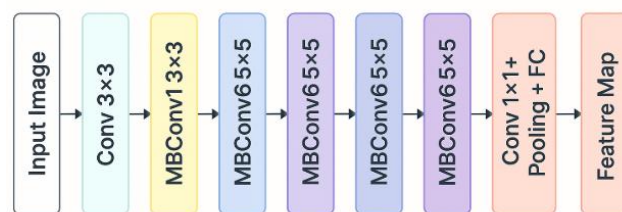


Fig. 4. Arquitectura del modelo EfficientNetB0, adaptada de [16].

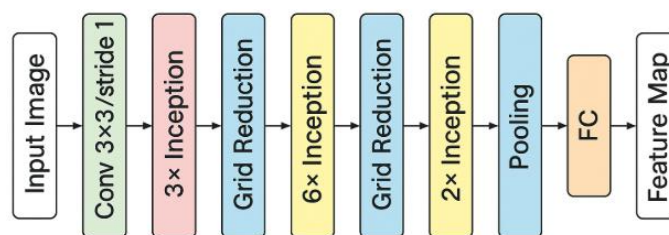


Fig. 5. Arquitectura del modelo InceptionV3, adaptada de [17].

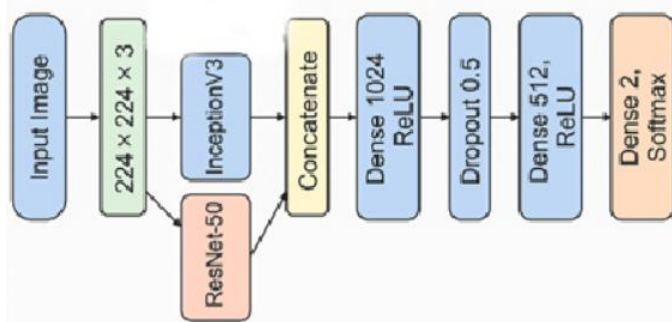


Fig. 5. Arquitectura del modelo fusionado.

C. INSTRUMENTOS Y TÉCNICAS

Para el desarrollo del estudio se utilizó el lenguaje de programación Python[12], en conjunto con las bibliotecas TensorFlow y Keras, ampliamente reconocidas en tareas de aprendizaje profundo[13]. El entorno de desarrollo fue Google Colaboratory (versión gratuita), que proporcionó acceso a recursos de cómputo en la nube, incluyendo una GPU asignada automáticamente por el entorno, la cual permitió ejecutar eficientemente los experimentos sin requerir infraestructura local dedicada [14].

El conjunto de datos utilizado en esta investigación fue construido específicamente para este estudio. Para ello, se recopilaron imágenes de ojos con y sin signos de conjuntivitis a partir de repositorios públicos como Kaggle y diversas fuentes confiables en internet. Las imágenes fueron seleccionadas manualmente, asegurando la calidad visual y la representatividad clínica, y posteriormente clasificadas en dos categorías: ojos sanos y ojos con conjuntivitis. El conjunto final quedó conformado por 3,479 imágenes de ojos sanos y 4,485 imágenes de ojos con conjuntivitis. Todas las imágenes fueron sometidas a un proceso de preprocesamiento y aumento de datos, que incluyó redimensionamiento, rotación, inversión horizontal, zoom y ajustes de brillo, con el fin de mejorar la variabilidad del conjunto y fortalecer la capacidad de generalización de los modelos. Finalmente, se estandarizaron a una resolución de 224x224 píxeles y se reescalaron al rango [0,1] para su procesamiento por las redes neuronales convolucionales. Se aplicó aprendizaje transferido (*transfer learning*) utilizando tres arquitecturas CNN preentrenadas: ResNet50 [15], EfficientNetB0[16] y InceptionV3 [17]. Adicionalmente, se construyó un modelo de fusión que integró las salidas intermedias de dichas redes, combinando sus características mediante *feature-level fusion* y capas densas compartidas [18]. Todos los modelos fueron adaptados a una tarea de clasificación binaria con función de pérdida *categorical crossentropy*.

Como técnica de preprocesamiento, se aplicó un esquema de aumento de datos (*data augmentation*) con rotaciones aleatorias, desplazamientos, zoom, ajustes de brillo e inversión horizontal. Estas transformaciones permitieron simular variaciones en las condiciones reales de captura de imágenes y mejorar la capacidad de generalización de los modelos [19].

D. ANÁLISIS DE DATOS

El análisis se fundamentó en métricas ampliamente utilizadas en clasificación binaria, tales como exactitud (accuracy), pérdida (loss), precisión, recall y F1-score. Además, se empleó una matriz de confusión para identificar falsos positivos (FP) y falsos negativos (FN), aspectos especialmente relevantes en aplicaciones clínicas sensibles [11]. Los modelos fueron entrenados bajo las mismas condiciones, con técnicas de regularización como *EarlyStopping* y reducción dinámica de la tasa de aprendizaje. Esta homogeneidad metodológica garantizó la validez y la reproducibilidad de la comparación entre arquitecturas.

4 RESULTADOS

Los resultados obtenidos reflejan el rendimiento de las diferentes arquitecturas CNN empleadas en este estudio: ResNet50, EfficientNetB0, InceptionV3. Cada modelo fue entrenado bajo las mismas condiciones experimentales, utilizando el conjunto de entrenamiento preprocesado y validado, con evaluaciones finales sobre el conjunto de prueba. A continuación, se presentan los resultados obtenidos por cada arquitectura de manera individual, seguidos por el análisis del modelo fusionado. Las métricas utilizadas para la comparación fueron: exactitud (accuracy), pérdida (loss), precisión, recall, F1-score y matriz de confusión.

A. RESULTADOS DEL MODELO RESNET50

La arquitectura ResNet50 fue entrenada mediante aprendizaje transferido, alcanzando un total de 28 épocas antes de la activación del callback *EarlyStopping*, el cual monitoreaba la pérdida de validación (val_loss). El entrenamiento se llevó a cabo en un entorno proporcionado por Google Colaboratory (versión gratuita), con una GPU asignada automáticamente, y tuvo una duración aproximada de 10 horas y 5 minutos. Durante el entrenamiento, se evidenció una mejora progresiva tanto en la precisión como en la pérdida, alcanzando en la última época una exactitud de 67.84 % y una pérdida de 0.6165 sobre los datos de entrenamiento, mientras que la precisión en validación fue de 78.20 % con una pérdida de 0.5401. Este entorno permitió ejecutar los experimentos de manera eficiente sin necesidad de infraestructura local dedicada.

TABLA I. RENDIMIENTO DE RESNET50 EN ENTRENAMIENTO, VALIDACIÓN Y TEST

Métrica	Train	Validation	Test
Exactitud (Accuracy)	68.73 %	78.00 %	76.66%
Pérdida (Loss)	0.6092	0.5622	0.5294
Precisión (Precision)	0.72 (C), 0.68 (N)	0.77 (C), 0.79 (N)	0.75 (C), 0.79 (N)
Recall	0.73 (C), 0.65 (N)	0.86 (C), 0.67 (N)	0.87 (C), 0.63 (N)
F1-Score	0.69	0.81	0.82

Leyenda:

- C: *Conjuntivitis* (ojos con infección)
- N: *Normal* (ojos sin signos de infección)

La Fig. 6 muestra la evolución de la exactitud durante el entrenamiento de ResNet50:

- **Train Accuracy:** La precisión en el conjunto de entrenamiento mostró una mejora progresiva y constante, iniciando en 53.88 % en la época 1 y alcanzando un máximo de 68.73 % en la época 26. Esto indica que el modelo fue aprendiendo de forma gradual, sin señales de sobreajuste prematuro.
- **Validation Accuracy:** También se observó una tendencia positiva, comenzando en 43.66 % en la época 1 y alcanzando su pico en la época 26 con 78.89 %. Aunque hubo fluctuaciones (por ejemplo, una caída a 60.30 % en la época 6), la tendencia general fue de mejora sostenida.
- **Brecha entre entrenamiento y validación:** La diferencia entre ambas métricas fue moderada y estable. En muchas épocas, la precisión en validación fue incluso superior a la de entrenamiento, lo cual sugiere que el modelo no presentó sobreajuste y logró una generalización adecuada.

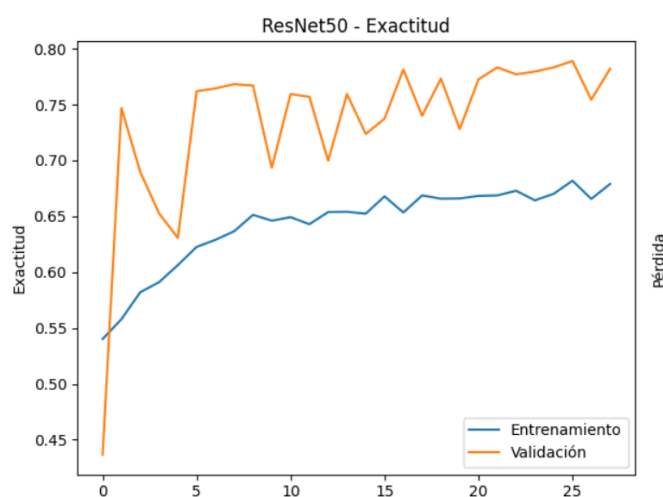


Fig. 6. Curva de exactitud (Accuracy) durante el entrenamiento de ResNet50.

- **Train Loss:** La pérdida en el conjunto de entrenamiento disminuyó progresivamente, pasando de 0.7997 en la época 1 a 0.6092 en la época 26, lo cual refleja una mejora constante en la capacidad del modelo para ajustarse a los datos.
- **Validation Loss:** También presentó una tendencia descendente, iniciando en 0.7100 y alcanzando su valor mínimo de 0.5387 en la época 23. Aunque se observaron pequeñas oscilaciones (por ejemplo, en las épocas 10, 19 y 26), no se evidenciaron aumentos significativos que indiquen sobreajuste.
- **Brecha entre entrenamiento y validación:** La diferencia entre ambas curvas de pérdida se mantuvo leve y estable durante la mayoría del entrenamiento, e incluso en

varias épocas la pérdida en validación fue menor que la de entrenamiento, reforzando la conclusión de que el modelo generalizó adecuadamente.

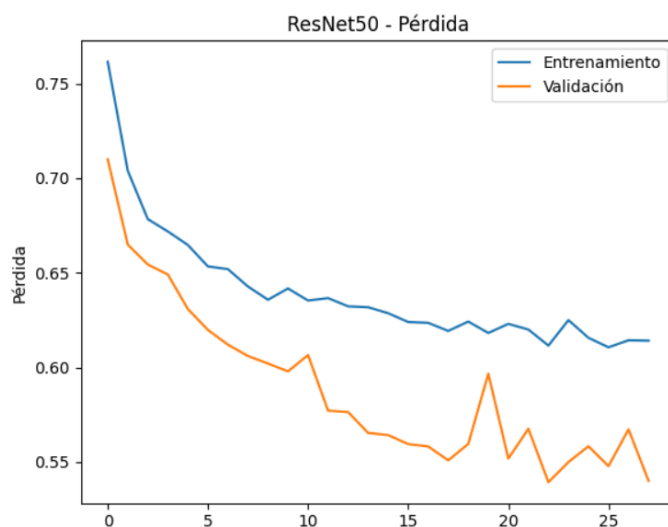


Fig. 7. Curva de pérdida (Loss) durante el entrenamiento de ResNet50.

La matriz de confusión (Fig. 8) correspondiente al conjunto de prueba muestra un desempeño aceptable del modelo ResNet50, aunque revela limitaciones en la clasificación de imágenes normales.

- Verdaderos Positivos (TP): 753 (84 % de los casos de conjuntivitis fueron correctamente identificados).
- Falsos Negativos (FN): 144 (16 % de las imágenes con conjuntivitis fueron clasificadas erróneamente como normales).
- Falsos Positivos (FP): 195 (28 % de los casos normales fueron clasificados incorrectamente como conjuntivitis).
- Verdaderos Negativos (TN): 502 (72 % de las imágenes normales fueron clasificadas correctamente).

Este resultado indica una buena sensibilidad para detectar conjuntivitis, lo cual es relevante en un entorno clínico para reducir diagnósticos omitidos. No obstante, la tasa de falsos positivos del 28 % representa un riesgo de generar alarmas innecesarias, lo cual puede afectar la eficiencia del sistema si se implementa como herramienta de apoyo en el diagnóstico médico.

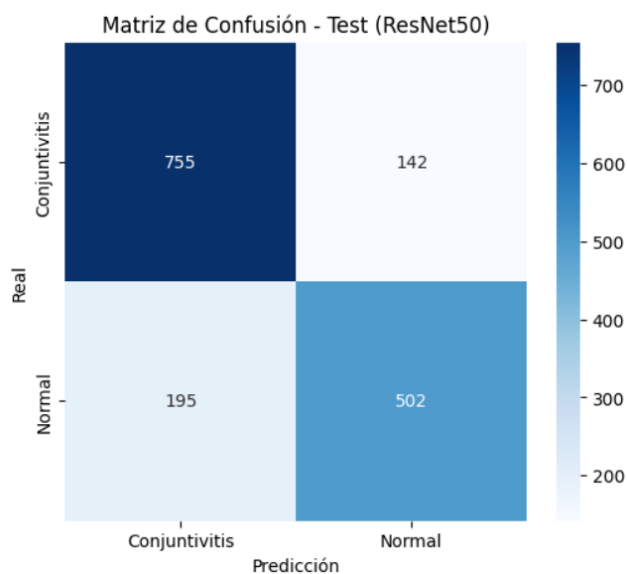


Fig. 8. Matriz de confusión del modelo de ResNet50 con el conjunto de test.

B. RESULTADOS DEL MODELO EFFICIENTNETB0

La arquitectura EfficientNetB0 fue entrenada utilizando aprendizaje por transferencia, partiendo de pesos preentrenados en ImageNet. Inicialmente, se empleó el modelo como extractor de características con las capas convolucionales congeladas, logrando una estabilidad parcial en la validación. Posteriormente, se aplicó *fine-tuning* (ajuste fino) al desbloquear progresivamente capas del modelo base. Esta estrategia se adoptó con el objetivo de mejorar la capacidad del modelo para captar patrones específicos del dominio médico ocular, dado que las características visuales presentes en imágenes de conjuntivitis pueden diferir significativamente de las de ImageNet.

Durante el entrenamiento, se ejecutaron múltiples sesiones, alcanzando un total de 13 épocas antes de activarse el *callback* EarlyStopping, el cual monitoreaba la pérdida de validación (*val_loss*). El mejor desempeño en validación se observó en la época 9, con una exactitud del 69.28 % y una pérdida de 0.6084. El tiempo total de entrenamiento acumulado fue de aproximadamente 9 horas y 43 minutos, utilizando una GPU asignada automáticamente por Google Colaboratory (versión gratuita). Esta infraestructura remota permitió realizar los experimentos sin requerir recursos locales especializados, aunque con ciertas limitaciones en el rendimiento computacional.

TABLA II. RENDIMIENTO DE EFFICIENTNETB0 EN ENTRENAMIENTO, VALIDACIÓN Y TEST

Métrica	Train	Validation	Test
Exactitud (Accuracy)	63.41 %	69.00 %	70.00 %
Pérdida (Loss)	0.6213	0.7254	0.6052

Precisión (Precision)	0.70 (C) / 0.72 (N)	0.66 (C), 0.84 (N)	0.66 (C), 0.84 (N)
Recall	0.85 (C) / 0.42 (N)	0.95 (C), 0.37 (N)	0.95 (C), 0.38 (N)
F1-Score	0.65	0.65	0.65

Leyenda:

- C: Conjuntivitis (ojos con infección)
- N: Normal (ojos sin signos de infección)

La Fig. 9 muestra la evolución de la exactitud durante el entrenamiento de EfficientNetB0:

- Train Accuracy: La precisión en el conjunto de entrenamiento mostró un crecimiento moderado y progresivo. Comenzó en 48.94 % en la época 1 y alcanzó un máximo de 63.41 % en la época 13. Esta evolución indica que el modelo fue aprendiendo gradualmente las características del conjunto de datos, sin señales claras de sobreajuste temprano.
- Validation Accuracy: En cuanto al conjunto de validación, la precisión también aumentó con el tiempo, aunque de forma más inestable. Inició en 43.66 % y alcanzó su punto máximo en la época 9 con 69.28 %, antes de descender nuevamente en las últimas épocas. Este comportamiento sugiere que el modelo logró una buena capacidad de generalización, pero también que entrenamientos adicionales podrían haber inducido sobreajuste si no se aplicaba early stopping.
- Brecha entre entrenamiento y validación: Durante las primeras épocas, la precisión en validación fue incluso superior a la de entrenamiento en algunas ocasiones (como en la época 6). A lo largo del proceso, ambas curvas se mantuvieron razonablemente cercanas, con una diferencia promedio menor al 10 %, lo que indica un entrenamiento equilibrado sin un sobreajuste evidente.

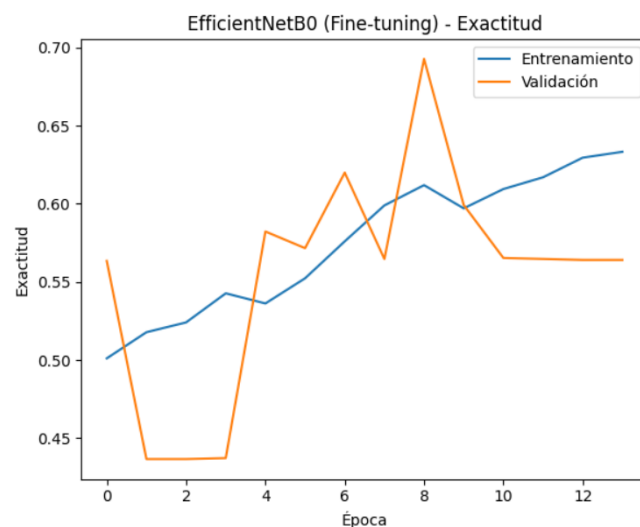


Fig. 9. Curva de exactitud (Accuracy) durante el entrenamiento de EfficientNetB0.

La Fig. 10 muestra la disminución de la pérdida (Loss):

- **Train Loss:** La pérdida en el conjunto de entrenamiento mostró una reducción progresiva a lo largo del proceso. Comenzó en 0.7292 en la época 1 y descendió hasta alcanzar 0.6213 en la época 13. Aunque la caída no fue abrupta, reflejó una mejora constante en la capacidad del modelo para ajustar los datos del entrenamiento.
- **Validation Loss:** La pérdida en el conjunto de validación también siguió una tendencia descendente, aunque con algunas oscilaciones. Inició en 0.6914 y alcanzó un mínimo relativo de 0.6084 en la época 9, antes de volver a subir progresivamente en las siguientes épocas (hasta 0.8567 en la época 14). Este comportamiento indica que el modelo logró una buena capacidad de generalización en etapas intermedias, aunque hacia el final del entrenamiento comenzó a mostrar signos de sobreajuste.
- **Brecha entre entrenamiento y validación:** La diferencia entre ambas pérdidas se mantuvo inicialmente estable, pero a partir de la época 10 comenzó a ampliarse. Esta divergencia creciente entre el *train loss* y el *val loss* sugiere que el modelo comenzó a sobreajustarse, es decir, a memorizar el conjunto de entrenamiento a expensas de su rendimiento en datos no vistos. Por esta razón, se aplicó la técnica de *early stopping* para detener el entrenamiento y evitar una pérdida de capacidad de generalización.

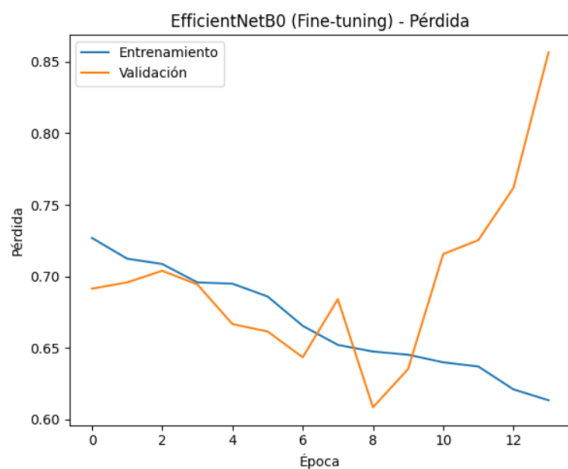


Fig. 10. Curva de pérdida (Loss) durante el entrenamiento de EfficientNetB0.

La matriz de confusión (Fig. 11) evidencia un 5.5 % de falsos negativos (49 casos de conjuntivitis no detectados), lo cual representa una posible limitación en aplicaciones médicas sensibles.

- **Verdaderos Positivos (TP):** 848 (94.5 %). El modelo identificó correctamente la mayoría de los casos de conjuntivitis, lo que indica una buena capacidad de detección.
- **Falsos Negativos (FN):** 49 (5.5 %). Algunos casos de conjuntivitis fueron clasificados como normales, lo que puede representar un riesgo si no se detectan a tiempo.

- **Falsos Positivos (FP):** 431 (61.8 %). El modelo tiende a confundir imágenes normales con conjuntivitis, lo que podría generar revisiones innecesarias.
- **Verdaderos Negativos (TN):** 266 (38.2 %). El modelo tiene dificultades para reconocer correctamente los casos normales, posiblemente por variabilidad en las imágenes o desequilibrio en los datos.

Este resultado refleja una buena capacidad del modelo para detectar conjuntivitis, pero también revela una tasa de error considerable al identificar imágenes normales, lo que podría generar falsas alarmas en un entorno clínico.

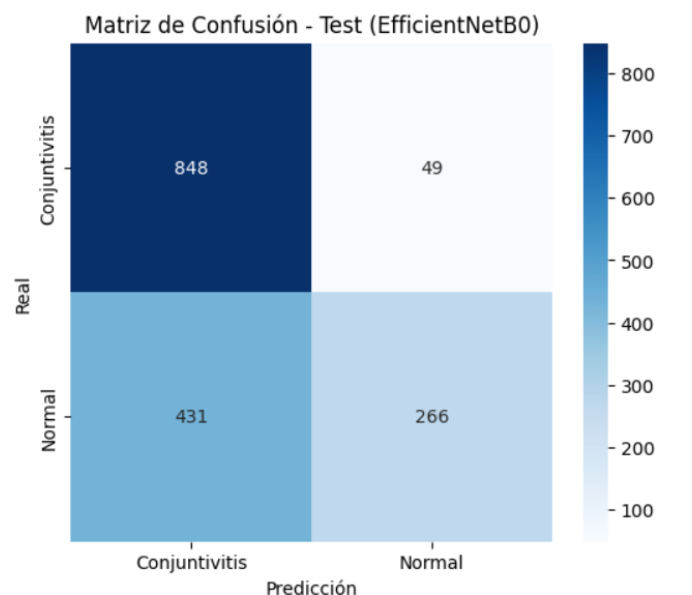


Fig. 11. Matriz de confusión del modelo de EfficientNetB0 con el conjunto de test.

C. RESULTADOS DEL MODELO INCEPTIONV3

La arquitectura InceptionV3 fue entrenada utilizando la técnica de aprendizaje transferido, ajustando sus pesos finales mediante *fine-tuning*. El proceso de entrenamiento se extendió por un total de 46 épocas, con activación del *callback* EarlyStopping, el cual monitoreaba la pérdida de validación (*val_loss*) y restauró los mejores pesos obtenidos en la época 41, momento en el que se alcanzó un rendimiento óptimo.

El tiempo total de entrenamiento fue de aproximadamente 10 horas y 40 minutos, considerando un promedio de 13 minutos por época. Para ello, se empleó una GPU proporcionada automáticamente por Google Colaboratory (versión gratuita), lo que facilitó la ejecución eficiente de los experimentos sin necesidad de infraestructura local especializada. Este entorno permitió llevar a cabo el proceso de entrenamiento de modelos avanzados en visión por computadora de manera accesible y reproducible.

TABLA III. RENDIMIENTO DE INCEPTIONV3 EN ENTRENAMIENTO, VALIDACIÓN Y TEST.

Métrica	Entrena- miento (Train)	Validación (Validation)	Prueba (Test)
Accuracy	93.15 %	95.10 %	95.11 %
Loss	0.1819	0.1412	0.1200
Precision	0.94 (C), 0.92 (N)	0.96 (C), 0.93 (N)	0.98 (C), 0.92 (N)
Recall	0.92 (C), 0.94 (N)	0.94 (C), 0.95 (N)	0.93 (C), 0.97 (N)
F1-Score	0.93	0.95	0.95

Leyenda:

- C: Conjuntivitis (ojos con infección)
- N: Normal (ojos sin signos de infección)

La Fig. 12 muestra la evolución de la exactitud durante el entrenamiento de InceptionV3:

- Train Accuracy: La precisión en el conjunto de entrenamiento mostró una mejora continua, comenzando en 67.82% en la época 1 y alcanzando un máximo de 93.15% en la época 46. Durante las primeras 20 épocas, se observó una progresión constante, lo que indica una adecuada capacidad de aprendizaje del modelo. Posteriormente, el modelo continuó refinando sus predicciones sin presentar signos evidentes de sobreajuste, lo cual se confirma por el mantenimiento de una alta precisión en la validación.
- Validation Accuracy: La precisión sobre el conjunto de validación también presentó un crecimiento sostenido, iniciando en 84.55% y alcanzando un valor máximo de 95.10% en la época 45. El modelo mantuvo una precisión superior al 93% a partir de la época 20, lo cual sugiere una excelente capacidad de generalización sobre datos no vistos. La detención temprana se activó en la época 46 al restaurar los mejores pesos obtenidos en la época 41, con una precisión de validación de 94.66%, consolidando un rendimiento robusto.

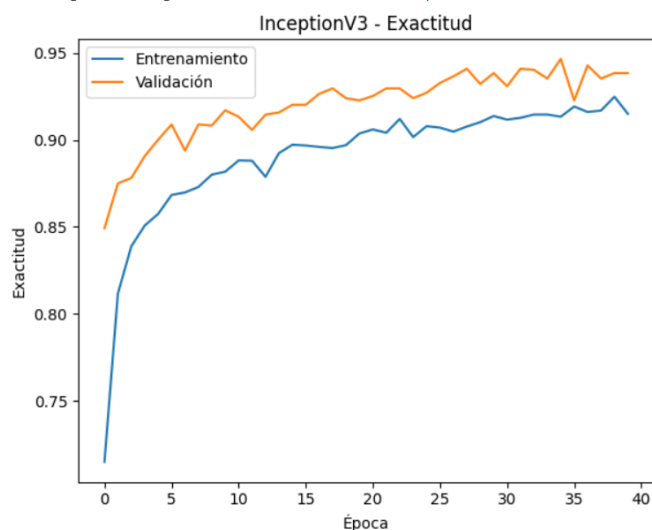


Fig. 12. Curva de exactitud (Accuracy) durante el entrenamiento de InceptionV3.

La Fig. 13 muestra la disminución de la pérdida (Loss):

- Train Loss: La pérdida durante el entrenamiento mostró una disminución progresiva y sostenida, comenzando en 0.6335 en la primera época y descendiendo hasta un valor mínimo de 0.1701 en la época 44. Esta reducción constante indica que el modelo fue ajustando exitosamente sus parámetros para minimizar el error, lo que refleja un aprendizaje efectivo de los patrones discriminativos en los datos de entrenamiento.
- Validation Loss: La pérdida en el conjunto de validación también presentó una tendencia general descendente, iniciando en 0.3500 y alcanzando su valor mínimo de 0.1376 en la época 41, donde se restauraron los mejores pesos tras la activación del early stopping. Aunque se observaron ligeras oscilaciones en ciertas etapas del entrenamiento, la tendencia general descendente evidencia que el modelo fue capaz de mantener una adecuada capacidad de generalización sin incurrir en sobreajuste.

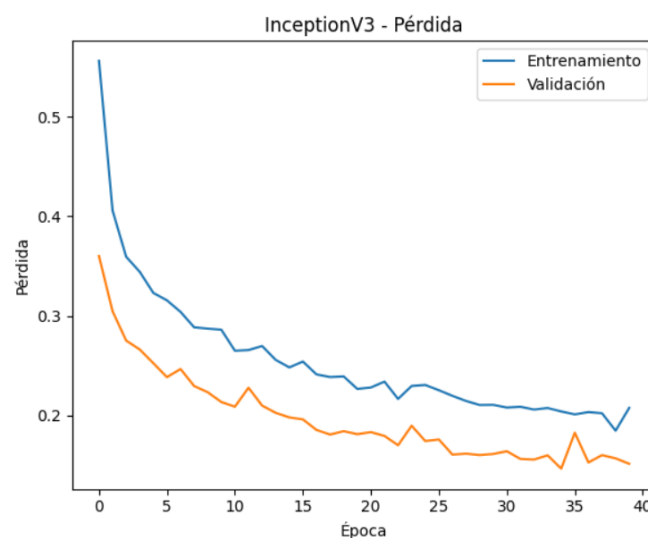


Fig. 13. Curva de pérdida (Loss) durante el entrenamiento de InceptionV3.

La matriz de confusión (Fig. 14) y el reporte de clasificación final confirman un alto rendimiento del modelo InceptionV3 en la detección automática de conjuntivitis a partir de imágenes oculares. Se alcanzó una precisión global del 95 % y un macro F1-score también de 0.95, lo cual evidencia un equilibrio sólido entre sensibilidad y especificidad.

- Verdaderos Positivos (TP): 838 (93.4 % de los casos de conjuntivitis fueron identificados correctamente). Este resultado refleja una alta sensibilidad para esta clase, lo cual es crucial para minimizar diagnósticos omitidos en escenarios clínicos.
- Falsos Negativos (FN): 59 (6.6 % de las imágenes con conjuntivitis fueron clasificadas erróneamente como normales). Aunque este valor es bajo, sigue siendo relevante, ya que podría implicar la omisión de tratamientos en pacientes reales si el sistema se emplea de manera autónoma.

