C&T Review Journal

Riqchary

Revista de Investigación en Ciencia y Tecnología

Volumen 06 - Número 1 -2024 ISSN: 2810-8124 (en línea)

ING. INFORMÁTICA Y SISTEMAS
UNAMBA

Online



Cube

Generado con Designer de Bing

DOI: https://doi.org/10.57166.riqchary.v6.n1.2024

UNIVERSIDAD NACIONAL MICAELA BASTIDAS DE APURÍMAC

VRIN

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 Arías Figueroa
  Micaela Bastidas Universidad Nacional de Apurímac, Perú
- Mag. Alejandrina Huaylla Quispe
  Universidad Tecnológica de los Andes Perú
- Mag. Betsabe M. Ccolqque Ruiz
  Micaela Bastidas Universidad Nacional de Apurímac, Perú
- Mag. Miguel Alfaro
  Micaela Bastidas Universidad Nacional de Apurímac, Perú

Periodicidad: Semestral
Vol. 6 Núm. 1 año (2024)
Áreas temáticas: Ciencias exactas, ingenierías y de materiales
Editorial: UNAMBA – Perú
Vice Rectorado de Investigación

https://doi.org/10.57166/riqchary/V6.n1.2024
ISSN: 2706-543X- **ISSN:** 2810-8124 (e)

Editorial

Con satisfacción presentamos esta edición dedicada al estudio del cube en inteligencia artificial, una herramienta clave para organizar y analizar grandes volúmenes de datos de manera eficiente. Su aplicación permite comprender información compleja desde distintas perspectivas y fortalecer la toma de decisiones en entornos tecnológicos.

En esta publicación se resalta la importancia de esta estructura dentro del desarrollo de sistemas inteligentes, promoviendo el uso de herramientas innovadoras orientadas al análisis de datos y la generación de conocimiento.

Esperamos que este trabajo motive a estudiantes e investigadores a seguir explorando el potencial de la inteligencia artificial como una disciplina capaz de aportar soluciones a los desafíos actuales de nuestra sociedad

Director: Dr. Ecler Mamani Vilca





C & T Riqchary

Revista de investigación - UNAMBA

Autores

- Smit Oscco Ccuiro
- Elias Huashuayo Miranda
- Mario Aquino Cruz
- Frank Michael Zuloaga Estacio
- Feliciano Escobedo Silva
- Neyka Escobedo Miranda
- Jinmer Bravo Apaza
- Anderson Nuñez Fernandez
- Flor de Cantuta Tello Sarmiento
- Manuel J. Ibarra Cabrera

- © 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 6 Número 1 – 2024
Se terminó de imprimir en El Misti EIRL
RUC 20450599108 - Cel: 950-509953, imprentamisti@hotmail.com

INDICE

13 *Pág.* → Modelo de Redes Neuronales Convolucionales para detectar enfermedades en las hojas del cultivo de Quinua (Chenopodium quinoa) en el Centro Agronómico K'ayra, San Jerónimo, Cusco 2023

20 *Pág.* → Estudio comparativo entre técnicas estadísticas multivariadas y redes neuronales artificiales para la optimización de la vigilancia de la calidad de agua para consumo humano en la red de salud Abancay 2022

31 *Pág.* → Globalización - Minería Ilegal - Pobreza, una relación Conflictiva en el Perú

37 *Pág.* → Evaluación de morteros con óxido de calcio y sulfato de calcio hidratado para uniones en mampostería

44 *Pág.* → Determinación del mejor modelo de Machine Learning para la predicción del California Bearing Ratio de suelos en Abancay, 2024

C&T Riqchary
Revista de Investigación en Ciencia y Tecnología
ISSN: 2810-8124 (en línea) / ISSN: 2706-543x

Vol. 6 Núm. 1 (2024)
Publicado: 22/03/24
https://doi.org/10.57166/riqchary.v6.n1.2024.117
Licencia: CC BY-NC-ND

Modelo de Redes Neuronales Convolucionales para detectar enfermedades en las hojas del cultivo de Quinua (Chenopodium quinoa) en el Centro Agronómico K'ayra, San Jerónimo, Cusco 2023

Convolutional Neural Network Model to detect diseases in the leaves of Quinoa (Chenopodium quinoa) crop at the Centro Agronómico K'ayra, San Jerónimo, Cusco 2023

Smit Oscco-Ccuiro[A], Elias Huashuayo-Miranda[B], Mario Aquino-Cruz[C].

ORCID: 0009-0009-9111-6124[A], ORCID: 0009-0006-6691-8756[B], ORCID: 0000-0002-2552-5669[C].

(Recepción: 15/04/2024 y aceptación 05/03/2024)

Resumen— En el mundo las enfermedades en los cultivos son la principal causa de reducción en la calidad de producción. Estas enfermedades afectan a los cultivos de quinua y cada año se producen una gran cantidad de pérdidas económicas. Es esencial identificar estas enfermedades a una etapa temprana para aumentar su producción. Una inspección visual es el método más común para identificar enfermedades, estos errores son comunes a través de la inspección visual. El tiempo es un factor clave para la detección de enfermedades y requiere de experiencia. Este estudio muestra cómo se puede utilizar el reconocimiento de imágenes para la detección de enfermedades. Este trabajo consistió en recolectar un conjunto de datos de imágenes para Mancha Foliar 1,120 imágenes, para Mancha Bacteriana 850 imágenes, para Mildiu 896 imágenes y 1,090 imágenes de sanas consta un total de 3,956 imágenes de las hojas de quinua del centro agronómico de K'ayra en el sector Leticia, San Jeronimo, Cusco del Perú, de los cuales se consideraron un 70% para entrenamiento, un 20% para validación y un 10% para pruebas. El modelo propuesto funcionó correctamente con una precisión 89.498%, lo cual permitirá a los agricultores de la quinua detectar tempranamente las enfermedades, esperando que conduzca a un aumento de la producción de la quinua en todo el mundo.

Palabra clave: Enfermedades de la quinua, Mancha Bacteriana, Mancha Foliar, Mildiu, Redes Neuronales Convolucionales (CNN).

Abstract— In the world, crop diseases are the main cause of reduction in production quality. These diseases affect quinoa crops and a large amount of economic losses occur each year. It is essential to identify these diseases at an early stage to increase production. A visual inspection is the most common method to identify diseases, these errors are common through visual inspection. Time is a key factor in disease detection and requires experience. This study shows how image recognition can be used for disease detection. This work consisted of collecting a data set of images for leaf spot 1,120 images, for bacterial spot 850 images, for downy mildew 896 images and 1,090 healthy images for a total of 3,956 images of quinoa leaves from the K'ayra agronomic center in the Leticia sector, San Jeronimo, Cusco, Peru, of which 70% were considered for training, 20% for validation and 10% for testing. The proposed model worked correctly with an accuracy of 89.498%, which will allow quinoa farmers to detect diseases early, hopefully leading to an increase in quinoa production worldwide.

Keyword: Quinoa diseases, Bacterial Spot, Leaf Spot, Downy Mildew, Convolutional Neural Networks (CNN).

A. *Smit Oscco Ccuiro, Escuela Académico Profesional de Ingeniería Informática y Sistemas de la Universidad Nacional Micaela Bastidas de Apurímac-Perú, smitosccoccuiro@gmail.com.*

B. *Elias Huashuayo Miranda, Escuela Académico Profesional de Ingeniería Informática y Sistemas de la Universidad Nacional Micaela Bastidas de Apurímac-Perú, huashuayo16@gmail.com.*

C. *Mario Aquino Cruz, .Escuela Profesional de Ingeniería Informática y Sistemas de la Universidad Nacional Micaela Bastidas de Apurímac-Perú, maquino@unamba.edu.pe.*

C&T Riqchary
Revista de investigación en ciencia y tecnología
Vol. 6 Núm. 1 (2024) - 22/03/2024
DOI https://doi.org/10.57166/riqchary.v6.n1.2024.117

1 INTRODUCCIÓN

Alrededor del mundo las enfermedades de los cultivos, son las que reducen la calidad en la producción; dentro de las causas más frecuentes para el cultivo agrícola son las plagas. Estas plagas se encuentran tanto en los tallos como en las hojas de cada planta.

Las enfermedades en los cultivos se han incrementado debido al desgaste de la capa de ozono de la tierra, a causa de la sobrepoblación que construyen las edificaciones en lugares de los cultivos [1,2]. El reconocimiento de enfermedades en los cultivos utilizando el método tradicional es ineficiente debido al elevado tiempo que se requiere para esta actividad. En tal sentido, es necesario sistemas de reconocimiento de enfermedades de precisión para la detección de enfermedades mediante las hojas en su etapa temprana, así de esta manera pueda proteger la calidad y cantidad del producto. Existe una variedad de plagas en la Quinua [3]. Las enfermedades más destacadas son: Mildiu, Mancha Foliar, Mancha Bacteriana [3,4,5], la calidad del producto se degrada bajo estas enfermedades. Según el ministerio de Agricultura INIA promueve acciones concretas para la quinua que exporta alrededor de 37,798 toneladas, sumando así sus ventas alrededor de US$ 94 millones [7], por otro lado, el mayor productor de esta especie en el Perú es el departamento de Puno, Cusco es uno de los departamentos promedio en cuanto en la producción de esta [8].

El método tradicional consiste en observar las plagas mediante el ojo humano, esto a su vez conlleva consultar a los expertos, sin embargo, el especialista en muchas ocasiones no está disponible a falta de tiempo, lo cual este proceso ocasiona un costo adicional al agricultor [9]. Es por ello que este método de identificar las plagas es ineficiente por las limitadas capacidades del ojo humano en grandes cantidades de sembrío de la quinua, al usar un modelo de Redes Neuronales Convolucionales para detectar enfermedades mediante las hojas de la Quinua ampliará investigaciones, así como en el campo de visión por computadoras y aprendizaje automático [10]. La mayoría de las enfermedades se da en las hojas de la quinua, por lo tanto, las hojas se convierten en una fuente portadora de información de las enfermedades que se encuentran en la quinua [8], así como se analiza [11], la detección temprana de enfermedades en los cultivos permitirá mejorar la calidad de este producto.

Los agricultores utilizan cada vez más modelos automatizados de diagnóstico de enfermedades debido a la llegada de las cámaras digitales y la tecnología informática. En los últimos tiempos, en diagnóstico de la gravedad de las enfermedades de las plantas se ha realizado mediante análisis automático basado en imágenes de aprendizaje profundo [12]. Sin embargo, el uso de técnicas de aprendizaje profundo para detectar enfermedades de las plantas quinua no es muy común en Cusco, a pesar de su uso generalizado en la agricultura inteligente a nivel mundial. En Alemania, se utilizó la arquitectura LeNet Convolucional Neural Netword (CNN) para clasificar enfermedades de la hoja de plantas como sigatoka y speckle [13].

En este estudio, propusimos una CNN para diagnosticar tres enfermedades frecuentes de la hoja de la quinua que incluyen Mancha Bacteriana, Mancha Foliar, Mildiu. El método propuesto CNN liviano para lograr mejor resultado.

Las principales aportaciones de este estudio son:

a) Conjunto de datos sobre las enfermedades de las manchas de las hojas de la quinua: recolectamos un conjunto de datos que incluye imágenes de hojas de la quinua con tres enfermedades principales de la hoja de quinua, incluidas Mancha Bacteriana, Macha Foliar y Mildiu, así como las hojas sanas el conjunto de datos se recopilo del Centro Agronómico K'ayra de la Universidad Nacional San Antonio Abad de Cusco.

b) Aplicación para la identificación de enfermedades de las hojas de la quinua: Desarrollamos una aplicación que pueda diagnosticar enfermedades de las hojas de la quinua a partir de imágenes reales de las hojas de la quinua. La aplicación permitirá a los agricultores tomar medidas de prevención para evitar la pérdida del rendimiento de los frutos de la quinua al detectar la propagación de enfermedades.

1.1. Trabajos relacionados

Introdujeron un sistema modificado de DLCNN basado en ResNet para identificar y clasificar enfermedades en las plantas del arroz utilizando imágenes de la base de datos de PlantVillage de las hojas del arroz, utilizaron ResNet-50 para entrenar múltiples características de imágenes con propiedades de textura, utilizaron MRDO para seleccionar características optimizadas tamaño de salida, a la vez utilizaron para el modelo de clasificar. Los resultados fueron en la comparación F1 del 99,73% y el 99,78%, para PlantVillage y del 99,68% y el 99.71% para Rice Plant [11,13].

Propusieron utilizar aprendizaje profundo y optimización bayesiana para detectar enfermedades de la hoja del plátano a partir de imágenes, para la clasificación, aplicaron BananaSqueezeNet, los resultados fueron satisfactorios con una precisión de 95,13% [15]. Propusieron utilizar redes residuales y mecanismos de atención de enfermedades a través de las hojas del arroz a partir de imágenes, para clasificación aplicaron GoogleNet con una precisión de 99,58 con una mejora de 1,72% con la evaluación de E-GoogleNet

[16].

Propuso utilizar red neuronal convolucional eficiente con aumento y reguladores de clasificación de plagas a través de imágenes, para la clasificación aplicaron Deep CNN con una precisión de 71,32% [17].

Propuso un aplicativo web para identificar deficiencia de Zinc de naranjo mediante las hojas a través de imágenes, para clasificación propusieron Redes Neuronales Convolucionales CGG16 con una precisión de 81,12% [18].

Propusieron dos ramas integradas ResNet y CBAM, para la clasificación de enfermedades mediante las hojas de la manzana, la experimentación se realizó con 4479 logrando así una precisión de 95.2% y 97.2%[10].

2 MATERIALES Y METODOLOGÍA

2.1. Materiales

- Cámara fotográfica canon de 16.0 megapíxeles.
- Micro USB para almacenamiento de la imagen.
- USB Kingston de 32GB.
- Computadora de marca Asus con una tarjeta gráfica NVIDIA GeForce de 4GB con un sistema operativo Ubuntu 20.04.
- Registro de imágenes de la quinua.
- Archivos con una extensión *.py.

2.2. Tipo y nivel de investigación

La investigación de tipo aplicada con un nivel descriptivo.

2.3. Modelo propuesto

En la figura 1 se muestra el modelo propuesto de nuestro trabajo para el procedimiento de la investigación.

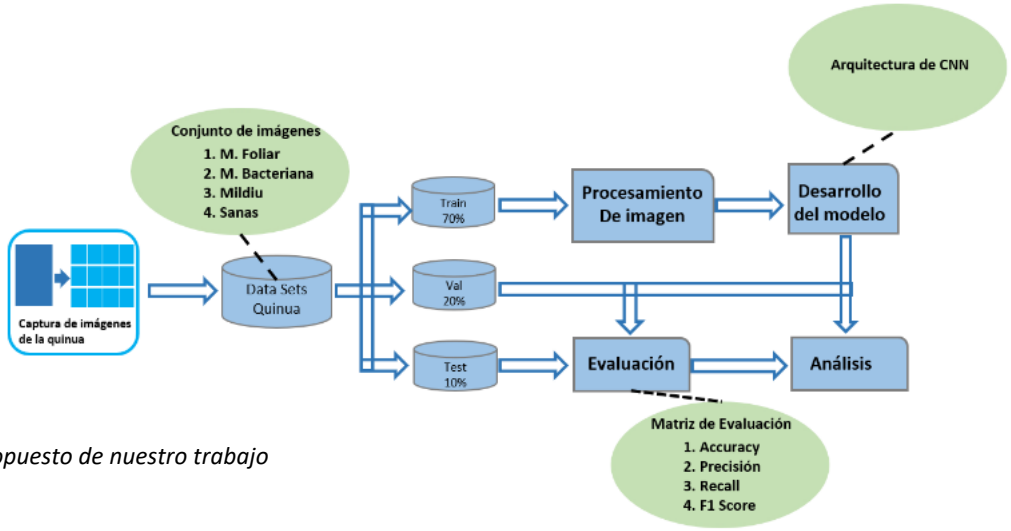


Fig. 1. Modelo Propuesto de nuestro trabajo

a. Recolección de imágenes.
b. Etiquetado y limpieza de imágenes.
c. Reducción de imágenes.
d. División y distribución.
e. Entrenamiento de reconocimiento de enfermedades en la quinua utilizando Redes Neuronales Convolucionales (CNN).
f. Pruebas.