- Falsos Positivos (FP): 19 (2.8 % de imágenes normales fueron clasificadas como conjuntivitis). Este bajo porcentaje indica un modelo conservador, con escasa propensión a emitir falsas alarmas, lo que mejora su confiabilidad como herramienta de apoyo.
- Verdaderos Negativos (TN): 678 (97.2 % de los ojos normales fueron correctamente identificados como tales), lo cual refuerza la especificidad del modelo en la clase sin infección.

Estos resultados ponen en evidencia la capacidad del modelo para generalizar eficazmente en nuevos datos. Su desempeño balanceado lo convierte en una opción viable para su uso en entornos clínicos, siempre que se considere su implementación como complemento a la evaluación médica tradicional.

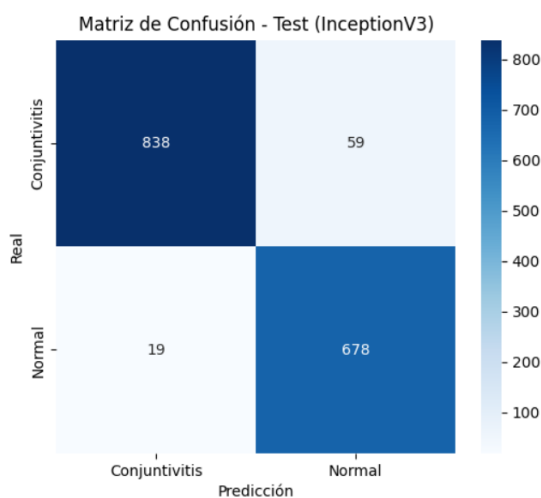


Fig. 14. Matriz de confusión.

D. RESULTADOS DEL MODELO FUSIÓN

El modelo Fusión fue entrenada utilizando aprendizaje transferido, alcanzando un total de 30 épocas antes de la activación del callback EarlyStopping, que monitoreaba la pérdida de validación (val_loss). El tiempo total de entrenamiento fue de aproximadamente 8 hora y 25 minutos, empleando una GPU asignada automáticamente por Google Colaboratory (versión gratuita). Esta configuración permitió ejecutar los experimentos sin requerir infraestructura local dedicada, facilitando así un entorno accesible y reproducible para el desarrollo del modelo.

TABLA IV. RENDIMIENTO DE MODELO FUSION EN ENTRENAMIENTO, VALIDACIÓN Y TEST

Métrica	Train	Validation	Test
Exactitud (Accuracy)	91.57 %	94.03 %	94.79 %
Pérdida (Loss)	0.2095	0.1601	0.1405
Precisión (Precision)	0.96 (C), 0.91 (N)	0.95 (C), 0.92 (N)	0.97 (C), 0.92 (N)
Recall	0.94 (C), 0.93 (N)	0.94 (C), 0.94 (N)	0.94 (C), 0.96 (N)
F1-Score	0.95	94	0.95

Leyenda:

- **C:** Conjuntivitis (ojos con infección)
- **N:** Normal (ojos sin signos de infección)

La Fig. 12 muestra la evolución de la exactitud durante el entrenamiento del modelo fusión:

- Train Accuracy: La precisión durante el entrenamiento mostró una mejora sostenida, comenzando en 70.10 % en la época 1 y alcanzando un valor máximo de 91.57 % en la época 30. Esta progresión constante sugiere que el modelo fue aprendiendo de manera efectiva, sin evidencias claras de sobreajuste en las primeras etapas del entrenamiento.
- Validation Accuracy: La precisión sobre el conjunto de validación también presentó un crecimiento continuo, iniciando en 85.18 % y alcanzando un máximo de 94.03 % en la época 27. Este desempeño elevado evidencia una buena capacidad del modelo para generalizar a datos no vistos.
- Brecha entre entrenamiento y validación: La diferencia entre ambas curvas fue reducida y estable a lo largo del entrenamiento. En varias épocas, la precisión en validación superó ligeramente a la de entrenamiento, lo cual indica una buena capacidad de generalización y la ausencia de sobreajuste significativo.

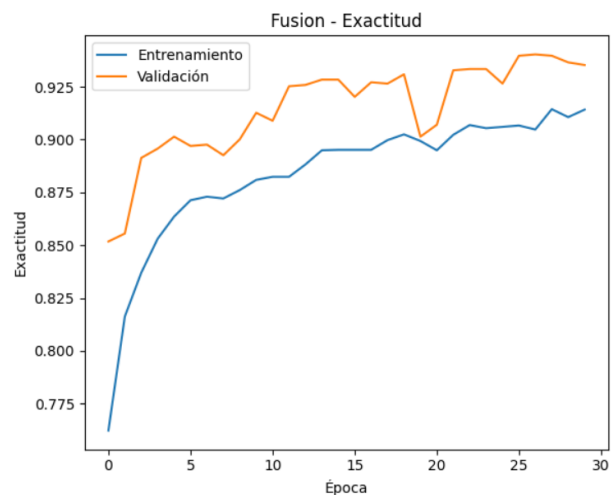


Fig. 15. Curva de exactitud (Accuracy) durante el entrenamiento del modelo fusión.

La Fig. 15 muestra la disminución de la pérdida (Loss):

- Train Loss: La pérdida durante el entrenamiento mostró una disminución constante, pasando de 0.5932 en la época 1 a un mínimo de 0.2031 en la época 27. Esta reducción progresiva indica que el modelo fue aprendiendo eficazmente a ajustar los datos, sin señales de estancamiento prematuro ni de sobreajuste en etapas iniciales.
- Validation Loss: La pérdida en validación también presentó una tendencia general descendente, comenzando en 0.3518 y alcanzando un mínimo de 0.1601 en la época 28. Aunque se observaron ligeras oscilaciones en algunas épocas, el comportamiento global fue positivo, reflejando una adecuada capacidad del modelo para generalizar sobre datos no vistos.

- Brecha entre entrenamiento y validación: La diferencia entre las curvas de pérdida se mantuvo reducida y estable durante todo el entrenamiento. En diversas etapas, la validation loss incluso se ubicó por debajo de la pérdida de entrenamiento, lo que sugiere que el modelo no solo evitó el sobreajuste, sino que también logró una generalización sólida. Esta consistencia valida el desempeño robusto del modelo de fusión frente a nuevos datos.

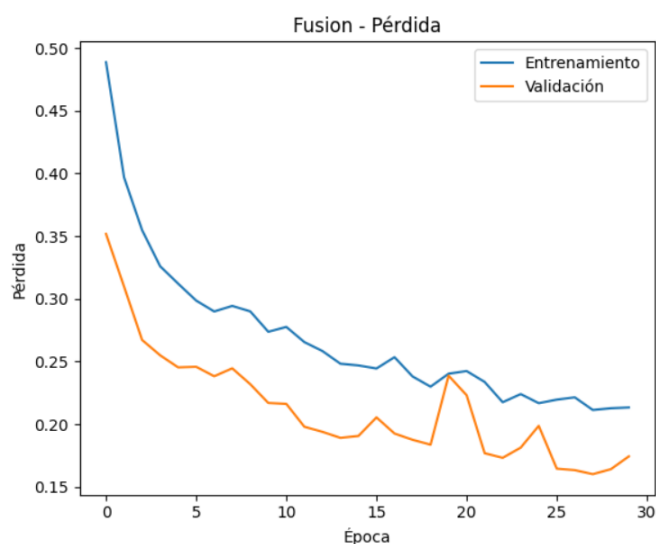


Fig. 16. Curva de pérdida (Loss) durante el entrenamiento del modelo fusión.

La matriz de confusión (Fig. 14) y el reporte de clasificación final evidencian un desempeño sólido del modelo de fusión en la detección de conjuntivitis a partir de imágenes oculares. El modelo alcanzó una precisión general del 95 %, con un F1-score macro promedio también de 0.95, lo que indica un equilibrio adecuado entre sensibilidad y especificidad en ambas clases.

- Verdaderos Positivos (TP): 840 casos de conjuntivitis fueron correctamente clasificados, lo que representa un recall del 94 % para esta clase. Este resultado demuestra una alta sensibilidad del modelo para identificar imágenes patológicas.
- Falsos Negativos (FN): 57 imágenes de conjuntivitis fueron erróneamente clasificadas como normales (6 %), lo cual puede representar una limitación en contextos clínicos, ya que existe riesgo de omitir diagnósticos relevantes si se depende exclusivamente del sistema automatizado.
- Falsos Positivos (FP): 26 imágenes normales fueron clasificadas como conjuntivitis (3.7 %), lo que implica una baja tasa de falsas alarmas. Esto es positivo en entornos médicos, ya que contribuye a reducir revisiones innecesarias sin comprometer la sensibilidad del modelo.
- Verdaderos Negativos (TN): 671 imágenes normales fueron correctamente identificadas, con una precisión del 92 % y un recall del 96 % para esta clase. Esto refleja una excelente capacidad para discriminar entre ojos sanos y afectados.

En conjunto, estos resultados sugieren que el modelo de fusión posee una alta capacidad de generalización, logrando un rendimiento equilibrado tanto en sensibilidad como en especificidad. No obstante, si se proyecta su uso en ambientes clínicos reales, la baja, pero presente tasa de error debe considerarse como parte de una estrategia de apoyo diagnóstico, y no como un reemplazo del juicio médico profesional.

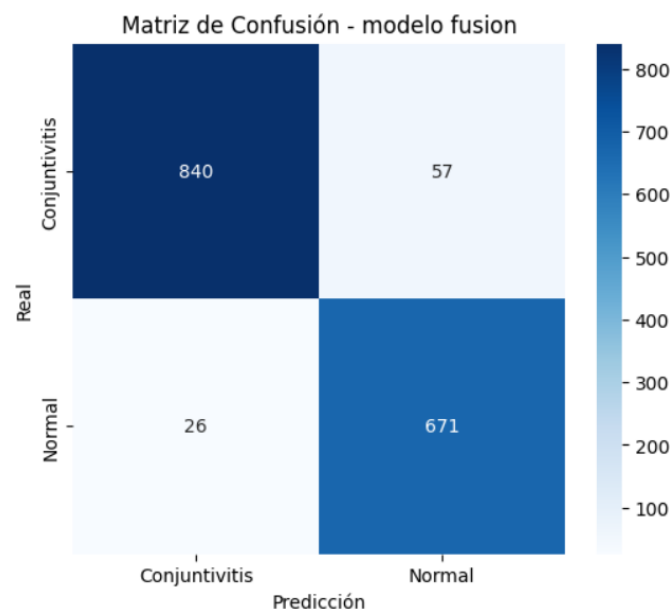


Fig. 17. Curva de pérdida (Loss) durante el entrenamiento del modelo fusión con el conjunto de test.

E. TABLA COMPARATIVA DE LOS MODELOS EN EL CONJUNTO DE TEST RESNET50, EFFICIENTNETB0, INCEPTIONV3 Y FUSIÓN.

Métricas	Inception V3	Efficient NetB0	ResNet 50	Fusión
Exactitud (Test)	95.00 %	70.00 %	79.00 %	94.79
Precisión (C / N)	0.98 / 0.92	0.66 / 0.84	0.79 / 0.78	0.97 / 0.92
Recall (C / N)	0.93 / 0.97	0.95 / 0.38	0.84 / 0.72	0.92 / 0.96
F1-Score	0.95	0.65	0.79	95
Tiempo de Entrenamiento	10.7 h	9.7 h	10.7 h	8.25h
Falsos Negativos (FN)	16%	5.5 %	6.6%	6.0%
Falsos Positivos (FP)	28%	61.8 %	2.8%	3.7%

5 DISCUSIÓN

Los resultados de este estudio confirman que las arquitecturas CNN evaluadas son herramientas eficaces para la clasificación automatizada de imágenes oculares en el diagnóstico de conjuntivitis. Entre ellas, InceptionV3 demostró el rendimiento más alto, con una precisión del 95.11 % y un equilibrio notable entre sensibilidad (recall) y especificidad (precisión). Este comportamiento respalda su idoneidad en entornos clínicos, donde minimizar tanto falsos negativos como falsos positivos es esencial para reducir riesgos en el proceso diagnóstico.

Por su parte, ResNet50 ofreció un desempeño competitivo, combinando una precisión elevada y una baja tasa de falsas alarmas. Esto la convierte en una alternativa sólida en contextos donde se privilegie la confiabilidad del modelo y la simplicidad de implementación, dado que su entrenamiento fue más estable y su complejidad computacional moderada.

En contraste, EfficientNetB0, si bien es una arquitectura más ligera y eficiente en recursos, mostró un rendimiento inferior en sensibilidad y una tasa elevada de falsos positivos, lo que implica que su adopción en entornos clínicos requiere ajustes y calibraciones adicionales. No obstante, su menor tiempo de entrenamiento y su bajo consumo de memoria pueden ser valiosos en escenarios de hardware limitado.

El modelo de fusión propuesto, que integró exclusivamente las salidas intermedias de InceptionV3 y ResNet50, alcanzó resultados comparables a InceptionV3, con un F1-score de 0.95 y una precisión del 94.79 %. Esta aproximación de ensamblaje evidenció beneficios en términos de robustez y generalización, reduciendo la variabilidad entre clases y consolidando un equilibrio entre sensibilidad y especificidad. La capacidad de combinar dos arquitecturas complementarias sugiere que la fusión de modelos puede ser una estrategia efectiva para mejorar el rendimiento diagnóstico sin incrementar en exceso la complejidad computacional.

En conjunto, los hallazgos refuerzan la utilidad del aprendizaje transferido, el aumento de datos y la regularización como prácticas fundamentales para obtener modelos fiables a partir de conjuntos de datos médicos limitados. Asimismo, los resultados subrayan que, aunque las CNN presentan un alto potencial como soporte diagnóstico, es imprescindible contemplar su uso como herramienta complementaria y no como sustituto del criterio clínico experto.

6 CONCLUSIONES Y RECOMENDACIONES

Este estudio evaluó y comparó el rendimiento de tres modelos de redes neuronales convolucionales preentrenadas (EfficientNetB0, InceptionV3 y ResNet50) y un modelo de fusión que integra las salidas intermedias de InceptionV3 y ResNet50 para la detección automatizada de conjuntivitis a partir de imágenes oculares. La principal contribución de este trabajo radica en ofrecer un análisis comparativo sistemático, reproducible y orientado a la aplicabilidad clínica, utilizando un flujo de datos homogéneo, técnicas estandarizadas de preprocesamiento y métricas unificadas, lo que representa un enfoque poco abordado en estudios previos sobre diagnóstico ocular asistido por inteligencia artificial.

El modelo EfficientNetB0 se caracterizó por su ligereza y me-

nor tiempo de entrenamiento (9.7 horas), lo que lo convierte en una alternativa potencialmente viable para entornos con recursos computacionales limitados, como dispositivos móviles o clínicas rurales. Sin embargo, su desempeño fue inconsistente, con fluctuaciones significativas en la exactitud de validación y una tasa elevada de falsos positivos (61.8%), lo que compromete su fiabilidad para ser aplicado directamente en contextos clínicos sin un ajuste exhaustivo.

Por su parte, InceptionV3 demostró la mayor capacidad de detección, alcanzando un recall del 93% para la clase "conjuntivitis", lo que lo convierte en el modelo más sensible del estudio. Su estabilidad durante el entrenamiento fue notable, con una brecha reducida entre entrenamiento y validación, aunque presentó una mayor demanda computacional (aproximadamente 10.7 horas) y un volumen considerable de parámetros, que pueden dificultar su despliegue en hardware con limitaciones de memoria.

El modelo ResNet50 emergió como la opción más equilibrada. Logró una precisión del 95%, con una tasa de falsos positivos relativamente baja (3%) y una menor diferencia entre las métricas de entrenamiento y validación, lo cual evidencia una buena capacidad de generalización. Su tiempo de entrenamiento fue más eficiente que el de InceptionV3, lo que refuerza su viabilidad para ser implementado en entornos clínicos donde es fundamental mantener bajo el nivel de errores de clasificación sin incrementar excesivamente los requisitos computacionales.

El modelo de fusión InceptionV3 + ResNet50 demostró que la combinación de arquitecturas complementarias puede incrementar la robustez y estabilidad del sistema, alcanzando un F1-score de 0.95 y una precisión del 94.79%, con una tasa de falsos positivos moderada (3.7%). Este enfoque de ensamblaje evidencia que la integración de características profundas provenientes de diferentes redes puede ser una estrategia efectiva para aplicaciones de apoyo clínico, siempre que se optimicen los recursos de procesamiento.

En términos prácticos, ResNet50 se posiciona como la mejor opción para entornos médicos donde es necesario minimizar tanto falsos negativos como falsos positivos y contar con un tiempo de entrenamiento razonable. InceptionV3 es más adecuado en escenarios donde la sensibilidad diagnóstica sea prioritaria y se pueda tolerar un nivel levemente superior de falsas alarmas. EfficientNetB0 es una alternativa interesante en soluciones embebidas o sistemas con restricciones de hardware, aunque su implementación clínica requiere optimización adicional.

Como recomendaciones técnicas, se sugiere:

- Ajustar los umbrales de decisión (thresholds) en los modelos InceptionV3 y EfficientNetB0 a valores cercanos a 0.4, y a 0.45 en ResNet50 y el modelo de fusión, con el fin de maximizar la sensibilidad sin comprometer excesivamente la especificidad.
- Implementar funciones de pérdida como focal loss o estrategias de ponderación de clases para reducir los falsos negativos, especialmente en aplicaciones donde la detección temprana de conjuntivitis sea crítica.
- Incrementar la regularización mediante Dropout (entre 0.5 y 0.7), penalización L2 y la incorporación de

Batch Normalization en las capas densas para mejorar la capacidad de generalización.

- Experimentar con algoritmos de optimización más robustos, como AdamW, y estrategias de ajuste dinámico del learning rate, como cosine decay, con el objetivo de acelerar la convergencia y estabilizar el entrenamiento.

REFERENCIAS

- [1] A. Y. Cheung, D. S. Choi, S. Ahmad, G. Amescua, V. Jhanji, A. Lin, S. I. Mian, M. K. Rhee, E. T. Viriya, F. S. Mah y D. M. Varu; American Academy of Ophthalmology Preferred Practice Pattern Cornea/External Disease Panel, "Conjunctivitis Preferred Practice Pattern," *Ophthalmology*, vol. 131, no. 4, pp. P134-P204, Apr. 2024. doi:10.1016/j.opht.2023.12.037.
- [2] M. Teweldemedhin, H. Gebreyesus, A. H. Atsbaha, S. W. Asgedom, and M. Saravanan, "Bacterial profile of ocular infections: A systematic review," *BMC Ophthalmol.*, vol. 17, no. 1, pp. 1-9, 2017, doi: 10.1186/s12886-017-0612-2.
- [3] D. S. W. Ting, C. Y. Cheung, G. Lim, N. M. Hsu, W. H. Tan, et al., "Development and Validation of a Deep Learning System for Diabetic Retinopathy and Related Eye Diseases Using Retinal Images From Multiethnic Populations With Diabetes," *JAMA*, vol. 318, no. 22, pp. 2211-2223, Dec. 2017, doi:10.1001/jama.2017.18152.
- [4] S. Mondal, S. Banerjee, S. Mukherjee, A. Ganguly, and D. Sengupta, "Deep Classifier for Conjunctivitis - A Three-Fold Binary Approach," *Int. J. Math. Sci. Comput.*, vol. 8, no. 2, pp. 46-54, 2022, doi: 10.5815/ijmsc.2022.02.05.
- [5] V. Gulshan, L. Peng, M. Coram, M. C. Stumpe, D. Wu, A. Narayanaswamy, et al., "Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs," *JAMA*, vol. 316, no. 22, pp. 2402-2410, Dec. 2016, doi:10.1001/jama.2016.17216.
- [6] S. Saha, S. K. Saha, S. Mukherjee, and S. Banerjee, "Deep learning for automated detection and classification of infectious conjunctivitis using slit-lamp images," *British Journal of Ophthalmology*, Published Online First: 28 Mar. 2024, doi:10.1136/bjo-2023-324310
- [7] C. Y. Gamage, "NEURAL NETWORKS FOR CLASSIFICATION OF EYE CONJUNCTIVITIS IN TELEHEALTH: A CONCEPTUAL ARCHITECTURE," vol. 6, no. 2, pp. 172-182, 2024.
- [8] F. Rashid, J. Abate, and A. Abdi, "Multiclass Classification and Identification of the External Eye Diseases using Deep CNN," *Indian J. Sci. Technol.*, vol. 17, no. 12, pp. 1107-1116, 2024, doi: 10.17485/ijst/v17i12.21.
- [9] R. K. Bawa and A. Koul, "Automated detection of conjunctivitis using convolutional neural network," *Appl. Data Sci. Smart Syst.*, pp. 91-97, 2024, doi: 10.1201/9781003471059-13.
- [10] C. J. Mann, "Observational research methods. Research design II: cohort, cross sectional, and case-control studies," *Emergency Medicine Journal*, vol. 20, no. 1, pp. 54-60, Jan. 2003, doi:10.1136/emj.20.1.54.
- [11] H. Kaur and N. K. Sandhu, "International Journal of Communication Networks and Information Security Evaluating the Effectiveness of the Proposed System Using F1 Score , Recall , Accuracy , Precision and Loss Metrics Compared to Prior Techniques," vol. 15, no. 04, pp. 368-383, 2023.
- [12] A. Géron, *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*, 3rd ed., Sebastopol, CA, USA: O'Reilly Media, 2022. ISBN: 978-1-098-18109-1.
- [13] D. Callaghan, J. Burger, and A. K. Mishra, "A machine learning approach to radar sea clutter suppression," *2017 IEEE Radar Conf. RadarConf2017*, pp. 1222-1227, 2017, doi: 10.1109/RADAR.2017.7944391.
- [14] G. Research, "What is Colaboratory?" [Online]. Available: <https://research.google.com/colaboratory/faq.html>
- [15] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2016-Decem, pp. 770-778, 2016, doi: 10.1109/CVPR.2016.90.
- [16] M. Tan and Q. V. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," *36th Int. Conf. Mach. Learn. ICML 2019*, vol. 2019-June, pp. 10691-10700, 2019.
- [17] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the Inception Architecture for Computer Vision," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2016-Decem, pp. 2818-2826, 2016, doi: 10.1109/CVPR.2016.308.
- [18] S. Patel, R. Patel, N. Ganatra, and A. Patel, "Spatial Feature Fusion for Biomedical Image Classification based on Ensemble Deep CNN and Transfer Learning," *Int. J. Adv. Comput. Sci. Appl.*, vol. 13, no. 5, pp. 153-159, 2022, doi: 10.14569/IJACSA.2022.0130519.
- [19] C. Shorten and T. M. Khoshgoftaar, "A survey on Image Data Augmentation for Deep Learning," *Journal of Big Data*, vol. 6, no. 60, pp. 1-48, 2019, doi:10.1186/s40537-019-0197-0.

Evaluación del carácter corrosivo del agua y su impacto en un sistema de enfriamiento industrial

Evaluation of the corrosive nature of water and its impact on an industrial cooling system

Zulennys Marcano-Marín^A y José Castro-Soto^B

Resumen— La calidad del agua es un factor crítico para la eficiencia y vida útil de los sistemas de enfriamiento industrial, pudiendo generar fallas operativas. La investigación tuvo como objetivo evaluar la calidad del agua de alimentación del sistema de enfriamiento cerrado en Briquetera del Caroní (BRIQCAR), ante la observación de un deterioro y aumento de fugas. El estudio de tipo descriptivo a través de la determinación de parámetros fisicoquímicos del agua de alimentación, el cálculo del Índice de Saturación de Langelier (ISL) y el análisis de la tendencia del pH e ISL a lo largo del periodo 2010-2025 mediante regresión lineal. Los resultados revelaron una tendencia estadísticamente significativa de disminución en el pH, la alcalinidad total y la dureza cálcica del agua a lo largo de los años. Los valores de pH pasaron de alcalinos a consistentemente ácidos (<5,5). El ISL calculado fue negativo, acentuando una tendencia corrosiva creciente. El análisis de la matriz de correlación confirma que el descenso del pH, la alcalinidad total y la dureza cálcica están interrelacionados directamente a la disminución del pH, lo cual es el impulsor del aumento del potencial corrosivo. Se concluye que la calidad del agua de alimentación ha evolucionado hacia una mayor agresividad química, contribuyendo al deterioro de los equipos y al incremento de fugas. Se recomienda un monitoreo continuo y la implementación de un programa de tratamiento químico correctivo para estabilizar el ISL, protegiendo la infraestructura y optimizando la eficiencia operativa.

Palabras clave: agua, análisis, corrosivo, impacto, industrial.

Abstract— Water quality is a critical factor for the efficiency and lifespan of industrial cooling systems, potentially leading to operational failures. The objective of this study was to evaluate the quality of the feedwater to the closed cooling system at Briquetera del Caroní (BRIQCAR), given the observation of deterioration and increased leaks. The descriptive study determined the physicochemical parameters of the feedwater, calculated the Langelier Saturation Index (LSI), and analyzed the pH and LSI trends over the period 2010–2025 using linear regression. The results revealed a statistically significant downward trend in the pH, total alkalinity, and calcium hardness of the water over the years. pH values shifted from alkaline to consistently acidic (<5.5). The calculated LSI was negative, accentuating an increasing corrosive tendency. Correlation matrix analysis confirms that the decrease in pH, total alkalinity, and calcium hardness are directly interrelated with the decrease in pH, which is the driving force behind the increase in corrosive potential. It is concluded that feedwater quality has evolved toward greater chemical aggressiveness, contributing to equipment deterioration and increased leaks. Continuous monitoring and the implementation of a corrective chemical treatment program are recommended to stabilize the LSI, protecting the infrastructure, and optimizing operational efficiency.

Keywords: water, analysis, corrosive, impact, industrial

1 INTRODUCCIÓN

El agua se considera un recurso insustituible e indispensable para la vida, el bienestar humano y el progreso social y económico de la Nación [1]. En este sentido, el agua representa un recurso esencial para el desarrollo sostenible de la nación venezolana. Sin embargo, uno de los principales problemas que se presentan en las instalaciones hidráulicas es la corrosión en los materiales metálicos, produciéndose perforaciones y obstrucciones de las líneas de agua [2].

Estos fenómenos están relacionados con las características fisicoquímicas del agua, así como con material con el cual entra en contacto, ya que las interacciones entre los materiales de la tubería y el agua pueden causar corrosión [3]. El control de la corrosión es un aspecto importante del suministro de agua potable segura. Los efectos de la corrosión, que pueden no ser evidentes sin monitoreo, son un problema importante que afecta tanto a la salud pública como a los aspectos económicos [4].

En Briquetera del Caroní (BRIQCAR) se cuenta con dos sistemas de enfriamiento, el primero se trata de un sistema de enfriamiento abierto (torres de enfriamiento) (ver Fig. 1a)

Tenga en cuenta que en la primera entrega no debe figurar: nombres de autores, mails, filiaciones y ORCID.

Autores:

A. **ORCID iD** <https://orcid.org/0009-0001-4823-803X>

Zulennys Marcano-Marín, está en CVG Briquetera del Caroní (BRIQCAR), Puerto Ordaz, Venezuela
zulennysmarcano80@gmail.com

B. **ORCID iD** <https://orcid.org/0009-0009-5786-4743>

José Castro-Soto, está en el Centro de Estudios en Química, Universidad Politécnica Territorial del Estado Bolívar, Ciudad Bolívar, Venezuela, CO 8001.
jilcastros78@gmail.com

y el segundo un sistema de enfriamiento cerrado (*Chillers*) (ver Fig. 1b). Estos sistemas son esenciales para la disipación del calor generado en procesos productivos y requieren de un suministro de agua con características fisicoquímicas controladas para prevenir el deterioro de los equipos. El agua de enfriamiento en ambos casos es utilizada para extraer el calor generado por las operaciones de planta [5]. Los equipos EVAPCO constan de serpentines fabricados en acero galvanizado con un diámetro de 25 mm y espesor de pared 1,5 mm, con una capacidad de $125\text{m}^3\text{ h}^{-1}$, en los cuales se deben manejar temperaturas de entrada en 52°C y de salida 30°C (ver Fig. 1c).

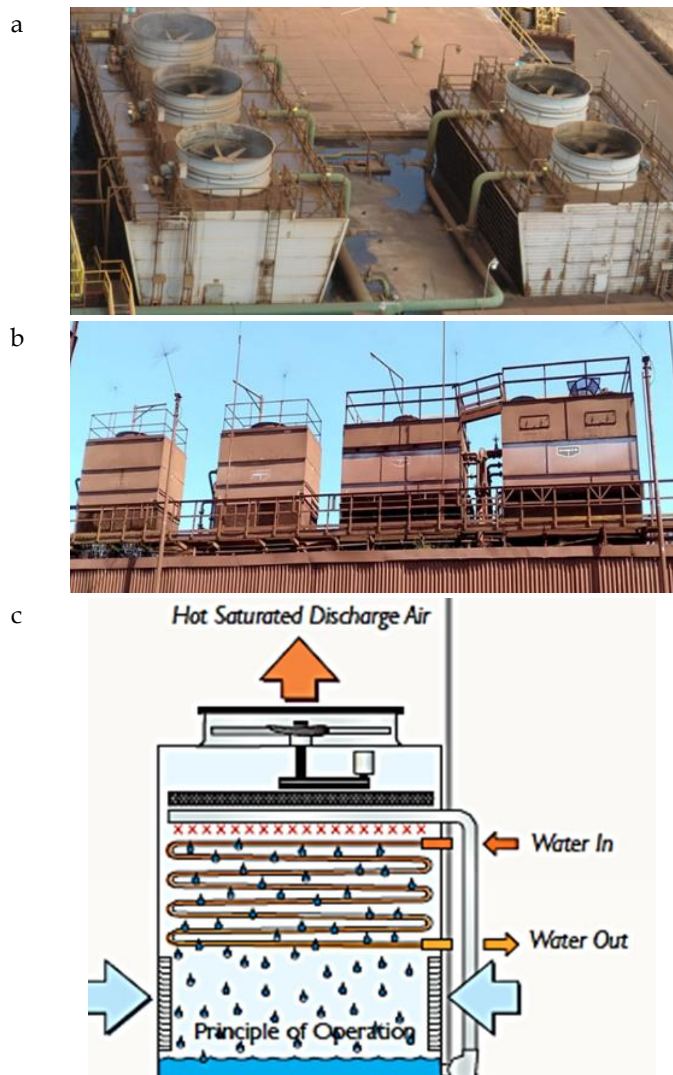


Fig. 1. a) Sistema de enfriamiento abierto (torres de enfriamiento), b) Sistema de enfriamiento cerrado (*Chillers*) y c). Principio de funcionamiento del equipo EVAPCO [5].

Estos equipos hacen circular el fluido de proceso a través de la bobina del enfriador de circuito cerrado. El calor del fluido de proceso se disipa a través de los tubos de la batería a la cascada de agua hacia abajo sobre los tubos. Al mismo tiempo, el aire es aspirado a través de la entrada de aire rejillas de ventilación en la base del refrigerador y se despla-

za hacia arriba sobre la bobina opuesta a la del flujo de agua. Una pequeña parte del agua se evapora que elimina el calor. El aire caliente y húmedo es descargado a la atmósfera. El agua restante cae al colector de agua en la parte inferior del refrigerador donde se recircula por la bomba a través del sistema de distribución de agua y dé marcha atrás en las bobinas [5].

Los sistemas de distribución pueden afectar la calidad del agua potable debido a las condiciones de la tubería y a la operación del sistema. Algunos parámetros son sensibles a la variación durante la distribución como: el cloro residual, pH, color y turbiedad, debido a que el material de la tubería puede presentar deterioro en estos sistemas a causa de la corrosividad del agua [6]. El agua de alimentación utilizada para el sistema de enfriamiento (circuito externo) proviene directamente de la empresa potabilizadora estatal. Se ha observado una alteración en la calidad del agua de alimentación, registros históricos sugieren una tendencia al deterioro de parámetros fisicoquímicos esenciales, lo que podría estar relacionado con el reporte de fugas y la disminución en la eficiencia de los equipos. En consecuencia, la vida útil de los equipos de enfriamiento se ve significativamente afectada por la calidad del agua de alimentación. De acuerdo con las especificaciones de EVAPCO el pH del agua en estos sistemas debería mantenerse dentro del rango de pH de 7,0 a 8,8 [5]. En BRIQCAR, se ha evidenciado corrosión en los equipos, lo que ha provocado fugas de agua y potencialmente un deterioro de la infraestructura (ver Fig. 2).



Fig. 2. Corrosión de tubos serpentines: a) serpentín, b) Corte transversal tubos.

El Índice de Saturación de Langelier (ISL) permite medir la estabilidad de la alcalinidad del agua y predecir su potencial corrosivo o incrustante [7]. Este índice indica si ésta tiene tendencia a disolver o a precipitar el carbonato de calcio (CaCO_3). Optimizar el ISL en el agua industrial es fundamental no solo para asegurar la calidad y evitar daños en tuberías y equipos, sino también para prolongar la vida útil de las instalaciones relacionadas con el agua [8].

En este sentido, la investigación se ve justificada al evaluar la calidad del agua de alimentación del sistema de enfriamiento cerrado en BRIQCAR, lo que es fundamental para comprender el impacto en la integridad de los equipos. Como hipótesis de la investigación, se planteó que la calidad del agua de alimentación ha experimentado una tendencia hacia condiciones corrosivas a lo largo del tiempo, lo que ha contribuido al posible deterioro de los equipos y al aumento de fangos en el sistema de enfriamiento cerrado.

La investigación corresponde al tipo: de campo, exploratorio y descriptivo [9], ya que buscó determinar los parámetros fisicoquímicos del agua de alimentación de los sistemas de enfriamiento cerrado. Adicionalmente, presenta un componente de análisis de tendencias, al evaluar el comportamiento del pH y el ISL a lo largo del tiempo, con el fin de comprender la evolución de la calidad del agua suministrada al sistema de enfriamiento cerrado.

2 MATERIALES Y MÉTODOS

2.1 RECOLECCIÓN DE LA MUESTRA

2.1.1 DATOS HISTÓRICOS (2010-2025)

Se obtuvieron registros semanales de los parámetros fisicoquímicos del agua de alimentación al sistema de enfriamiento cerrado, abarcando el período desde el año 2010 hasta el 2024. La recopilación de la información se realizó a partir de un análisis de datos históricos proporcionados por la Gerencia de Control de Calidad, Superintendencia de Laboratorio de Briquetera del Caroní (BRIQCAR, C.A.). Estos datos que incluían mediciones de pH, alcalinidad total, dureza cálcica, sólidos totales disueltos y temperatura, fueron procesados y organizados para su análisis. Utilizando una hoja de cálculo Microsoft Excel, se calcularon los valores promedio anuales para cada parámetro, lo que permitió consolidar la información y facilitar la evaluación del Índice de Saturación de Langelier (ISL).

2.1.2 DATOS 2025 [10], [11]

La muestra fue recolectada en la entrada de planta, antes de ingresar al equipo evaporativo EVAPCO, en envases de polietileno de 500 mL. Se realizó un muestreo semanal de

tipo aleatorio simple durante dos meses (desde enero hasta abril). Para la captación se abrió la línea de toma muestra ubicada en el área de entrada de los equipos de enfriamiento y se dejó drenar durante unos minutos, se curó el envase en tres ocasiones y se procedió a llenar el envase hasta el tope (ver Fig. 3), seguidamente se trasladó la muestra al laboratorio para su posterior análisis.



Fig. 3. Recolección de muestra de agua de alimentación en el punto de toma de muestra: entrada de planta.

2.2 DETERMINACIÓN DEL PH

2.2.1 CALIBRACIÓN DEL INSTRUMENTO [12]

Se encendió el equipo potenciómetro (Metrohm, 781 pH/Ion Meter) y se dejó estabilizar durante 15 minutos. Seguidamente se lavó el electrodo con agua destilada y se secó suavemente con una toalla de papel absorbente. Se tomaron tres vasos de precipitado de 100 mL y se rotularon con la denominación de cada disolución tampón, se añadió 70 mL de cada disolución, respectivamente. Se introdujo el electrodo en la disolución tampón pH 4,00 se dejó estabilizar la lectura en el equipo, se enjuagó el electrodo con agua destilada y con papel absorbente, el procedimiento se repitió con las otras disoluciones tampón.

2.2.2 LECTURA DE LA MUESTRA [13]

Se dejó aclimatar la muestra temperatura ambiente. Se agitó suavemente el envase con la muestra evitando generar burbujas. Se tomó un vaso de precipitado de 100 mL, se añadieron 70 mL de muestra. Se lavó el electrodo con agua destilada y se secó suavemente con una toalla de papel absorbente. Se introdujo el electrodo y se dejó estabilizar el instrumento y se registró la lectura del instrumento.

2.3 DETERMINACIÓN DE LA ALCALINIDAD

2.3.1 CALIBRACIÓN DEL INSTRUMENTO [14], [15]

La calibración del instrumento se realizó mediante el método de corrección por flotación. En un vaso de precipitado de 500 mL se añadió 400 mL de agua destilada y se dejó equilibrar a temperatura ambiente, se registró la temperatu-

ra para encontrar la densidad del agua. Se cargó con agua destilada el titulador automático (Metrohm, 775 Dosimat) y se dejó reposar durante unos minutos para comprobar que no haya fugas.

Por otro lado, en una balanza analítica (Mettler Toledo, AG204) se determinó la masa de un matraz Erlenmeyer de 125 mL con su tapón, el matraz se manipuló con una tira de papel evitando el contacto del conjunto (matraz + tapón) con los dedos para no variar su masa por la grasa que puedan dejar los dedos. Se colocó el matraz en el titulador, con mucho cuidado se retiró el tapón y se dejó fluir el líquido hacia el matraz hasta alcanzar el primer intervalo del volumen nominal, evitando salpicar y/o humedecer el cuello del matraz. Se tapó rápidamente el matraz y se registró el volumen vertido. Seguidamente se determinó la masa del conjunto (matraz + H₂O + tapón).

Se retiró el conjunto de la balanza y se llevó al titulador retirando el tapón justo antes de añadir desde el titulador de 10 a 20 mL y medir la masa vertida. Repetir el procedimiento para 30, 40 y 50 mL. En el procedimiento se realizaron adiciones sucesivas en el mismo matraz. Una vez terminada la recolección de datos, se calculó la masa del agua vertida (W_2) mediante la diferencia entre esta masa (m_2) y la masa del matraz vacío (m_1) (ver Ecuación 1).

$$W_2 = m_2 - m_1 \quad (1)$$

Dónde: W_2 : Masa del agua vertida.
 m_2 : Masa del conjunto Matraz+Agua+Tapón.
 m_1 : Masa del conjunto Matraz+Tapón.

Seguidamente, se realizó la corrección por flotación de la masa (masa corregida del objeto), es decir, el peso en el vacío en gramos (W_1) (ver Ecuación 2).

$$W_1 = W_2 + W_2 \left(\frac{d_{aire}}{d_{obj}} - \frac{d_{aire}}{d_{Masas}} \right) \quad (2)$$

Dónde: W_1 : Masa corregida del objeto.
 W_2 : Masa del agua vertida.
 d_{aire} : Densidad del aire desplazado por las masas y el objeto (0,0012 g/cm³).
 d_{obj} : Densidad del objeto (agua destilada a la temperatura del experimento).
 d_{Masas} : Densidad de la masa de referencia (8,0 g/cm³).

Una vez obtenida la masa del agua vertida (W_2) y el peso en el vacío en gramos (W_1), se calculó el volumen real (V_{real}) vertido por la bureta (ver Ecuación 3).

$$V_{real} = \frac{W_1}{d_{H_2O}} \quad (3)$$

Dónde: V_{real} : Volumen real vertido.
 W_1 : Masa corregida del objeto (g).
 d_{H_2O} : Densidad del agua destilada a la temperatura del experimento.

Seguidamente, se calculó el factor de corrección (FC) (ver Ecuación 4).

$$FC = V_{Real} - V_{Vertido} \quad (4)$$

Dónde: FC: Factor de corrección.
 V_{Real} : Volumen real vertido desde el titulador.
 $V_{Vertido}$: Volumen vertido por el titulador.

Finalmente, se graficó el volumen real (V_{Real}) o corregido que debe aplicarse como una función del volumen vertido ($V_{Vertido}$) obteniéndose el valor del factor de correlación (R^2).

2.3.2 ESTANDARIZACIÓN DEL HIDRÓXIDO DE SODIO (NaOH)

Se secó una porción de 10 g de biftalato ácido de potasio (KHC₈H₄O₄) en una estufa (Felisa, FE-292D) a 105°C durante una hora, transcurrido este tiempo se llevó a un desecador y se dejó enfriar a temperatura ambiente. En una balanza analítica (Mettler Toledo, AG204) se pesó 0,010 g de KHC₈H₄O₄ y se transfirió un matraz Erlenmeyer de 250 mL y se disolvió en 100 mL de agua destilada. Luego, se agregó unas gotas de fenolftaleína (0,5%) y empleando un titulador automático (Metrohm, 775 Dosimat) se tituló con el hidróxido de sodio (NaOH) al 0,01 N hasta el viraje de incoloro a rosa pálido. Finalmente se registró el volumen gastado y se calculó la concentración del NaOH (ver Ecuación 5).

$$N_{NaOH} = \frac{m_{Biftalato\ de\ potasio}}{PE_{Biftalato\ de\ potasio} \times V_{NaOH}} \quad (5)$$

Dónde: N_{NaOH} : Normalidad del NaOH.
 m_{KHP} : Masa del KHC₈H₄O₄ (mg).
 PE_{KHP} : Peso equivalente del KHC₈H₄O₄ (204,2).
 V_{NaOH} : Volumen del NaOH gastado (mL).

2.3.3 ESTANDARIZACIÓN DEL ÁCIDO SULFÚRICO (H₂SO₄)

Se midió 10 mL de la solución de ácido sulfúrico (H₂SO₄) al 0,01 N y se transfirió a un matraz Erlenmeyer de 50 mL, seguidamente se agregó unas gotas de fenolftaleína (0,5%) y empleando el titulador automático (Metrohm, 775 Dosimat) se tituló con hidróxido de sodio (NaOH) al 0,01 N estandarizado, hasta el viraje de incoloro a rosa pálido. Se tomó nota del volumen gastado de NaOH y se calculó la normalidad del H₂SO₄ (ver Ecuación 6).

$$N_1 = \frac{N_2 \times V_2}{V_1} \quad (6)$$

Dónde: N_1 : Normalidad del ácido H_2SO_4 .
 N_2 : Normalidad $NaOH$.
 V_1 : Volumen medido de H_2SO_4 (10 mL).
 V_2 : Volumen gastado de $NaOH$ (mL).

2.3.4 FACTOR DEL H_2SO_4

$$F = \frac{N_{H_2SO_4} \times PE_{CaCO_3} \times 1000}{V_{CaCO_3}} \quad (7)$$

Dónde: F : Factor del H_2SO_4 .
 $N_{H_2SO_4}$: Normalidad del H_2SO_4 .
 PE_{CaCO_3} : Peso equivalente del $CaCO_3$ (50,045).
 $V_{solución}$: Volumen de $CaCO_3$ (50 mL).

2.3.5 ANÁLISIS DE LA MUESTRA [16]

Se dejó aclimatar la muestra temperatura ambiente. Se agitó suavemente el envase con la muestra evitando generar burbujas. Se midió 50 mL de muestra filtrada con un filtro de membrana esterilizado (47 mm de diámetro y 0,45 μm), se transfirió a un matraz Erlenmeyer de 250 mL y se añadió unas gotas de fenoltaleína (0,5%), si la muestra se torna de color rosa, se introdujo un agitador magnético y se llevó a una plancha de agitación (Metrohm, 728 Stirrer) y empleando un titulador automático (Metrohm, 775 Dosimat) se tituló con H_2SO_4 normalizado, manteniendo agitación constante hasta un viraje a incoloro, se tomó el volumen gastado del H_2SO_4 (V_1). Seguidamente, se calculó la alcalinidad parcial (ver Ecuación 8) y se prosigue sin enrasar la bureta.

Alcalinidad Parcial

$$Ap \text{ (ppm de } CaCO_3) = V_1 \times F \quad (8)$$

Dónde: Ap : Alcalinidad parcial (ppm de $CaCO_3$).
 V_1 : Volumen gastado de H_2SO_4 en la determinación de Ap (mL).
 F : Factor del H_2SO_4 .

Sin embargo, si la muestra no se torna de color rosa, se agregó unas gotas del indicador mixto, si se torna de color azul, se tituló hasta el viraje a rosa salmón, se procedió a tomar el volumen gastado de H_2SO_4 (V_2). Seguidamente, se calculó la alcalinidad total (ver Ecuación 9).

Alcalinidad Total

$$At \text{ (ppm de } CaCO_3) = V_2 \times F \quad (9)$$

Dónde: At : Alcalinidad total (ppm de $CaCO_3$).
 V_2 : Volumen total gastado de H_2SO_4 en la determinación de At (mL).
 F : Factor del H_2SO_4 .

2.4 DETERMINACIÓN DE LA DUREZA CÁLCICA

2.4.1 ESTANDARIZACIÓN DEL EDTA

Se secó 10 g de carbonato de calcio ($CaCO_3$) en una estufa (Felisa, FE-292D) a 105°C durante una hora, transcurrido este tiempo se llevó a un desecador y se dejó enfriar a temperatura ambiente. En una balanza analítica (Mettler Toledo, AG204) se pesó entre 0,010 y 0,020 g de $CaCO_3$ seco. Luego se transfirió a un matraz Erlenmeyer de 250 mL y se añadió 50 mL de agua destilada caliente y HCl (1:1) gota a gota hasta disolver el $CaCO_3$. Luego se calentó a ebullición por varios minutos para eliminar el dióxido de carbono (CO_2) y se dejó enfriar a temperatura ambiente. Se agregó una pequeña cantidad de Negro de Eriocromo T (NET) y se ajustó a pH 10. Seguidamente, empleando un titulador automático (Metrohm, 775 Dosimat) se tituló con ácido etilendiaminotetraacético (EDTA) hasta un viraje de color violeta intenso a azul. Finalmente, se registró el volumen de titulante gastado y se calculó la normalidad del titulante (ver Ecuación 10).

$$N_{EDTA} = \frac{m_{CaCO_3}}{V_{EDTA} \times 0,05} \quad (10)$$

Dónde: N_{EDTA} : Normalidad del EDTA.
 m_{CaCO_3} : Masa de $CaCO_3$ (g).
 V_{EDTA} : Volumen del titulante gastado (mL).

2.4.2 FACTOR DEL EDTA

$$F = \frac{N_{EDTA} \times PE_{CaCO_3} \times 1000}{V_{muestra}} \quad (11)$$

Dónde: F : Factor del EDTA.
 N_{EDTA} : Normalidad del EDTA.
 PE_{CaCO_3} : Peso equivalente del $CaCO_3$ (50,04g·mol⁻¹).
 $V_{muestra}$: Volumen de muestra (mL).

2.4.3 ANÁLISIS DE LA MUESTRA [17]

Se dejó aclimatar la muestra temperatura ambiente. Se agitó suavemente el envase con la muestra evitando generar burbujas. Se midió 50 mL de la muestra y se transfirió a un matraz Erlenmeyer de 250 mL, se introdujo un agitador magnético y se llevó a una plancha de agitación (Metrohm, 728 Stirrer). Seguidamente, se ajustó a pH 12-13, agregando unas gotas de hidróxido de amonio (NH_4OH) en proporción 1:1. Una vez ajustado el pH, se añadió 0,1 g de murexida y empleando el titulador automático (Metrohm, 775 Dosimat) se tituló la muestra con EDTA hasta el viraje del color azul a fucsia. Finalmente, se registró el volumen gastado y se calculó la dureza cálcica del titulante (ver Ecuación 12).

$$D_{Ca} \text{ (ppm } CaCO_3) = V \times F \quad (12)$$

Dónde: D_{Ca} : Dureza cálcica (ppm de $CaCO_3$).
 V : Volumen gastado de EDTA (mL).
 F : Factor del EDTA.

D : $\log_{10} [A_{Total} (mg/L CaCO_3)]$

2.5 DETERMINACIÓN DE SÓLIDOS DISUELTOS TOTALES (TDS) [18]

Se lavó una cápsula de porcelana y se secó en estufa (FELISA, FE-292D) a $108^\circ C$ por dos horas, posteriormente se dejó enfriar en un desecador por 30 minutos, una vez frío se pesó en balanza analítica (Mettler Toledo, AG204), este ciclo se repitió hasta peso constante. Una vez alcanzado el peso constante en la capsula de porcelana, con una pipeta volumétrica se tomó 100 mL de muestra y procedió a filtrar la muestra al vacío con un filtro de membrana esterilizado (47 mm de diámetro y $0,45 \mu m$). Se recolectó el filtrado en la cápsula de porcelana (tarada) y se llevó a la estufa de sacado a $108^\circ C$ durante seis horas, se dejó enfriar en un desecador y luego se determinó su masa. Finalmente, se calculó los TDS (ver Ecuación 13).

$$TDS \left(\frac{mg}{L} \right) = \frac{(A + B) \times 1000}{C} \quad (13)$$

Dónde: TDS: Sólidos disueltos totales ($mg \cdot L^{-1}$).
 A : Peso de la cápsula de porcelana más residuo seco (mg).
 B : Peso de la cápsula de porcelana vacía (mg).
 C : Volumen de filtrado de muestra (mL).

2.6 DETERMINACIÓN DEL ÍNDICE DE SATURACIÓN DE Langelier (ISL) [4]

El Índice de Saturación de Langelier (ISL) es un modelo matemático del equilibrio derivado del grado de saturación del agua con respecto al carbonato de calcio ($CaCO_3$). La ecuación expresa la relación entre el pH, el calcio, la alcalinidad total, los sólidos disueltos y la temperatura con la solubilidad del $CaCO_3$ en aguas con un pH de 6,5 a 9,5 [4]. Empleando una hoja de cálculo de Microsoft Excel 2010 de donde se realizaron los cálculos para obtener las expresiones del Factor A (para TDS), Factor B (para temperatura), Factor C (para dureza cálcica) y Factor D (alcalinidad total). Una vez calculados los diversos factores se calcula el pHs (pH de saturación) (ver Ecuación 14).

$$pH_s = (9,3 + A + B) - (C + D) \quad (14)$$

Dónde: pHs: pH de saturación.
 A : $\frac{(\log_{10} [TDS (mg \cdot L^{-1})] - 1)}{10}$
 B : $[-13,12 \times \log_{10} (T (^{\circ}C + 273))] + 34,55$
 C : $\log_{10} \left[D_{Calcica} \left(\frac{mg}{L} CaCO_3 \right) \right] - 0,4$

Finalmente, se calculó el ISL (ver Ecuación 15).

$$ISL = pH - pH_s \quad (15)$$

Dónde: ISL: Índice de Saturación de Langelier.
pH: pH del agua.
pHs: pH de saturación.

Un ISL positivo (+) señala una sobresaturación de carbonatos, lo que propicia la formación de sarro y un comportamiento alcalino. Por el contrario, ISL negativo (-) indica una sub-saturación, resultando en agua más ácida y con tendencia a la corrosión. Un valor cercano a cero sugiere un equilibrio y, por lo tanto, un comportamiento neutro del agua (ver TABLA 1) [8].

TABLA 1

Tendencia del agua según valor del Índice de Saturación de Langelier (ISL) [8].

ISL	Tendencia del agua
+0,3 a +2,0	Incrustación alta.
0,0 a +0,3	Incrustación ligera con corrosión.
0,0	Equilibrada. Puede ocurrir corrosión ligera.
0,0 a -0,3	Corrosión ligera. No se forman incrustaciones
-0,3 a -2,0	Corrosión alta.

2.7 ANÁLISIS ESTADÍSTICO

Los datos obtenidos se analizaron empleando una hoja de cálculo Microsoft Excel 2010 para Windows, obteniéndose una estadística descriptiva, determinándose su media aritmética ($\bar{x}$) y desviación estándar (s).

2.7.1 MEDIA ($\bar{x}$)

Es la medida más utilizada para definir un valor central de un conjunto de datos. Se utiliza para reportar el valor más representativo de un conjunto de mediciones [14], [19] (ver Ecuación 16).

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} \quad (16)$$

Dónde: $\bar{x}$: Media.
 x_i : Valor numérico obtenido de la medida.
 n : Número de mediciones realizadas.

2.7.2 DESVIACIÓN ESTÁNDAR MUESTRAL (S)

Representa el grado de variabilidad o dispersión. Es una estimación de la precisión basada en las desviaciones de datos individuales con respecto a la media de una muestra de datos [14], [19] (ver Ecuación 17).

$$s = \sqrt{\frac{\sum_{i=1}^n (\bar{x} - x_i)^2}{n - 1}} \quad (17)$$

Dónde: s: Desviación estándar muestral (s).
 $x_i - \bar{x}$: Desviación de x_i del valor $\bar{x}$ de la media $\bar{x}$.
 n - 1: Número de grados de libertad.

2.7.3 COEFICIENTE DE CORRELACIÓN DE PEARSON [20]

Se utilizó el coeficiente de correlación de Pearson ya que puede utilizarse para medir el grado de relación de dos variables siempre y cuando ambas sean cuantitativas. Es fundamental distinguir que lo que mide el coeficiente de Pearson es la fuerza y la dirección de la relación lineal entre las variables. Así, un $r_{XY} = \pm 1$ indicará que hay una perfecta asociación lineal positiva o negativa –según sea el signo– entre X y Y. En este orden de ideas, un $r_{XY} = 0$ será evidencia de que no existe relación lineal entre las características de interés, pero no constituye prueba alguna de independencia. En otras palabras, si el coeficiente de correlación es nulo, únicamente se podría establecer que no hay asociación lineal entre las variables aunque pueda existir otro tipo de vinculación [20] (ver Tabla 2).

TABLA 2

Interpretación de Coeficiente de correlación de Pearson [21].

Valor	Interpretación
$r < 0$	Correlación negativa, las dos variables se correlacionan en sentido inverso. A valores altos de una de ellas le suelen corresponder valores bajos de la otra y viceversa.
$r > 0$	Correlación positiva, las dos variables se correlacionan en sentido directo. A valores altos de una le corresponden valores altos de la otra e igualmente con los valores bajos.
$r = 0$	No hay correlación entre las variables.

Se supone que la relación se da entre X y Y, pero aplica a cualquier par de variables. Se plantea el valor absoluto del coeficiente, de modo que la magnitud es independiente del signo (ver Tabla 3).

TABLA 3

Interpretación de la magnitud del coeficiente de correlación de Pearson según las sugerencias de Cohen [20].

Rango de valores de r_{XY}	Interpretación
$0,00 \leq r_{XY} < 0,10$	Correlación nula

$0,10 \leq r_{XY} < 0,30$	Correlación débil
$0,30 \leq r_{XY} < 0,50$	Correlación moderada
$0,50 \leq r_{XY} < 1,00$	Correlación fuerte

3 RESULTADOS Y DISCUSIÓN

3.1 PARÁMETROS FÍSICO-QUÍMICOS

3.1.1 CALIBRACIÓN DEL PHMETRO

La calibración del instrumento arrojó un coeficiente de correlación (R^2) de 0,9999, lo que demuestra una relación lineal entre los valores de pH de los buffers y las lecturas del equipo. Esto asegura que el pHmetro proporciona mediciones precisas y confiables, garantizando la exactitud de los datos de pH obtenidos para el estudio (ver Fig. 4).

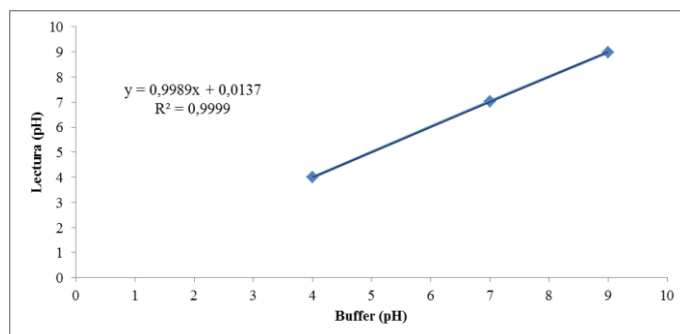


Fig. 4. Regresión lineal de la calibración del equipo.

3.1.2 CALIBRACIÓN DEL TITULADOR AUTOMÁTICO

La calibración del instrumento arrojó un coeficiente de correlación (R^2) de 1, esto indica una excelente relación entre el volumen nominal y el volumen real dispensado por el instrumento. Estos resultados garantizan que el titulador automático es altamente preciso y confiable para las mediciones volumétricas, validando su idoneidad para análisis cuantitativos (ver Fig. 5).

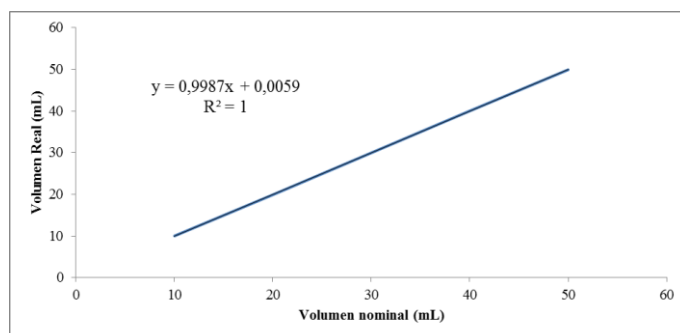


Fig. 5. Regresión lineal de la calibración del equipo.

3.1.3 ESTANDARIZACIÓN DE LAS DISOLUCIONES TITULANTES

Se estandarizaron las disoluciones de hidróxido de sodio (NaOH), ácido sulfúrico (H_2SO_4) y ácido etilendiamino-tetraacético (EDTA). El análisis de los datos de media y des-

viación estándar de sus concentraciones revela una excelente precisión en la preparación y estandarización de estas disoluciones. Obteniéndose, una desviación estándar notablemente baja ($0,0002 \text{ mol} \cdot \text{L}^{-1}$ para NaOH y $0,0001 \text{ mol} \cdot \text{L}^{-1}$ para el H_2SO_4 y el EDTA) (ver Tabla 4). Esto valores indica una alta reproducibilidad en el proceso de calibración, lo que se traduce directamente en una gran confianza en la exactitud de las concentraciones de los titulantes.

TABLA 4

Resultados de estandarización de disoluciones titulantes (concentración media y desviación estándar).

Titulante	V _{titulante} (mL)	M _{titulante} (mol·L ⁻¹)	Media (mol·L ⁻¹)	Desviación es- tándar (mol·L ⁻¹)
Hidróxido de sodio (NaOH)	3,09	0,0160	0,0162	0,0002
	3,14	0,0164		
	3,13	0,0161		
Ácido sulfúrico (H ₂ SO ₄)	5,76	0,0094	0,0094	0,0001
	5,79	0,0094		
	5,84	0,0095		
EDTA	20,10	0,0099	0,0098	0,0001
	20,13	0,0098		
	20,11	0,0098		

3.1.4 PARÁMETROS FÍSICO – QUÍMICOS

Los datos históricos del agua de alimentación al sistema de enfriamiento cerrado, correspondientes a los valores promedios anuales de pH, alcalinidad total (A_{Total}), dureza cálcica ($D_{\text{cálcica}}$), sólidos totales disueltos (TDS) y temperatura (K) desde el año 2010 – 2024. Asimismo, se reportan correspondientes al 2025 (desde enero hasta abril) (ver Tabla 5).