2.4. Procedimiento

Primero se realizó la recolección de datos el cual consta de imágenes de hojas de la quinua de cuatro categorías. Estas son hojas afectadas de Mildiu, hojas afectadas de Mancha Foliar, hojas afectadas de Mancha Bacteriana y hojas Sanas, las imágenes se obtuvieron del Centro Agronómico K'ayra de la Universidad Nacional San Antonio Abad de Cusco. Se utilizó una cámara fotográfica digital para tomar las imágenes. Existe un total de 3,974 imágenes de 4 clases: sanas, mildiu, mancha bacteriana y mancha foliar.

Segundo se etiquetó cuidadosamente las imágenes, con ayuda de un científico de datos se realizó la limpieza de datos para eliminar las imágenes inconsistentes.

Tercero se preprocesaron las imágenes utilizando diferentes técnicas de preprocesamiento que incluyen cambio de tamaño y de normalización. Las imágenes constan de 3 canales de tamaño 238 x 246 píxeles sin perder la calidad de cada imagen, para ello se utilizó la siguiente fórmula.

$$V_{flip}(m,n) = V_I\left(m,\left(n_{\max} - n\right)+1\right)$$

Donde V_{flip} es la matriz de entrada (output) de la imagen que corresponde a la transformación de que produce un movimiento de la columna m, filas n a la columna m y fila $(n_{max} - n) + 1$, para n_{max} de la dirección de n.

Cuarto se dividió el conjunto de datos en pruebas, validación y entrenamiento, y se realizó la distribución de las imágenes para entrenamiento el 70%, para validación el 20% y para pruebas el 10%. Para el conjunto de entrenamiento se contabilizaron 763 imágenes de hojas sanas, 627 imágenes de hojas afectadas por mildiu, 595 imágenes de hojas afectadas por la enfermedad mancha bacteriana y 784 imágenes de hojas afectadas por la enfermedad de mancha foliar. Para el conjunto de validación 218 imágenes de hojas sanas, 179 imágenes de hojas afectadas por la enfermedad mildiu, 170 imágenes de hojas afectadas por la enfermedad de mancha bacteriana y 224 imágenes de hojas afectadas de la enfermedad de mancha foliar. El conjunto de prueba tiene 109 imágenes de hojas sanas, 90 imágenes de hojas afectadas por la enfermedad de mildiu, 85 imágenes de hojas afectadas por la enfermedad de mancha bacteriana y 112 imágenes de hojas afectadas por la enfermedad de mancha foliar, como se muestra en la Figura 2 el cual ilustra el conjunto que conforman imágenes.

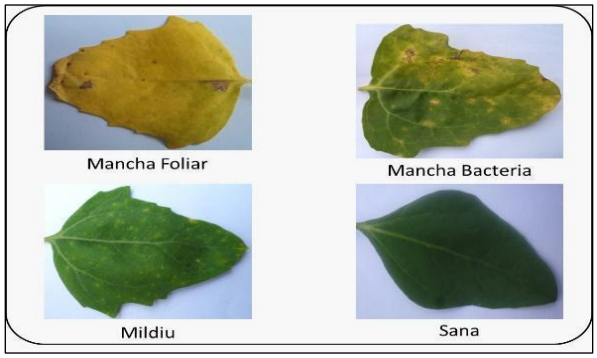


Fig. 2. Conjunto de datos de imágenes de la Quinua.

Estas son hojas sanas, hojas afectadas por mancha foliar, hojas afectadas por mancha bacteriana y hojas afectadas de mildiu.

La tabla 1 ilustra la cantidad de imágenes de las diferentes clases en el conjunto de entrenamiento, validación y prueba.

Tabla 1. Distribución de clases de imágenes en el conjunto de datos de Quinua

Clase	Conjunto de datos		
	Train	Val	Test
Mildiu	627	179	90
M-Bacteriana	595	170	85
M-Foliar	784	224	112
Saludables	763	218	109
Total	2769	791	396

Quinto se realizó el entrenamiento en Python usando PyTorch [19], el entrenamiento se desarrolló en una computadora de 64 bit con un sistema operativo Ubuntu 20.04. En el entrenamiento se utilizó 40 épocas y demoró alrededor de 3:30 Hrs (tres horas con treinta minutos), se utilizaron las siguientes herramientas y librerías con sus distintas versiones: Python==3.8, Numpy==1.24, Pandas==2.0, Matplotlib==3.7, Scikit-Learn==1.3, Scipy==1.10, Torch==1.12+cu116, Torchvision==0.12+cu116, Zipp==3.17, OpenCV-Python==4.8.

La arquitectura generada tiene cuatro capas convolucionales. Y que cada capa subsiguiente tiene el doble de número anterior, cada una de las capas convolucionales va seguida de la actividad de ReLU y Max-Pool, tenemos una capa lineal con cuatro números de clases en el conjunto de datos, las primeras capas convolucionales tienen núcleos de 5x5, luego el siguiente de 3x3 y la última nuevamente de 5x5 de núcleos.

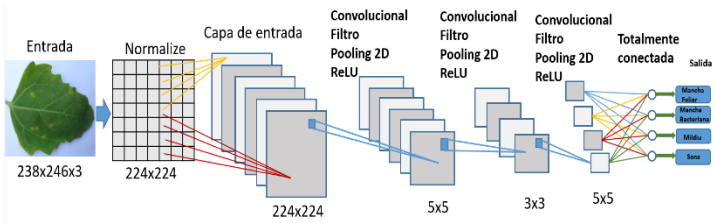


Fig. 3. Arquitectura de Modelo de Red Neuronal Convolucional.

De la figura 3, Se utilizó la red neuronal convolucional (CNN) para realizar la tarea de clasificación. CNN tiene dos tipos de capas: capas de extracción de características y capas de clasificación. Las capas de extracción de características extraen las características de la imagen, mientras que la capa de clasificación realiza la clasificación de la imagen en función de las características extraídas. Tamaño del modelo CNN debe ser pequeña ya que pretendemos implementar identificar para distintos dispositivos para que los agricultores puedan usarlo en el campo. Por esta razón, entrenamos por el modelo de CNN liviano [20]. El beneficio de utilizar este modelo liviano es que se pueden implementar en dispositivos integrados, ya que son de tamaño muy pequeño.

Por último, se realizaron las pruebas con las imágenes separadas para este proceso.

2.5. Medición de desempeño

Se han utilizado varias métricas de rendimiento para medir el rendimiento de nuestro modelo, incluida la exactitud, la precisión, la recuperación, la puntuación. Se utilizaron las siguientes ecuaciones para calcular la exactitud,

la precisión, la recuperación y la puntuación.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$F1 = 2x\frac{Precision * Recall}{Precision + Recall}$$

La cantidad de muestras verdaderos positivos, verdaderos negativos, falsos positivos y falsos negativos están representados por TP, TN, FP y FN, respectivamente.

3 RESULTADOS

3.1. Resultados de evaluación

En la figura 4 se muestra la matriz de confusión para nuestro modelo. Las imágenes sanas fueron clasificadas con un 89.92% de precisión, tres imágenes de mildiu y cinco de mancha bacteriana fueron clasificadas erróneamente como sanas. Las imágenes de mildiu se clasificaron al 80.03% de precisión, una imagen como mancha foliar y una como mancha bacteriana fueron clasificadas erróneamente. Las sanas también tienen la tasa promedio de falsos positivos. El modelo tuvo problemas para clasificar las imágenes de mancha foliar.

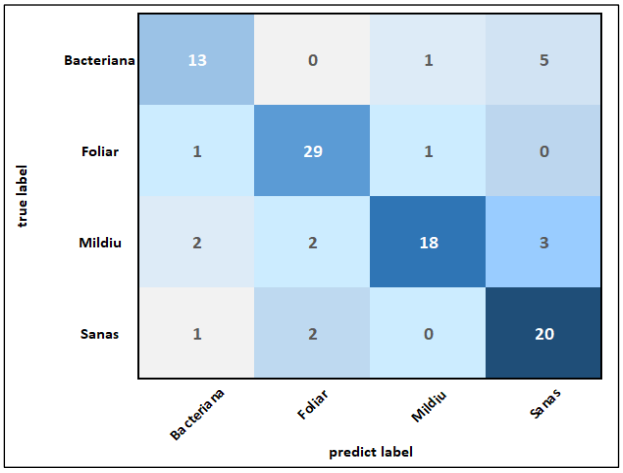


Fig. 4. Matriz de confusión del conjunto de datos de entrenamiento

El entrenamiento de nuestra red neuronal convolucional muestra parámetros que impactan significativamente en el rendimiento del modelo. Los optimizadores determinan cómo actualizar los pesos del modelo durante el entrenamiento, mientras que la función de pérdida determina cómo se penalizará el modelo cuando se realicen predicciones incorrectas. La cantidad de épocas determina la cantidad de veces que el modelo encontrara todo el conjunto de datos de entrenamiento. En este trabajo utilizamos el optimizador Adam, el número de épocas se estableció en 40. Durante los experimentos se utilizó aprendizaje por transferencia para mejorar el entrenamiento en el rendimiento.

La figura 5 muestra el cambio de precisión con el número de épocas para nuestro modelo de arquitectura. A medida que se aumenta el número de épocas, la precisión no aumenta pues tiene una curva de aprendizaje accidentada.

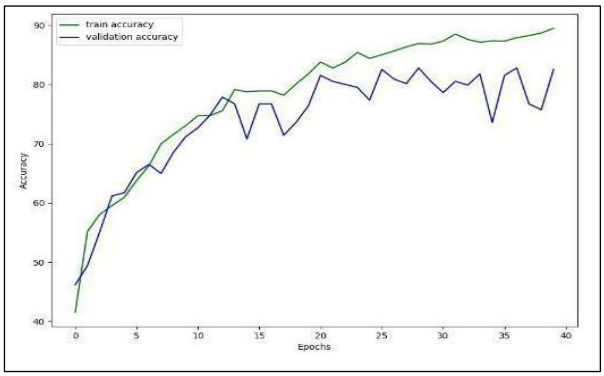


Fig. 5. Análisis de Precisión del Modelo Propuesto

En la figura 6 muestra la pérdida, disminuye hasta alcanzar la saturación para el modelo por lo que logró una curva de aprendizaje más fluida.

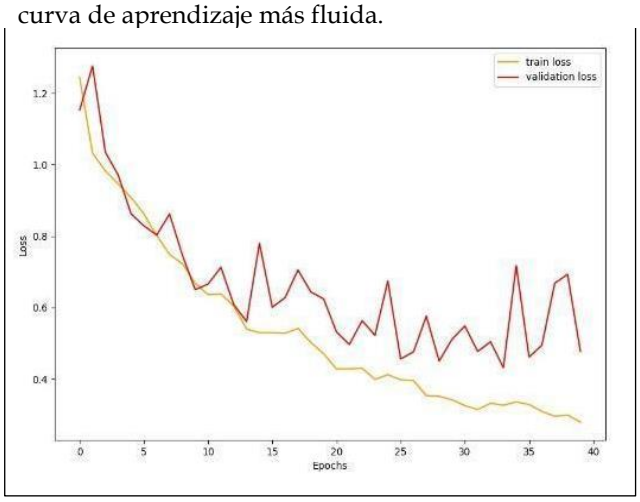


Fig. 6. Análisis de pérdida del modelo propuesto.

La tabla 2 muestra la matriz de confusión para las enfermedades de la quinua en la cual se observa los resultados en nuestro modelo propuesto de entrenamiento CNN, con un accuracy bastante viable. En la figura 6 muestra los resultados de la predicción en las enfermedades de la

quinua, en la que se puede observar que nuestro modelo propuesto obtiene un valor de 89.50.

Tabla 2 Matriz de confusión para enfermedades de la quinua

	Precisión	Recall	F1-Score	Support
0	0.87	0.81	0.84	16
1	0.91	0.73	0.80	20
2	0.64	1.00	0.78	9
3	0.82	0.88	0.82	12
Accuracy			89.50	40

La matriz de confusión del modelo propuesto, presenta una precisión por clase que se encuentra entre 89.50%.

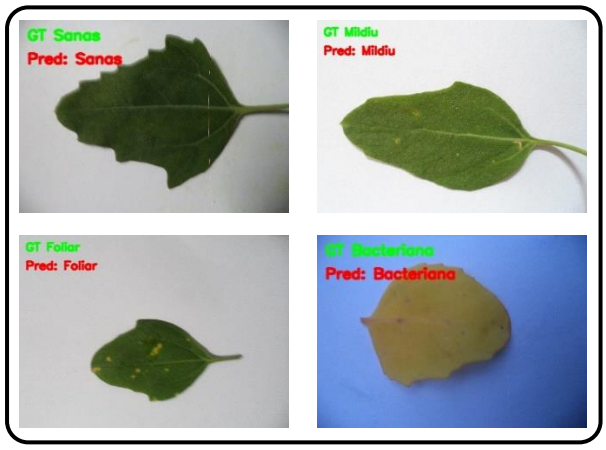


Fig. 7. Predicción para las enfermedades de la quinua.

4 Conclusión y Recomendaciones

4.1. Conclusión

En este trabajo, se demostró cómo se puede detectar a partir de imágenes diferentes enfermedades que afectan la hoja, y por consiguiente el fruto de la quinua. Recopilamos un conjunto de datos de imágenes de hojas de quinua del Centro Agronómico K'ayra de la Universidad Nacional San Antonio Abad de Cusco, se implementó una arquitectura de CNN que puede identificar enfermedades a partir de imágenes de hojas de la quinua con una alta precisión de 89.50%.

4.2. Recomendaciones

Este modelo se puede implementar en una plataforma web para el libre acceso de los agricultores y partes interesadas, para que puedan utilizar el modelo para realizar el diagnóstico temprano de las enfermedades de la hoja de la quinua y para tomar medidas de precaución.

En futuras investigaciones se mejore la eficacia del modelo para identificar enfermedades en las hojas de la quinua. También se debe instruir a los agricultores sobre qué medidas tomar en función de la gravedad de la enfermedad.

Agradecimientos

Los reconocimientos a la Universidad Nacional San Antonio Abad del Cusco, Facultad de Agronomía y Zootecnia por facilitarnos acceso a los cultivos de su centro Agronómico en el sector Leticia de K'ayra San Jeronimo para la recopilación de las hojas de quinua para poder conformar nuestro DataSets, el cual se usó para el entrenamiento de nuestra red neuronal y posteriormente para las pruebas.