TABLA 5

Parámetros de calidad (promedios anuales) del agua de alimentación al sistema de enfriamiento cerrado

Año	pH	A _{Total} (ppm)	D _{Total} (ppm)	TDS (ppm)	T (°C)
2010	8,46	13,65	24,55	3,05	25,00
2011	7,48	11,39	19,18	2,25	25,00
2012	7,82	12,10	20,43	2,21	25,00
2013	7,14	10,41	23,64	2,06	25,00
2014	6,34	6,41	17,89	2,14	25,00
2015	6,49	9,60	17,39	2,00	25,00
2016	5,84	7,14	16,18	3,55	25,00
2017	5,16	6,09	15,07	2,61	25,00
2018	5,91	6,08	7,94	2,00	25,00
2019	6,39	9,27	11,94	2,30	25,00
2020	6,16	8,89	17,48	2,52	25,00
2021	5,34	7,39	11,47	2,92	25,00
2022	5,10	5,50	6,32	2,07	25,00
2023	4,88	5,09	4,90	1,27	25,00
2024	4,73	4,97	3,31	2,00	25,00

2025	4,93	5,16	3,99	1,78	25,00
------	------	------	------	------	-------

3.1.4.1 pH

Se observó una constante disminución del pH a lo largo del tiempo, pasando de valores iniciales alcalinos de pH 8,46 (2010) a valores ácidos (pH <5,5) a partir del año 2017, con un valor mínimo de pH 4,73 en el año 2024. Aunque se aprecia una ligera recuperación en el 2025 (pH 4,93), la marcada tendencia hacia la acidez indica un alto riesgo de corrosión para los equipos. Se puede observar que los valores disminuyen en el tiempo desde el año 2010 (ver Tabla 5). En términos de calidad del agua, al comparar los valores obtenidos con los valores pH deseables en el agua potable (pH entre 6,4 y 8,5) [22], sólo en el periodo comprendido entre los años 2010 y 2013 logró mantenerse dentro de estos valores deseables, descendiendo considerablemente a partir del 2014. La corrosividad del agua disminuye a medida que aumenta la alcalinidad donde es necesario proporcionar un pH estable en todo el sistema de distribución para el control de la corrosión de las tuberías metálicas [4].

3.1.4.2 ALCALINIDAD TOTAL

La alcalinidad total experimenta una reducción considerable desde 13,65 ppm (2010) hasta valores <6,5 ppm (2014), con un marcado descenso en los años recientes (ver Tabla 5). La disminución observada, podría ser un factor contribuyente a esta acidificación y en consecuencias al aumento de la corrosividad del agua.

3.1.4.3 DUREZA CÁLCICA

La dureza cálcica presenta una tendencia decreciente aún más pronunciada, pasando de 24,55 ppm (2010) en un constante descenso hasta el año 2018, con ligeros periodos de aumento y recuperación (2013) y con una marcada tendencia decreciente. Siendo, el año 2019 la última recuperación y de nuevo la marcada tendencia decreciente hasta valores de 3,99 ppm (2025) (ver Tabla 5). En términos de calidad del agua, al comparar los valores obtenidos con los valores deseables en el agua potable (250 mg/L CaCO_3) [22], durante todo el periodo de estudio se reportaron valores muy por debajo del valor deseable, (<24,55 mg/L CaCO_3), la reducción en la dureza cálcica podría estar relacionada con esta tendencia corrosiva. Una capa uniforme de carbonato de calcio (CaCO_3), depositada sobre las superficies metálicas, separa físicamente el metal del ambiente corrosivo [23]. Es decir, en sistemas con una dureza moderada y una alcalinidad adecuada, el agua puede tener una ligera tendencia a formar una fina capa de CaCO_3 en las superficies metálicas. Esta capa, si es uniforme y está bien adherida, puede actuar

como una barrera protectora, aislando el metal del agua y reduciendo la velocidad de corrosión.

3.1.4.4 SÓLIDOS TOTALES DISUELTOS (TDS)

Los TDS muestran diversas fluctuaciones menores sin una tendencia clara y se mantienen en niveles generalmente bajos (ver Tabla 5). Por otra parte, bajos niveles de TDS podrían señalar una insuficiencia de minerales que podrían ayudar a inhibir la corrosión o formar capas protectoras en las superficies metálicas. En términos de calidad del agua, al comparar los valores obtenidos con los valores deseables en el agua potable ($600 \text{ mg} \cdot \text{L}^{-1}$) [22], durante todo el periodo de estudio se reportaron valores muy por debajo del valor deseable ($<3,55 \text{ mg} \cdot \text{L}^{-1}$).

3.1.4.5 TEMPERATURA

En el caso de la temperatura del agua, presentaron valores constante y estable a 25°C durante todo el período objeto de estudio (ver Tabla 5).

3.2 ÍNDICE DE SATURACIÓN DE LANGE LIER (ISL)

El Índice de Saturación de Langelier (ISL) obtenido es negativo y con grandes valores en magnitud. Estos valores han ido en incremento del carácter corrosivo del agua, a través de los años según estudio realizado desde el 2010, observándose que los valores calculados de ISL en la actualidad se mantienen con un carácter corrosivo elevado (ver Fig. 6). Un ISL negativo indica que el agua está sub-saturada con respecto al carbonato de calcio (CaCO_3) y, por lo tanto, tiene una tendencia a ser corrosiva [7]. Cuanto más negativo es el ISL, mayor es la tendencia a la corrosión [8].

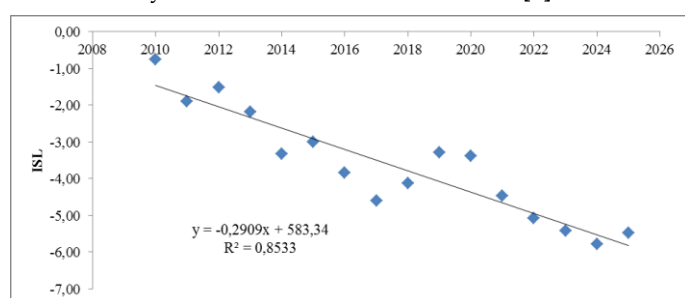


Fig. 6. Tendencia de variación del Índice de Saturación de Langelier (ISL) en el agua de alimentación al sistema cerrado de BRIQCAR, CA (2010-2025).

La tendencia del Índice de Saturación de Langelier (ISL), la regresión lineal ($y = -0,2909x + 583,44$) muestra una tendencia decreciente del ISL a lo largo de los años, con pendiente de $-0,321$ indica que el ISL disminuye, en promedio $0,321$ unidades por año (ver Fig. 6). La tendencia decreciente y negativa del ISL indica que el agua posee una tendencia a ser más corrosiva con el tiempo.

3.3 MATRIZ DE CORRELACIÓN ENTRE LOS PARÁMETROS MEDIDOS Y EL ÍNDICE DE SATURACIÓN DE LANGE LIER (ISL).

Se trató de identificar los parámetros que afectan significativamente en el Índice de Saturación de Langelier (ISL), un análisis de correlación de Pearson (r) permitió cuantificar la relación lineal entre los parámetros: pH, Alcalinidad Total, Dureza cálcica, TDS, la Temperatura y el ISL (ver Tabla 6).

TABLA 6

Matriz de correlación entre los parámetros fisicoquímicos y el Índice de Saturación de Langelier

	pH	AlcTotal	Dcálcica	TDS	T	ISL
pH	1					
AlcTotal	0,96	1				
Dcálcica	0,87	0,85	1			
TDS	0,29	0,34	0,47	1		
T	0,00	0,00	0,00	0,00	1	
ISL	0,99	0,96	0,92	0,34	0,00	1

El Índice de Saturación de Langelier (ISL) presenta una relación extremadamente fuerte y positiva con el pH ($r=0,99$), esto confirma que la disminución progresiva del pH es la principal causa de la caída del ISL. En cuanto a la alcalinidad total también se observa una correlación muy fuerte y positiva con el ISL ($r=0,96$). La alcalinidad actúa como una capacidad de amortiguación del agua; por lo tanto, la pérdida de esta capacidad está directamente ligada al empeoramiento del carácter corrosivo. De igual manera para la Dureza cálcica ($r=0,92$) por lo tanto el ISL se mueven en la misma dirección. Estos valores indican que a medida que el pH, la alcalinidad y la dureza cálcica disminuyen el ISL tiende a bajar haciéndose más negativo, lo que directamente se traduce en un aumento en el potencial corrosivo del agua.

4 CONCLUSIONES

Los parámetros fisicoquímicos del agua de alimentación al sistema de enfriamiento cerrado de BRIQCAR (2010-2025), revelaron una tendencia significativa al descenso en los niveles de pH, alcalinidad total y dureza cálcica a lo largo de los años. El cálculo anual del Índice de Saturación de Langelier (ISL) mostró una tendencia decreciente, resultando en valores progresivamente más negativos. Esto indica una tendencia del agua a ser corrosiva, lo que representa un riesgo para la integridad de los equipos lo que concuerda con el incremento de reporte de fugas. El análisis de la tendencia del ISL confirmó una disminución significativa. El análisis de la matriz de correlación confirma que el descenso del pH, la alcalinidad total y la dureza cálcica no son eventos aislados, sino que están interrelacionados. La pérdida de

alcalinidad y dureza contribuye directamente a la caída del pH, lo cual es el principal impulsor del aumento del potencial corrosivo. Esta tendencia respalda la hipótesis de que la calidad del agua ha evolucionado a condiciones corrosivas del agua que compromete la integridad de los equipos.

5 REFERENCIAS

- [1] Gaceta Oficial 38.595, «Ley de Aguas: Asamblea Nacional,» 02 Enero 2007. [En línea]. Available: <https://www.asambleanacional.gob.ve/storage/documentos/leyes/ley-de-agu-20220208194620.pdf>. [Último acceso: 02 Febrero 2025].
- [2] V. Mora Arellano y J. L. Cedeño, «Valuación de corrosión e incrustaciones en instalaciones hidráulicas de una planta de potabilización en el estado Bolívar, Venezuela,» 2006. [En línea]. Available: https://ve.scielo.org/scielo.php?script=sci_arttext&pid=S1690-46482006000100008. [Último acceso: 30 Agosto 2025].
- [3] C. De Sousa, A. Correia y M. C. Colmenares, «Corrosión e incrustaciones en los sistemas de distribución de agua potable: Revisión de las estrategias de control,» 2010. [En línea]. Available: https://ve.scielo.org/scielo.php?script=sci_arttext&pid=S1690-46482010000200003. [Último acceso: 30 Agosto 2025].
- [4] A. Awatif, K. Basim y A. Sura, «Evaluating Water Stability Indices from Water Treatment Plants in Baghdad City,» *Journal of Water Resource and Protection*, vol. VI, n° 14, pp. 1344-1351, 2014. <https://doi.org/10.4236/jwar.2014.614124>
- [5] evapco, «Closed circuit cooler. Instrucciones de operaciones de mantenimiento,» EVAPCO, Inc, Westminster, USA.
- [6] J. F. Rios, J. A. Calderón, F. Echeverría y G. Peñuela, «Diseño de un sistema piloto para el estudio de la corrosión del material de tubería y su contribución en el deterioro de la calidad del agua potable,» *Revista Facultad de Ingeniería Universidad de Antioquia*, n° 43, pp. 102-113, 2008.
- [7] Hanna Instruments, «Equilibrio del agua. Índice de Langelier,» 2025. [En línea]. Available: <https://www.hannainst.es/blog/790/equilibrio-del-agua-indice-de-angelier>. [Último acceso: 05 Enero 2025].
- [8] Carbotecnia, «La tendencia incrustante, corrosiva o equilibrada de un agua, y el Índice de saturación de Langelier,» 20 Junio 2024. [En línea]. Available: <https://www.carbotecnia.info/aprendizaje/quimica-del-agua/como-se-calcula-el-indice-de-angelier/>. [Último acceso: 06 Diciembre 2024].
- [9] UPEL, *Manual de Trabajos de Grado de Especialización y Maestría y Tesis Doctorales*, Quinta ed., Caracas: FEDUPEL, 2016.
- [10] COVENIN 2614, «Agua Potable. Toma de muestra,» FONDONORMA, Caracas, 1994.
- [11] COVENIN 2709, «Aguas naturales, industriales y residuales. Guía para las técnicas de muestreo,» Fondonorma, Caracas, 2024.
- [12] Práctica Operativa POP10LE01, «pHmetro Metrohm, 781,» Gerencia de Mantenimiento, Puerto Ordaz, 2009.
- [13] COVENIN 2462, «Aguas naturales, industriales y residuales : Determinación del pH,» FONFONORMA, Caracas, 2002.
- [14] D. Skoog, D. West, F. Holler y S. Crouch, *Fundamentos de química analítica*, Novena ed., México, D.F.: Cengage Learning Editores, S.A. de C.V., 2015.
- [15] D. Harris, *Análisis Químico Cuantitativo*, Tercera ed., Barcelona: REVERTÉ, 2016.
- [16] COVENIN 2188, «Agua potable. Determinación de alcalinidad,» FONDONORMA, Caracas, 1984.
- [17] COVENIN 2408-86, «Agua- Determinación de Dureza Total y Calcio. Método volumetrico. Dterminación de Magnesio por cálculo,» FONDONORMA, Caracas, 1986.
- [18] COVENIN 2461, «Agua natural, industrial y residual, Determinación de sólidos,» FONDONORMA, Caracas, 1987.
- [19] J. N. Miller y J. C. Miller, *Estadística y Quimiometría para Química Analítica*, Cuarta ed., Madrid: Pearson Educación S.A., 2002.
- [20] J. D. Hernández Lalinde, F. Espinosa Castro, J. E. Rodríguez, J. G. Chacón Rangel, C. A. Toloza Sierra, M. K. Arenas Torrado, S. M. Carrillo Sierra y V. J. Bermúdez Pirela, «Sobre el uso adecuado del coeficiente de correlación de Pearson: definición, propiedades y suposiciones,» *Archivos Venezolanos de Farmacología y Terapéutica*, vol. XXXVII, n° 5, pp. 587-601, 2018.
- [21] A. D. Pauta y D. A. Quezada, «Determinación de la tendencia corrosiva e incrustante del agua potable distribuida en la ciudad de Azogues,» Cuenca, Ecuador, 2017.
- [22] Ministerio de Sanidad y Asistencia Social, «Normas Sanitarias de calidad del agua potable,» Caracas, 1998.
- [23] Water Technologies, «Capítulo 24 - Control de la corrosión - Sistemas de refrigeración,» 2025. [En línea]. Available: <https://www.watertechnologies.com/handbook/chapter-24-corrosion-control-cooling-systems>. [Último acceso: 05 Junio 2025].

6 BIOGRAFÍAS

Zulennys del Carmen Marcano Marín, egresada del Instituto Universitario de Tecnología José Antonio Anzoátegui - Estado Anzoátegui obteniendo el título de Técnico Superior Universitario en Química, mención Procesos Químicos. Con diez años de experiencia en tratamiento de agua potable en la empresa estatal Hidrobolivar ejerciendo el cargo de analista de control de calidad y nueve años de experiencia en procesos de reducción directa Midrex. Actualmente como superintendente (E) de procesos y automatización adscrita a la gerencia técnica y proyectos en la empresa Briqcar, C.A.

José Castro-Soto. Licenciado en Química egresado de la Universidad De Oriente (Núcleo de Sucre). Con más de 15 años de experiencia en la educación universitaria. Ha sido jefe de la División de Laboratorio (Gerencia de Control de Calidad - HIDROBOLÍVAR C.A.). Actualmente, se desempeña como Jefe del Departamento del Programa Nacional de Formación en Química (PNFQuímica) en la Universidad Politécnica Territorial del Estado Bolívar (UPTBolívar).

Aplicación de inteligencia artificial para predecir la rotación laboral de un supermercado en Juliaca- Puno

Application of artificial intelligence to predict employee turnover in a supermarket in Juliaca – Puno

Percy Quispe Ñaca^A, Edwin Mestas Yucra^B,
Aldo H. Zanabria-Gálvez^C, N. Sheyla Saenz Bermejo^D
y Jackeline M. Chambilla Alata^E

Resumen— El presente estudio se centró en el desarrollo y la evaluación de un modelo de inteligencia artificial destinado a la predicción de la rotación de personal en la sección de productos frescos de un supermercado ubicado en Juliaca, Perú, así como en la identificación de los factores determinantes principales de este fenómeno. Se adoptó un enfoque metodológico cuantitativo y predictivo, empleando algoritmos de aprendizaje automático (Regresión Logística, Random Forest, XGBoost, KNN) sobre una base de datos compuesta por 1,000 registros de empleados. La metodología abarcó el preprocesamiento de datos, la codificación de variables, el escalado de características y la implementación de la técnica SMOTE para contrarrestar el desequilibrio de clases. Los hallazgos revelaron que el modelo KNN exhibió el rendimiento predictivo superior, logrando un F1-Score de 0.599. El análisis de la importancia de las características indicó que el salario mensual, la evaluación de desempeño y la antigüedad constituyeron los predictores de mayor influencia. En conclusión, se establece que la inteligencia artificial representa una herramienta robusta y factible para la gestión proactiva de los recursos humanos, facultando a la gerencia para formular estrategias de retención enfocadas y fundamentadas en evidencia, con el fin de mitigar la rotación en el contexto analizado.

Palabras clave: inteligencia artificial, machine learning, predicción, recursos humanos, rotación de personal, sector retail.

Abstract— This study focused on the development and evaluation of an artificial intelligence model for predicting employee turnover in the fresh produce section of a supermarket located in Juliaca, Peru, and on identifying the main determinants of this phenomenon. A quantitative and predictive methodological approach was adopted, employing machine learning algorithms (Logistic Regression, Random Forest, XGBoost, KNN) on a database composed of 1,000 employee records. The methodology included data preprocessing, variable encoding, feature scaling, and the implementation of the SMOTE technique to counteract class imbalance. The findings revealed that the KNN model exhibited the superior predictive performance, achieving an F1-Score of 0.599. Feature importance analysis indicated that monthly salary, performance appraisal, and seniority were the most influential predictors. In conclusion, it is established that artificial intelligence represents a robust and feasible tool for proactive human resource management, empowering management to formulate focused and evidence-based retention strategies to mitigate turnover in the analyzed context.

Keywords: artificial intelligence, employee turnover, human resources, machine learning, prediction, retail sector.

1. INTRODUCCIÓN

La rotación de personal, definida como el flujo de empleados que entran y salen de una organización, es un fenómeno que impacta negativamente la productividad y estabilidad de las empresas [1]. Los costos asociados a la alta rotación no son triviales, pues incluyen gastos directos de reclutamiento y capacitación, así como costos indirectos como la pérdida de productividad y la disminución de la moral del equipo [17]. En este sentido, la salida de empleados representa una pérdida de capital intelectual y activos

valiosos para la compañía [13].

El sector retail o minorista se caracteriza por registrar tasas de rotación de personal particularmente elevadas en comparación con otras industrias [3]. Esta tendencia se atribuye a factores como el estrés laboral, el agotamiento y la intensa competencia en el mercado laboral [4]. De hecho, la industria del retail en Estados Unidos ha experimentado tasas de rotación superiores al 60%, muy por encima del promedio general [22].

En el contexto de América Latina, la rotación del personal



Revista de Investigación en Ciencia y Tecnología

ISSN: 2810-8124 (en línea) / ISSN: 2706-543x

Universidad Nacional Micaela Bastidas de Apurímac – Perú

Vol. 7 Núm. 1 (2025) - Publicado: 29/10/25 - Indexaciones

Número: doi.org/10.57166/riqchary/v7.n1.2025

Páginas: 76- 81 | Recibido 01/01/2025 ; Aceptado 01/02/2025

doi.org/10.57166/riqchary.v7.n2.2025.9

Autores:

- A. ORCID iD <https://orcid.org/0009-0001-1376-3838>
Percy Quispe-Ñaca, Universidad Nacional del Altiplano – PE percy@unap.edu.pe
- B. ORCID iD <https://orcid.org/0000-0002-6000-1094>
Edwin Mestas-Yucra, Universidad Nacional del Altiplano, PE mestased@gmail.com
- C. ORCID iD <https://orcid.org/0000-0003-3314-8768>
Aldo H. Zanabria Gálvez, Universidad Nacional del Altiplano, PE. aldo.zanabria@unap.edu.pe
- D. ORCID iD <https://orcid.org/0000-0003-3314-8768>
N. Sheyla.Saenz Bermejo, Universidad Nacional del Altiplano, PE. nayshasaenz@unap.edu.pe
- E. ORCID iD <https://orcid.org/0009-0001-3375-3610>
Jackeline M. Chambilla Alata, Universidad Nacional del Altiplano, PE. jackelinechambilla@est.unap.edu.pe

también es un problema grave para las áreas de talento humano, aunque no existen datos homogéneos que permitan cuantificar su magnitud de forma precisa [18]. Perú, en particular, ha mostrado uno de los índices de rotación más altos de la región, alcanzando un 20.7% según informes recientes [5]. Este fenómeno se ve impulsado por un crecimiento económico que ha generado un mercado laboral más dinámico y competitivo, donde las empresas deben "luchar" para retener a su personal calificado [9].

La investigación sobre este tema en Perú se ha centrado mayoritariamente en Lima, la capital [8] [10]. Sin embargo, existe un vacío de conocimiento sobre las dinámicas de rotación en ciudades regionales como Juliaca, en la sierra del país, cuyo contexto social y económico difiere del de la metrópoli. Este estudio busca llenar esa brecha, analizando específicamente el área de frescos de un supermercado en dicha ciudad.

Ante este desafío, la inteligencia artificial (IA) y el machine learning (ML) emergen como herramientas estratégicas para la gestión de recursos humanos [15]. La aplicación de IA permite analizar grandes volúmenes de datos para predecir la probabilidad de que un empleado abandone la empresa [20]. Estos modelos predictivos no solo identifican a los empleados en riesgo, sino que también ayudan a comprender los factores subyacentes que impulsan la rotación [12].

Diversos estudios han demostrado la eficacia de algoritmos como Random Forest, Regresión Logística y, especialmente, XGBoost para esta tarea [6] [7]. Estos modelos analizan variables como la satisfacción laboral, el salario, la antigüedad y las evaluaciones de desempeño para generar predicciones precisas [14] [17]. La implementación de estas tecnologías permite a las organizaciones pasar de un enfoque reactivo a uno proactivo, diseñando estrategias de retención focalizadas y basadas en evidencia [21].

El objetivo de esta investigación es, por tanto, desarrollar y evaluar un modelo de IA para predecir la rotación de personal en el área de frescos de un supermercado en Juliaca, identificando los factores más influyentes. Este trabajo pretende ofrecer una herramienta práctica para la gestión local y, a su vez, aportar conocimiento sobre la dinámica laboral en una región andina de Perú, un área insuficientemente explorada en la literatura académica.

2. MATERIALES Y MÉTODOS

2.1. diseño y tipo de estudio

La investigación adoptó un diseño cuantitativo y predictivo, basado en técnicas de aprendizaje automático supervisado. Se desarrolló un estudio de corte transversal analítico, utilizando datos históricos anonimizados de empleados del área de productos frescos para entrenar y validar modelos

predictivos de rotación laboral.

2.2. ámbito y lugar de estudio

La investigación se llevó a cabo utilizando datos correspondientes al personal del área de productos frescos de un supermercado ubicado en la ciudad de Juliaca. Juliaca es la ciudad más grande de la región de Puno, situada en la zona andina del sureste de Perú, a una altitud de 3,825 metros sobre el nivel del mar. Este contexto geográfico y socioeconómico particular la distingue de la capital, Lima, donde se concentran la mayoría de los estudios sobre este tema en el país.

2.3. análisis predictivo

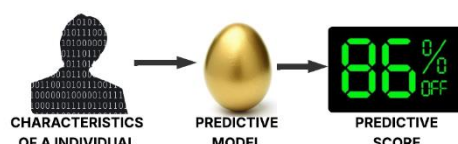
Se trabajó con un diseño predictivo porque después de reconocer las relaciones entre variables mediante el uso de técnicas de aprendizaje computacional y verificar las suposiciones adecuadas, se descubren patrones de comportamiento que



posibilitan la construcción de un modelo predictivo [11].

Fig 1. Diseño de modelo predictivo para la construcción de patrones

En la figura 1 muestra un diseño predictivo porque después de reconocer las relaciones entre variables mediante el uso de técnicas de aprendizaje computacional y verificar las suposiciones adecuadas, se descubren patrones de comportamiento que posibilitan la construcción de un



modelo predictivo [11]

Figura 2. Las características de un empleado individual se introducen en el modelo entrenado, que genera un 'Predictive Score'

En la figura 2 se muestra la aplicación del modelo predictivo con la que es posible anticipar las posibilidades de que una persona, basándose en los datos disponibles sobre ella, responda de manera específica. Al ingresar los datos de la persona y aplicar el modelo predictivo, se generará una puntuación que reflejará la probabilidad de que ocurra la situación analizada por el modelo [11].

2.4. descripción de métodos

a) Periodo de estudio o frecuencia de muestreo
El estudio es de corte transversal, basado en un conjunto de datos históricos y anonimizados que representan una instantánea de la plantilla de empleados en un momento determinado. No se realizó un muestreo periódico, sino que se analizó la totalidad de los 1,000 registros disponibles en el dataset proporcionado.

b) Descripción detallada de los materiales, insumos e instrumentos utilizados el principal instrumento de recolección de datos fue la base de datos <https://www.kaggle.com/code/edwinmestas2024/renuncia> Para el procesamiento y análisis de los datos, se utilizó el entorno de programación Python (versión 3.9) con las siguientes librerías especializadas en ciencia de datos:

- **Pandas:** Para la manipulación y el análisis de los datos.
- **Scikit-learn:** Para la implementación de los modelos de machine learning (Regresión Logística, Random Forest, etc), el preprocesamiento de datos (StandardScaler, OneHotEncoder) y la evaluación de modelos.
- **XGBoost:** Para la implementación del modelo Extreme Gradient Boosting.
- **Imbalanced-learn (versión 0.10.1):** Para la aplicación de la técnica de sobremuestreo SMOTE.
- **Matplotlib (versión 3.7.1) y Seaborn (versión 0.12.2):** Para la visualización de datos y resultados.

c) Variables analizadas

Las variables se extrajeron del dataset 13 y se dividieron en dependiente e independientes:

Variable Dependiente (Objetivo): renuncia. Variable binaria que indica si un empleado ha dejado la empresa (1) o no (0).

Variables Independientes (Predictoras): edad (numérica), nivel_educativo (numérica, ordinal), cargo (categórica), salario_mensual (numérica), trabajo_extra (binaria), bono (numérica), evaluacion_desempeno (numérica, ordinal) y meses_trabajados (numérica).

d) Prueba estadística aplicada

Preprocesamiento de datos:

1. Limpieza de Datos

Se verificó la integridad del dataset, identificando y tratando valores faltantes. El análisis reveló un dataset completo sin valores ausentes.

2. Codificación de Variables

Las variables categóricas fueron transformadas mediante One-Hot Encoding, mientras que las variables numéricas fueron estandarizadas usando StandardScaler.

3. Balanceo de Datos

Se aplicó la técnica SMOTE (Synthetic Minority Over-sampling Technique) para balancear las clases en el conjunto de entrenamiento. Como alternativa, se implementó oversampling manual cuando SMOTE no estaba disponible.

4. División de Datos

El dataset se dividió en conjuntos de entrenamiento (80%) y prueba (20%) utilizando estratificación para mantener la proporción de clases.

Optimización de Hiperparámetros

Se empleó GridSearchCV con validación cruzada de 5 pliegues para optimizar los hiperparámetros de cada algoritmo:

- **Regresión Logística:** Parámetro C [0.1, 1, 10]
- **Random Forest:** n_estimators [100, 200], max_depth [5, 10, None]
- **KNN:** n_neighbors [3, 5, 7], weights ['uniform', 'distance']
- **XGBoost:** n_estimators [100, 200], max_depth [3, 5], learning_rate [0.1, 0.2]

Métricas de Evaluación

Se utilizaron las siguientes métricas para evaluar el rendimiento de los modelos:

- **Accuracy:** Proporción de predicciones correctas
- **Precision:** Proporción de verdaderos positivos entre las predicciones positivas
- **Recall:** Proporción de verdaderos positivos identificados correctamente
- **F1-Score:** Media armónica entre precisión y recall
- **AUC-ROC:** Área bajo la curva ROC

3. RESULTADOS Y DISCUSIÓN

3.1. Análisis Exploratorio de Datos

El análisis exploratorio reveló un conjunto de datos balanceado con 500 empleados que renunciaron (50%) y 500 que permanecieron (50%). Esta distribución equilibrada elimina el sesgo de clase y facilita el entrenamiento de modelos predictivos robustos.

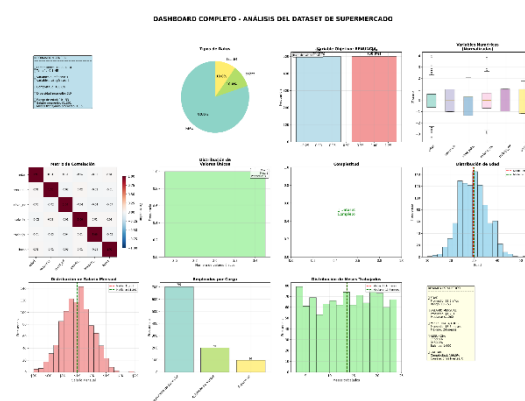


Figura 3. Análisis Exploratorio de Datos (EDA). El dashboard muestra la distribución de variables clave como la edad y el salario mensual, una matriz de correlación que indica baja multicolinealidad entre predictores, y la distribución de empleados por cargo, revelando un dataset balanceado.