Los reconocimientos a nuestros Co-Asesores: Julio Cesar Llo-clli Champi, guiándonos en sus conocimientos en modelos de clasificación de nuestra DataSets con redes neuronales, a Leonardo Ccuyro Quispe por brindarnos su conocimiento en enfermedades de las hojas de quinua especialista en plagas en las plantas.

Referencias

[1] H. Duan, T. R. Miller, G. Liu, and V. W. Y. Tam, "Construction debris becomes growing concern of growing cities," Waste Manag., vol. 83, pp. 1–5, Jan. 2019, doi: 10.1016/J.WASMAN.2018.10.044.

[2] M. Bravo, M., Reyna R., J., Gómez Sánchez, l., & Huapaya H., "ESTUDIO QUÍMICO Y NUTRICIONAL DE GRANOS ANDINOS GERMINADOS DE QUINUA (CHENOPODIUM QUINOA) Y KIWICHA (AMARANTUS CAUDATUS)." p. 16, 2013. [Online]. Available: https://revistasinvestigacion.unmsm.edu.pe/index.php/quim/article/view/6558

[3] "INIAP -Estación Experimental Santa Catalina," 16-18, p. 8, 1985, [Online]. Available: http://181.112.143.123/bitstream/41000/2827/1/iniapsc322est.pdf

[4] D. Solveig and T. Ames, "EL MILDIU (Peronospora farinosa) DE LA QUINUA (Chenopodium quinoa) en la zona andina. Manual práctico para el estudio de la enfermedad y del patógeno," Cent. Int. la papa, p. 38, 2000, [Online]. Available: http://cipotato.org/wp-content/uploads/2014/10/AN60198.pdf

[5] G. McAvoy, "Una solución para la Mancha foliar por Alternaria," 2017. https://www.hortalizas.com/author/gmcavoy/

[6] M. F. Yañez-Yazlle, N. Romano-Armada, M. M. Acreche, V. B. Rajal, and V. P. Irazusta, "Halotolerant bacteria isolated from extreme environments induce seed germination and growth of chia (Salvia hispanica L.) and quinoa

(Chenopodium quinoa Willd.) under saline stress," Ecotoxicol. Environ. Saf., vol. 218, p. 112273, Jul. 2021, doi: 10.1016/J.ECOENV.2021.112273.

[7] Fairlie.A, La quinua en el perú. 2016. [Online]. Available: http://repositorio.pucp.edu.pe/index/bitstream/handle/123456789/54092/Nro_6_Fairlie_quinua_Perú.pdf?sequence=1&isAllowed=y

[8] M. de D. A. y Riego, "El Perú se consolida como el primer productor y exportador mundial de quinua," 2021. https://www.gob.pe/institucion/midagri/noticias/324394-el-peru-se-consolida-como-el-primer-productor-y-exportador-mundial-de-quinua

[9] E. Efraín and J. Lepe, "CENTRO DE INVESTIGACIÓN Y DE ESTUDIOS SUPERIORES DEL IPN A beginner ' s tutorial for CNN," pp. 1–35, 2017.

[10] R. Ding et al., "Improved ResNet Based Apple Leaf Diseases Identification," IFAC-PapersOnLine, vol. 55, no. 32, pp. 78–82, 2022, doi: 10.1016/j.ifacol.2022.11.118.

[11] S. R. G. Reddy, G. P. S. Varma, and R. L. Davuluri, "Resnet-based modified red deer optimization with DLCNN classifier for plant disease identification and classification," Comput. Electr. Eng., vol. 105, no. November 2022, p. 108492, 2023, doi: 10.1016/j.compeleceng.2022.108492.

[12] G. Wang, Y. Sun, and J. Wang, "Automatic Image-Based Plant Disease Severity Estimation Using Deep Learning," Comput. Intell. Neurosci., vol. 2017, 2017, doi: 10.1155/2017/2917536.

[13] J. Amara, B. Bouaziz, and A. Algergawy, "A deep learning-based approach for banana leaf diseases classification," Lect. Notes Informatics (LNI), Proc. - Ser. Gesellschaft fur Inform., vol. 266, pp. 79–88, 2017.

[14] R. E. Castañeda Valdivieso, J. R. Guerrero Meza, B. E. Renteros Parra, and J. A. Villanueva Mejía, "Detección de nutrientes del suelo y planta, y pestes en campos de cultivo de banano orgánico con Machine Learning," Pirhua, 2021, [Online]. Available: https://hdl.handle.net/11042/5204

[15] M. A. B. Bhuiyan, H. M. Abdullah, S. E. Arman, S. Saminur Rahman, and K. Al Mahmud, "BananaSqueezeNet: A very fast, lightweight convolutional neural network for the diagnosis of three prominent banana leaf diseases," Smart Agric. Technol., vol. 4, no. February, p. 100214, 2023, doi: 10.1016/j.atech.2023.100214.

[16] L. Yang et al., "GoogLeNet based on residual network and attention mechanism identification of rice leaf diseases," Comput. Electron. Agric., vol. 204, p. 107543, Jan. 2023, doi: 10.1016/J.COMPAG.2022.107543.

[17] A. Setiawan, N. Yudistira, and R. C. Wihandika, "Large scale pest classification using efficient Convolutional Neural Network with augmentation and regularizers," Comput. Electron. Agric., vol. 200, p. 107204, Sep. 2022, doi: 10.1016/J.COMPAG.2022.107204.

[18] F. D. E. I. Y. Arquitectura, "Shiane Lizceth Farfan Vergara," 2021.

[19] David Beazley and B. K. Jones, Python Cookbook: 3rd Edition. 2013.

[20] F. Coban, "The Role of the Media in International Relations: From the CNN Effect to the Al –Jazeere Effect," J. Int. Relations Foreign Policy, vol. 4, no. 2, pp. 45–61, 2016, doi: 10.15640/jirfp.v4n2a3.

BIOGRAFÍA

Smit Oscco Ccuiro, bachiller en Ingeniería Informática y Sistemas de la Universidad Nacional Micaela Bastidas de Apurímac.

Elias Huashuayo Miranda, bachiller en Ingeniería Informática y Sistemas de la Universidad Nacional Micaela Bastidas de Apurímac.

Mario Aquino Cruz, Docente en la Universidad Nacional Micaela Bastidas de Apurímac – Perú, MSc. In Informática, Investigador en las áreas de informática educativa, IoT, Inteligencia artificial y ciberseguridad.

C&T Riqchary
Revista de Investigación en Ciencia y Tecnología
ISSN: 2810-8124 (en línea) / ISSN: 2706-543x

Vol. 6 Núm. 1 (2024)
Publicado: 22/03/24
https://doi.org/10.57166/riqchary.v6.n1.2024.118
Licencia: CC BY-NC-ND

Estudio comparativo entre técnicas estadísticas multivariadas y redes neuronales artificiales para la optimización de la vigilancia de la calidad de agua para consumo humano en la red de salud Abancay 2022

Comparative study between multivariate statistical techniques and artificial neural networks for the optimization of the surveillance of water quality for human consumption in the Abancay health network 2022

Frank Michael Zuloaga Estacio[A], Mario Aquino Cruz[B]

ORCID:0009-0003-8946-5030[A], ORCID:0000-0002-2552-5669[B]

(Recepción: 13/02/2024 y aceptación 10/03/2024)

Resumen— Actualmente en la mayoría de instituciones, incluyendo la red de salud Abancay, se utiliza estadística tradicional con el fin de determinar tendencias sobre un conjunto de datos que se centra en una sola variable, es complicado aplicar este tipo de análisis a conjunto de datos multivariados, que son los que usualmente se obtienen en los programas de control de calidad de agua y, que excluye diferencias entre las variables analizadas y sus relaciones. El objetivo del estudio es realizar una comparación entre las diferentes técnicas estadísticas multivariables y redes neuronales artificiales, con la finalidad de relacionar y clasificar las variables. Para realizar esto se eligió dos técnicas de estadística multivariada, análisis de componentes principales (ACP) y análisis discriminante (AD) y dos tipos de redes Neuronales artificiales, de aprendizaje no supervisado, hebbiano (RNAH), y de aprendizaje supervisado, perceptrón multicapa (RNAPM), el tipo de investigación que se usará en el estudio será investigación aplicada de enfoque cuantitativo, con un nivel de investigación explicativo y con un diseño transversal. Dada la comparación entre el análisis de componentes principales y la red neuronal artificial de tipo Hebbiano, se obtuvo que las redes neuronales pudieron asociar mejor las variables que el análisis de componentes principales. En la segunda comparación entre el análisis discriminante y la red neuronal artificial perceptrón multicapa los resultados fueron buenos para el análisis discriminante debido a que obtuvo un 95% de clasificación correcta, mientras que la red neuronal artificial obtuvo un 74.6%, sin embargo, debido a las limitaciones del análisis discriminante, se infiere que la red neuronal artificial perceptrón multicapa es un mejor modelo a escoger.

Palabra clave: análisis de componentes principales, análisis discriminante, red neuronal artificial de tipo hebbiano, red neuronal artificial perceptrón multicapa.

Abstract— Currently, in most institutions, including the Abancay health network, traditional statistics are used in order to determine trends on a data set that focuses on a single variable, it is complicated to apply this type of analysis to multivariate data sets, which are those usually obtained in water quality control programs, and which exclude differences between the variables analyzed and their relationships. The objective of the study was to make a comparison between different multivariate statistical techniques and artificial neural networks, in order to relate and classify variables. To do this, two multivariate statistical techniques were chosen, principal component analysis (PCA) and discriminant analysis (DA) and two types of artificial neural networks, unsupervised learning, hebbian (RNAH), and supervised learning, multilayer perceptron (RNAPM), the type of research I used in the study will be applied research of quantitative approach, with an explanatory level of research and with a cross-sectional design. Given the comparison between the principal component analysis and the Hebbian type artificial neural network, I obtained that the neural networks were able to associate the variables better than the principal component analysis. In the second comparison between the discriminant analysis and the multilayer perceptron artificial neural network, the results were good for the discriminant analysis because it obtained 95% of correct classification, while the artificial neural network obtained 74.6%, however, due to the limitations of the discriminant analysis, I inferred that the multilayer perceptron artificial neural network is a better model to choose.

Keyword: discriminant analysis, hebbian-type artificial neural network, multilayer perceptron artificial neural network, principal component analysis.

A. *Frank Michael Zuloaga Estacio[A], Escuela Profesional de Ingeniería Informática y Sistemas de la Universidad Nacional Micaela Bastidas de Apurímac-Perú, frankzuloaga3@gmail.com*

B. *Mario Aquino Cruz[B] Escuela Profesional de Ingeniería Informática y Sistemas de la Universidad Nacional Micaela Bastidas de Apurímac-Perú, maquino@unamba.edu.pe*

1 INTRODUCCIÓN

El análisis de componentes principales es una técnica estadística general que reduce una tabla de datos variables a sus características básicas, llamadas componentes principales. Los componentes principales son ciertas combinaciones lineales de las variables originales que explican la varianza máxima de todas las variables. [1]. En el proceso, el método proporciona una aproximación de la tabla de datos original utilizando solo estos pocos componentes principales.

Su objetivo es extraer información importante de datos estadísticos y representarla como un nuevo conjunto de variables ortogonales llamadas componentes principales y mostrar el patrón de similitudes entre observaciones y variables en forma de puntos en un mapa de puntos. Matemáticamente, está determinado por vectores propios y valores propios. Los vectores propios y los valores propios son números y vectores asociados con matrices cuadradas. Juntos proporcionan una descomposición matricial adecuada para analizar la estructura de esa matriz, como una matriz de correlación, covarianza o producto. [2]. Si bien, se han listado diferentes tipos de matrices para el estudio se eligió la matriz de correlación.

El análisis de función discriminante es un análisis estadístico que se utiliza para analizar datos cuando la variable dependiente o el resultado es categórico y la variable independiente o predictor es paramétrico. Este es un método paramétrico para determinar qué pesos de variables continuas o predictores diferencian mejor dos o más tipos de variables dependientes y lo hacen mejor que el azar. El análisis discriminante se utiliza para determinar la precisión de un sistema de clasificación determinado al predecir una muestra dentro de un grupo en particular. El análisis discriminante implica desarrollar funciones discriminantes para cada muestra y obtener una puntuación umbral utilizada para clasificar la muestra en diferentes grupos. [3].

Los algoritmos de aprendizaje no supervisado de carácter hebbiano se basan en el siguiente postulado, formulado por Donald O. Hebb en 1949: "Cuando un axón de una celda A esta suficientemente cerca para conseguir excitar una celda B y repetida o persistentemente toma parte en su activación, algún proceso de crecimiento o cambio metabólico tiene lugar en una o ambas celdas, de tal forma que la eficiencia de A, cuando la celta activa es B, aumenta". De esta forma, identificando las celdas con neuronas fuertemente conectadas [4]. La regla de Hebb es de tipo no supervisado, pues la modificación de los pesos depende de los estados (salidas) de las neuronas obtenidos tras la presentación de un estímulo determinado, con independencia de que coincidan o no con las deseadas. De esta forma, en el aprendizaje hebbiano múltiples neuronas de salida pueden activarse simultáneamente.

La arquitectura de un perceptrón multicapa es variable, pero en general consiste en varias capas de neuronas, puede tener una o más capas ocultas y finalmente una capa de salida. El perceptrón multicapa se ha aplicado a una amplia variedad de tareas, todas las cuales pueden clasificarse como predicción, aproximación de funciones o clasificación de patrones. La predicción implica pronosticar tendencias futuras en una serie temporal de datos dadas las condiciones actuales y anteriores. La aproximación de funciones se ocupa de modelar la relación entre variables. La clasificación de patrones implica clasificar datos en clases discretas. Todas estas aplicaciones están estrechamente relacionadas [5].

Este artículo presenta los resultados de la aplicación y comparación entre el análisis de componentes principales con la red neuronal de tipo hebbiano y el análisis discriminante con una red neuronal perceptrón multicapa. Los datos analizados corresponden a las variables de calidad de agua correspondientes a cloro, pH, conductividad, turbiedad, debido a que solo estas estaban presentes en los datos obtenidos, además se añadió una variable periodo para realizar la comparación entre el análisis discriminante y la red neuronal perceptrón multicapa debido a que la predicción de datos no se puede realizar si no existe una variable de salida, la cual se obtuvo al analizar los datos de precipitación en la zona durante el mes de enero a diciembre desde 1991 al 2021.

2 METODOLOGÍA

El tipo de investigación que se realiza en el presente estudio es una investigación aplicativa con un enfoque cuantitativo, el nivel de investigación será explicativo y su diseño es transversal.

2.1 ASOCIACIÓN DE VARIABLES

Para determinar la asociación entre las variables fisicoquímicas que hacen parte de los datos que se adquieren comúnmente en los programas de monitoreo de calidad de agua se emplearon técnicas de análisis de componentes principales y las redes neuronales artificiales de tipo Hebbiano.