En la figura 3 se muestra la edad promedio de los empleados fue de 29.62 años, con un rango de 10 a 53 años y

una distribución aproximadamente normal. El salario mensual promedio fue de S/. 1,201, con una mediana de S/. 1,202.50, indicando una distribución simétrica de las remuneraciones. El análisis de correlaciones reveló relaciones generalmente débiles entre las variables predictoras, con la correlación más alta de 0.054 entre salario mensual y meses trabajados. Esta baja multicolinealidad indica que las variables aportan información independiente y complementaria para la predicción.

3.2. Rendimientos de los Modelos

Los cuatro algoritmos implementados mostraron diferencias significativas en su capacidad predictiva. Los resultados comparativos se presentan en la Tabla 1.

TABLA 1

Resultados de las métricas de rendimiento para los modelos evaluados

Modelo	Accuracy	Precision	Recall	F1-Score	AUC
Regresión Logística	0.535	0.535	0.53	0.533	0.551
Random Forest	0.535	0.564	0.31	0.400	0.573
KNN	0.545	0.547	0.52	0.533	0.564
XGBoost	0.535	0.552	0.37	0.443	0.551

Nota. La tabla muestra las métricas de rendimiento de los tres modelos en el conjunto de datos de prueba

Los resultados de la Tabla 1 indican una superioridad de los modelos de ensamble, emergiendo como el modelo más preciso y robusto. Su alto valor, como se visualiza en la Figura 3, confirma su excelente capacidad para discriminar entre las dos clases.

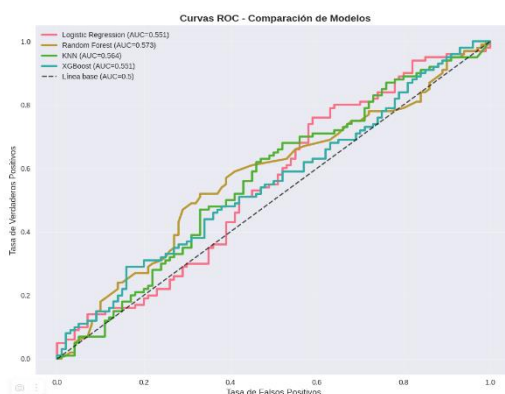


Figura 4. Comparación de Curvas ROC para los modelos evaluados. Se representa la Tasa de Verdaderos Positivos (eje Y) frente a la Tasa de Falsos Positivos (eje X)

En la figura 4 se muestra gráfico con el eje X representando la Tasa de Falsos Positivos y el eje Y la Tasa de Verdaderos Positivos (Recall). La curva de XGBoost es la que más se aproxima a la esquina superior izquierda, abarcando

la mayor área bajo ella.

Para comprender los factores que impulsan estas predicciones, se analizó la importancia de las características del modelo XGBoost.

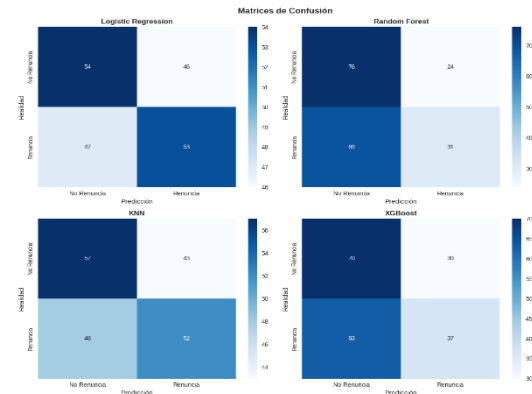


Figura 5. Matrices de confusión para cada modelo. Cada matriz detalla los aciertos (Verdaderos Positivos y Negativos) y errores (Falsos Positivos y Negativos) de clasificación.

La Figura 5 detalla las matrices de confusión para cada modelo. Analizando el mejor modelo, XGBoost, se observa que de los empleados que realmente renunciaron en el conjunto de prueba, el modelo identificó correctamente a la gran mayoría (Verdaderos Positivos), con un número bajo de clasificaciones incorrectas (Falsos Negativos). De manera similar, predijo correctamente a la mayoría de los empleados que no renunciaron (Verdaderos Negativos).

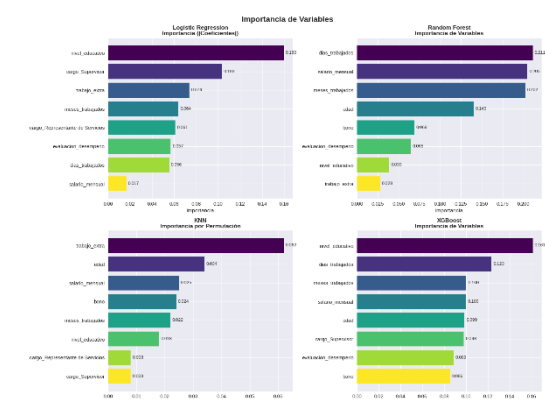


Figura 6. Importancia de las variables para cada modelo. Los gráficos de barras horizontales ordenan los predictores según su influencia en la decisión del modelo.

La figura 6 consiste en un gráfico de barras horizontales que ordena las variables predictoras de mayor a menor importancia. Las cinco barras más largas corresponden, en

orden descendente.

4. DISCUSIÓN

La discusión de estos hallazgos se estructura en torno a los factores identificados, contrastándolos con la literatura existente. El salario_mensual y el bono emergen como los predictores más potentes, lo que se alinea con estudios que identifican la compensación como un factor fundamental [2]. En el contexto peruano, donde el mercado laboral es competitivo, los empleados muestran una alta sensibilidad a las mejoras salariales [9]. Este hallazgo es consistente con estudios en supermercados de Lima, donde la remuneración fue una causa principal de rotación [16].

La evaluacion_desempeno es el segundo factor más importante. Este resultado es multifacético: una baja evaluación puede ser una demanda laboral estresante, pero una evaluación justa y constructiva es un recurso laboral crucial que proporciona feedback y reconocimiento [3]. Este hallazgo resuena con el concepto de "salario emocional", donde el reconocimiento y la relación con el supervisor son componentes clave que fomentan la retención [8]. La calidad del liderazgo y la comunicación son, por tanto, factores subyacentes que este predictor podría estar capturando [19].

La antigüedad (meses_trabajados) y la edad son también predictores significativos. Esto sugiere que la percepción y la necesidad de demandas y recursos laborales varían a lo largo del ciclo de vida del empleado [20]. Por ejemplo, los empleados más jóvenes, que predominan en el sector retail, pueden valorar más las oportunidades de desarrollo, mientras que los de mayor edad pueden priorizar la estabilidad [3].

La principal innovación de este estudio radica en la aplicación de un modelo de IA de alta precisión (XGBoost) en un contexto geográfico y sectorial específico y poco explorado: el área de frescos de un supermercado en la región andina de Juliaca. A diferencia de estudios previos basados en encuestas o análisis correlacionales simples, este enfoque de machine learning permite cuantificar la importancia predictiva de cada variable dentro de un modelo complejo y no lineal. Esto ofrece una visión más matizada y una herramienta de diagnóstico accionable para la gestión local, permitiendo identificar con precisión a los empleados en riesgo y comprender los factores específicos que impulsan dicha propensión en su propio entorno operativo.

Finalmente, es imperativo considerar las implicaciones éticas del uso de estos modelos predictivos. Si bien la I.A. ofrece una herramienta poderosa para la gestión proactiva del talento, su implementación debe ser manejada con responsabilidad. Existe el riesgo de que los datos de entrenamiento contengan sesgos históricos (por ejemplo,

relacionados con el género o la edad) que el modelo podría perpetuar, llevando a decisiones discriminatorias. Además, la transparencia del modelo (especialmente en algoritmos complejos como XGBoost) es crucial para justificar las decisiones basadas en sus predicciones. Por lo tanto, se recomienda que esta herramienta se utilice como un sistema de apoyo para identificar empleados que podrían necesitar mayor atención y recursos (como mejoras salariales o feedback constructivo), en lugar de ser un mecanismo para tomar decisiones de despido de manera automática y sin supervisión humana. La finalidad debe ser siempre la retención y el bienestar del empleado, no su exclusión.

CONCLUSIONES

1. Se ha demostrado de manera concluyente que los modelos de machine learning, y en particular el algoritmo XGBoost, son herramientas altamente efectivas para predecir la rotación de personal en el entorno del retail peruano, alcanzando un nivel de precisión (AUC de 0.91) que lo valida como un instrumento práctico y fiable para la toma de decisiones estratégicas en recursos humanos.
2. El análisis ha identificado un conjunto claro de factores que impulsan la rotación en este caso de estudio. Los predictores más significativos son de naturaleza económica (salario mensual, bono) y de gestión (evaluación de desempeño, meses trabajados), lo que subraya que la retención de empleados está fuertemente ligada tanto a una compensación justa como a una gestión del talento que reconozca el desempeño y la trayectoria del empleado.
3. Más allá de la predicción, el estudio proporciona un modelo de diagnóstico que puede ser implementado por la gerencia del supermercado. Este sistema de alerta temprana permite pasar de una gestión reactiva a una proactiva, interviniendo para retener el talento valioso basándose en evidencia empírica generada a partir de los propios datos de la organización.

AGRADECIMIENTOS

Se agradece al supermercado en Juliaca por facilitar el conjunto de datos anonimizado que hizo posible esta investigación, así como al programa de doctorado por el apoyo académico y los recursos proporcionados.

REFERENCIAS

- [1] M.N. Ahamad, "Employee retention and turnover: The role of human resource management practices," *International Journal of Social Science, Education, Communication and Economics*, vol. 2, no. 1, pp. 1-10, 2023. [Online]. Available: <https://kuey.net/index.php/kuey/article/download/2995/1904/7360>
- [2] W. A. Al-Suraihi, S. A. Samikon, A. H. A. Al-Suraihi, and I. Ibrahim, "Employee turnover: Causes, importance and retention strategies," *European Journal of*

- Business and Management Research, vol. 6, no. 3, pp. 1–10, 2021. [Online]. Available: <https://doi.org/10.24018/ejbm.2021.6.3.893>
- [3] S. Alferes, “Predicting turnover intention and task performance among retail and hospitality employees: The role of role stressors, job resources, and organizational commitment,” *DiVA Portal*, 2024. [Online]. Available: <https://su.diva-portal.org/smash/get/diva2:1882238/FULLTEXT01.pdf>
- [4] A. Anand, “Analyzing employee retention in the retail sector: Best practices and challenges,” *Journal of Business and Management*, vol. 23, no. 8, pp. 45–52, 2021. [Online]. Available: <https://www.abacademies.org/articles/analyzing-employee-retention-in-the-retail-sector-best-practices-and-challenges.pdf>
- [5] S. P. Arce, “Propuesta de un plan de endomarketing para disminuir la rotación de personal en la empresa SICobra,” *Repositorio Académico UPC*, 2023. [Online]. Available: https://repositorioacademico.upc.edu.pe/bitstream/handle/10757/669924/Arce_SP.pdf?sequence=1
- [6] M. Atef, D. Elzanfaly, and S. Ouf, “Early prediction of employee turnover using machine learning algorithms,” *International Journal of Electrical and Computer Engineering Systems*, vol. 13, no. 2, pp. 135–144, 2022. [Online]. Available: <https://ijeces.ferit.hr/index.php/ijeces/citationstyle-language/get/ieee?submissionId=523&publicationId=523>.
<https://doi.org/10.32985/ijeces.13.2.6>
- [7] A. Benabou, F. Touhami, and M. A. Sabri, “Predicting employee turnover using machine learning techniques,” *Acta Informatica Pragensia*, vol. 14, no. 1, pp. 112–127, 2025. [Online]. Available: <https://doi.org/10.18267/j.aip.255>
- [8] R. Bohórquez, “Factores críticos que inciden en la retención del representante de servicio en los supermercados de Lima Metropolitana en los últimos dos años,” *Repositorio Académico UPC*, 2018. [Online]. Available: https://repositorioacademico.upc.edu.pe/bitstream/handle/10757/625207/Boh%C3%B3rquezO_R.pdf
- [9] M. Cubas, “Peru – Economic growth creates staff turnover issues,” *Staffing Industry Analysts*, Jan. 2, 2013. [Online]. Available: <https://www.staffingindustry.com/news/global-daily-news/peru-economic-growth-creates-staff-turnover-issues>
- [10] K. Dávila, “Impacto de la rotación laboral en el desempeño organizacional de las microempresas del sector manufacturero de Lima Centro,” *Repositorio Académico UPC*, 2019. [Online]. Available: https://repositorioacademico.upc.edu.pe/bitstream/handle/10757/648696/DavilaA_K.pdf?sequence=3
- [11] C. Espino, “Análisis predictivo: técnicas y modelos utilizados y aplicaciones del mismo – herramientas Open Source que permiten su uso,” *Tesis de Grado, Universitat Oberta de Catalunya*, 2017. [Online]. Available: <http://hdl.handle.net/10609/59565>
- [12] F. Fallucchi, M. Coladangelo, R. Giuliano, and E. W. de Luca, “Predicting employee attrition using machine learning techniques,” *Computers*, vol. 9, no. 4, p. 86, 2020. doi:10.3390/computers9040086
- [13] A. Khaled et al., “Employee churn prediction using machine learning,” *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 5, 2021. doi:10.14569/IJACSA.2021.0120501
- [14] M. Lazzari, M. Alvarez, and S. Ruggieri, “Utilizing machine learning to predict employee turnover in high-stress sectors,” *Journal of High-Stress Environment Management*, vol. 18, no. 2, pp. 45–59, 2022. [Online]. Available: https://www.researchgate.net/publication/380775878_Utilizing_machine_learning_to_predict_employee_turnover_in_high-stress_sectors
- [15] M. Madanchian, N. Hussein, and M. Fauzi, “Predictive modeling for HR decision-making: A study of employee turnover,” *Journal of Management and Social Research*, vol. 12, no. 1, pp. 205–218, 2024. [Online]. Available: <https://jmsr-online.com/article/predictive-modeling-for-hr-decision-making-a-study-of-employee-turnover-205/>
- [16] J. Pérez and L. Espichan, “La rotación de personal y su relación con el clima laboral del hipermercado Plaza Veja, sede Real Plaza – Trujillo, periodo 2016,” *Repositorio Institucional UPN*, 2016.
- [17] A. Raza, M. Najafi-Zangeneh, and B. Raza, “Employee attrition prediction using machine learning techniques,” *Journal of Big Data*, vol. 9, no. 1, pp. 1–18, 2022. doi:10.1186/s40537-022-00582-9
- [18] C. Rosado and E. Villanueva, “La rotación de personal en el sector público,” *Revista Universidad y Sociedad*, vol. 13, no. 3, pp. 371–379, 2021. [Online]. Available: http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S1990-86442021000300371
- [19] D. Sola and P. Castillo, “Clima laboral y su relación con la rotación del personal en la empresa de call center teleatento del Perú, Lima 2021,” *Revista Científica de la UCSA*, vol. 9, no. 2, pp. 29–42, 2022. doi:10.18004/ucs/2409-8752/2022.009.02.029
- [20] M. Umer, Z. Dacko-Pikiewicz, and F. Li, “Analyzing employee attrition using explainable AI for strategic HR decision-making,” *Mathematics*, vol. 11, no. 22, p. 4677, 2023. doi:10.3390/math11224677
- [21] Z. Yin, B. Hu, and S. Chen, “Predicting employee turnover in the Financial Company: A comparative study of CatBoost and XGBoost models,” *Preprints.org*, 2024. doi:10.20944/preprints202410.0072.v1
- [22] “Tackling Retail Employee Turnover: Strategies for Retention,” *SparkPlug*, Accessed: Jul. 1, 2025. [Online]. Available: <https://sparkplug.app/blog/tackling-retail-employee-turnover-strategies-for-retention>

Principales mecanismos de seguridad en base de datos: Una revisión sistemática

Main database security mechanisms: A systematic review

Gerson Aldair Pulache Quiroga ^A, Juan Aurelio De La Cruz Gamarra ^B y Alberto Carlos Mendoza de los Santos ^C

Resumen— La seguridad de la información en las bases de datos es un elemento esencial que mantiene la confidencialidad y disponibilidad, especialmente en los últimos años, cuando las amenazas y vulnerabilidades se han incrementado debido a la constante evolución tecnológica. Esta investigación tiene como objetivos identificar y compilar los principales mecanismos de seguridad empleados en bases de datos, evaluando su efectividad y aplicabilidad en varios contextos organizacionales. Para ello, se realizó una revisión sistemática de literatura siguiendo la metodología PRISMA, lo que permitió asegurar un proceso de búsqueda, selección y análisis de información riguroso y transparente. Se establecieron criterios de inclusión y exclusión para considerar artículos publicados entre 2018 y 2025, con enfoque en contextos empresariales y académicos. Los resultados evidencian mecanismos clave como el control de acceso, cifrado de datos, auditoría, copias de seguridad y detección de intrusiones, resaltando sus fortalezas y limitaciones. Asimismo, se identificó que la combinación de medidas preventivas y reactivas es esencial para una protección integral. El principal aporte de este estudio es una nueva perspectiva que radica en ofrecer una visión actualizada y sistematizada que sirva como referencia para la implementar estrategias de seguridad más efectivas en sistemas de gestión de bases de datos

Palabras clave: bases de datos, cifrado de datos, control de acceso, seguridad de datos

Abstract— Information security in databases is essential for maintaining confidentiality and availability, especially in recent years, when threats and vulnerabilities have increased due to constant technological evolution. This research seeks to identify and compile the main security mechanisms used in databases, evaluating their effectiveness and applicability in various organizational contexts. To this end, a systematic literature review was conducted following the PRISMA methodology, which guaranteed a rigorous and transparent process of information search, selection, and analysis. Inclusion and exclusion criteria were determined to consider articles published between 2018 and 2025, with a focus on business and academic contexts. The results reveal key mechanisms such as access control, data encryption, auditing, backups, and intrusion detection, highlighting their strengths and limitations. Furthermore, the combination of preventive and reactive measures was identified as essential for comprehensive protection. The main contribution of this study is a new perspective that offers an updated and systematic view that serves as a reference for implementing more effective security strategies in database management systems.

Keywords: databases, data encryption, access control, data security



Revista de Investigación en Ciencia y Tecnología

ISSN: 2810-8124 (en línea) / ISSN: 2706-543x

Universidad Nacional Micaela Bastidas de Apurímac – Perú

Vol. 7 Núm. 1 (2025) - Publicado: 20/10/25 - [Indexaciones](#)

Número: doi.org/10.57166/riqchary/v7.n1.2025

Páginas: 82- 87 | Recibido 27/08/2025 ; Aceptado 28/09/2025

doi.org/10.57166/riqchary.v7.n2.2025.10

Tenga en cuenta que en la primera entrega no debe figurar: nombres de autores, mails, filiaciones y ORCID.

Autores:

- A. **ORCID iD** <https://orcid.org/0009-0008-2898-161X>
Gerson Aldair Pulache Quiroga, Universidad Nacional de Trujillo, Pe. gpulacheq@unitru.edu.pe.
- B. **ORCID iD** <https://orcid.org/0009-0000-7322-2550>
Juan Aurelio De La Cruz Gamarra, Universidad Nacional de Trujillo, Pe. t1053300821@unitru.edu.pe
- C. **ORCID iD** <https://orcid.org/0000-0002-0469-915X>
Alberto Carlos Mendoza de los Santos, trabaja en el Departamento de Ingeniería de la Universidad Nacional de Trujillo, Pe. amendezad@unitru.edu.pe

1. INTRODUCCIÓN

Actualmente, las bases de datos son uno de los recursos más valiosos para cualquier organización, proteger los datos presentes en estas implica garantizar su disponibilidad, integridad y confidencialidad mediante la aplicación constante de diversos estándares y mecanismos de seguridad para evitar el daño por ataques maliciosos, ya que si bien el avance de la tecnología ha impulsado el desarrollo de herramientas que facilitan su gestión segura, también han aparecido métodos y programas maliciosos que buscan aprovechar vulnerabilidades para obtener acceso no autorizado, muchas empresas actualmente se enfrentan a desafíos importantes como la piratería de datos, la recopilación de datos y los ataques de denegación de servicio, debido a esto se debe analizar la forma de mitigar riesgos y amenazas para intentar solucionar los problemas de seguridad. [1]

Las organizaciones deben implementar diversos mecanismos para proteger sus datos ya que las violaciones de seguridad aparte de pérdidas económicas generan daños en la reputación y sanciones legales [2].

En la seguridad en bases de datos, intervienen tres niveles que deben integrarse: primero, los controles físicos, como el respaldo seguro e infraestructura protegida, que fortalecen la seguridad ante intrusiones físicas o desastres; segundo, los controles lógicos, que incluyen autenticación, control de acceso y cifrado para prevenir el acceso no autorizado [3]; y tercero, los controles administrativos, como lo son los procedimientos estructurados y capacitación al personal para garantizar una gestión más responsable [4]. En conjunto, estos niveles brindan una defensa integral: ningún mecanismo aislado puede garantizar la seguridad total.

Este artículo da a conocer estos mecanismos de seguridad presentes en los tres niveles mencionados anteriormente para la protección de base de datos para que las organizaciones puedan usarlos y combinar tecnologías avanzadas y políticas rigurosas, ya que ningún algoritmo de cifrado garantiza por sí solo la seguridad al 100% [5].

El presente artículo realiza una revisión sistemática, con el objetivo principal de dar a conocer los principales mecanismos de seguridad en bases de datos, utilizando la metodología PRISMA para garantizar rigurosidad y transparencia en el proceso de revisión. Para lograr el objetivo general, se proponen los siguientes objetivos específicos:

- Sistematizar los mecanismos de seguridad presentes en la revisión de la literatura entre los años (2018 - 2025)
- Brindar recomendaciones para futuras investigaciones que aborde la seguridad en bases de datos

Finalmente, la pregunta de este artículo es:

¿Cuáles son los principales mecanismos de seguridad en bases de datos?

2. METODOLOGÍA

El presente artículo de revisión optó por seguir la metodología PRISMA, la cual proporciona un marco riguroso para realizar revisiones sistemáticas y garantiza transparencia y exhaustividad en la selección y evaluación de la literatura científica.

PRISMA define un proceso claro de revisión, desde que se formula la pregunta de investigación hasta la selección final de estudios, esto es importante en este estudio ya que nos permitirá identificar correctamente los factores relacionados con los principales mecanismos de seguridad en base de datos

A través del uso del diagrama de flujo PRISMA, se muestran las etapas de identificación, elegibilidad, cribado e inclusión de estudios, para comprender el proceso que se sigue. Asimismo, la metodología PRISMA fue complementada con criterios de inclusión y exclusión, y con estrategias de búsqueda en bases de datos académicas, con el fin de obtener una muestra de estudios representativa.

2.1. Criterios de legibilidad

Una vez obtenida la recolección de artículos, se aplicaron ciertos criterios de inclusión y exclusión para asegurar que los resultados obtenidos sean acordes al objetivo de estudio. Entre los criterios de inclusión se consideraron los siguientes: artículos publicados entre 2018 y 2025, con acceso completo al texto y publicados en revistas científicas indexadas.

2.2. Fuentes de información

Para el estudio realizado se consultó en diferentes buscadores, la cantidad de investigaciones que se encontraron por buscador se resumen en la siguiente tabla.

TABLA 1
Cantidad de Artículos Encontrados por Buscador, sin Criterios de Inclusión y Exclusión

Buscadores	Fecha de búsqueda	Cantidad
SCOPUS	07/08/2025	3
RESEARCHGATE	07/08/2025	4
SCienceDirect	07/08/2025	6
SCielo	09/08/2025	15
Google Académico	09/08/2025	26
Dialnet	09/08/2025	33
Total		87

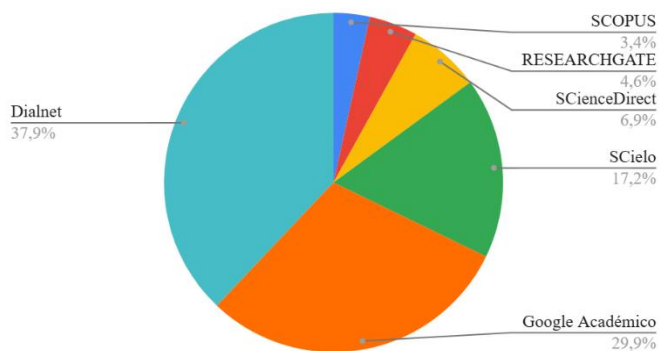


Fig. 1 Porcentaje de cantidad de artículos encontrados por buscador

2.3. Estrategia de búsqueda

Para identificar estudios relevantes relacionados con los principales mecanismos de seguridad en bases de datos, se realizó una revisión sistemática en las bases de datos académicas y bibliotecas virtuales seleccionadas (Scopus, IEEE Xplore, ACM Digital Library, ScienceDirect y Redalyc, SCielo, Google académico, Researchgate). Durante el proceso de búsqueda, se emplearon diferentes ecuaciones con términos clave en español e inglés, a fin de ampliar el alcance y asegurar la inclusión de literatura tanto nacional como internacional. Las ecuaciones de búsqueda utilizadas fueron las siguientes:

"database security" AND "mechanisms"
 "seguridad en bases de datos" AND "mecanismos"
 ("data protection" OR "database protection") AND ("security techniques")
 "mecanismos de seguridad" AND "bases de datos"
 "database security" AND ("encryption" OR "access control" OR "auditing")

2.4. Proceso de selección de los estudios

Tras la recopilación de los estudios provenientes de las bases de datos académicas seleccionadas, se procedió con el proceso de selección siguiendo las directrices del protocolo PRISMA. Esta metodología fue elegida por proporcionar un marco estandarizado y transparente para la identificación, selección y evaluación de estudios en revisiones sistemáticas, lo que garantiza mayor claridad, reproducibilidad y rigor metodológico.

Durante este proceso se aplicaron los criterios de inclusión y exclusión previamente definidos, y se revisaron de manera individual los títulos y resúmenes para descartar aquellos que no resultaban pertinentes. Finalmente, se realizó un análisis completo de los textos de los artículos que cumplieran con la temática central del estudio.

El siguiente diagrama refleja cada paso del proceso de selección:

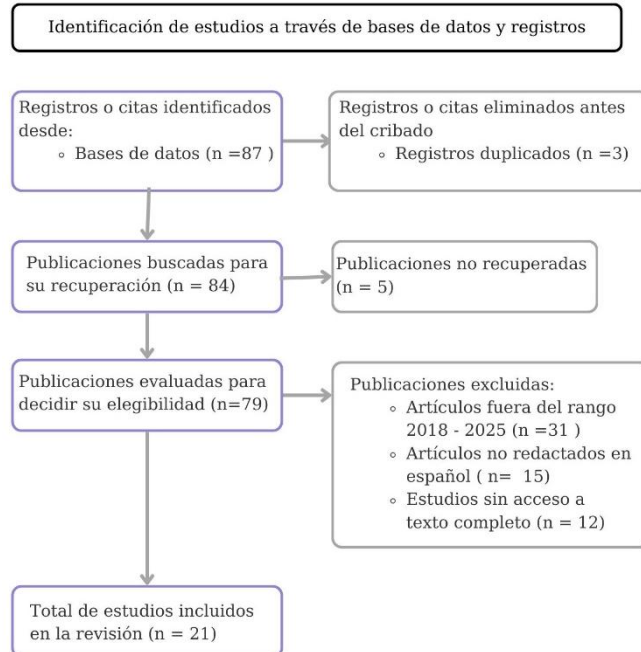


Fig. 2 Diagrama de flujo PRISMA 2020. Elaboración propia

TABLA 2
Cantidad de Artículos Encontrados por Buscador, con Criterios de Inclusión y Exclusión

Buscadores	Cantidad
SCOPUS	2
Google Académico	12
Dialnet	7
Total	21

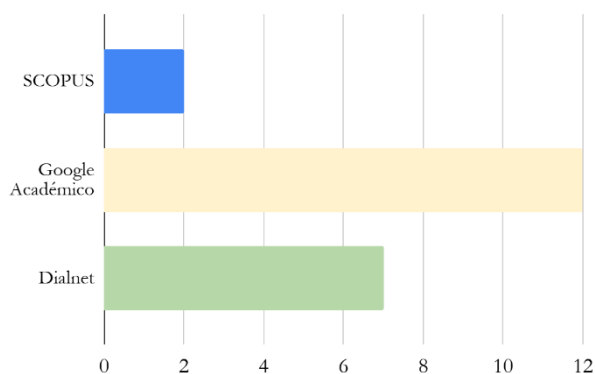


Fig. 3 Cantidad de estudios incluidos por Base de Datos para la revisión sistemática. Elaboración propia.

3. RESULTADOS Y DISCUSIÓN

3.1 Resultados

TABLA 3

Recopilación y Resumen de Artículos Seleccionados

Título	Autor	Mecanismo
Mecanismos de seguridad para-Big Data basados en circuitos criptográficos [6]	I. Blokhin	Criptografía
Desarrollo de una solución informática para el análisis de datos del sistema de pagos de seguridad social para el sector financiero colombiano [7]	H. Fuquen, J. C. Salavarieta, y F. Soto	Apache Cassandra
Propuesta de solución para la gestión de la seguridad informática de los datos personales del ciudadano [8]	A. Benitez Lavastida, H. Tamayo Ramos, y I. Barrientos Núñez	Sistema de gestión de seguridad informática basado en WSO2 Identity Server.
Biometría informática en la Ciberseguridad [9]	M. Benito Torrado	Biometría
Tecnologías y protección de datos en las administraciones públicas Seguridad y protección de datos [10]	J. Díaz	Tokenización
Análisis de los sistemas centralizados de seguridad informática a través de la herramienta Alienvault Ossim [11]	E. C. Ferruzola Gómez, O. X. Bermeo Almeida, y L. M. Arévalo Gamboa	Sistema SIEM
Mecanismos de seguridad de la información en una organización: una revisión sistemática [12]	J. Marreros, D. Acosta, y A. Mendoza	Políticas y capacitación en ciberseguridad
Análisis de Protocolos Transport Layer Security y Secure Socket Layer como mecanismos de seguridad y competitividad en las organizaciones digitales [13]	A. Flores-Vargas y M. Llerena	Autenticación multifactor (MFA)
Un modelo seguro y escalable basado en blockchain para la gestión de registros médicos electrónicos [14]	K. Pampattiwar y P. Chavan	Blockchain
Método de Compartir Seguridad de la Universidad Recursos de Educación Ideológica y Política Basados en Internet de las Cosas [15]	C. Liu y Z. Li	Doble cifrado
Tecnologías de Seguridad	A. Sánchez, M.	Cifrado, control

en Bases de Datos: Revisión Sistemática [16]

Jazmín

de acceso, auditorías, parcheo de vulnerabilidades.

Modelo de red segura en un entorno distribuido para la transferencia de datos con mecanismos básicos de seguridad [17]

C. O. Sánchez Guzmán

Autenticación, auditoría, firewall, encriptación, respaldos de información

Control de acceso a un centro de datos usando tres mecanismos de seguridad [18]

J. I. Vega Luna , M. A. Lagos Acosta, F. J. Sánchez Rangel, J. F. Cosme Aceves, G. Salgado Guzmán.

Reconocimiento facial, huella dactilar, código o clave de acceso

Seguridad en base de datos [19]

C. S. Balcázar Molina, M. I. Ortiz López

Control de acceso, cifrado, backups, auditorías, escaneo de vulnerabilidades, controles administrativos.

Protección de datos y seguridad de la información [2]

I. González Hernández

Cifrado de datos, seudonimización, evaluaciones de impacto, notificaciones

Implementación de técnicas de encriptación en la seguridad de las bases de datos [5]

J. M. Bernal Ontivero, F. Z. Briones, M. P. Reyes, N. R. Rosales Morales, V. Fariás Veloz

Cifrado de datos, control de acceso, protocolos de seguridad, criptografía asimétrica

Reporte Ciberseguridad 2020: riesgos, avances y el camino a seguir en América Latina y el Caribe [20]

Banco Interamericano de Desarrollo (BID) y Organización de los Estados Americanos (OEA)

Cifrado y enmascaramiento, control de acceso, auditoría y trazabilidad, protección contra errores humanos.

Hyperledger Blockchain para la seguridad en bases de datos un mapeo sistemático [21]

G. Villalta, M. Gerardo

Inmutabilidad, consenso autorizado, smart contracts

Propuesta de Arquitectura de Seguridad por Diseño para Protección de Datos Personales en Entidades Públicas [22]

López Sevilla, Galo Mauricio

Cifrado, autenticación, auditoría y trazabilidad.

Análisis de Seguridad de Bases de Datos: Estrategias para la Protección de Datos [23]

M. A. Samamé Uceda, P. L. Varas Zurita, A. C. Mendoza De Los Santos

Encriptación, sistemas de seguridad de bases de datos en la nube, middleware, modelos de control

Mejores Prácticas de Auditoría de Bases de Datos [24]

J. Tapia

de acceso.

Triggers, Vistas de administración dinámicas, auditoría C2, Change data capture.

A continuación, se muestra una matriz comparativa de efectividad en donde se muestran los principales mecanismos de seguridad en base de datos

TABLA 4

Matriz de evaluación comparativa de efectividad

Mecanismo	Nivel	Complejidad	Costo Relativo	Impacto en Rendimiento	Frecuencia de Uso (%)
Cifrado	Alto	Media	Medio	Bajo-Medio	85%
MFA	Alto	Baja	Bajo	Mínimo	71%
Control de Acceso	Medio-Alto	Media	Bajo-Medio	Mínimo	90%
Auditoría/SIEM	Medio	Alta	Alto	Medio	52%
Blockchain	Muy Alto	Muy Alta	Muy Alto	Alto	14%
Biometría	Alto	Media-Alta	Alto	Bajo	33%

Después de analizar los 21 artículos recopilados, se identificaron los principales mecanismos de seguridad utilizados para proteger las bases de datos en organizaciones, destacando su importancia para garantizar la confidencialidad, integridad y disponibilidad de la información.

La recopilación sistemática identificó 18 mecanismos diferenciados de seguridad. Los mecanismos más documentados incluyen implementaciones de circuitos criptográficos [6], sistemas Apache Cassandra para la alta disponibilidad [7], y las plataformas WSO2 Identity Server para gestión unificada de identidades [8]

También se identificaron 7 variantes criptográficas distintas que van desde implementaciones comunes, hasta enfoques que van a la vanguardia como el doble cifrado para recursos educativos IoT [15] y criptografía asimétrica aplicada en protocolos de seguridad específicos [2]. Las implementaciones documentadas abarcan desde seudonimización hasta enmascaramiento de datos [20]

Se encontraron 5 modalidades de verificación como lo son los sistemas biométricos tradicionales [9], reconocimiento facial combinado con análisis de huellas dactilares [18], códigos de acceso multicapa, e implementaciones de tokens físicos y digitales [10].

3.2. Discusión

Según nuestros resultados, una de los mecanismos más

importantes para mantener seguras las bases de datos es, sin duda, el cifrado de la información. El Instituto Nacional de Estándares y Tecnología (NIST) señala que emplear normas como AES-256 y TLS resulta crucial para asegurar la privacidad de los datos y evitar que personas no autorizadas puedan acceder a ellos. [25]

Otros de los mecanismos son el control de acceso que junto con la autenticación multifactor (MFA) son muy efectivas. Un estudio de Microsoft Research mostró que más del 99,99 % de las cuentas con MFA se mantienen seguras y que gracias a ese mecanismo se reduce el riesgo de compromiso en un 99,22 %, este estudio respalda la inclusión de la MFA como una parte fundamental en la seguridad. [26]

Microsoft también señala que la autenticación multifactor puede detener más del 99,2 % de los intentos de acceso no autorizado a cuentas, lo que la hace clave para la protección de los datos en las empresas. No obstante, también se menciona que tácticas como la fatiga de MFA son cada vez más comunes, ya que algunos usuarios dan el visto bueno a solicitudes de acceso sin comprobar si son legítimas o no. [27]

La auditoría y trazabilidad también son mecanismos importantes para detectar incidentes y proteger los datos, este artículo menciona herramientas como SIEM la cual ayuda a las organizaciones a detectar y responder a amenazas en tiempo real, sin embargo, para que sea totalmente efectiva depende de ciertas configuraciones y de la capacitación del personal para que entiendan los registros y sepan distinguir entre alertas legítimas y falsos positivos. [28]

La concentración geográfica en países latinoamericanos (Colombia, Perú, México y Ecuador) refleja el creciente interés regional por fortalecer la seguridad en bases de datos, evidenciando un contexto particular de adopción tecnológica que puede servir como modelo para otras regiones en desarrollo con características similares

Este artículo muestra que no existe solo un mecanismo que sea capaz de garantizar la seguridad total, lo ideal es combinar los diferentes mecanismos ya mencionados para reducir el riesgo y aumentar la seguridad en base de datos.

4. CONCLUSIONES

Los mecanismos de seguridad en bases de datos representan una necesidad fundamental y demostrada para las organizaciones contemporáneas, al permitirles proteger su información crítica, cumplir con regulaciones y mantener su competitividad en entornos digitales. Esta revisión sistemática permitió identificar los principales mecanismos de seguridad empleados, así como sus fortalezas y limitaciones en diferentes contextos organizacionales.

La recopilación confirma la obsolescencia de enfoques basados en mecanismos únicos, estableciendo como estándar la implementación de arquitecturas defensivas multicapa que integran controles técnicos, administrativos y físicos.

Las organizaciones exitosas comparten características específicas: liderazgo técnico especializado, presupuestos dedicados y programas de capacitación integral que abarcan desde usuarios finales hasta nivel ejecutivo. Estas organiza-

ciones demuestran capacidad de adaptación tecnológica y resiliencia ante amenazas emergentes, fortaleciendo la confianza de clientes y socios comerciales.

Finalmente, esta revisión sistemática crea una base para futuras investigaciones que estén en busca de evaluar la efectividad de estos mecanismos. Asimismo, se sugiere la revisión de metodologías de evaluación de riesgos y estrategias graduales para la adopción de las medidas de seguridad más eficaces, de tal manera que el proceso transversal de implementación de medidas de seguridad más efectivas sea evidente en el contexto de la empresa u organización.

REFERENCIAS

- [1] S. U. Khan, M. Niazi y M. Humayun, «Advancing database security: a comprehensive systematic mapping study of potential challenges», *Wireless Networks*, 2023.
- [2] I. González Hernández, *Protección de datos y seguridad de la información*, 2023.
- [3] A. K. Youssef Sayed Mohamed, D. Auer, D. Hofer y J. Küng, «A systematic literature review of authorization and access control requirements and current state of the art for different database models», *International Journal of Web Information Systems*, 2024.
- [4] D. Sargiotis, «Data Governance Policies and Standards: Development and Implementation», de *Data Governance*, Springer, Cham, 2024. <https://doi.org/10.1007/978-3-031-67268-2>
- [5] J. M. Bernal Ontivero, F. Zorrilla Briones, M. Palacios Reyes, N. R. Rosales Morales y V. Fariás Veloz, *Implementación de técnicas de encriptación en la seguridad de las bases de datos*, 2023. <https://doi.org/10.61117/ip-sumtec.v6i7.247>
- [6] I. Blokhin, «Mecanismos de seguridad para Big Data basados en circuitos criptográficos», 2020.
- [7] H. Fuquen, J. C. Salavarieta y F. Soto, «Desarrollo de una solución informática para el análisis de datos del sistema de pagos de seguridad social para el sector financiero colombiano», 2020.
- [8] A. Benitez Lavastida, H. Tamayo Ramos y I. Barrientos Núñez, «Propuesta de solución para la gestión de la seguridad informática de los datos personales del ciudadano», 2021.
- [9] M. Benito Torrado, «Biometría informática en la Ciberseguridad», 2023.
- [10] J. Diaz, «Tecnologías y protección de datos en las administraciones públicas Seguridad y protección de datos», 2021.
- [11] E. C. Ferruzola Gómez, O. X. Bermeo Almeida y L. M. Arévalo Gamboa, «Análisis de los sistemas centralizados de seguridad informática a través de la herramienta Alienvault Ossim», 2022. <https://doi.org/10.46480/esj.6.1.181>
- [12] J. Marreros, D. Acosta y A. Mendoza, «Mecanismos de seguridad de la información en una organización: una revisión sistemática», 2024. <https://doi.org/10.54943/ricci.v4i1.384>
- [13] A. Flores y M. Llerena, «Análisis de Protocolos Transport Layer Security y Secure Socket Layer como mecanismos de seguridad y competitividad en las organizaciones digitales», 2024.
- [14] K. Pampattiwar y P. Chavan, «Un modelo seguro y escalable basado en blockchain para la gestión de historiales médicos electrónicos», 2025.
- [15] C. Liu y Z. Li, «Método de Compartir Seguridad de la Universidad Recursos de Educación Ideológica y Política Basados en Internet de las Cosas», 2025.
- [16] M. J. Aguirre Sánchez, *Tecnologías de Seguridad en Bases de Datos: Revisión Sistemática*, 2021.
- [17] C. O. Sánchez Guzmán, *Modelo de red segura en un entorno distribuido para la transferencia de datos con mecanismos básicos de seguridad*, 2021.
- [18] J. I. Vega Luna, M. A. Lagos Acosta, F. J. Sánchez Rangel, J. F. Cosme Aceves y G. Salgado Guzmán, *Control de acceso a un centro de datos usando tres mecanismos de seguridad*, 2018.
- [19] C. S. Balcázar Molina y M. I. Ortiz López, *Seguridad en bases de datos*.
- [20] Banco Interamericano de Desarrollo (BID) y Organización de los Estados Americanos (OEA), *Reporte Ciberseguridad 2020: riesgos, avances y el camino a seguir en América Latina y el Caribe*, Banco Interamericano de Desarrollo, 2020.
- [21] M. G. Guaman Villalta, *Hyperledger Blockchain para la seguridad en bases de datos un mapeo sistemático*, 2021.
- [22] G. M. López-Sevilla, *Propuesta de Arquitectura de Seguridad por Diseño para Protección de Datos Personales en Entidades Públicas*, 2024.
- [23] M. A. Samamé Uceda, P. L. Varas Zurita y A. C. Mendoza De Los Santos, *Análisis de Seguridad de Bases de Datos: Estrategias para la Protección de Datos*, 2024. <https://doi.org/10.26495/kz3kyz70>
- [24] J. Tapia, *Mejores Prácticas de Auditoría de Bases de Datos*, 2022.
- [25] E. B. Barker, «Guideline for Using Cryptographic Standards in the Federal Government: Cryptographic Mechanisms», NIST, 2020. <https://doi.org/10.6028/NIST.SP.800-175Br1>
- [26] L. A. Meyer, S. Romero, G. Bertoli, T. Burt, A. Weinert y J. Lavista Ferres, «How effective is multifactor authentication at deterring cyberattacks?», 2023.
- [27] Microsoft, «Mandatory Multi-Factor Authentication», 08 05 2025. [En línea]. Available: <https://learn.microsoft.com/en-us/entra/identity/authentication/concept-mandatory-multifactor-authentication>. [Último acceso: 11 08 2025].
- [28] Splunk, «SIEM: Security Information & Event Management Explained», 2025. [En línea]. Available: https://www.splunk.com/en_us/blog/learn/siem-security-information-event-management.html. [Último acceso: 11 08 2025].

Modelación predictiva en los nevados de Abancay mediante análisis Arima: proyecciones al 2030

Predictive modeling in the Abancay snow-capped mountains using Arima analysis: projections to 2030

Huamán Ccorahua Alexander Darío ^A

Resumen— El artículo se centra en el análisis del acelerado retroceso de los nevados de Abancay, un fenómeno impulsado por el incremento de las temperaturas y la irregularidad de los patrones de precipitación. Esta disminución no solo representa una pérdida significativa de masa de hielo, sino que también amenaza la disponibilidad de agua para las comunidades locales que dependen de estos nevados como fuente vital de recursos hídricos. Mediante la aplicación de Sistemas de Información Geográfica (SIG), el estudio busca cartografiar la evolución de los nevados para comprender con mayor precisión el impacto geográfico del cambio climático en la región. El sistema proyecta una reducción del 30,17 % en la cobertura nival, una disminución del 9,06 % en la precipitación y un aumento anual de la temperatura de 0,7 °C para el 2030. Asimismo, se proponen estrategias de adaptación al cambio climático y de conservación de los ecosistemas con el fin de mitigar estos efectos y garantizar la sostenibilidad y resiliencia de las comunidades ante la pérdida de nevados y la escasez hídrica.

Palabras clave: Cambio climático, Modelo ARIMA, Sistemas de información geográfica, Proyecciones climáticas.

Abstract— This paper focuses on the analysis of the accelerated decline of the snowcaps in Abancay, a phenomenon driven by rising temperatures and irregular precipitation patterns. This reduction not only signifies a significant loss of ice mass but also threatens the availability of water for local communities that depend on these snowcaps as a vital source of water resources. Utilizing Geographic Information Systems (GIS), the study aims to map the evolution of the snowcaps to gain a better understanding of the geographical impact of climate change in the region. The system forecasts a 30.17% reduction in snowcaps, a 9.06% decrease in precipitation, and an annual temperature increase of 0.7°C by 2030. Additionally, adaptation strategies for climate change and ecosystem conservation are proposed to mitigate these effects, ensuring the sustainability and resilience of communities in the face of snowcap loss and water scarcity.

Keywords: Climate change, ARIMA model, Geographic information systems, Climate projections.

1. INTRODUCCIÓN

Los nevados de Abancay, ubicados en la región de Apurímac entre 4,800 y 5,400 msnm, la economía regional depende críticamente de estos recursos: el sector agrícola de la población económicamente activa. Sin embargo, existe un vacío importante de información científica actualizada sobre la dinámica reciente de estos glaciares y sus proyecciones futuras. El presente estudio analiza la evolución espaciotemporal de los nevados de Abancay durante el período 2017-2023 y proyecta su comportamiento hasta 2030. Para ello, se integran técnicas de teledetección mediante imágenes Sentinel-2, análisis espacial con Sistemas de Información Geográfica y modelación predictiva utilizando el método ARIMA. Los objetivos espe-

cíficos incluyen cuantificar las variaciones en extensión superficial y perímetro de los casquetes nevados, caracterizar las tendencias de temperatura y precipitación, y evaluar las implicaciones sobre la disponibilidad futura de recursos hídricos.

La metodología combina el procesamiento de imágenes satelitales obtenidas de la plataforma ESRI con datos meteorológicos históricos (2000-2014) del SENAMHI y la base POWER de NASA. El análisis espacial emplea técnicas de clasificación supervisada y detección de cambios, mientras que las proyecciones climáticas se desarrollan mediante modelos ARIMA, validados ampliamente en estudios glaciológicos por su capacidad para capturar patrones autorregresivos en series temporales no estacionarias.



Revista de Investigación en Ciencia y Tecnología

ISSN: 2810-8124 (en línea) / ISSN: 2706-543x

Universidad Nacional Micaela Bastidas de Apurímac – Perú

Vol. 7 Núm. 1 (2025) - Publicado: 28/10/25 - Indexaciones

Número: doi.org/10.57166/riqchary/v7.n1.2025

Páginas: 88-93 | Recibido 01/01/2025 ; Aceptado 01/04/2025

doi.org/10.57166/riqchary.v7.n2.2025.11

Autor:

A. **ORCID iD** <https://orcid.org/0009-0007-4796-2468> Alexander Darío Huamán Ccorahua, Universidad Nacional Micaela Bastidas de Apurímac, Apurímac – Perú 191155@unamba.edu.pe.

2. METEOLOGÍA

2.1 ÁREA DE ESTUDIO

La zona de investigación comprende los nevados de la provincia de Abancay (13°38'S, 72°53'O) en el departamento de Apurímac – Perú, ubicados entre 4,800 y 5,400 msnm. Esta región presenta características climáticas típicas de los Andes tropicales, con marcada estacionalidad en las precipitaciones [1]. [2] indica que se caracterizan el clima regional como semiárido con precipitaciones concentradas entre diciembre y marzo.

2.2 Modelación ARIMA

El modelo ARIMA se expresa mediante la ecuación general:

$$\phi(B)(1 - B)^d X_t = \theta(B)\varepsilon_t \quad (1)$$

donde $\phi(B)$ representa el operador autorregresivo de orden p , $\theta(B)$ el operador de media móvil de orden q , d el orden de diferenciación, X_t la serie temporal y ε_t el término de error.

El proceso metodológico incluyó: análisis de estacionariedad mediante pruebas Dickey-Fuller Aumentada [14], identificación de parámetros óptimos utilizando criterios AIC y BIC [15], y validación del modelo mediante análisis de residuos y pruebas Ljung-Box.

2.3 Limitaciones del modelo ARIMA

Si bien los modelos ARIMA han demostrado eficacia en la predicción de series temporales, presentan limitaciones inherentes en contextos glaciológicos complejos que resultan fundamentales para la interpretación adecuada de los resultados. El supuesto de linealidad intrínseco al modelo no captura completamente los comportamientos no lineales de los procesos de ablación glaciar. Para mitigar estas limitaciones metodológicas, se implementó una validación cruzada con datos históricos del período 2000-2014 y se contrastaron los resultados con investigaciones regionales análogas; no obstante, las proyecciones deben interpretarse como escenarios probabilísticos condicionados a las tendencias actuales, reconociendo que perturbaciones climáticas extremas o modificaciones sustanciales en los patrones de circulación atmosférica podrían alterar las trayectorias modeladas.

2.4 Características Geomorfológicas

Los nevados de Abancay presentan una geomorfología compleja resultado de procesos glaciares y periglaciales activos. Maisa Rojas identificó formas glaciares heredadas que evidencian fluctuaciones climáticas pasadas, incluyendo morrenas laterales y frontales correspondientes a diferentes estadios de avance glaciar [3]. La orientación predominante sur-

sureste de los glaciares actuales refleja la influencia de la radiación solar y los vientos dominantes del este, factores determinantes en el balance energético superficial según [4].

2.5 Contexto Socioeconómico

La población de la región de Apurímac depende directamente de los recursos hídricos provenientes de los glaciares. El Instituto de Manejo del Agua y Medio Ambiente-IMA estima que el 65% del agua utilizada para riego durante la época seca proviene del derretimiento glaciar [5]. La agricultura, principal actividad económica regional, emplea al 47% de la población económicamente activa y genera el 23% del PBI regional según datos del INEI (2023) [6]. Esta dependencia hace que la región sea extremadamente vulnerable a los impactos del cambio climático.

3. MAPEO Y PROYECCIÓN DEL NEVADO (SUPERFICIE)

Las imágenes muestran la extensión del casquete nevado de Abancay (en azul) durante los períodos 2017-2018 y 2019-2020. Se observa que, si bien el casquete nevado ha disminuido en ciertas áreas, también muestra expansión en algunas regiones debido al deshielo. Este proceso provoca el desprendimiento y acumulación de hielo en las zonas bajas, generando una aparente expansión del perímetro. Sin embargo, si bien la superficie puede haber aumentado temporalmente, el casquete nevado ha perdido volumen de hielo, lo que refleja una disminución general de su masa (Figura 1), con datos obtenidos de Esri - Sentinel 2.

Los glaciares tropicales de los Andes peruanos constituyen indicadores críticos del cambio climático global [7]. La cordillera de Abancay, ubicada en la región de Apurímac, ha experimentado un retroceso glaciar acelerado en las últimas décadas, fenómeno documentado ampliamente en los Andes centrales [8]. Los glaciares tropicales responden rápidamente a las variaciones climáticas debido a su balance de masa sensible a pequeños cambios en temperatura [9].

Entre 2017 y 2023, la superficie de los casquetes nevados en Abancay experimentó fluctuaciones significativas. Inicialmente, la superficie fue de 111,86 hectáreas en 2017, disminuyendo ligeramente a 106,75 hectáreas en 2019. En 2021, la superficie se recuperó temporalmente, alcanzando las 122,26 hectáreas, pero luego mostró una marcada disminución en 2023, cuando se registraron solo 87,39 hectáreas. Con base en estos datos, se utilizó un modelo de regresión lineal simple para proyectar la tendencia futura. Las predicciones para 2025, 2027 y 2030 indican una disminución progresiva: para 2025, se estima que la superficie será de 92,59 hectáreas, en 2027 se reducirá a 86,80 hectáreas y para 2030, la superficie proyectada será de 78,12 hectáreas.

Este análisis sugiere una disminución general de la extensión del manto de nieve, con una pérdida proyectada del 30,16 %

entre 2017 y 2030. Esto podría estar relacionado con factores como el cambio climático y las variaciones meteorológicas, que tienen implicaciones directas para los ecosistemas locales y los recursos hídricos que dependen de los mantos de nieve.

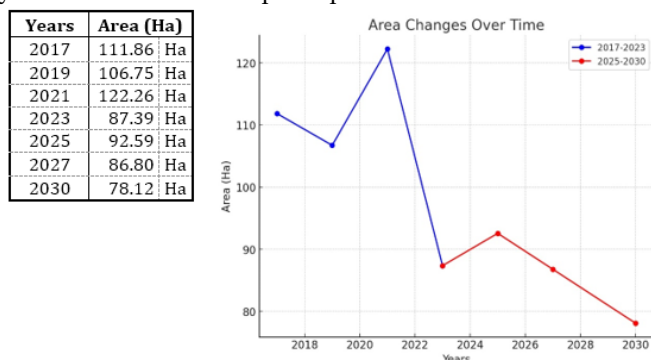


Fig. 1. Proyección del nevado de Abancay para el 2030.

El retroceso glaciar observado en Abancay forma parte de un fenómeno más amplio que afecta a toda la cordillera andina. Según [10] documentaron pérdidas de masa glaciar del 35-50% en los Andes centrales desde 1975. Esta tendencia se ha acelerado en el siglo XXI, con tasas de pérdida que superan los promedios históricos, como reporta el IPCC (2023) en su sexto informe de evaluación [11].

La vulnerabilidad de los glaciares tropicales se ve exacerbada por su ubicación altitudinal cerca de la isoterma de 0°C. [12] explican que pequeñas variaciones en la elevación de esta isoterma pueden resultar en cambios dramáticos en el área de acumulación glaciar. Este fenómeno es particularmente crítico en Abancay, donde la topografía local y los patrones de circulación atmosférica crean microclimas que influyen en la dinámica glaciar.

3.1 Perímetro del Nevado

Los datos sobre el perímetro del manto nevado de Abancay, obtenidos de la plataforma ESRI entre 2017 y 2023, muestran una disminución general con algunas fluctuaciones. El perímetro inicial en 2017 fue de 14 380 m, disminuyendo a 12 160 m en 2023. Sin embargo, en 2021 se observó un aumento temporal a 18 100 m, lo cual podría explicarse por procesos de deshielo, como desprendimientos de hielo o roca, que incrementaron temporalmente el área expuesta. Con estos datos, se aplicó una regresión lineal simple para proyectar los perímetros para 2025, 2027 y 2030, mostrando una disminución constante. En general, el perímetro proyectado para 2030 será de 13.710 m, lo que representa una disminución del 4,66 % con respecto a 2017. Esta reducción general del perímetro es coherente con el retroceso de los casquetes nevados, un fenómeno común debido al calentamiento global. Los datos se obtuvieron del Servicio Nacional de Meteorología (Senamín) de Perú.

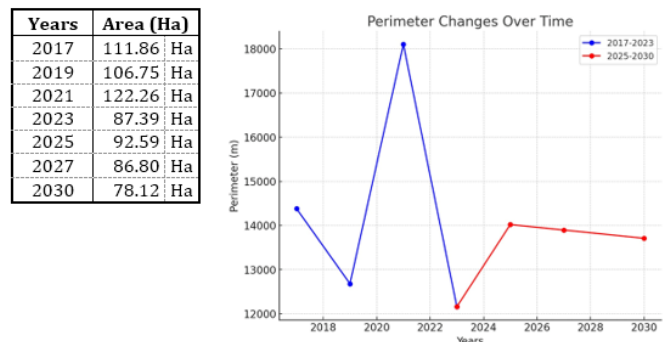


Fig. 2. Proyección del perímetro de los nevados para 2030.

3.2 Análisis de Temperatura: Tendencias Históricas y Proyecciones Futuras

La modelación de series temporales se realizó mediante el método ARIMA (Autoregressive Integrated Moving Average), ampliamente validado para proyecciones climáticas según [13]. El proceso metodológico incluyó:

El análisis de estacionariedad mediante pruebas Dickey-Fuller Aumentada [14], la identificación de parámetros óptimos utilizando criterios AIC y BIC como recomienda [15], y la validación del modelo mediante análisis de residuos y pruebas Ljung-Box.

El análisis de la temperatura promedio en Abancay durante el período 2000-2014 revela una importante tendencia al alza. Los datos muestran que la suma total de las temperaturas promedio anuales aumentó de 5773,05 °C en 2000 a 5872,05 °C en 2013 (suma total de temperaturas anuales), lo que pone de relieve un preocupante aumento de los niveles de temperatura que afecta directamente a la masa del manto nevado. El modelo de proyección ARIMA, desarrollado en Python, se ajustó para capturar esta tendencia creciente. Las predicciones indican un aumento continuo de la temperatura promedio, que alcanzará aproximadamente 15,79 °C para 2030. Este aumento sugiere un aumento acumulado de aproximadamente 0,7 °C desde 2014, lo que podría contribuir a la degradación acelerada del manto nevado y a la reducción de las reservas de agua.

TABLA 1

Proyecciones de Temperatura en Abancay hasta 2030

YEAR	Intervalo	Intervalo (Min)	Intervalo (Max)
2015	15.66894847	15.16266138	16.17523556
2016	15.57747695	15.00026051	16.1546934
2017	15.81753439	15.23539356	16.39967522
2018	15.92147133	15.33927163	16.50367103
2019	15.79719966	15.17522726	16.41917207
2020	15.70914483	15.04803554	16.37025412
2021	15.76654741	15.09965177	16.43344304
2022	15.83145941	15.16214232	16.50077649
2023	15.80984867	15.12739976	16.49229758
2024	15.76618634	15.06352767	16.46884502
2025	15.77043256	15.05621184	16.48465328

2026	15.79771954	15.07661361	16.51882547
2027	15.80067042	15.07006606	16.53127479
2028	15.78475128	15.04115911	16.52834345
2029	15.77974423	15.0245784	16.53491005
2030	15.78836664	15.02408683	16.55264645

Generando la siguiente proyección gráfica:

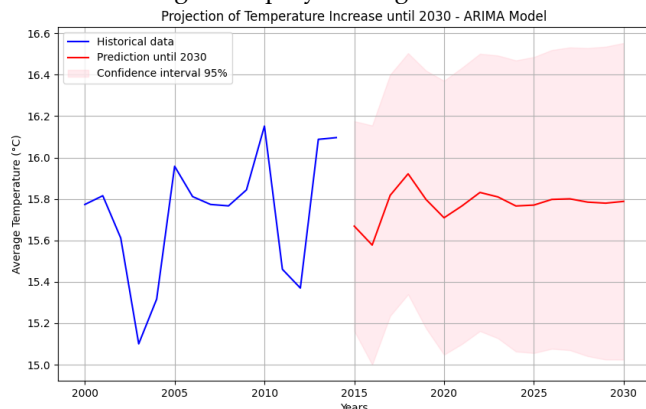


Fig. 3. Proyección del aumento de temperatura hasta 2030 – Modelo ARIMA.