2.1.1 ANÁLISIS DE COMPONENTES PRINCIPALES

Inicialmente se elaboró histogramas para poder observar si la distribución presentada por las variables se trataba de una distribución simétrica o normal o si se requería el uso de ciertos parámetros para evitar que los resultados se distorsionen por ejemplo haciendo uso de la transformación

logarítmica para disminuir su rango y posteriormente realizar una estandarización de los datos como paso preliminar para la determinación de los componentes principales.

En este caso se observó que ninguna de las variables tenía una distribución normal (cloro, conductividad, pH, turbiedad)

Como se puede observar en el valor de significancia ninguna de las variables presento una distribución normal usando el método de Kolmogorov-Smirnov debido a la cantidad de datos, por lo cual se realizó transformación logarítmica a todas las variables para evitar que los datos distorsionen los resultados del análisis estadístico.

Realizada la transformación logarítmica. Se procedo a estandarizar los datos previos a la determinación de los componentes principales (CP). El análisis de CP arrojo como resultados un conjunto de vectores propios que corresponden a un nuevo espacio en donde se proyectan los datos originales. Con estos, se pudo calcular las coordenadas de los datos en este nuevo espacio, para efecto de visualización y para determinar asociaciones de variables presentes en los datos. Esto último se pudo estudiar mediante la matriz de coeficientes de correlación entre los CP y las variables originales.

En la figura 1. podemos observar que no existe una pendiente aparente por lo cual se pueden utilizar todos los componentes para el estudio.



Fig. 1. Gráfico de Sedimentación de los componentes principales

Así mismo, con el fin de determinar las asociaciones entre las variables fisicoquímicas estudiadas se calculó la matriz de coeficientes de correlación entre los CP a utilizar en el estudio y las variables originales.

2.1.2 Red neuronal artificial de tipo hebbiano

Todo el procedimiento computacional de las redes se realizó con las funciones diseñadas en Matlab.

El tipo de red neuronal utilizada fue la red neuronal de tipo hebbiano con entrenamiento hebbiano generalizado (aprendizaje no supervisado)

El procedimiento seguido para determinar los componentes principales con RNAH fue el siguiente.

1. Estandarización de los datos en SPSS
2. Cargar los datos estandarizados en Matlab.
3. Se realizo el entrenamiento de red con los siguientes parámetros.
4. 4 variables de entrada
5. 4 variables de salida
6. La tasa de aprendizaje fue igual a $1x10^{-6}$
7. Numero de épocas de entrenamiento igual a 1000

A esta red no se le determino el coeficiente de correlación, puesto que es una red de entrenamiento no supervisado, y no se contó con una salida deseada con la cual se pudiera comparar.

2.2 Clasificación de variables

Para realizar este tipo de análisis usando las funciones discriminantes y las RNAPM se hizo necesario etiquetar las muestras para que los métodos aplicados pudiesen predecir la pertenencia de una variable o parámetro a un determinado grupo. Los grupos establecidos fueron dos: seco y húmedo, de acuerdo a la fecha de toma de las muestras analizadas. Esto se hizo con la finalidad de establecer si el periodo hidrológico afectaba de forma significativa los valores de las concentraciones de los parámetros analizados.

Para realizar el ejercicio de clasificación, no se utilizó ninguna técnica específica para definir los grupos, el investigador lo determino según el análisis de la información de precipitación del área de estudio, como herramienta de apoyo en la selección de periodos secos y húmedos se graficó la precipitación promedio mensual, en las estaciones ubicadas en el embalse, así como otras estaciones pluviométricas ubicadas en los alrededores de la zona de estudio. A los datos de precipitación se les realizo un tratamiento estadístico para su análisis, ya que en este caso particular lo que interesaba era el patrón de variación temporal y la determinación de la ocurrencia de periodos secos y húmedos. Esta parte del análisis se encuentra motivada por la variabilidad que han presentado los parámetros hidrológicos en los últimos años. A continuación, en la tabla 1, se presenta el histograma de la precipitación promedio mensual entre 1991 y 2021.

TABLA 1
Tabla de climatología

Mes	Medida		
	Temperatura media (°C)	Precipitación (mm)	Días lluviosos (días)
Enero	10.5	225	21
Febrero	10.4	205	19

Marzo	10.3	185	21
Abril	10	100	17
Mayo	9.5	35	8
Junio	8.7	18	3
Julio	8.4	18	4
Agosto	9.2	29	6
Septiembre	10	50	12
Octubre	10.6	108	18
Noviembre	11.1	132	18
Diciembre	10.7	185	20

Como base a un análisis visual de la tabla, se determinó, por criterio del investigador, que todos los datos por debajo de 60 mm de precipitación serian periodos secos y por encima de estos serían periodos húmedos.

2.2.1 Análisis discriminante

Para poder correr la función discriminante fue necesario conseguir que los datos fueran de distribución normal, para lograr eso se realizó transformación de distribución inversa sobre los datos de una de las estaciones con la intención de utilizar el modelo en una muestra más amplia para comprobar su validez. Sin embargo, debido a la cantidad de datos anómalos en diferentes variables, se decidió utilizar solo una variable, la variable de precipitación la cual permitirá que se cumpla la prueba de homogeneidad de matrices de covarianza M de Box cuyo valor de significancia fue de 0.589 la cual prueba que el método se puede usar sobre los datos.

Al correr la función, se determinaron los puntos medios y un centroide a cada periodo, cuyos valores son presentados en la tabla 2.

TABLA 2
Funciones de centroide

Variable	Centroide
Periodo Húmedo	1.316
Periodo Seco	-2.057

Esto significa que el punto medio donde se definirá si un elemento es del periodo húmedo o seco será en el valor -0.3705, teniendo en cuenta esto si una muestra tiene un valor de coordenada menor a -0.3705 se clasificara como periodo seco, y en caso contrario como periodo húmedo.

2.2.1 Red neuronal artificial perceptrón multicapa

Para realizar el Análisis discriminante se utilizó una RNA de entrenamiento supervisado con retropropagación de error y alimentación hacia adelante (perceptrón multicapa). Esta red se seleccionó debido a que es la red más sencilla para hacer un análisis equivalente al análisis discriminante.

La cantidad de datos escogidos es el total donde se definió que 571 datos se encontraban en el periodo húmedo y 400 en el periodo seco los cuales son mostrados en la tabla 3.

TABLA 3
Datos escogidos

	Precipitación			
Periodos	Frecuencia	Porcentaje	Porcentaje Válido	Porcentaje Acumulado
Periodo Húmedo	571	58.8	58.8	58.8
Periodo Seco	400	41.2	41.2	100.0
Total	971	100.0	100.0	

Los pasos que se realizaron para el diseño del modelo de red neuronal han sido los siguientes:

Selección de variables, procesamiento de datos, creación de grupos y selección y construcción del modelo de red neuronal.

1. Selección de variables: las variables seleccionadas para confeccionar la muestra fueron: cloro, conductividad, pH y turbiedad. Son 4 variables en total, con lo que se usaran 4 neuronas de entrada y 1 de salida, ya que se busca si la muestra se tomó en periodo húmedo o seco. La función usada en las neuronas de la capa oculta es una función tangente hiperbólica. En cambio, en la neurona de salida se aplicará una función sigmoidal, ya que se realiza una tarea de predicción y para estas tareas se recomienda el uso de esta función, ya que los valores no son continuos, sino que dentro de los valores dados (0 y 1), la respuesta se puede desplazar de forma no continua en ellos. En la capa oculta se usará una función tangente hiperbólica, ya que, al implementarla en una red neuronal, se pueden lograr salidas simétricas y balanceadas, lo que permite modelar relaciones no lineales entre los datos con mayor precisión.
2. Procesamiento de datos: Las 4 variables que se han tenido en cuenta a la hora de conformar la muestra se transformaron y estandarizaron para establecerla en una nueva distribución comprendida.
3. Creación de los grupos de entrenamiento, validación y reserva: La muestra de 971 datos se han dividido en tres grupos: entrenamiento, el cual se usará en la fase de aprendizaje, validación, usado en la fase de funcionamiento, aplicado en la etapa para demostrar el funcionamiento de la red. Para ello, el grupo de entrenamiento

consta del 61.5% de los datos para que la red sea capaz de ajustar los pesos de forma iterativa. Luego el grupo de validación estará compuesto por el 27.6% y el grupo de reserva está compuesto por el 10.9% de los datos.

4. Selección y construcción del modelo de red neuronal: Se usará el modelo de red de perceptrón multicapa, ya que es la red de interés del trabajo y que además se ajusta al objetivo del mismo, el cual es la clasificación de un grupo de muestras para saber si se recogieron en un periodo húmedo o seco. Para encontrar el modelo más eficaz se creó diferentes modelos modificando los siguientes valores: número de neuronas ocultas y tasa de aprendizaje. El primer valor a modificarse ha sido el número de neuronas ocultas para encontrar el modelo con el mejor desempeño, el número de neuronas que se han aplicado han sido 1, 2, 3, 4 y 5. Después se ha modificado el valor de la tasa de aprendizaje a los valores 0.4, 0.2, 0.1, 0.05, sin embargo, se decidió no incluir un cambio en el número de capas ocultas debido a que dichos modelos no mostraron que él porcentaje de acierto varíe sustancialmente. Para ello se han creado 20 modelos de red neuronal, se observó el modelo que tuvo un desempeño más equilibrado y mayor en las fases de entrenamiento y validación, para que su desempeño en la fase de test sea más estable. Siguiendo estos criterios el modelo que se seleccionó para pasar a la fase de test fue el modelo número 17 debido a que aparte de ser el modelo con mejor porcentaje de acierto de clasificación en la fase de entrenamiento no existe mucha diferencia con el porcentaje de acierto de la fase de validación. Una vez seleccionado el modelo, se pasó a la fase final, la fase test, para obtener el rendimiento real de la red, la arquitectura del modelo 17 es representado gráficamente en la figura 2.

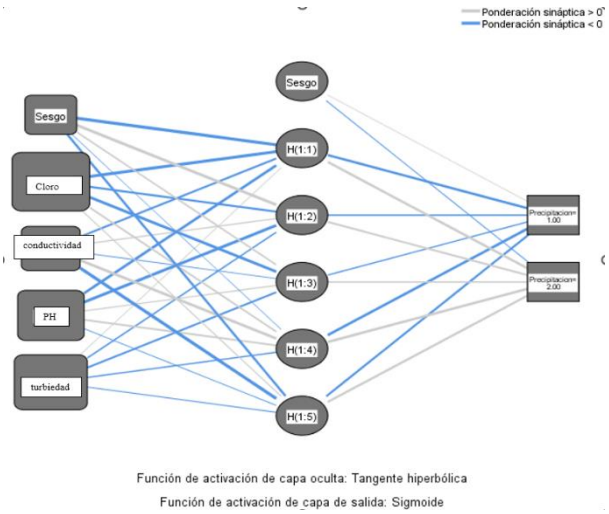


Fig. 2. Arquitectura del modelo de red neuronal escogido.

5. Evaluación de rendimiento del modelo: El análisis se hizo a través de la curva de ROC obtenida a través del programa SPSS. Se usó este método de análisis, ya que nos encontramos ante un problema de clasificación de dos variables y este es la medida precisa y válida para evaluar la precisión diagnostica del modelo seleccionado.

3 Resultados y discusión

3.1 Resultados del análisis de componentes principales

En la siguiente tabla se puede observar la relación entre los componentes principales escogidos para el estudio y las variables originales, para ello se utilizó el método de Spearman debido a que la distribución no es normal.

TABLA 4
Matriz de correlaciones ACP

Variable	Correlaciones			
	Componente 1	Componente 2	Componente 3	Componente 4
Cloro	0.494	-0.849	0.702	0.353
Conductividad	0.744	0.300	-0.501	0.307
pH	0.131	0.569	-0.008	-0.039
Turbiedad	-0.797	0.124	0.007	0.468

Siguiendo el análisis de la matriz de correlación, se observó que la mayoría de los valores de dichos coeficientes resultaron tener valores suficientemente altos. Para efectos del presente trabajo se considera que la relación entre dos variables es significativa si es mayor a 0.4 en valor absoluto.

En la tabla 4 se muestra la correlación entre las variables fisicoquímicas identificadas a partir del Análisis de Componentes Principales. Como se evidencia de la inspección de dicha tabla, existe muy poca relación entre el componente 1 y la variable pH mientras que el componente 2 tiene muy poca relación con la variable conductividad y turbiedad, el componente 3 tiene muy poca relación con pH y turbiedad mientras que el componente 4 tiene muy poca relación con pH y tiene apreciable relación con cloro y conductividad Sin embargo, todas las variables se relacional al menos con un componente por lo cual se puede decir que existe una correlación considerable entre las variables analizadas en el presente estudio.

3.2 Resultados de la red neuronal artificial de tipo hebbiano

En la tabla 5, se puede observar que existe asociación significativa entre las variables estudiadas y los componentes principales obtenidos mediante la red neuronal artificial Hebbiana.

TABLA 5
Matriz de correlaciones RNAH

Variable	Correlaciones			
	Componente 1	Componente 2	Componente 3	Componente 4
Cloro	-0.258	0.258	0.258	0.775
Conductividad	0.600	0.800	0.400	0.400
pH	0.000	0.400	0.800	0.800
Turbiedad	-0.200	-0.400	-0.800	0.000

De esta se puede ver que el componente 1 agrupa a las variables Conductividad, el componente 2 agrupa a todas las variables menos Cloro, la tercera agrupa a todas las variables menos Cloro, la cuarta a todas menos a turbiedad.

De lo anterior se infiere que en los componentes 1 no se encuentra relación significativa entre las variables originales, sin embargo, en el componente 2 y 3 se agrupa 2 variables químicas y una física y en el componente 4 se agrupan 3 variables químicas.

3.2 Análisis de resultados de la comparación entre red neuronal artificial de tipo hebbiano y análisis de componentes principales

Como se observó en los resultados de cada metodología, las RNA de tipo hebbiano obtuvieron un mejor desempeño al poder asociar más variables entre sí. Lo que indica que la no linealidad de la RNA permitió encontrar asociaciones entre variables más concretas, que las de la metodología convencional.

Podríamos partir del hecho que la dinámica físico-química que se presenta en el embalse es compleja, y por lo tanto la identificación de los procesos específicos resulta de vital importancia. La metodología de ACP convencional fue menos útil al determinar las asociaciones entre las variables, que son las que permiten identificar procesos fisicoquímicos; mientras que con las RNA de tipo hebbiano las asociaciones se pudieron interpretar de una manera más clara.

3.3 Resultados del análisis discriminante

Los resultados obtenidos por la función se presentan en la tabla 6 donde podemos comparar los datos reales con lo modelado, además se considera que el periodo húmedo tiene el valor de 1 y el periodo seco el valor de 2.

TABLA 6
Matriz de resultado del análisis discriminante

Nombre de la estación	Fecha	Periodo Real	Periodo Modelado
Quisapata alta	22-01-2021 09:34 AM	1	1
Quisapata alta	22-01-2021 09:34 AM	1	1
Quisapata alta	22-01-2021 09:34 AM	1	1
Quisapata alta	28-02-2021 04:03 PM	1	1
Quisapata alta	28-02-2021 04:03 PM	1	1
Quisapata alta	28-02-2021 04:03 PM	1	1
Quisapata alta	28-02-2021 04:03 PM	1	1
Quisapata alta	22-03-2021 12:34 PM	1	1
Quisapata alta	22-03-2021 12:34 PM	1	1
Quisapata alta	22-03-2021 12:34 PM	1	1
Quisapata alta	22-03-2021 12:34 PM	1	1
Quisapata alta	24-05-2021 10:43 AM	2	2
Quisapata alta	24-05-2021 10:43 AM	2	2
Quisapata alta	24-05-2021 10:43 AM	2	2
Quisapata alta	24-05-2021 10:43 AM	2	2
Quisapata alta	26-06-2021 08:26 PM	2	2
Quisapata alta	26-06-2021 08:26 PM	2	2
Quisapata alta	26-06-2021 08:26 PM	2	2
Quisapata alta	26-06-2021 08:26 PM	2	2
Quisapata alta	25-07-2021 08:33 PM	2	2
Quisapata alta	25-07-2021 08:33 PM	2	2
Quisapata alta	25-07-2021 08:33 PM	2	2
Quisapata alta	25-07-2021 08:33 PM	2	2
Quisapata alta	24-09-2021 11:55 AM	2	2
Quisapata alta	24-09-2021 11:55 AM	2	2
Quisapata alta	24-09-2021 11:55 AM	2	2
Quisapata alta	24-09-2021 11:55 AM	2	2
Quisapata alta	31-10-2021 06:35 PM	1	1
Quisapata alta	31-10-2021 06:35 PM	1	1
Quisapata alta	31-10-2021 06:35 PM	1	1
Quisapata alta	31-10-2021 06:35 PM	1	1
Quisapata alta	21-11-2021 04:54 PM	1	1
Quisapata alta	21-11-2021 04:54 PM	1	1
Quisapata alta	21-11-2021 04:54 PM	1	1
Quisapata alta	21-11-2021 04:54 PM	1	1
Quisapata alta	24-11-2021 11:09 AM	1	1
Quisapata alta	15-12-2021 04:04 PM	1	1
Quisapata alta	15-12-2021 04:04 PM	1	1
Quisapata alta	15-12-2021 04:04 PM	1	1
Quisapata alta	15-12-2021 04:04 PM	1	1

A simple vista se puede observar que el periodo modelado y el periodo real fueron clasificados correctamente en el 100% de los casos, esto lo podemos ver en la tabla 7 de validación cruzada.

TABLA 7
Tabla de validación cruzada análisis discriminante

Validación Cruzada		Pertenencia a grupos pronosticada		
	Precipitación	Periodo Húmedo	Periodo Seco	Total
Recuento	Periodo Húmedo	25	0	25
	Periodo Seco	0	16	16
%	Periodo Húmedo	100.0	0.0	100.0
	Periodo Seco	0.0	100.0	100.0

100% de los casos agrupados fueron clasificados correctamente.

Si probamos este modelo con todos los datos el valor de significancia que obtendremos en la prueba de homogeneidad m de box será de 0 debido a que existe algún valor que muestre disparidad, sin embargo, podríamos calcular un estimado de lo que sería una validación cruzada con todos los datos disponibles para probar que el modelo funciona, lo cual se muestra en la tabla 8.

TABLA 8
Tabla de validación cruzada análisis discriminante con todos los datos

Validación Cruzada		Pertenencia a grupos pronosticada		
	Precipitación	Periodo Húmedo	Periodo Seco	Total
Recuento	Periodo Húmedo	523	48	571
	Periodo Seco	0	400	400
%	Periodo Húmedo	91.6	8.4	100.0
	Periodo Seco	0.0	100.0	100.0

Como se puede observar que el 91.6% de los datos fueron clasificados correctamente en el Periodo húmedo y el 100% en el Periodo seco, teniendo esto en cuenta podemos decir que el 95.1% de los casos fueron clasificados correctamente.

Estos resultados indican que el análisis discriminante convencional puede realizar una clasificación satisfactoria del periodo hidrológico a partir de los valores de las muestras.

Sin embargo, como fue mencionado anteriormente el modelo es inestable si se ejecuta en una cierta cantidad de datos debido a que la significancia del M de Box será 0.

3.3 Resultados de la red neuronal artificial perceptrón multicapa

Teniendo en cuenta el modelo ya escogido los resultados son los siguientes:

1. Grupo de entrenamiento: el grupo de entrenamiento, formado por 597 sujetos se obtuvo que 358 datos se encontraban en el periodo húmedo y 239 en el periodo seco. Pero de esta clasificación, la red consiguió clasificar correctamente el 74.5% de los patrones. De los 358 datos en el periodo húmedo 279 son correctos y 79 fueron discriminados incorrectamente. Al igual en los datos de periodo seco, en donde 166 fueron clasificados correctamente y 73 no lo están. Estos datos aparecen en la tabla 9.

TABLA 9
Tabla de clasificación de datos de entrenamiento

Entrenamiento Observado	Pertenencia a grupos pronosticada		
	Periodo Húmedo	Periodo Seco	Porcentaje Correcto
Periodo Húmedo	279	73	79.3%
Periodo Seco	79	166	67.8%
Porcentaje Global	60.0%	40.0%	74.5%

2. Grupo de validación: En cuanto al grupo de validación, formado por 268 datos, de los cuales 157 fueron clasificados como periodo húmedo y 111 como periodo seco. Del grupo de periodo húmedo 116 son verdaderos positivos y 41 falsos positivos. En el grupo de periodo seco 79 fueron clasificados correctamente y 32 no. Por lo tanto, el 72.8% de los patrones se encuentran bien clasificados, como se muestra en la tabla 10.

TABLA 10
Tabla de clasificación de datos de validación

Validación	Pertenencia a grupos pronosticada		
Observado	Periodo Húmedo	Periodo Seco	Porcentaje Correcto
Periodo Húmedo	116	32	78.4%
Periodo Seco	41	79	65,8%
Porcentaje Global	58.6%	41.4%	72.8%

3. Grupo de reserva: En cuanto al grupo de reserva de 106 participantes, el cual es el que nos interesa se muestra el verdadero rendimiento de la red. De este obtenemos que el modelo pudo predecir correctamente 85 (80%) patrones. De los cuales 57 de ellos son del periodo húmedo y 28 del periodo seco. De los 21 patrones mal predichos, 7 son del

periodo húmedo y 14 del periodo seco, la cual es reflejada en la tabla 11.

TABLA 11
Tabla de clasificación de datos de reserva

Reserva	Pertenencia a grupos pronosticada		
Observado	Periodo Húmedo	Periodo Seco	Porcentaje Correcto
Periodo Húmedo	57	14	80.3%
Periodo Seco	7	28	80.0%
Porcentaje Global	60.4%	39.6%	80.2%

Podemos observar que el porcentaje de acierto del grupo de reserva es muy elevado por lo que podemos decir que el modelo es bueno.

La curva ROC muestra que el área bajo la curva es de 0.795. El cual es un valor próximo a uno, por lo que se puede decir que el modelo tiene una buena capacidad de clasificación, como se muestra en la figura 3.

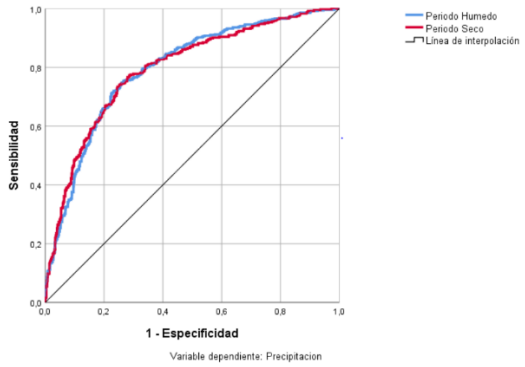


Fig. 3. Curva ROC del modelo

Si ejecutamos una prueba con todos los datos que tenemos disponibles podremos observar cómo se comporta el modelo en una prueba real los resultados que se obtuvieron se muestran en la tabla 12:

TABLA 12
Tabla de clasificación de datos de test con todos los datos

Test	Pertenencia a grupos pronosticada		
Observado	Periodo Húmedo	Periodo Seco	Porcentaje Correcto
Periodo Húmedo	459	112	80.4%
Periodo Seco	135	265	66.3%
Porcentaje Global	61.2%	38.8%	74.6%

Como se puede observar 459 datos fueron clasificados correctamente en el periodo húmedo y 265 en el periodo seco así mismo el porcentaje que obtenemos es menor al que teníamos en la prueba anterior, sin embargo, al comparar la curva ROC y su área 0.801 podemos notar que el modelo establecido funciona, lo cual es reflejado en la figura 4.

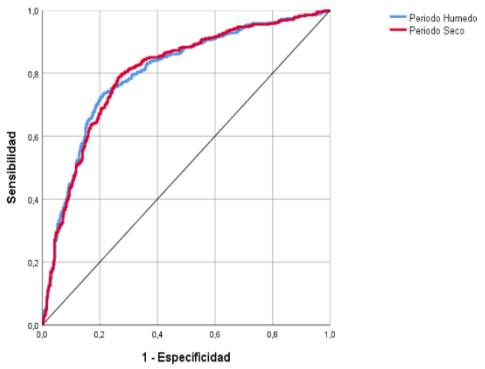


Fig. 4. Curva ROC de la prueba con todos los datos

Revisando los resultados de esto, observamos que los porcentajes de clasificación correcta son altos, lo que indica que el modelo usado con la red es confiable y logro capturar la relación de dependencia presentes en los conjuntos de datos analizados. El porcentaje de clasificación correcta para el periodo húmedo fue del 80.4%, mientras que para el periodo seco fue de 66.3%, los porcentajes de clasificación errónea fueron de 19.6% y 33.7% respectivamente.

3.3 Análisis de resultados de la comparación entre la red neuronal artificial perceptrón multicapa y el análisis discriminante

Los resultados de clasificación obtenidos por los métodos son distintos debido a que para el análisis discriminante se utilizó solo una variable debido a las restricciones de esta, ya que es necesario que se cumpla la significancia de M de Box, sin embargo, si deseamos probar la misma cantidad de variables y compararlas con la RNA el resultado se muestra en la tabla 13.

TABLA 13
Tabla de validación cruzada de análisis discriminante con todos los datos y variables

Validación Cruzada	Pertenencia a grupos pronosticada			
	Precipitación	Periodo Húmedo	Periodo Seco	Total
Recuento	Periodo Húmedo	350	221	571
	Periodo Seco	113	287	400
%	Periodo Húmedo	61.3	38.7	100.0
	Periodo	28.3	71.8	100.0

Validación Cruzada	Pertenencia a grupos pronosticada			
	Precipitación Seco	Periodo Húmedo	Periodo Seco	Total

Como podemos ver los casos clasificados correctamente fueron reducidos desde 95% a 65.8%, pero este valor nos permite observar que el modelo se vuelve inestable debido a que ninguna variable cumple con la prueba de homogeneidad de m de box.

La ventaja que puede tener las RNA sobre el método tradicional son aparentes pues esta puede aprender y mejorar poco a poco siempre y cuando existan más datos para utilizar en el proceso de aprendizaje y validación.

Como pudimos observar en el modelo inicial del análisis discriminante este posee una mejor clasificación correcta que las RNA, pero si lo comparamos usando la misma cantidad de variables y obviando los problemas de significancia del m de box podríamos decir que el RNA es más eficiente debido a que pudo clasificar mejor el Periodo Húmedo y el Periodo Seco.

Si utilizamos el primer modelo de análisis discriminante para realizar la comparación podemos observar que hay una gran distancia entre el 95% y el 74.6% obtenido por la red neuronal. Además, de que los datos que se recogieron para el entrenamiento, validación, y reserva fueron de forma aleatoria con base en 60%, 30% y 10% por lo que no se probó con la misma cantidad de datos cada modelo.

4 Conclusiones

En el caso de la comparación entre el análisis de componentes principales y la red neuronal artificial de tipo hebbiano, el método tradicional si consiguió una reducción de dimensionalidad en la información de calidad de agua estudiada, ya que permitió asociar las variables en grupos que fueran claramente interpretables. Respecto a las redes neuronales artificiales de tipo hebbiano, esta técnica permitió identificar asociaciones con sentido físico/químico entre las variables analizadas, además los grupos de variables encontradas se ubicaron en todos los componentes a excepción de una en la que solo se encontró una variable. El primer componente solo pudo asociarse con la conductividad, mientras que el segundo, tercer y cuarto componente pudo asociarse con 3 variables.

En conclusión, ambos métodos obtuvieron buenos resultados, pero la RNA de tipo hebbiano pudo encontrar más relación entre las variables estudiadas.

En el caso de la comparación entre el análisis discriminante y la red neuronal artificial perceptrón multicapa El método tradicional pudo demostrar mucho mejor la clasificación de datos debido a que se utilizó sobre una variable, la limitación de este método, a pesar de obtener un valor de 95%, es que se necesitan cumplir todos los requisitos para poder ejecutarla si no los datos obtenidos no son estables. En comparación con la red neuronal perceptrón multicapa que es más flexible, también obtiene una clasificación correcta de un porcentaje menor, pero que puede mejorar si se emplean un entrenamiento con un número mayor de patrones, además de que es un sistema tolerante a fallos.

De acuerdo a los resultados obtenidos en este trabajo, los procedimientos de análisis multivariado empleados comúnmente en el análisis de la información de calidad de agua pueden brindar resultados que pueden ser poco interpretables, o muy sensibles a la presencia de valores anómalos. Estos problemas pueden ser resueltos con la utilización de metodologías no lineales como las redes neuronales artificiales de perceptrón multicapa, las cuales permiten construir información que dependen de las relaciones entre el conjunto de datos, y las cuales no ponen restricciones con respecto a la información original. Así mismo, estos modelos son robustos frente a la presencia de valores anómalos.

Cabe mencionar que, en el análisis exploratorio de los datos encontrados en la red de salud Abancay. se observó variables con datos ausentes en su totalidad cuyos fueron, bacterias coliformes totales, bacterias coliformes fecales, bacterias heterotróficas, por lo cual se optó a utilizar solo las variables que contaban con datos en todos los campos las cuales fueron cloro, conductividad, pH, turbiedad, además de la variable periodo que se utilizó de acuerdo a lo mencionado anteriormente en el estudio. Por lo que es recomendable para la red de salud Abancay, conseguir equipamiento que permita recoger más datos que expresen las demás variables físico-químicas.

Es necesario realizar más investigaciones utilizando métodos convencionales y las redes neuronales artificiales para poder determinar que herramienta es más eficaz para asociar y clasificar datos, teniendo en cuenta, que en este último existen una variedad de redes por aplicar.