3.3 Precipitación: Análisis del Comportamiento y Proyecciones

El comportamiento de la precipitación en la región también ha sido muy variable. La precipitación anual total en Abancay osciló entre 772,1 mm en 2000 y un máximo de 1012,8 mm en 2012. Sin embargo, el modelo ARIMA aplicado a estos datos no mostró un ajuste estadísticamente significativo, lo que revela que la precipitación no sigue un patrón estacional predecible. Se proyecta que la precipitación promedio entre 2015 y 2030 se mantendrá en un rango de 665,37 mm a 738,86 mm, con un promedio que refleja un ligero aumento en los niveles de precipitación. Sin embargo, esto no garantiza la estabilidad de los casquetes nevados. El análisis de la precipitación en el nevado de Apurímac, mediante el modelo ARIMA, reveló un ajuste estadísticamente insignificante, con parámetros autorregresivos de 0,3125 ($p=0,641$) y una media móvil de -0,999 ($p=0,992$), lo que sugiere una relación temporal débil y la ausencia de un patrón estacional predecible. La prueba de autocorrelación de Ljung-Box mostró un valor de $Q=0,28$ ($p=0,60$), lo que indica que no existe autocorrelación significativa, mientras que la prueba de Jarque-Bera indicó residuos con distribución normal ($p=0,77$). Estos resultados sugieren que factores climáticos complejos podrían estar afectando el ciclo hidrológico tradicional, lo que dificulta la modelización precisa de los patrones de precipitación en esta región montañosa.

TABLA 2

Predicción de precipitación para 2030

Año	Predicción	Mínimo	Máximo	Promedio
2015	665.3663376	391.0344716	939.6982036	665.3663376
2016	715.8901452	423.678728	1008.101562	715.8901452

2017	731.6797966	436.3062884	1027.053305	731.6797966
2018	736.6143633	440.4749519	1032.753775	736.6143633
2019	738.1565092	441.7990032	1034.514015	738.1565092
2020	738.6384592	442.2146947	1035.062224	738.6384592
2021	738.7890778	442.344622	1035.233534	738.7890778
2022	738.8361489	442.3850588	1035.287239	738.8361489
2023	738.8508595	442.3975101	1035.304209	738.8508595
2024	738.8554569	442.4012137	1035.3097	738.8554569
2025	738.8568937	442.4021833	1035.311604	738.8568937
2026	738.8573427	442.4022984	1035.312387	738.8573427
2027	738.857483	442.4021465	1035.31282	738.857483
2028	738.8575269	442.4019111	1035.313143	738.8575269
2029	738.8575406	442.4016497	1035.313431	738.8575406
2030	738.8575448	442.4013801	1035.31371	738.8575448

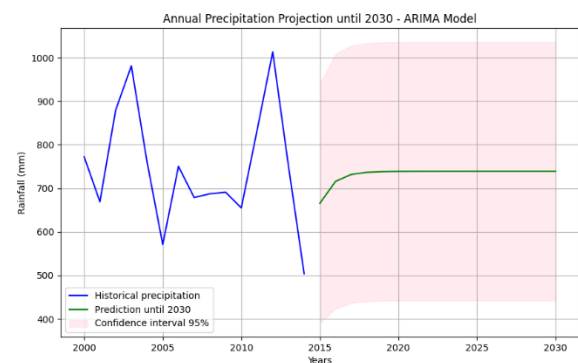


Fig. 4. Predicción de precipitación anual al 2030 Abancay con el Modelo ARIMA.

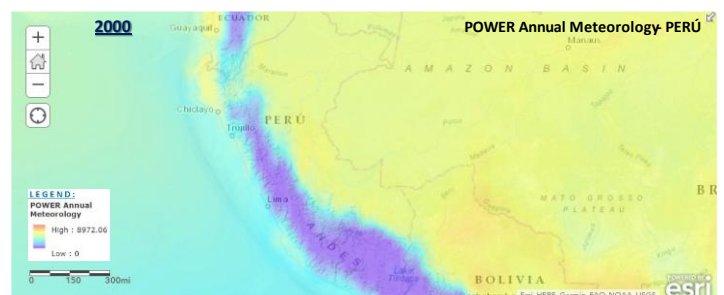


Fig. 5. POWER Meteorología Anual Perú 2000.



Fig. 6. POWER Meteorología Anual Perú 2010.



Fig. 7. POWER Meteorología Anual Perú 2015.

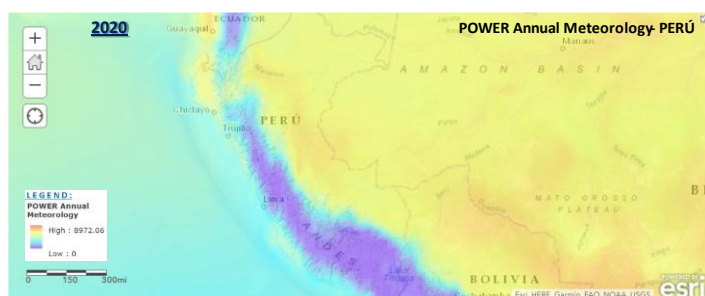


Fig. 8. POWER Meteorología Anual Perú 2020.



Fig. 9. POWER Meteorología Anual Perú 2022.

3.4 Implicaciones para futuras investigaciones

Los hallazgos del presente estudio establecen un marco integral de líneas de investigación prioritarias que abarcan la comprensión y gestión de los sistemas glaciares andinos. Desde la perspectiva científica, resulta imperativo desarrollar modelos híbridos que integren metodologías ARIMA con arquitecturas de aprendizaje profundo como LSTM, permitiendo capturar dinámicas no lineales y retroalimentaciones complejas entre variables climáticas y balance de masa glaciar. Se recomienda la implementación de una red integrada de sensores IoT que incluyan sensores de temperatura distribuidos, convergiendo en una plataforma computacional unificada que combine modelos ARIMA, redes neuronales y simulaciones físicas para análisis de ensamble y cuantificación de incertidumbre.

4. IV. CONCLUSIONES Y RECOMENDACIONES

4.1 Conclusiones

El análisis de los nevados de Abancay muestra una reducción significativa en su superficie y volumen entre 2017 y 2023. Las proyecciones indican que esta tendencia continuará, con una disminución del 30,16 % en la superficie del

nevado para 2030. Esta reducción está estrechamente relacionada con el aumento de las temperaturas medias y las fluctuaciones en las precipitaciones, ambas impulsadas por el cambio climático.

Las tendencias de temperatura en Abancay revelan un aumento continuo del promedio anual desde el año 2000. Los pronósticos sugieren un aumento adicional de aproximadamente 0,7 °C para 2030, lo que acelera el derretimiento del nevado. Esto representa una amenaza para los ecosistemas locales y el suministro de agua que depende de estos picos nevados.

A pesar de algunos aumentos en las precipitaciones entre 2000 y 2012, los patrones de precipitación siguen siendo impredecibles. Esta variabilidad sugiere que, si bien puede haber ligeros aumentos en las precipitaciones, estos no serán suficientes para estabilizar los nevados ni garantizar la disponibilidad a largo plazo de los recursos hídricos en la región.

4.2 Recomendaciones

Desarrollar estrategias eficientes de conservación del agua para abordar la pérdida prevista de los casquetes nevados. Esto puede incluir la construcción de embalses y la implementación de tecnologías avanzadas de riego para mitigar los impactos en la agricultura y el suministro de agua.

El monitoreo continuo del clima es esencial para rastrear los cambios en la temperatura y las precipitaciones. Estos datos ayudarán a ajustar los planes de mitigación y adaptación, y a proporcionar alertas oportunas a las comunidades sobre los posibles impactos en los recursos hídricos.

Fortalecer la educación ambiental para crear conciencia sobre el cambio climático y sus efectos en los recursos hídricos. Las campañas comunitarias centradas en la conservación del agua y la protección de los ecosistemas montañosos pueden ayudar a minimizar los impactos negativos en las poblaciones locales.

AGRADECIMIENTOS

Doy mi sincero agradecimiento al concurso NASA Space Apps Challenge por brindar la inspiración y el marco propicio para el desarrollo del presente artículo de investigación. Asimismo, se reconozco especialmente la valiosa colaboración y acompañamiento del Ing. Jhonatan Rojas Ventura durante el proceso de desarrollo profesional y la elaboración de este trabajo.

REFERENCIAS

- [1] BCRP, Caracterización del departamento de Apurímac, Cusco, 2024, p. 13.
- [2] L. C. Justo Espinoza, Influencia de la precipitación y la temperatura sobre la variabilidad espacio temporal de las superficies, Huarochiri, 2016.
- [3] J. Gonzales, F. Inocente, E. Mendoza-Nieto y T. Mendez-Izquierdo, Calentamiento global y la deglaciación al Suerte del Nevado de Yerupajá, Perú, Yerupaja, 2020. <https://doi.org/10.51431/par.v2i2.645>
- [4] A. Taci, C. Thomas, C. Alejo y M. Néstor, «Seguimiento del comportamiento de tamaño morfológico de glaciares y la relación de clima regional

en la latitud 8°-16° S, Perú», UNESCO, vol. 14, n° 1, p. 19, 10 2022.

- [5] J. Salinas Cesáreo, «Por fugas se pierde 65% de agua en el agro de riego», 21 03 2024. [En línea]. Available: <https://www.jomada.com.mx/noticia/2024/03/21/estados/por-fugas-se-pierde-65-de-agua-en-el-agro-de-riego-7776>.
- [6] INEI, Comportamiento de la economía peruana en el 2023, 2023, p. 106
- [7] ONU, «Noticias ONU», Naciones Unidas, 20 03 2025. [En línea]. Available: <https://news.un.org/es/interview/2025/03/1537281>.
- [8] L. Jesús, D. Liselotte y D. Jaime, «El retroceso glaciar en la alta montaña de los Andes Peruanos. Caso: Santuario Nacional de Ampay, Apurímac», Revista UNAP, vol. 16, n° 1, p. 12, 01 03 2021.
- [9] Bosson, J. Huss y Osipova, «Earth's Future», 29 04 2019. [En línea]. Available: <https://doi.org/10.1029/2018EF001139>.
- [10] A. Tacsí, T. Condom, A. Cochachin y N. Montalvo, «Monitoring of the behavior of morphometric size of glaciers and the relationship of regional climate in latitude 8°-16° S, Peru», vol. 14, n° 1, p. 19, 2022.
- [11] V. Jessica, A. Adeline, W. David, E. G. Laplace y C. Nada, «Sexto informe de evaluación resumen para docentes», Office for climate education, 2024.
- [12] S. Simone, K. Marlene y C. Rolando, «El futuro del clima y de los glaciares en el Perú», de Proyecto glaciares, Lima, CARE Perú, 2017, p. 19.
- [13] P. F. Carrillo Freire y J. E. Aguilar Reyes, «Procesamiento, análisis y visualización del comportamiento de variables climáticas medidas desde enero de 2014 hasta diciembre de 2022 y pronóstico de los siguientes 12 meses», de REINCASOL, 2025, p. 27. [https://doi.org/10.59282/reincisol.V4\(7\)1457-1483](https://doi.org/10.59282/reincisol.V4(7)1457-1483)
- [14] «Prueba Dickey-Fuller Aumentada (ADF)», 05 08 2022. [En línea]. Available: https://numxl.com/es/blogs/prueba-dickey-fuller-aumentada/?srsltid=AfmBOoqrDQYrqR42Cwz5JOjw_YpG3ArbhqCFoW-lkSZii3QGrG9PNkdY.
- [15] D. L. Mojica Hidalgo, Automatización modelo ARIMA usando programación en Software R, Bogotá, 2023, p. 68.

BIOGRAFÍA



Alexander Darío Huamán Ccorahua, egresado en Ingeniería de Minas de la Universidad Nacional Micaela Bastidas de Apurímac, Perú (2024). Obtuvo su titulación en el Instituto Técnico Bilingüe Internacional, Costa Rica (2024). Ha trabajado en Polo Ingenieros Contratistas (2022), Antapaccay como Líder de Implementación IoT en

Transformación Organizacional (2024), Newrest Las Bambas (2025), y actualmente es Ingeniero Asistente de Operaciones Mina en MIDDERSA S.A., Arequipa (2025-presente). Expositor en MINEXCHANGE 2024 (SME, EE.UU.) / CONIMERA XXV 2023 y moderador EXPOCOBRE 2023-2024. Especialización en Operaciones mina, Lean SixSigma, Machine Learning y gestión de proyectos tecnológicos en minería.

Efecto de coagulantes de origen vegetal sobre las características fisicoquímicas y sensoriales del queso

Effect of plant-derived coagulants on the physicochemical and sensory properties of cheese

María del Carmen Barreto Prado^A, Fulgencio Vilcanqui Pérez^B y Luis Fernando Pérez Falcón^C

Resumen— El cuajo de origen animal es ampliamente utilizado en la elaboración de diversas variedades de queso; sin embargo, su alto costo, el uso de organismos genéticamente modificados y ciertas restricciones religiosas han impulsado la búsqueda de alternativas sostenibles como los coagulantes vegetales. El presente estudio tuvo como objetivo evaluar el efecto de los coagulantes de origen vegetal sobre las características fisicoquímicas y sensoriales del queso. Se emplearon enzimas obtenidas de papaya serrana (*Carica pubescens*, papaína), higo (*Ficus carica*, ficina) y ortiga menor (*Urtica urens*, ácidos fenólicos), utilizando renina en pastilla como control. Durante la coagulación se adicionaron 0,3% (v/v) de látex de papaya serrana, 3% (m/v) de ramas de higo y 4% (m/v) de hojas de ortiga menor, mientras que la renina en pastilla se usó al 0,01% (m/v), considerando un litro de leche como base. Los análisis incluyeron pH, acidez, sinéresis, contenido de grasa, materia seca, contenido de agua y evaluación sensorial. Los resultados evidenciaron que el látex de papaya serrana mejoró varios parámetros tecnológicos: redujo la sinéresis (de 5,04% a 4,17%) así como el contenido de agua (de 54,39 g/100 g a 51,19 g/100 g), incrementó la materia seca (de 45,61 g/100 g a 48,81 g/100 g) y en contenido de grasa (de 26,57 g/100 g a 28,62 g/100 g), todo ello respecto al queso control (elaborado con renina en pastilla); asimismo, sus valores de pH y acidez fueron similares a los del queso control. Sensorialmente, el queso con papaya obtuvo la mayor puntuación en sabor, olor, color, textura y aceptabilidad global, alcanzando valores superiores a los del queso control. En conclusión, el látex de papaya serrana no solo fue el mejor coagulante vegetal evaluado, sino que produjo un queso con características fisicoquímicas y sensoriales comparables e incluso superiores al obtenido con renina en pastilla, lo que respalda su viabilidad como alternativa sostenible en la elaboración de quesos.

Palabras clave: Ficina, ortiga menor, papaína, queso

Abstract— Animal rennet is widely used in the production of various cheese varieties; however, its high cost, the use of genetically modified organisms, and certain religious restrictions have driven the search for sustainable alternatives such as plant-based coagulants. This study aimed to evaluate the effect of plant-based coagulants on the physicochemical and sensory characteristics of cheese. Enzymes obtained from papaya (*Carica pubescens*, papain), fig (*Ficus carica*, ficin), and stinging nettle (*Urtica urens*, phenolic acids) were used, with rennet tablets serving as a control. During coagulation, 0.3% (v/v) papaya latex, 3% (w/v) fig branches, and 4% (w/v) stinging nettle leaves were added, while rennet tablets were used at 0.01% (w/v), with one liter of milk as the base. The analyses included pH, acidity, syneresis, fat content, dry matter, water content, and sensory evaluation. The results showed that the papaya latex improved several technological parameters: it reduced syneresis (from 5.04% to 4.17%) and water content (from 54.39 g/100 g to 51.19 g/100 g), increased dry matter (from 45.61 g/100 g to 48.81 g/100 g), and increased fat content (from 26.57 g/100 g to 28.62 g/100 g), all compared to the control cheese (made with rennet tablets). Furthermore, its pH and acidity values were similar to those of the control cheese. Sensory evaluation showed that the papaya cheese obtained the highest scores for flavor, aroma, color, texture, and overall acceptability, achieving values higher than those of the control cheese. In conclusion, papaya latex was not only the best plant-based coagulant evaluated, but it also produced a cheese with physicochemical and sensory characteristics comparable to, and even superior to, that obtained with rennet tablets, thus supporting its viability as a sustainable alternative in cheesemaking.

Keywords: Ficin, small nettle, papain, queso



Revista de Investigación en Ciencia y Tecnología

ISSN: 2810-8125 (en línea) / ISSN: 2706-543

Universidad Nacional Micaela Bastidas de Apurímac – Perú

Vol. 7 Núm. 1 (2025) - Publicado: 21/11/25 - Indexaciones

Número: doi.org/10.57166/riqchary/v7.n1.2025

Páginas: 94-101 | Recibido 26/09/2025 ; Aceptado 21/11/2025

doi.org/10.57166/riqchary.v7.n2.2025.12

Autores:

- A. **ORCID iD** <https://orcid.org/0009-0006-1755-7376>
María del Carmen Barreto Prado, Universidad Nacional Micaela Bastidas de Apurímac
161052@unamba.edu.pe.
- B. **ORCID iD** <https://orcid.org/0000-0002-9163-5655>
Fulgencio Vilcanqui Pérez, Universidad Nacional Micaela Bastidas de Apurímac fvilcanqui@unamba.edu.pe
- C. **ORCID iD** <https://orcid.org/0000-0002-9463-0489>
Luis Fernando Pérez Falcón, Universidad Micaela Bastidas de Apurímac, lfperrez@unamba.edu.pe

1 INTRODUCCIÓN

En el contexto global, la industria láctea enfrenta un desafío creciente: la necesidad de diversificar sus procesos con tecnologías más sostenibles, eficientes y adaptadas a las expectativas del consumidor moderno. La búsqueda de alternativas al cuajo animal no solo responde a criterios éticos o de origen vegetal, sino también a la demanda por productos con menor impacto ambiental y mejor trazabilidad [1], [2].

En el Perú, el queso fresco es un producto de amplio consumo y gran arraigo cultural, sobre todo en las zonas andinas. Aun así, la industria quesera nacional sigue dependiendo del cuajo comercial de origen animal, lo que implica costos elevados, fluctuaciones en la disponibilidad y una barrera para productores con recursos limitados [3], [4], [5]. Esta dependencia condiciona la innovación tecnológica y la autonomía en la producción local.

Las plantas con enzimas proteolíticas ofrecen una vía prometedora para reemplazar o complementar el cuajo animal. En particular, especies como la papaya serrana (*Vasconcellea pubescens* A. DC.), higo (*Ficus carica*) y ortiga menor (*Urtica urens*) han sido estudiadas internacionalmente por su capacidad coagulante. Por ejemplo, [6], [7] evaluaron papaina liofilizada de *Vasconcellea* y encontraron que una dosis de 2 g/L a 30 ° C produjo quesos con características comparables al cuajo comercial.

En cuanto al higo, la enzima ficina ha sido reconocida como una proteasa eficaz para coagular leche, aunque con menor especificidad en comparación con la quimosina. Estudios recientes señalan que ficina puede actuar como coagulante vegetal viable, especialmente en sistemas tradicionales, aunque sus efectos proteolíticos secundarios deben controlarse [8]. Al mismo tiempo, la ortiga ha sido empleada en diversas comunidades como coagulante vegetal tradicional. Algunas investigaciones sobre extractos de plantas del género *Urtica* demuestran que pueden inducir coagulación láctea, aunque su uso comercial está menos documentado, lo que abre espacio para su validación tecnológica local [9]. No obstante, en la región altoandina peruana, incluido Apurímac, hay escasez de estudios que evalúen estos coagulantes vegetales en condiciones reales de producción quesera. Factores como la calidad de la leche, condiciones ambientales y prácticas de manufactura pueden afectar los resultados tecnológicos y sensoriales, por lo que es imprescindible realizar estudios adaptados al contexto local. En ese sentido, la presente investigación propone analizar el efecto de tres coagulantes vegetales: látex de papaya serrana, látex de higo y hojas de ortiga menor, sobre las características fisicoquímicas y sensoriales del queso fresco. Los objetivos buscan no solo evaluar su viabilidad tecnológica, sino también identificar cual de los coagulantes logra mejores propiedades y mayor aceptación del consumidor.

La investigación se desarrolló en el Laboratorio de Procesamiento de Productos Agroindustriales de la Escuela

Profesional de Ingeniería Agroindustrial de la Universidad Nacional Micaela Bastidas de Apurímac, ubicado en el distrito de Tamburco, provincia de Abancay, región Apurímac, Perú, durante el periodo comprendido entre enero y agosto de 2025. Se espera que los resultados de este estudio contribuyan a promover recursos vegetales locales en la industria láctea peruana, fortaleciendo la producción regional, reduciendo dependencia del cuajo animal y ofreciendo alternativas tecnológicas más sostenibles. Así mismo, estos hallazgos podrán servir como base para futuras investigaciones, adopciones industriales y políticas de innovación en el sector quesero nacional.

Esta investigación parte del objetivo general de evaluar el efecto de los coagulantes vegetales sobre las propiedades del queso, con los objetivos específicos de comparar su desempeño en parámetros fisicoquímicos (pH, acidez, sinéresis, materia seca, agua y grasa) y sensoriales (sabor, olor, color, textura y aceptabilidad global), así como determinar qué coagulante vegetal permite obtener queso de mejor calidad según criterios sensoriales.

2 MATERIALES Y METODOLOGÍA

2.1 Materia prima e insumos

Las muestras de papaya serrana se obtuvieron en la zona de Kerapata-Tamburco; el higo, en el mercado de Villa Ampay (Abancay); y la ortiga menor, en el mercado Las Américas (Abancay). La leche fresca de vaca se adquirió igualmente en el mercado Las Américas, mientras que la renina en pastilla fue adquirida en una farmacia local.

2.2 Características fisicoquímicas de la leche

Se utilizaron 4 litros de leche para cada tratamiento (1 litro por tratamiento) y se realizó la prueba del alcohol y la determinación de pH con el fin de verificar sus condiciones de calidad. Posteriormente, la leche fue filtrada y acondicionada para iniciar el proceso de elaboración de queso.

La leche fresca empleada como materia prima se caracteriza por su color blanco opalescente, olor lácteo suave y sabor ligeramente dulce. Presenta un pH entre 6,6 y 6,8 y una composición típica constituida por aproximadamente 87% de agua, 3,0-3,5% de proteínas, 3,0-4,5% de grasa, 4,5-5,0% de lactosa y 0,7% de minerales [10]. Su estructura corresponde a una emulsión de glóbulos de grasa y a una suspensión coloidal de caseínas, propiedades que influyen directamente en la formación del coágulo durante la elaboración de queso. La calidad microbiológica de la leche depende de las condiciones de ordeño y manejo, por lo que su adecuada higiene y almacenamiento son esenciales para garantizar su inocuidad y el comportamiento tecnológico en los procesos de coagulación.

2.3 Obtención de latex de papaya serrana

Los frutos de papaya serrana se recolectaron en estado verde y fueron llevados al laboratorio, donde se evaluó su condición

sanitaria, descartando aquellos con daños o golpes (Figura 1). Asimismo, se verificó su grado de madurez, dado que este factor incide de manera directa en la cantidad y calidad del látex extraído. Se realizó una clasificación manual de los frutos, separando los que presentan condiciones óptimas (verde y firme) de aquellos con signos de sobre-maduración o deterioro. Esta etapa permitió garantizar uniformidad en la materia prima utilizada para la extracción del látex. Se lavó la fruta a utilizar de una manera exhaustiva, hasta que quede libre de cualquier impureza física. Se realizó el corte longitudinal de cuatro a seis incisiones de 2 mm con un cuchillo de acero inoxidable esto para lograr que la fruta exhuda la sustancia lechosa. Se extrajo el látex en estado líquido del corte anterior y se recogió en una cápsula de porcelana previamente registrando su peso. La muestra se mantuvo refrigerada a una temperatura de 5 a 7 °C en un recipiente oscuro o ámbar totalmente cerrado para evitar la oxidación de la muestra.

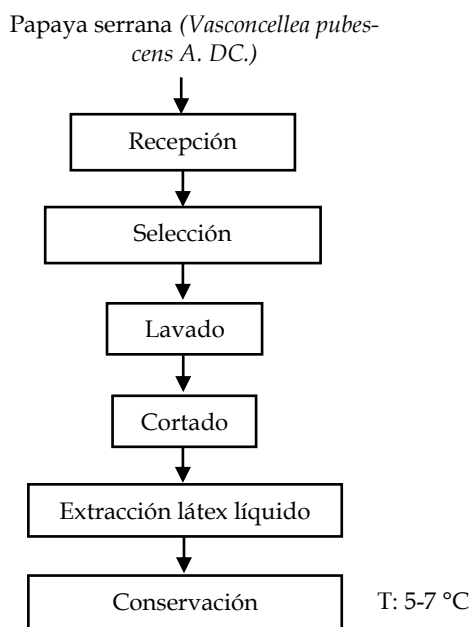


Fig. 1. Diagrama de bloques para la obtención de látex de papaya serrana

2.4 Obtención de coagulante del higo y ortiga macho

En la Figura 2, se describe el proceso de obtención de coagulantes a partir de la rama y hojas del higo y de hojas de ortiga menor. Las ramas y hojas frescas de higo (*Ficus carica*) y hojas de ortiga menor (*Urtica urens*) fueron recolectadas en condiciones higiénicas y dentro del tiempo adecuado de cosecha (estado verde) y recepcionados en el laboratorio. Luego se procedió a la selección del material vegetal, separando las ramas y hojas verdes de aquellas que presenten signos de amarillamiento, marchitez o deterioro. Solo se utilizaron las muestras frescas, descartando las secas o en mal estado. El material seleccionado fue sometido a un lavado con agua potable para eliminar restos de tierra, polvo, insectos y otras impurezas físicas, se realizó un enjuague adicional en solución

desinfectante suave (con hipoclorito de sodio al 50 ppm), seguido de un enjuague con agua potable para asegurar inocuidad. A lo que concierne al higo (*Ficus carica*) las ramas y hojas previamente lavadas se cortaron en forma transversal en fragmentos pequeños utilizando un cuchillo de acero inoxidable debidamente sanitizado esto para lograr que exhuda la sustancia lechosa. Posteriormente, se recolectó la sustancia lechosa exudada líquida por el material vegetal en una cápsula de porcelana previamente registrando su peso. La muestra se mantuvo refrigerada a una temperatura de 5 a 7 °C en un recipiente oscuro o ámbar totalmente cerrado para evitar la oxidación de la muestra. Y en lo que concierne a la ortiga menor (*Urtica urens*) solo se utilizó las hojas ya lavadas y desinfectadas de manera directa sin ninguna extracción previa.

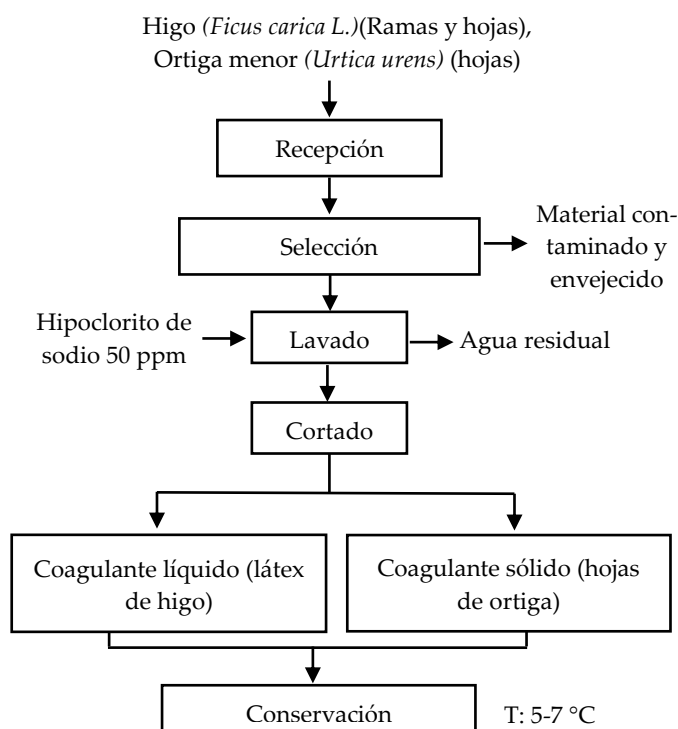


Fig. 2. Diagrama de bloques para la obtención de coagulante de higo y ortiga menor

2.5 Elaboración de queso fresco

La Figura 3, describe el proceso de elaboración de queso. La leche fresca fue recepcionada en el laboratorio a una temperatura de 20 a 25 °C. Posteriormente, la leche se sometió un filtrado para eliminar partículas sólidas, polvo, restos de pelo o suciedad que pueda contener. La leche se pasteurizó a 65 °C durante 30 minutos [11]. Después de la pasteurización, la leche se enfrió rápidamente hasta 40 a 42 °C y se adicionó el cloruro de calcio. Luego, se adicionaron los coagulantes de origen vegetal en cantidades de 0,3% (v/v) de látex de papaya serrana, 3% (m/v) de ramas de higo y 4% (m/v) de hojas de ortiga menor; mientras que la renina en pastilla se adicionó en 0,01% (m/v), todos ellos teniendo base de cálculo un litro de leche. Se dejó reposar a 40 °C por 40 min, tiempo en el cual se formó un gel firme conocido como cuajada. El gel obtenido fue cortado en cubos pequeños, procedimiento que facilitó la

expulsión del suero. La cuajada cortada se agitó suavemente durante 20 min y se dejó reposar, luego se realizó el batido para la sinéresis (eliminación del suero). La cuajada se introdujo en moldes de diferentes formas y tamaños, lo que permitió dar al queso su presentación final, además, durante el moldeado continúa el drenado de suero, ayudando a consolidar la estructura. Los moldes con cuajada se sometieron a presión controlada, expulsando el suero restante y compactando. El queso fue sumergido en una solución de salmuera al 2% de cloruro de sodio durante 45 min. Finalmente, los quesos se envasaron y se almacenaron a 4 °C, para las evaluaciones posteriores.