Referencias

[1] M. Greenacre, P.J.F. Groenen, T. Hastie, et al, "Principal component analysis", Nat Rev Methods Primers vol. 2, pp. 100, 2022, https://doi.org/10.1038/s43586-022-00184-w

[2] S. Mishra, U. Sarkar. et al, "Principal Component Analysis," International Journal of Livestock Research, 2017, https://doi.org/10.5455/ijlr.20170415115235

[3] D. Dhamnetiya, MK. Goel, RP. Jha, S. Shalini, K. Bhattacharyya, "How to Perform Discriminant Analysis in Medical Research? Explained with Illustrations," J Lab Physicians, vol. 14(4), pp. 511-520, 2022 https://doi.org/10.1055/s-0042-1747675

[4] R. Flóres y J. Fernández, "Las Redes Neuronales Artificiales: Fundamentos teóricos y aplicaciones prácticas," La Coruña, NETBIBLO. 2008 S. L, 2008. ISBN: 978-84-9745-246-5.
[5] M. W. Gardner and S. Dorling, "Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences," Atmospheric environment, vol. 32, pp. 2627–2636, 1998
[6] R. Aroca, "Redes neuronales para el análisis y control de calidad del aceite de oliva virgen extra," Universidad Complutense de Madrid, Madrid, 2019.
[7] N. Rodríguez, "Pronóstico de Demanda de Agua potable mediante Redes Neuronales," Universidad Técnica Federico Santamaria, Valparaiso, 2016.
[8] S. Gour y M Gour, "Neural Network Approach In Water Quality Data Analysis For The River Narmada,", Binary Journal of Data Mining & Networking, vol. 4, pp. 49-53, 2014.
[9] S. Heddam, A. Bermad y N. Dechime, "Applications of Radial-Basis Function and Generalized Regression Neural Networks for Modeling of Coagulant Dosage in a Drinking Water-Treatment Plant: Comparative Study," Journal of Environmental Engineering, vol. 137, pp. 1209-1214, New York, 2011, https://doi.org/10.1061/(ASCE)EE.1943-7870.0000435
[10] J.C. Rodríguez, "Estudio Comparativo de Técnicas Estadísticas Multivariadas versus las Redes Neuronales Artificiales en el análisis de datos de Calidad de Agua," Pontificia Universidad Javeriana, Bogotá, 2008.
[11] J. Valencia, "Estudio Estadístico de la Calidad de las aguas en la cuenca hidrográfica del río Ebro," Universidad Politécnica de Madrid, Madrid, 2007.
[12] R. Padilla, "Modelo de Red Neuronal para mejorar la Dosificación de Cloro Gas en la planta de tratamiento de agua potable de la municipalidad provincial de Tayacaja," Universidad Nacional del Centro del Perú, Huancayo, Perú, 2021.
[13] A. Peña, L. Flores del Pino, "Redes neuronales para el tratamiento de agua potable en zona de altitud del Perú," Ambiente y Desarrollo vol. 18, pp. 109-116, 2014, https://doi.org/10.11144/Javeriana.AyD18-35.rnta
[14] R. Álvarez, "Estadistica mutivariante y no parametrica con SPSS. Aplicacion a las ciencias de la salud," Ediciones Diaz de Santos, S. A., Madrid, 1995. ISBN: 978-84-7978-180-4.
[15] M. Andersson, J. Palm, "Forecasting the Stock Market - A Neural Network Approach," MÄLARDALEN UNIVERSITY, Västerås, 2009.
[16] L. Díaz, "Estadística Multivariada: Inferencia y Métodos", Bogota, 2007. ISBN 978-958-701-195-1.
[17] D. Hahs-Vaughn, "Applied Multivariate Statical Concepts," New York, 2017. ISBN: 978-1-315-81668-5.
[18] J. Hilera. y V. Martínez. "Redes Neuronales Artificiales: Fundamentos, modelos y aplicaciones," Madrid, 1995. ISBN: 84-7897-155-6.
[19] P. Isasi y I. Galván, "Redes de Neuronas Artificiales: Un enfoque práctico," Madrid, 2004. ISBN: 84-205-4025-0.
[20] R. Johnson y D. Wichern, "Applied Multivariate Statistical Analysis," vol. 6a, New Jersey, 2007. ISBN 13: 9780131877153.
[21] J. Rodríguez, M. Ferreras y A. Núñez, "Inferencia Estadística, niveles de precisión y diseño muestral," Madrid, 1991.
[22] C. Sierra, "CALIDAD DEL AGUA – Evaluación y diagnóstico," Medellín, 2011. ISBN: 978-958-8692-06-7.
[23] Y. Vidaurre, "Aplicación de las redes neuronales artificiales para el pronóstico de la demanda de agua potable en la empresa EPSEL S.A de la ciudad de Lambayeque," Universidad Católica Santo Toribio de Mogrovejo, Chiclayo, Perú, 2012.
[24] F. Murtaghl, "The Multilayer Perceptron for Discriminant Analysis: Two Examples," Springer, Berlin, Heidelberg, 1992 doi: https://doi.org/10.1007/978-3-642-46757-8_32
[25] L. Heazlewood, J. Walsh, M. Climstein, J. Kettunen, K. Adams, M. DeBeliso, "A Comparison of Classification Accuracy for Gender Using Neural Networks Multilayer Perceptron (MLP), Radial Basis Function (RBF) Procedures Compared to Discriminant Function Analysis and Logistic Regression Based on Nine Sports Psychological Constructs to Measure Motivations to Participate in Masters Sports Competing at the 2009 World Masters Games," Advances in Intelligent Systems and Computing, vol. 392, Springer, Cham, 2016, doi: 10.1007/978-3-319-24560-7_12
[26] E.I. Altman, G. Marco, F. Varetto, "Corporate distress diagnosis: Comparisons using linear discriminant analysis and neural networks (the Italian experience)," Journal of Banking & Finance, vol. 18, pp. 505-529, 1994, https://doi.org/10.1016/0378-4266(94)90007-8
[27] K. Diamantaras, "Robust hebbian learning and noisy principal component analysis," International Journal of Computer Mathematics, vol. 67, pp. 5-24, 1998, https://doi.org/10.1080/00207169808804649

BIOGRAFÍA

Frank Michael Zuloaga Estacio, bachiller en Ingeniería Informática y Sistemas de la Universidad Nacional Micaela Bastidas de Apurímac.

Mario Aquino Cruz, Docente en la Universidad Nacional Micaela Bastidas de Apurímac - Perú, MSc. en Informática, investigador en las áreas de informática educativa, IoT, inteligencia artificial y ciberseguridad.

C&T Riqchary
Revista de Investigación en Ciencia y Tecnología
ISSN: 2810-8124 (en línea) / ISSN: 2706-543x

Vol. 6 Núm. 1 (2024)
Publicado: 22/03/24
https://doi.org/10.57166/riqchary.v6.n1.2024.119
Licencia: CC BY-NC-ND

Globalización - Minería Ilegal - Pobreza, una relación Conflictiva en el Perú

Globalization - ilegal minig - poverity, a conflictive relationship in Peru

Feliciano Escobedo-Silva[a], ORCID: 0000-0002-5740-1935 y Neyka Escobedo-Miranda[b] ORCID: 009-000-7461-1160

(Recepción: 08/05/2024 y aceptación28/06/2024)

Resumen. El presente manuscrito analiza la Globalización, influencia del auge de la minería ilegal y el incremento de la pobreza en el Perú. Al incrementar la demanda de recursos naturales, facilitar la difusión de nuevas tecnologías, permitir la movilidad de capital y personas, contribuye indirectamente a su crecimiento de la minería ilegal. La motivación es plantear un constructo epistémico con la finalidad de someter a debate en los foros académicos las consecuencias que dicha relación generará en un futuro muy próximo el debilitamiento a la estructura del estado por la no tributación, daños al medio ambiente y la masificación de la pobreza. Los resultados indican que la Globalización influye en el desarrollo de la minería ilegal por la mayor demanda internacional del oro, aplicación de tecnologías nuevas que permiten la explotación de minerales de baja ley, disponibilidad de capitales e inversores, factores socioeconómicos (Falta de empleos y pobreza) y la aplicación de políticas económicas neoliberales. Existen evidencias de la vinculación al lavado de activos del narcotráfico y del contrabando de oro hacia Bolivia por zonas fronterizas como Tilali y Cobija de donde son exportados como oro boliviano. El avance de la minería ilegal es es un peligro para la seguridad nacional.

Palabras clave: Globalización, Minería ilegal y pobreza.

Abstract: This manuscript analyzes Globalization, the influence of the rise of illegal mining and the increase in poverty in Peru. By increasing the demand for natural resources, facilitating the dissemination of new technologies, allowing the mobility of capital and people, it indirectly contributes to the growth of illegal mining. The motivation is to propose an epistemic construct with the purpose of submitting to debate in academic forums the consequences that this relationship will generate in the very near future, the weakening of the structure of the state due to non-taxation, damage to the environment and massification. of poverty. The results indicate that Globalization influences the development of illegal mining due to the greater international demand for gold, application of new technologies that allow the exploitation of low-grade minerals, availability of capital and investors, socioeconomic factors (Lack of jobs and poverty) and the application of neoliberal economic policies. There is evidence of links to drug trafficking money laundering and gold smuggling into Bolivia through border areas such as Tilali and Cobija from where they are exported as Bolivian gold. The advance of illegal mining is a danger to national security.

Keywords: Globalization, Illegal mining and poverty.

1 INTRODUCCIÓN

Frente a los alarmantes efectos de la minería ilegal materializados en la desforestación de miles de hectáreas en la selva de Madre de Dios y el avance incontenible de la minería ilegal en todas zonas mineras del país con el aval del congreso de la república, cabe hacer muchas preguntas de reflexión y asombro. ¿Qué es lo que está generando el vertiginoso desarrollo de la minería ilegal?; ¿Por qué el congreso peruano ha aprobado la ampliación del proceso de formalización minera? Sabiendo que la minería ilegal está socavando la estructura socio económica del estado. ¿La Globalización incentiva el crecimiento de la minería ilegal?; ¿A quiénes beneficia la minería ilegal?;, ¿La minería Ilegal genera empleos?

En la actualidad ningún hecho socio económico y político están desvinculados, precisamente la globalización es el hilo conductor de las tendencias como el incremento del precio del oro a nivel internacional está ocasionando que depósitos de oro de bajas leyes y que no eran rentables en tiempos pasados ahora con explotables. Debido a esto ha surgido una nueva generación de "Empresarios" de la peor calaña, sin moral ni ética, de mentalidad Neoliberal en su mayoría (Fujimoristas) dedicados a la minería que se refieren que son mineros informales y que en realidad son mineros ilegales.

Las consecuencias de sus actividades económicas no caben en ninguna lógica filosófica razonable que se pueda entender, aparte de evadir impuestos, aprovecharse de los incentivos tributarios de la zona de ceja de selva, laboran con trabajadores semi esclavizados sin contar con ningún seguro de salud y pensiones; Generan daños irreparables al medio ambiente.
Dar respuesta a estas problemáticas son los objetivos de la investigación, la población analizada se limita a la disponibilidad de la información y se resume a las las regiones de Madre de Dios y Puno.

[a] Docente Facultad de Ingeniería UNAMBA, fescobe@unamba.edu.pe
[b] Estudiante Facultad educación y Ciencias Sociales UNAMBA, 192288@unamba.edu.pe

C&T Riqchary
Revista de investigación en ciencia y tecnología
Vol. 6 Núm. 1 (2024) - 22/03/2024
DOI https://doi.org/10.57166/riqchary.v6.n1.2024.119

1.1Globalización

Es un fenómeno socio económico cultural y tecnológico que trasciende fronteras y contextos, y afecta fundamentalmente la comprensión de la identidad, el conocimiento, la ética, la política y la cultura. sus implicaciones para ofrecer una comprensión más profunda y crítica de este proceso global de la economía Neoliberal [1].

1.2 Minería Ilegal

Según el DL N° 105EM, Es toda actividad minera que se realizan Personas naturales o jurídicas en espacios prohibidos (Riberas de los ríos, lagunas, cabeceras de cuenca, zonas de amortiguamiento de áreas naturales protegidas; Tambien son ilegales los que usan maquinaria y equipos que no corresponda a la categoría de pequeña minería y finalmente aquellos que no cumplan con las exigencias administrativas, técnica, sociales y ambientales de la ley que se realizan en las zonas prohibidas.

1.3 La pobreza

El adjetivo hace referencia a las personas que no tienen lo necesario para vivir dignamente. Se considera que la población se encuentra en condición de pobreza, si el gasto per cápita del hogar está por debajo del valor de la canasta total, compuesta por alimentos y no alimentos; y en condición de pobreza extrema cuando el gasto per cápita del hogar es menor al valor de la canasta de alimento [2].

2 METODOLOGÍA

Para la realización de la presente manuscrito se aplicó las siguientes estrategias: Búsqueda automatizada en base de datos; Revisión en revistas especializadas de Globalización, Minería- pobreza e identificación de autores referentes.

3 ANÁLISIS

La industria minera contribuye aporta el 9 % del PBI nacional al crecimiento económico del país mientras que los precios de los minerales sean altos seguirá siendo una fuente importante de empleo y crecimiento debido al efecto multiplicador de la actividad.

La minería ilegal no es un patrimonio del Perú, también es un fenómeno globalizado y se encuentra en auge en Sudamérica [3]. De acuerdo con la UIF (Unidad de información financiera), el delito más grave en Perú es la minería ilegal, que ha generado el mayor monto acumulado en los reportes de operaciones sospechosas emitidos entre enero de 2013 y junio de 2023, con un valor de USD 8.216 millones. Esta cantidad es doce veces mayor que la cantidad reportada por tráfico ilícito de drogas (640 millones de dólares) y también supera otros delitos.



Fig. 1. Mapa de minería legal en el Perú: [4]

3.1Casos emblemáticos de Minería ilegal

En Madre **de Dios, a**ctualmente, la producción de oro en Perú depende de la minería informal e ilegal, que abarca entre el 22 % y el 28 % del total. Además, en Madre de Dios, donde el 90% de las actividades mineras son ilegales o informales, la minería representa más del 66 % de sus exportaciones.
Se calcula que la producción de 412.000 onzas troy al año es ilegal y representa el 41% del PBI regional. Además, las actividades relacionadas tienen un impacto económico del 70%, y 7 de cada 10 soles están relacionados con la minería.
Se estima que la minería aluvial representa alrededor del 50% de la economía de toda la región. Sin embargo, debido a las condiciones extremadamente difíciles en las que se desarrolla, Esta actividad tiene graves consecuencias ambientales, de seguridad y de salud pública. Aproximadamente 100,000 hectáreas de bosque en la región Madre de Dios se han perdido solo en las últimas tres décadas, y cerca del 78% de los adultos de Puerto Maldonado tiene niveles de mercurio por encima de los estándares internacionales debido a las malas prácticas mineras.

Los resultados de la investigación de Prevenir Amazonia-Usaid son los siguientes, se aclara que la información satelital se tomado en base a los equipos y maquinarias con que extrae el oro.

Consecuencias de la minería ilegal en la Región Madre de Dios. Tan solo en la cuenca del rio Madre de Dios han sido devastadas 50 mil hectáreas. Aproximadamente 4.500 mujeres son explotadas sexualmente en los restaurantes y bares cercanos a los campamentos mineros ilegales. De ellas, el 78% es menor de edad y el 52% proviene de la región de Cusco.

Las tratantes suelen ser mujeres de entre 30 y 40 años de edad, y que los varones son captados con fines de

explotación laboral. 8 de cada 10 pobladores de Puerto Maldonado presentan niveles de mercurio 3 veces más altos' que el máximo permisible para la salud [6].

En Pampa Blanca-La Rinconada Puno, l**a extracción de** oro a través de la minería informal e ilegal en tres áreas de la cuenca alta del río Ramis (CARR) es una de las actividades económicas más perjudiciales para el medio ambiente y la sociedad en Perú. A pesar de que existen reglas específicas destinadas a formalizarlo y regularlo, sus efectos perjudiciales han aumentado significativamente en los últimos doce años. Se emplearon tecnologías de teledetección, que son herramientas útiles y económicas para evaluar y analizar la problemática ambiental de un área tan amplia como el ámbito de interés. De esta manera, se pudo observar la evolución de la minería ilegal de oro en esta región. La zona de La Rinconada y Pampa Blanca es la cabecera de la cuenca del río Ramis, donde se explota de manera agresiva el oro aluvial sin criterios técnicos adecuados, lo que contamina toda la cuenca endorreica del lago Titica Caca. Cuando no se cumplen los estándares de calidad ambiental (ECA), el límite máximo permisible (LMP) o cualquier norma laboral, no está sujeta a inspecciones por parte de las autoridades reguladoras.

El período de análisis fue de treinta y un años en total. Este período se dividió en dos secciones: la primera, que se llevó a cabo cada dos años entre 1984 y 2000, y la segunda, que se llevó a cabo anualmente entre 2001 y 2015.



Fig. N° 05. Panorama de la cordillera de Carabaya y Pampa Blanca.

La fig, N° 05, se obtuvo de Cuarto Poder.14 nov. 2022. América TV, Ilegal Mining that moves millons of soles in Perú.

La extensión explotada por la minería informal en la cuenca alta del río Ramis se expandió más allá de las cien hectáreas entre 2009 y 2011, con una superficie de 907,49 hectáreas, lo que representa aproximadamente el 39,6% de la superficie alterada.

4 Consecuencias de la minería ilegal

La minería ilegal está afectando negativamente la estructura socioeconómica del Perú debido a los efectos perjudiciales que tiene en el ecosistema y en las personas que participan en ella, como la contaminación por mercurio, el alcoholismo, el trabajo infantil, la prostitución, la explotación sexual de menores, el sicariato y las matanzas por control de la actividad, como sucede en Pataz (La libertad).

4.1 Producción ilegal de oro, se lava la "Cara"

Como resultado, la producción ilegal e informal sumó 262,2 toneladas durante ese período, lo que representa una producción anual promedio de 21,8 toneladas. La producción ilegal e informal aumentó en un 41.0% en comparación con las 186.0 toneladas de producción informal de Madre de Dios, o 15.5 toneladas al año, estimadas/registradas por el MINEM entre 2003 y 2014. Si dividimos la producción total de oro del país en producción formal e ilegal o informal, la producción informal de Madre de Dios, estimada o registrada por el MINEM, sumó 448.1 toneladas entre 2003 y 2014, lo que representa una producción promedio anual de 37.3 toneladas. en comparación con la producción formal total de 1,893 toneladas. declaraciones de 15,777 millones de dólares.

4.2 Contrabando de Oro ilegal a Bolivia

En el año 2012, se informó que una parte del oro extraído ilegalmente e informalmente en el Perú fue transportado hacia la Paz-Bolivia. Además, el oro proveniente del sur del país comenzó a ser transportado como contrabando a través de la frontera de Puno (Desaguadero y Tilali, cerca del lago Titicaca, y la localidad minera de La Rinconada y Suches) y Madre de Dios y Cobija-Bolivia. Era transportado primero por tierra hasta la capital de ese país, luego exportado por vía aérea a Miami.

4.3 Estimación de la Evasión Tributaria

Lavado de activos por la minería ilegal en el Perú, el GAFI (Grupo de Acción Financiera Internacional) define el lavado de activos como "el proceso a través del cual se encubre el origen de los fondos generados mediante el ejercicio de algunas actividades ilegales o criminales".

5 Tipificación del delito de minería ilegal

Es el acto de explorar, extraer, explotar u otras actividades similares con recursos minerales, metálicos o no metálicos, sin la autorización de la entidad administrativa competente, que cause o pueda causar daño, alteración o daño al ambiente o sus componentes, la calidad ambiental o la salud. (Código penal,1991). Las formas agravadas de Minería ilegal son:

a) En áreas naturales protegidas, o en tierras de comunidades nativas, campesinas o indígenas.
b) Utilizando dragas, artefactos u otros instrumentos similares.
c) Si el agente emplea instrumentos u objetos capaces de poner en peligro la vida, la salud o el patrimonio de las personas.
d) Si se afecta sistemas de irrigación o aguas destinados al consumo humano
e) Si el agente se aprovecha de su condición de funcionario o servidor público g.-Si el agente emplea

para la comisión del delito a menores de edad u otra persona inimputable.

Delitos de financiamiento de la minería ilegal (artículo 307-C).

1. Delito de obstaculización de la fiscalización administrativa (Artículo 307-D).
2. Delito de tráfico ilícito de insumos químicos y maquinaria destinada a minería ilegal (artículo 307-E)
3. Responsabilidad de funcionario público por otorgamiento ilegal de derecho (artículo 314). [9].
4. La falta de empleos adecuados, tanto en áreas urbanas como rurales, hace que ciertos grupos de personas opten por una actividad que les genera ingresos significativos, como la minería ilegal o informal.

Según un análisis del Instituto Peruano de Economía (IPE), Madre de Dios ocupa el tercer lugar de los ingresos más altos (1 640 soles), solo por debajo de regiones como Moquegua y Lima, lo que la convierte en una región atractiva para la población de escasos recursos en la región selvática. El Estado Peruano tiene una debilidad institucional preocupante que resulta en una capacidad limitada de control y fiscalización en los territorios. El indicador INCORE sobre el desempeño de la institucionalidad nacional ha experimentado una disminución, pasando de 5.24 puntos en 2014 a 5.12.

De cada 100 toneladas de oro producidas en Perú entre 2003 y 2014, poco más de 19 provienen de la minería ilegal o informal. Además, durante 2010-2013, esta producción representó en promedio poco más de la cuarta parte de la producción total de oro en Perú. El valor total de la producción ilegal o informal de oro en Perú, tanto registrada como no registrada, ascendió a US$ 15.777 millones.

Estas estadísticas sobre la actividad minera ilegal o informal coinciden con las predicciones de algunos estudios que indican que las exportaciones ilegales aumentaron entre 2010 y 2011. [10].

6 Influencia de la minería ilegal

En medio del aumento de la minería ilegal y el aumento de los precios de los metales, se ha observado un aumento en la presencia e influencia de mineros dedicados a esta actividad, tanto en sus áreas de origen como en algunos territorios nacionales. La alta rentabilidad de la explotación del oro provocó una acumulación en el sector de la pequeña minería, la minería artesanal e ilegal nunca antes vista, y los procesos productivos comenzaron a incorporar maquinaria pesada, lo que provocó un aumento en el volumen de extracción de oro mientras aumentaban los efectos ambientales y sociales. Además, la minería informal e ilegal se convirtió en una práctica en varias áreas, como en Madre de Dios [11].

El Registro Integral de Formalización Minera (REINFO), que a menudo retrasa la fecha límite de inscripción, lo que no ayuda a fomentar la formalización. Según los datos de la plataforma REINFO, al 27 de octubre de 2023, había 87 204 inscripciones para la formalización minera, de las cuales 16 992 aún estaban vigentes (19.5%), lo que significa que 8 de cada 10 inscritos están suspendidos por no cumplir con la normativa. Finalmente, las entidades principales responsables de abordar la minería ilegal son: Gobiernos regionales y Ministerio del ambiente.

El número de asociaciones y federaciones aumentó junto con el número de afiliados. Las federaciones más destacadas incluyen la Confederación Nacional de Mineros Artesanales y Pequeños Productores Mineros del Perú (CONAMI), la Federación Nacional de Mineros Artesanales del Perú (FENAMARPE), la Sociedad Nacional de Minería en Pequeña Escala (SONAMIPE) y el Consejo Nacional de Mineros del Perú, así como la poderosa e influyente Federación Minera de Madre de Dios (FEEMIN). [10].

7 Agradecimientos

Expreso mi agradecimiento al **Dr. Silberth Solórzano Gutierrez**, docente del curso: Seminario de gestión de Negocios internacionaes de la EPG. UNSAAC, del doctorado en Administración, quién nos asigno como tarea la realización de este articulo que constituye un modesto amporte al debate académico y pone en alerta un fenenómeno socioeconómico.

8 Conclusiones

Existen 3 factores que están impulsando el avance demoledor de la MINERIA ILEGAL, la primera es la Globalización, ya que ha generado el alza del precio del oro en el mercado internacional como respaldo monetario de las economías emergentes.

El segundo factor del auge de la minería ilegal se debe a la destructuralización y debilitamiento del estado peruano, la normatividad del Ministerio de Energía y Minas y del Ambiente está desfasado y obsoleto, sus procedimientos de control no funcionan amparado por la corrupción de funcionarios; hace que los inversionistas se camuflen como mineros informales y exploten el oro sin escrúpulo alguno y son financiados por los bancos privados, como evidencia no se explicaría el conjunto de maquinarias utilizadas (Excavadoras, Volquetes, plantas de lavado) cuyo costo supera un millón de dólares, es imposible el acceso de personas sin capital.

El tercer factor es la falta de empleo en el Perú, la tasa de desempleo es superior al 5.4 % (INEI 2023) muchos peruanos ven a la minería como la única opción de empleo.

Los pactos políticos del Fujimorismo y la derecha peruana con los mineros informales (ilegales) y el congreso de la

república hacen se ha ampliado la formalización (DL N° 31388) hasta diciembre 2024.Influencias políticas y económicas poderosas de los mineros ilegales, ha pesado más que los intereses del país.

No existe una política del estado ni plan que ponga orden y autoridad en las zonas mineras liberadas por la minería ilegal (La pampa, Rinconada, Pataz y otros).

Los daños ocasionados a la naturaleza por la explotación del oro aluvial son Irreparables; Genera desertización: deforestación, erosión, pérdida de suelo fértil. Modificación del relieve, impacto visual, alteración de la dinámica de los procesos de ladera, cambio de relieve aumento de la escorrentía, erosión, alteraciones en el nivel freático, perdida de hábitat de especies animales y forestales.

El actual método de extracción aurífera "Tierra Arrasada "es ilógico, no es ético ni moral, contradice a la razón humana, aplica una filosofía depredadora. No se aplica el método de la "Transferencia" que la recomienda la Academia, para la explotación de los depósitos tipo placeres en la selva baja.

La minería Ilegal, no es ejercida por personas pobres. Son mafias organizadas por empresarios asociados a capitales del sector financiero, corruptos y delincuentes que sin mayores escrúpulos realizan sus actividades. Se camuflan en la minería informal para evadir impuestos, pagar menores sueldos y destruir la naturaleza.

El supuesto crecimiento económico 3.8 % anual (INEI 2020) cuyo aporte también es de la minería ilegal, no compensa con el daño ocasionado al medio ambiente [9].

8.8 La minería ilegal no contribuye a la solución de la pobreza, los trabajadores laboren en condiciones de semi esclavitud no cuentan con el amparo de leyes labores vigentes, seguro social ni de jubilación son sobreexplotados a abandonados a su suerte en casos de accidentes.

9. BIBLIOGRAFÍA

[1] A. Romero, Globalizacion y pobreza, Pasto-Colombia: Ediciones Uñarino, 2002.

[2] I. Goldin y K. A. Reinert, Glogalizacion y pobreza, Bogotá Colombia: Banco Mundial-Ediciones AlfaOmega, 2005.

[3] J. R. Kuramoto, «Mineria Artesanal e informal en el Perú,» GRADE, Lima-Perú, 2001.

[4] Sociedad Nacional de Minería, Petroleó y Energía., «Mineria ilegal: Un flagelo en su nivel mas crítico,» Desde Adentro Vol 2 N° 272 Nov.2, p. 10, 2022.

[5] PREVENIR AMAZONIA-USAID, Estimación de la población minera informal e ilegal en el departamento de Madre de Dios a partir de imagénes satelitales sub métricas, 2022.

[6] MINAM, «Dialogos Ambientales.Minería ilegal,» MINNAM, Lima-Peru, 2013.

[7] V. Torres Cuzcano, Mineria Ilegal e informal en el Perú.Impaco socioeconómico, Lima Perú, 2015. https://doi.org/10.15381/pc.v19i2.11113

[8] SPDA (Sociedad Peruana de Derecho Amazónico), «Las rutas del oro ilegal.Estudio de caso en cinco paises Amazónicos,» SPDA, Lima Perú, 2015.

[9] DAS-UIF-SBS, «Ánalisis de la minería ilegal como delito precedente de lavado de activos,» SBS, Lima-Perú, 2022.

[10] J. De Echade, Mineria Ilegal en el Perú, entre la informalidad y el delito, Lima-Perú: UPCP, 2016.

[11] J. De Echade, R. Hutmer y M. Polaris, «Mineeria y Territorio en el Perú.Conflictos,resistencias propuestas en tiempos de Globalizacion,» Cooper accion,UNMSM, Lima-Perú, 2009.

[12] P. C. Fabián Vera y E. M. Fernandez Vasquez, «Crecimiento económico y su incidencia en la contaminación ambiental en el Perú 1960-2020,» UPN, Trujillo-Perú, 2022.

[13] MINJUSDH-INDAGA, la mineria ilegal en la Amazñia Peruana, Lima-Perú: MINJUSDH, 2021.

[14] M. A. Munsibay Muñoa y L. L. Cavero Espinoza, «Analiisis de mineria informal en la economia peruana periodo 2018 a 2022,» Innova, Lima Perú, 2022. https://doi.org/10.33890/innova.v7.n3.1.2022.2141

[15] U. F. Giraldo Malca y P. G. Vásquez Ruesta, «Evaluacion y Medición de la expansion territorial de la minería informal en la cuenca del río Ramis-Puno,usando imagenes satelitales,» Revista Espacio y Desarrollo UPCP, Puno Peru, 2019. https://doi.org/10.18800/espacioydesarrollo.201902.001

[16] The Global initiative Against Transtational organized crime, «El crimen organizado y la Minería Ilegal de oro en América Latina,» Berna Suiza, 2016.

[17] H. Manrique y C. A. Sanborn, La minéria en el Perú.Un balance y prespectiva de cinco décadas de Investigación, Lima-Perú: Universidad del Pacífico, 2021. https://doi.org/10.21678/978-9972-57-458-0

18] O. Plaza, «Los mercados: Un acercamiento desde la sociología,» de Mercados Globales y (Des)articulaciones internas, DESCO (Centro de estudios y promoción del desarrollo), 2007.

[19] E. Vasquez Huamán, «El perú de los pobres no visibles para el Estado:La inclusion social pendiente a julio del 2012,» Universidad del Pacífico, Lima-Perú, 2012.

[20] P. Spicker, «Definiciones de pobreza:Doce significados,» 1999.

[21] J. Esquival y M. Valdivia, «Perú:Hacia una estrategía de desarrollo para sierra sur rural.,» Lima, 2004.

[22] J. Coasaca P., V. Valcarcel A. y Y. Burgos G., La globalizacion:Analisis e Impacto en el Perú, Lima-Perú : UNMSM, 2003.

[23 T. Piketty, La economia de la desigualdades, Como implementar una redistribución justa y eficaz de la riqueza, Paris Francia, 2008.

[24] A. Zapata y R. Rojas, ¿Desiguales desde siempre?

Miradas históricas sobre la desigualdad, IEP Lima Perú, 2013.

[25] Oficina de las Naciones Unidas contra las drogas y el delito (UNODC), «Respuesta de minería ilegal y el tráfico de metales y minerales,» ONU, Nueva York, 2023.