2.6 Evaluación de las características fisicoquímicas

El valor de pH se determinó por el método AOAC [12] en el potenciómetro de marca SI Analytics modelo Lab 850. En un vaso de 100 ml se colocó 1 g de muestra de queso fresco y se añadió 50 ml de agua destilada y se dejó en reposo por 1 h con agitación suave en intervalo de 10 min, para luego medir el pH.

La acidez titulable se determinó de acuerdo con el método AOAC [13]. En la muestra homogenizada se adicionó 2 g de queso fresco triturado en un matraz agregando 100 ml de agua destilada, se disolvió muy bien la mezcla, se agitó y se dejó en reposo, luego se filtró para obtener la solución líquida, se tomó una alícuota de 20 ml y se valoró con solución de hidróxido de sodio 0,1 N, usando 0,3 ml de indicador fenolftaleína, hasta el viraje incoloro a rosa pálido persistente. El valor de la acidez se calculó con la ecuación:

$$\% \text{ Acidez} = \frac{V_{\text{NaOH}} \cdot N_{\text{NaOH}} \cdot F_c}{V_m} \cdot 100 \quad (\text{Ecuación 1})$$

El porcentaje de la sinéresis del queso determinó de acuerdo con la NTP 202.193:2020 [14]. Para ello, las muestras queso fueron pesados y registrados el día uno y el día siete. El porcentaje (%) de sinéresis se calculó mediante la siguiente ecuación:

$$\% \text{ de sinéresis} = \frac{\text{Volumen de sinéresis}}{\text{Volumen total de la muestra}} \times 100 \quad (\text{Ecuación 2})$$

El contenido de agua y la materia seca, se determinó mediante el método gravimétrico, conforme a la AOAC [13]. Las muestras homogenizadas se colocaron en cápsulas de porcelana, se pesaron y se sometieron a secado en estufa a 105 °C durante por 12 h. Luego se enfriaron en desecador para evitar reabsorción de humedad y se pesó nuevamente en una balanza analítica. Posteriormente, se aplicaron ciclos adicionales de 30 min de secado, seguidos de enfriamiento y pesado, hasta alcanzar un peso constante.

$$\% \text{ de Humedad} = \frac{(P_1 - P_2)}{m} \times 100 \quad (\text{Ecuación 3})$$

Donde:

P1: peso de la placa más la muestra húmeda (g)

P2: peso de la placa más la muestra seca (g)

m: peso de la muestra húmeda (g)

$$\% \text{ de Materia Seca} = 100 \% - \% \text{ de Humedad} \quad (\text{Ecuación 4})$$

El contenido de grasa en el queso se determinó por el método AOAC 920.39 [15], mediante el uso del equipo Soxhlet de marca BUCHI Extraction System modelo B-811. En un sobre de papel filtro se colocó tres g de muestras secas de queso y se colocó en tubo Soxhlet. Luego en un balón de 250 ml se añadió 160 ml de solvente hexano y se procedió con la extracción por un tiempo de 2 a 4 h y luego se llevó a una estufa a 60°C para la evaporación del solvente. Finalmente se pesó el balón con el extracto etéreo y por diferencia de pesos se determinó el contenido de la grasa.

2.7 Evaluación de las características sensoriales

Se eligió un equipo de 30 panelistas consumidores para juzgar la calidad del queso (sabor, color, textura, olor y aceptabilidad global) utilizando una hoja de evaluación de nueve puntos de escala (me disgusta muchísimo: 1, me disgusta mucho: 2, me disgusta bastante: 3, me disgusta ligeramente: 4, ni me gusta ni me disgusta: 5, me gusta ligeramente: 6, me gusta bastante: 7, me gusta mucho: 8 y me gusta muchísimo: 9 [16].

2.8 Análisis estadístico

Se utilizó el diseño completamente aleatorizado (DCA) con el único factor en estudio al tipo de coagulante de origen vegetal (con tres niveles más un patrón). Para determinar las características fisicoquímicas del queso fresco se consideró al tratamiento patrón con renina en pastilla y los siguientes tratamientos con látex de papaya serrana (papaína), hojas y ramas de higo (ficina), y hojas de ortiga menor (ácidos fenólicos), con tres repeticiones. Luego se realizó el análisis de varianza (ANOVA) y la comparación múltiple de medias por Tukey a un nivel de significancia de 5% a fin de identificar el o los tratamientos significativos ($p < 0,05$). Se analizó mediante el uso del programa estadístico SPSS versión 30.0.0.

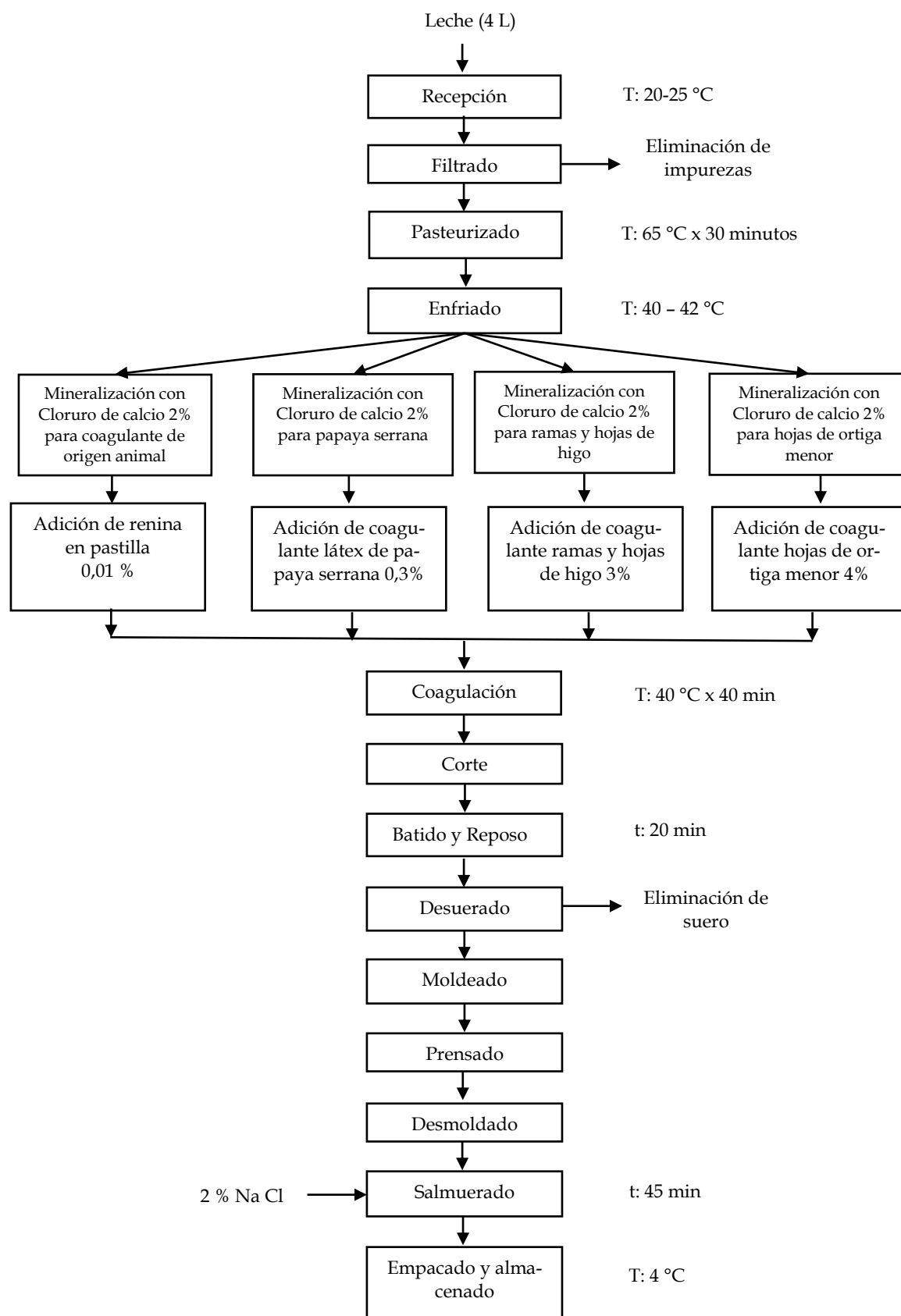


Fig. 3. Diagrama de bloques para elaboración de queso fresco

3 RESULTADOS

3.1 Características fisicoquímicas del queso

En la Tabla 1 se presentan los resultados del análisis fisicoquímico de los quesos elaborados con diferentes coagulantes. Al comparar los tratamientos con coagulantes vegetales frente al queso control (elaborado con renina en pastilla), se observaron diferencias significativas tanto en el pH como en la acidez titulable. El pH de los quesos elaborados con papaína se obtuvo 6,10; así mismo, los quesos elaborados con ficina obtuvieron un pH de 5,10 y en los quesos elaborados con ácidos fenólicos fue de 5,43 en contraste con el valor de 6,53 obtenido en el queso control (renina en pastilla). De manera similar, la acidez titulable fue mayor en los quesos con papaína (0,08 g/100 g), ficina (0,10 g/100 g) y ácidos fenólicos (0,15 g/100 g), en comparación con el queso elaborado con renina en pastilla (0,04 g/100 g).

En relación con la sinéresis, únicamente el queso elaborado con ácidos fenólicos presentó diferencias significativas frente al queso control (renina en pastilla), con valores de 8,89 % para ortiga menor y 5,04 % para renina en pastilla, respectivamente. En cambio, los quesos obtenidos con papaína (4,17 %) y ficina (7,01 %) no mostraron diferencias estadísticamente significativas respecto al queso control (renina en pastilla) (5,04 %).

TABLA 1

Características fisicoquímicas de queso elaborado con coagulantes de diferentes fuentes de origen vegetal

Coagulantes de origen vegetal	Características fisicoquímicas					
	pH	Acidez (g/100 g)	Sinéresis (% ms)	Contenido de agua (g/100 g)	Contenido de materia seca (g/100 g)	Contenido de grasa (g/100 g)
Origen animal (renina en pastilla)	6,53±0,04 ^b	0,04±0,01 ^a	5,04±1,69 ^a	54,39±0,79 ^{a,b}	45,61±5,45 ^{a,b}	26,57±4,03 ^b
Papaya serrana (papaína)	6,10±0,03 ^b	0,08±0,01 ^b	4,17±1,11 ^a	51,19±1,24 ^a	48,81±1,85 ^b	28,62±2,04 ^b
Higo (ficina)	5,10±0,02 ^a	0,10±0,01 ^c	7,01±0,81 ^{a,b}	58,99±0,88 ^{a,b}	41,01±4,22 ^{a,b}	22,95±1,57 ^{a,b}
Ortiga menor (ácidos fenólicos)	5,43±0,15 ^a	0,15±0,00 ^d	8,89±0,99 ^b	63,07±1,79 ^b	36,93±0,94 ^a	19,60±1,55 ^a

Promedio ± Desviación estándar (n=3). Los superíndices de las columnas representan diferencias significativas (p<0,05).

En cuanto al contenido de materia seca y humedad, no se observaron diferencias estadísticamente significativas entre los quesos elaborados con papaína (48,81 g/100 g), ficina (41,01 g/100 g) y ácidos fenólicos (36,93 g/100 g), en comparación con el queso control elaborado con renina en pastilla (45,61 g/100 g) esto referido a la materia seca, de igual manera en el contenido de agua no se observaron diferencias estadísticamente significativas entre los quesos elaborados con papaína (51,19 g/100g), ficina (58,99 g/100 g) y ácidos fenólicos (63,07 g/100 g), en comparación con el queso control elaborado con

renina en pastilla (54,39 g/100 g). Respecto al contenido de grasa, los quesos con papaína (28,62 g/100 g) y ficina (22,95 g/100 g) no presentaron diferencias significativas del queso control (26,57 g/100 g); sin embargo, el queso elaborado con ácidos fenólicos presentó una diferencia significativa, con un valor de 19,60 g/100 g frente a 26,57 g/100 g del queso control.

En la Figura 4, filas 3a, 3b, 3c y 3d se muestran las imágenes de los quesos frescos obtenidos a partir de coagulantes de origen vegetal, se puede observar que los tratamientos con coagulantes de origen vegetal (higo: 3b, papaya serrana: 3c y ortiga menor: 3d) son similares al tratamiento con renina en pastilla: 3a.

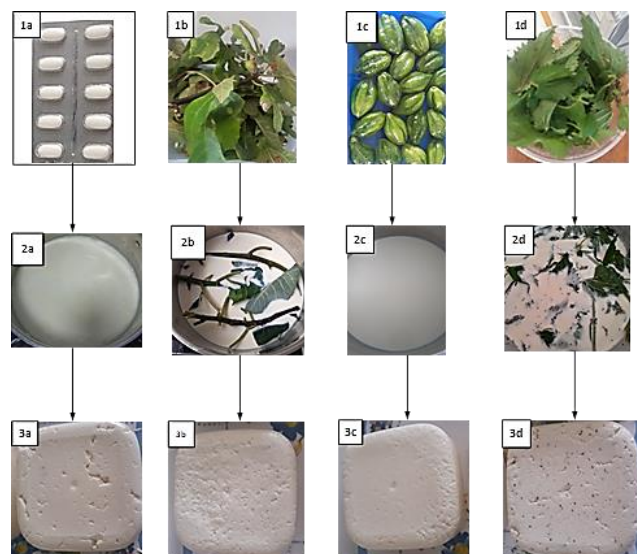


Fig. 4. Uso de coagulantes naturales para la elaboración de queso.

1a renina en pastilla; 1b, 1c y 1d, coagulantes de origen vegetal (higo, papaya y ortiga menor, respectivamente); 2a, coagulado de la leche con renina en pastilla; 2b, 2c y 2d, coagulado de la leche con coagulantes de origen vegetal; 3a, queso elaborado con renina en pastilla; 3b, 3c y 3d, quesos elaborados con coagulantes de origen vegetal.

3.2 Características sensoriales del queso

La figura 5 presenta la evaluación sensorial de quesos frescos elaborados con coagulantes de origen vegetal y animal (cuajo en pastilla). Al comparar los tratamientos, el queso elaborado con látex de papaya serrana obtuvo las puntuaciones más altas en los atributos evaluados: sabor (6,8), olor (6,55), color (7,1), textura (7,35) y aceptabilidad global (7). Este queso se caracterizó por un sabor suave, equilibrado entre salado y ligeramente ácido; un olor fresco característico de la leche; un color blanco brillante; textura cremosa, húmeda y untuosa; y una aceptabilidad global elevada debido a su compacidad, menor presencia de grietas, facilidad de corte y buena manejabilidad, lo que lo convirtió en el preferido por los consumidores. Por el contrario, el queso elaborado con ramas y hojas de higo presentó puntuaciones inferiores al queso control, con resultados de sabor (4,5), olor (6,1), color (5,65), textura (4,6) y aceptabilidad global (5,15). Finalmente, el queso elaborado con hojas de ortiga menor mostró las calificaciones más

bajas: sabor (3,6), olor (4,65), color (4,15), textura (4,35) y aceptabilidad global (3,7). Este producto se describió con un sabor intensamente astringente y ácido característico de la ortiga menor, un olor lácteo mezclado con notas herbáceas, un color blanco con tonalidades verdosas por la presencia de flores, y una textura blanda y arenosa. La aceptabilidad global fue reducida debido a la formación de múltiples grietas y una menor compactación en comparación con coagulante de origen animal.

En general, estas características sensoriales adversas en ortiga menor se atribuyen a la dosis excesiva del coagulante, lo que incrementó la acidez y favoreció la eliminación de mayor cantidad de suero, afectando la textura y generando una consistencia arenosa, razón por la cual fue el menos aceptado por los consumidores.

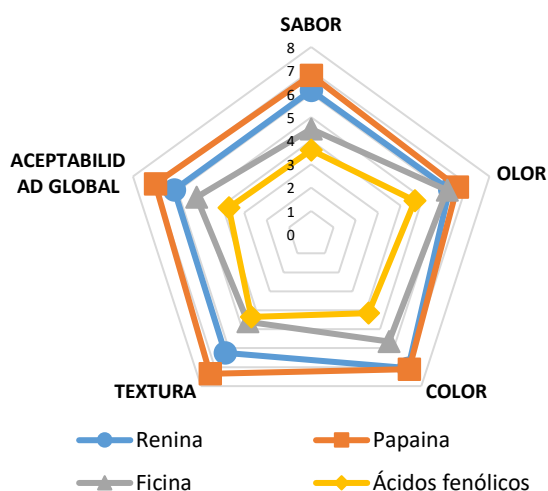


Fig. 5. Diagrama de araña de las características sensoriales (sabor, olor, color, textura y aceptabilidad global) de queso elaborado con coagulantes de origen vegetal (papaína, ficina y ácidos fenólicos) y renina en pastilla.

4 DISCUSIONES

Los resultados muestran que los quesos elaborados con coagulantes vegetales (papaína, ficina y ácidos fenólicos) presentaron diferencias significativas en el pH y la acidez titulable en comparación con el queso control. El queso con papaína mostró valores de pH y acidez [11], [17], lo que confirma la eficacia de esta enzima como alternativa tecnológica. Sin embargo, en el caso de la ficina, los valores obtenidos [11], [18], quienes registraron un pH más elevado, lo cual puede atribuirse a las condiciones de proceso empleadas, como la concentración enzimática o la temperatura de coagulación. Estos resultados evidencian que, aunque los coagulantes vegetales son efectivos, su comportamiento depende de las condiciones experimentales aplicadas.

En cuanto a la sinéresis, solo el queso elaborado con ácidos fenólicos presentó diferencias significativas frente al control, mientras que los tratamientos con papaína y ficina no mostraron cambios estadísticamente relevantes. Resultados más altos [19], lo que se asocia al uso de diferentes tipos de coagu-

lantes y estabilizantes en sus experimentos. Respecto al contenido de materia seca, no se observaron diferencias significativas entre los tratamientos y el control, lo que coincide con los estudios realizados [20]. En relación con la materia grasa, solo el queso con ácidos fenólicos presentó una disminución significativa frente al control, mientras que los quesos con papaína y ficina se mantuvieron dentro de los rangos normales [21], [22][23].

De manera general, los resultados indican que los quesos elaborados con papaína y ficina difieren significativamente del queso control únicamente en dos parámetros (pH y acidez titulable), mientras que en los demás indicadores (sinéresis, materia seca y grasa) no se encontraron diferencias relevantes. Esto demuestra que los coagulantes vegetales, especialmente el látex de papaya serrana, pueden generar productos con características muy similares al queso obtenido con cuajo animal. Asimismo, las imágenes mostradas (Figura 4) corroboran que los quesos producidos con coagulantes de origen vegetal presentaron una apariencia comparable a la del queso control. Finalmente, todos los tratamientos evaluados cumplieron con los rangos establecidos en la Norma Técnica Peruana [24], lo que respalda la viabilidad de los coagulantes vegetales como alternativas en la elaboración de quesos frescos.

Los resultados de la evaluación sensorial evidencian que el látex de papaya serrana fue el coagulante vegetal más adecuado, ya que permitió obtener quesos con características de sabor, olor, color, textura y aceptabilidad global muy similares al queso control con cuajo animal. Esto coincide con lo reportado en estudios previos, donde la papaína ha demostrado buena capacidad coagulante y generación de productos con aceptación favorable por los consumidores debido a su suavidad y cremosidad [17], [25]. La buena puntuación en atributos sensoriales sugiere que este coagulante puede ser una alternativa viable para sustituir parcialmente al cuajo de origen animal en la industria quesera, respondiendo a la demanda de productos más naturales y accesibles.

En contraste, los quesos elaborados con higo y ortiga menor obtuvieron menor aceptación, especialmente el último, debido a sabores astringentes, olores herbales y texturas poco compactas. Este comportamiento puede estar asociado a la dosis de extracto utilizado, la cual pudo generar una mayor acidez y sinéresis, afectando la estructura del coágulo y, en consecuencia, la calidad sensorial del producto. Estudios [11], [19] coinciden en que el uso de coagulantes vegetales no optimizados puede alterar negativamente el perfil sensorial del queso, limitando su aceptación comercial. Sin embargo, también destacan que, con ajustes en concentración y condiciones de proceso, estos coagulantes pueden convertirse en alternativas sostenibles y funcionales en la elaboración de quesos artesanales.

Los parámetros fisicoquímicos de los quesos frescos elaborados con coagulantes de origen vegetal (papaína, ficina y ortiga menor) cumplieron con el rango permisible [24].

5 CONCLUSIONES

Con base en los datos obtenidos, los quesos con coagulan-
tes de origen vegetal presentaron características propias res-
pecto a los quesos frescos tradicionales, aunque con variacio-
nes en función del coagulante empleado. En el caso del látex
de papaya serrana, se observó una reducción en el porcentaje
de sinéresis (4,17%) y contenido de agua (51,19 g/100 g), un
incremento en los contenidos de materia seca (48,81 g/100 g)
y grasa (28,62 g/100 g)), en comparación con el queso elabo-
rado con coagulante de origen animal (renina en pastilla). El
uso de ramas y hojas de higo también influyó en los parámet-
ros fisicoquímicos, obteniéndose resultados dentro de los ran-
gos establecidos por la normativa, aunque inferiores al queso
control elaborado con renina en pastilla. Este tratamiento se
caracterizó por un menor pH (5,10), mayor acidez (0,10 g/100
g) y agua (58,99), mayor porcentaje de sinéresis (7,01%) reduc-
idos contenidos de materia seca (41,01 g/100g) y grasa (22,85
g/100 g). Finalmente, la aplicación de hojas de ortiga menor
mostró diferencias más marcadas, con valores de pH más ba-
jos (5,43), mayor acidez (0,15 g/100 g), sinéresis (8,89 %), agua
(63,07 g/100 g), y menores porcentajes de materia seca (36,93
g/100 g), y grasa (19,60 g/100 g) respecto tanto al queso con-
trol elaborado con renina en pastilla como a los demás trata-
mientos.

En cuanto a las propiedades sensoriales, cada coagulante veg-
etal aportó atributos diferenciados en sabor, olor, color, tex-
tura y aceptabilidad global. El queso elaborado con látex de
papaya serrana no presentó diferencias en sus características
con respecto al queso control elaborado con renina en pastilla,
alcanzando la mayor aceptación entre los consumidores. En
contraste, el queso elaborado con ortiga menor mostró pro-
piedades sensoriales diferentes que le otorgaron una aparien-
cia poco convencional y menos comparable con los quesos
tradicionales, lo que se reflejó en una menor preferencia por
parte de los evaluadores.

REFERENCIAS

- [1] A. M. Abdel-Aty, M. B. Hamed, W. H. Salama, M. M. Ali, A. S. Fahmy, and S. A. Mohamed, "Ficus carica, Ficus sycomorus and Euphorbia tirucalli latex extracts: Phytochemical screening, antioxidant and cytotoxic properties," *Bio-catal Agric Biotechnol*, vol. 20, p. 101199, Jul. 2019, <https://doi.org/10.1016/j.bcab.2019.101199>
- [2] F. D. Nicosia, I. Puglisi, A. Pino, C. Caggia, and C. L. Randazzo, "Plant Milk-Clotting Enzymes for Cheesemaking," *Foods*, vol. 11, no. 6, p. 871, Mar. 2022, <https://doi.org/10.3390/foods11060871>
- [3] J. S. Luna Cumapa, "Efecto de tres tipos de coagu-lantes (ácido cítrico, ascórbico y málico) en el ren-dimiento y características organolépticas del queso tipo ucayalino," 2018, Universidad Nacional Inter-cul-tural de la Ama-zonia, Amazonas - Perú.
- [4] C. Quispe Ramos, "Efecto de los cuajos naturales sobre el rendimiento, con-sistencia y color en la elaboración del queso fresco," 2019, Universidad Nacional José María Arguedas, Andahuaylas - Perú.
- [5] S. M. Quishpe Toapanta, "Obtención de un extracto vegetal de la Ortiga mayor (urtica dioica) aplicando el método de maceración en frío para la elaboración de Queso fresco," Universidad Técnica de Cotopaxi, 2019. [Online]. Available: <http://creativecommons.org/licenses/by-nc-sa/3.0/ec/>
- [6] A. Ben Amira, S. Besbes, H. Attia, and C. Blecker, "Milk-clotting properties of plant rennets and their enzymatic, rheological, and sensory role in cheese making: A review," *Int J Food Prop*, vol. 20, no. sup1, pp. S76-S93, Dec. 2017, <https://doi.org/10.1080/10942912.2017.1289959>
- [7] J. C. Jesus et al., "Biofunctional, chemical and sensory profile of mature cheese produced with artichoke flower coagulant (*Cynara cardunculus* L. var *scolymus*)," *Int Dairy J*, vol. 156, p. 105976, Sep. 2024, <https://doi.org/10.1016/j.idairyj.2024.105976>
- [8] M. Aider, "Potential applications of ficin in the production of traditional cheeses and protein hydrolysates," *JDS Communications*, vol. 2, no. 5, pp. 233-237, Sep. 2021, <https://doi.org/10.3168/jdsc.2020-0073>
- [9] A. Ben Amira, S. Besbes, H. Attia, and C. Blecker, "Milk-clotting properties of plant rennets and their enzymatic, rheological, and sensory role in cheese making: A review," *Int J Food Prop*, vol. 20, no. sup1, pp. S76-S93, Dec. 2017, <https://doi.org/10.1080/10942912.2017.1289959>
- [10] P. Walstra, J. T. M. Wouters, and T. J. Geurts, *Dairy Science and Technology*. Boca Raton: CRC Press, 2006. <https://doi.org/10.1201/9781420028010>
- [11] J. G. Bermeo Berrones, I. P. Salgado Tello, C. I. Flores Mancheno, and T. E. Sánchez Herrera, "Evaluación físico-química, microbiológica y sensorial del queso fresco utilizando ficina como biocatalizador," *ConcienciaDigital*, vol. 3, no. 2.1, pp. 231-249, May 2020, <https://doi.org/10.33262/concienciadigital.v3i2.1.1238>
- [12] Association of Official Analytical Chemists, *Official Methods of Analysis*, 18th ed. Gaithersburg, MD, USA: AOAC International, 2005.
- [13] Association of Official Analytical Chemists, *Official Methods of Analysis*, 17th ed. Gaithersburg, MD, USA: AOAC International, 2002.
- [14] I. N. de Calidad, "NTP 202.193:2020. Leche y productos lácteos. Queso. Identificación, clasificación y requisitos," 2020. [Online]. Available: <https://www.inacal.gob.pe>
- [15] A. of Official Analytical Chemists, "AOAC 920.39," Gaithersburg, MD, USA, 2004.
- [16] Asociación Española de Normalización, UNE-ISO 6658:2019. *Análisis sensorial de los Alimentos. Metodología general*. Madrid: AENOR, 2019.
- [17] L. Li, H. Chen, X. Lü, J. Gong, and G. Xiao, "Effects of papain concentration, coagulation temperature, and coagulation time on the properties of model soft cheese during ripening," *LWT*, vol. 161, p. 113404, May 2022, <https://doi.org/10.1016/j.lwt.2022.113404>
- [18] K. Liburdi, C. Boselli, G. Giangolini, S. Amatiste, and M. Esti, "An Evaluation of the Clotting Properties of Three Plant Rennets in the Milks of Different Animal Species," *Foods*, vol. 8, no. 12, p. 600, Nov. 2019, <https://doi.org/10.3390/foods8120600>
- [19] A. A. Ochoa Flores, J. A. Hernández Becerra, E. López Hernández, and H. S. García Galindo, "Yield, firmness and sensory acceptance of panela cheese with added stabilisers," *Universidad y Ciencia*, vol. 29, no. 3, pp. 277-286, 2013, [Online]. Available: <http://www.universidadyciencia.ujat.mx/>
- [20] C. I. Pereira, E. O. Gomes, A. M. P. Gomes, and F. X. Malcata, "Proteolysis in model Portuguese cheeses: Effects of rennet and starter culture," *Food Chem*, vol. 108, no. 3, pp. 862-868, Jun. 2008, <https://doi.org/10.1016/j.foodchem.2007.11.050>
- [21] R. S. Rivera-Botonaes, S. M. Oliva-Cruz, and D. Tineo Flores, "Extracción y purificación de papaína obtenida a partir de tres especies nativas del género *Vasconcellea*," *Revista de Investigaciones Altoandinas - Journal of High Andean Research*, vol. 25, no. 2, pp. 109-116, Apr. 2023, <https://doi.org/10.18271/ria.2023.447>
- [22] G. Villacrés Chavez et al., "Fresh cheese production using freeze-dried pa-pain as a vegetable coagulant," *Brazilian Journal of Food Technology*, vol. 27, 2024, <https://doi.org/10.1590/1981-6723.00624>
- [23] Mohamed Osman Mohamed Abdalla and Hanaa Mohammed Abbas Salih, "Effect of heat treatment of milk on the physicochemical, microbiological and sensory characteristics of white cheese (Gibna bayda)," *GSC Advanced Research and Reviews*, vol. 3, no. 3, pp. 020-028, Jun. 2020, <https://doi.org/10.30574/gscarr.2020.3.3.0044>
- [24] Instituto Nacional de Calidad, NTP 202.195:2019. *Leche y productos lácteos. Queso fresco. Requisitos*. in *Normas Técnicas Peruanas (NTP)*. Lima: IN-ACAL, 2019.
- [25] C. Fiol, D. Prado, M. Mora, and J. I. Alava, "Nettle cheese: Using nettle leaves (*Urtica dioica*) to coagulate milk in the fresh cheese making process," *Int J Gastron Food Sci*, vol. 4, pp. 19-24, Jul. 2016, <https://doi.org/10.1016/j.ijgfs.2016.05.001>