[26] OEA/DDOT (Departamento contra la delincuencia organizada transnacional), «Tras el dinero del oro ilícito Perú. Fortaleciendo la lucha contra las finanzas de la nueva ilegal,» OEA, Nueva York, 2021.

C&T Riqchary
Revista de Investigación en Ciencia y Tecnología
ISSN: 2810-8124 (en línea) / ISSN: 2706-543x

Vol. 6 Núm. 1 (2024)
Publicado: 25/06/2024
https://doi.org/10.57166/riqchary.v6.n1.2024.120
Licencia: CC BY-NC-ND

Evaluación de morteros con óxido de calcio y sulfato de calcio hidratado para uniones en mampostería

Evaluation of mortars with calcium oxide and hydrated calcium sulfate for masonry joints

Jinmer Bravo-Apaza [A], Anderson Nuñez-Fernandez [B]

0000-0002-0217-7075 [A], 0000-0002-6565-0392 [B]

(Recepción: 15/05/2024 y aceptación 25/06/2024)

Resumen— El uso de cal y yeso en morteros es crucial para el desempeño mecánico de edificaciones tradicionales. El objetivo fué evaluar la influencia del óxido de calcio y sulfato de calcio hidratado en la resistencia a la compresión, el tipo de fisuración y la densidad del mortero utilizado para uniones en mampostería. Se elaboraron 21 briquetas con siete tratamientos: T1 (Testigo), T2-T4 con dosificaciones de óxido de calcio y T5-T7 con sulfato de calcio hidratado, mezclados con cemento y arena en relación de 1:5 (v/v). los resultados mostraron que la resistencia a la compresión en T5 superó a T1. Los tratamientos T1, T2 y T5 presentaron fisuración cónica en 100%. La densidad fue mayor en T4, seguido por T1 y T5. La adición de sulfato de calcio hidratado mejora la resistencia a la compresión y en cuanto a la fisuración columnar y por corte se debe a imperfecciones en los materiales. Se demostró que el sulfato de calcio hidratado aumenta la resistencia a la compresión del mortero alcanzando 144.47 kg/cm² en comparación con 139.87 kg/cm² del testigo, pero una adición excesiva reduce la resistencia. Estas dosificaciones también influyen en el tipo de fisuración, sin embargo, no afectan significativamente la densidad.

Palabras clave: Mampostería, óxido de calcio, sulfato de calcio hidratado.

Abstract— The use of lime and gypsum in mortars is crucial for the mechanical performance of traditional buildings. The objective was to evaluate the influence of calcium oxide and hydrated calcium sulfate on the compressive strength, type of cracking, and density of mortar used for masonry joints. Twenty-one briquettes were prepared with seven treatments: T1 (Control), T2-T4 with calcium oxide dosages, and T5-T7 with hydrated calcium sulfate, mixed with cement and sand in a 1:5 (v/v) ratio. The results showed that the compressive strength in T5 exceeded that of T1. Treatments T1, T2, and T5 exhibited conical cracking in 100% of samples. The density was highest in T4, followed by T1 and T5. The addition of hydrated calcium sulfate improves compressive strength, and conical and shear cracking are due to material imperfections. It was demonstrated that hydrated calcium sulfate increases the compressive strength of mortar to 144.47 kg/cm² compared to 139.87 kg/cm² for the control, but excessive addition reduces strength. These dosages also influence the type of cracking; however, they do not significantly affect density.

Keywords: Masonry, calcium oxide, hydrated calcium sulfate

1 INTRODUCCIÓN

La composición de morteros con cemento Portland y cal hidráulica afecta significativamente la resistencia a la compresión. Una mezcla con cemento Portland en proporción 1:3 alcanza resistencias de 19.2 MPa, mientras que mezclas con cal muestran resistencias menores. La adición de cal al concreto también afecta significativamente su resistencia a la compresión. Añadir entre 10-15% de cal reduce la resistencia, mientras que una adición del 5% es aceptable y económicamente beneficiosa [1]. Además, la adición de cenizas volantes al concreto incrementa su resistencia a la compresión y disminuye su permeabilidad, alcanzando la máxima resistencia con un 6% de cenizas volantes [2]. La mezcla de concreto ternario con residuos de mampostería y cal hidratada muestra comportamiento puzolánico favorable, mejorando la resistencia a compresión a largo plazo [3].

Los muros de mampostería combinada y confinada demuestran capacidad para soportar cargas sísmicas la-

A. *Departamento de Ingeniería, Universidad Nacional Micaela Bastidas de Apurímac, Abancay, Perú. Correo electrónico: jbravo@ unamba.edu.pe.*
B. *Departamento de Ingeniería, Universidad Nacional Micaela Bastidas de Apurímac, Abancay, Perú. Correo electrónico: Anuñez@unamba.edu.pe.*

terales al utilizar morteros con cal y mampostería adherida [4]. Se ha propuesto un enfoque normativo para el diseño sísmico de edificaciones de mampostería confinada, basado en la resistividad y desempeño sísmico, considerando las lecciones aprendidas de terremotos anteriores [5]. Los ciclos de congelación-descongelación deterioran las propiedades físicas y mecánicas de los materiales de mampostería, por ello la necesidad de buscar materiales con resistencias optimas [6]. Los residuos de fosfoyeso mejora las propiedades mecánicas y de durabilidad del mortero estructural sin cemento, incorporado escoria de alto horno granulada molida y óxido de calcio (CaO). La adición óptima mejora la resistencia a la compresión y flexión [7].

Los morteros estructurales se utilizan para unir unidades individuales de mampostería. Los morteros de cal y yeso fueron los más extendidos desde la antigüedad hasta principios del siglo XIX. Los morteros de cal prevalecieron sobre la construcción en seco cuando los constructores romanos comenzaron a experimentar con aglutinantes de alta calidad [8]. Ademas el mortero de revestimiento y uniones protege y sella el edificio, sufre diversos mecanismos de degradación, especialmente relacionados con el cemento [9]. Por otro lado, los morteros de cal hidráulica acumulan humedad y cuando son secados en una solución de cloruro de sodio, la acumulación de sal produce un material más denso, mejorando la resistencia a la compresión y la flexión [10]

El óxido de calcio se obtiene mediante la calcinación de piedra caliza, produciendo cal viva (CaO), que al exponerse al agua se convierte en cal apagada (hidróxido de calcio). La cal apagada es una base fuerte que puede contener hasta un 5% de óxido de magnesio y se caracteriza por endurecerse incluso en presencia de agua. El mortero de cal viva se utiliza en revestimientos y suelos, mientras que el mortero hidráulico es resistente a agentes agresivos [11]. Ademas en morteros pusolanicos, la cal, cal hidraulica incrementa la resistencia a la traccion [12].

El mortero para juntas de mampostería se elabora con un aglutinante como cemento o con una mezcla de cal y arena a la cual se añade cemento en el sitio de uso. La cal de construcción se clasifica según su composición química: la cal de dolomita, la cal aérea y la cal hidráulica. La cal cálcica, o cal viva, se obtiene calcinando calizas con más del 95% de carbonato cálcico ($CaCO_3$) a unos 900°C. La cal hidratada se produce al hidratar la cal viva con agua. La cal se usa en morteros para revestimientos y juntas de albañilería debido a su buena plasticidad, trabajabilidad y elasticidad sin sufrir contracción por humedad. Permite alta permeabilidad al vapor, evitando condensación y eflorescencias, y es un excelente aislante térmico y acústico [13].

La resistencia a la compresión es el principal indicador de calidad del mortero, crucial para soportar cargas y resistir deformaciones. En aplicaciones como muros de carga y de contención, se usan morteros de alta resistencia a la compresión para garantizar la seguridad estructural y la integridad de los muros durante la vida útil del edificio [14]. La durabilidad de los especímenes de mortero es crucial para evaluar su resistencia a largo plazo. Se recomienda que los especímenes curados en agua no potable alcancen al menos el 90% de la durabilidad de los curados en agua potable [15].

La mampostería es un material de construcción compuesto por partes de diferentes materiales conectadas por juntas de mortero. Aunque es heterogénea, puede considerarse homogénea en muros con muchas partes macroscópicas. Las juntas de mortero actúan como superficies débiles, haciendo que el material sea ortótropo. El comportamiento no lineal del muro proviene del comportamiento no lineal de los materiales tras la deformación y del comportamiento de la interfaz de desechos bajo esfuerzo cortante [16].

El mortero es una mezcla de cemento Portland, cal, arena y agua. Añadir cal al mortero de cemento mejora su manejabilidad. Los morteros se clasifican en mortero de cal, mortero de cemento Portland, y mortero de cemento y cal. Se utiliza para unir elementos de mampostería, nivelarlos y mejorar la estética de la estructura. El mortero de cal gana resistencia lentamente, aumentando su resistencia a la compresión con el tiempo [17]. El mortero de cal puro, la mezcla de arena y áridos arcillosos tambien mejoran la capacidad con el mismo volumen. El mortero de cal puro también mejora la plasticidad y ductilidad y es adecuado para estructuras [18]. Asi mismo el mortero de yeso proporciona mayor resistencia a la compresión y resistencia al corte diagonal en muros de mampostería [19]. Ademas en cuanto a la influencia de la cal en el mortero, en cuanto a su viscosidad esta dado mas por la morfologia de la cal que por su composicion quimica [20].

La resistencia de adherencia del mortero es crucial para el desempeño mecánico de la mampostería de ladrillo [21]. Se debe tener en cuenta tambien la prolongación de la vida útil de las construcciones ya que es crucial para reducir los costos de mantenimiento y la huella de carbono en la industria de la ingeniería civil [22]

La relación de mezcla del hormigón debe corresponder a la calidad requerida (resistencia a cargas e influencias químicas y a las condiciones de instalación en el sitio. La dosificación es correcta si cumple ambos requisitos con el máximo ahorro de material [23]. El hormigón es una mezcla de productos compuesta por un aglomerante y áridos, que pueden ser gruesos o finos. Las propiedades del hormigón dependen de su composición y de sus propiedades físicas y químicas [24].

C&T Riqchary
Revista de investigación en ciencia y tecnología
Vol. 6 Núm. 1 (2024) - publicado:25/06/2024
DOI https://doi.org/10.57166/riqchary.v6.n1.2024.120

2. METODOLOGÍA

La investigación evaluó la influencia del óxido de calcio y del sulfato de calcio hidratado en las características mecánicas y físicas del mortero mediante la elaboración de 21 briquetas. Se diseñaron siete tratamientos: T1 (Testigo), T2-T4 con diferentes dosificaciones de óxido de calcio, y T5-T7 con distintas dosificaciones de sulfato de calcio hidratado, todos mezclados con cemento y arena en una relación volumen a volumen (v/v) fija de 01:05.

Tabla 01: Diseño de mezcla para morteros con adicion de oxido de calcio y Sulfato de calcio hidratado.

Tratamiento	Dosificación de mezcla (v/v)		
	M1	M2	M3
T1	0.00:01:05	0.00:01:05	0.00:01:05
T2	0.25:01:05	0.25:01:05	0.25:01:05
T3	0.50:01:05	0.50:01:05	0.50:01:05
T4	0.75:01:05	0.75:01:05	0.75:01:05
T5	0.25:01:05	0.25:01:05	0.25:01:05
T6	0.50:01:05	0.50:01:05	0.50:01:05
T7	0.75:01:05	0.75:01:05	0.75:01:05

La medición de oxido de calcio y sulfato de calcio hidratado se realizo con probeta graduada para examinar con precisión las variables de adición de óxido de calcio: cemento: arena de 0.25-0.75:01:05 v/v y sulfato de calcio hidratado: cemento: arena de 0.25-0.75:01:05 v/v. El requerimiento de agua incrementó conforme se aumenta el volumen de óxido de calcio o sulfato de calcio hidratado en la mezcla del mortero.

2.1 PROCEDIMIENTO

Los componentes de la mezcla fueron dosificados en envases de plástico de 18 litros, utilizando una probeta graduada de PVC y una espátula para el llenado. Las proporciones de las muestras T1-T7 y sus réplicas se midieron y vertieron en un recipiente de plástico.

La mezcla se vertió en un recipiente de plástico para medir las muestras M1-M4 y sus réplicas por triplicado. Se utilizó una espátula para llenar la cubeta con cal, cemento y arena, luego se transfirió a un balde de 18 litros para la mezcladora.

El mortero se mezcló en un equipo de concreto de 0.80 HP y 125 litros a 20 rpm durante 05 minutos. Luego, se vertió en un balde de 18 litros para colocarlo en las briqueteras de pvc.

El mortero se colocó en briqueteras de pvc cilíndricas en tres capas, compactadas con 25 golpes de una varilla y tres golpes laterales.

Transcurrido un día, se desencofraron las briquetas utilizando un inflador para facilitar el proceso.

Posteriormente, las briquetas se curaron sometieron al curado recubrieron con agua para asegurar su hidratación adecuada.

Para determinar la resistencia a la compresión, se retira las briquetas de la cama de curado después de 28 días, se mide el diámetro con un vernier y se calcula el área transversal. Luego, se limpian las superficies de contacto, se colocan las briquetas en la prensa y se aplica una carga continua de 1.20 kN/s.



Fig. 01. Imagenes que muestra la rotura de briquetas con prensa a 1.20 kN/s

Después de aplicar la compresión, se observa el tipo de falla en la muestra. Las fallas comunes incluyen la falla cónica y la rotura de la columna, esta última puede deberse a superficies de carga convexas, falta de material, abolladuras o convexidades en las placas de carga.

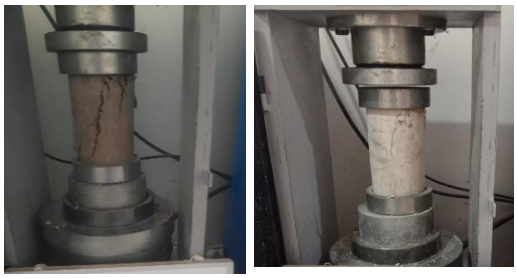


Fig. 02. Imagenes que muestra la fisuración – tipo de falla en la briqueta que se sometió a esfuerzo de compresión.

La densidad de las briquetas se determinó según la norma ASTM C62, midiendo su masa aparente tras 28 días de curado y aplicando la fórmula de densidad basada en la relación masa-volumen.

2.2 POBLACIÓN Y MUESTRA

La población de investigación consistió en briquetas cilíndricas elaboradas en el laboratorio de suelos, agregados y concreto ConChipa E.I.R.L. en Abancay, Apurímac. La muestra total incluyó 21 briquetas, se sometieron a pruebas de resistencia a la compresión, con 3 muestras por tratamiento (óxido de calcio, sulfato de calcio hidratado y muestra testigo). Los resultados se analizaron estadísticamente tras la rotura de las briquetas.

3 RESULTADOS Y DISCUSIONES

3.1 Resistencia a la compression

Tabla 02. Datos observados en laboratorio para determinar la resistencia a la compresión.

Resistencia a compresión (kg/cm²)	T1	T2	T3	T4	T5	T6	T7
R1	139.00	125.60	124.30	121.70	145.10	136.60	132.10
R2	140.50	125.00	124.80	123.50	144.80	138.10	131.00
R3	140.10	126.50	125.30	124.00	143.50	137.50	131.70
Total	419.60	377.10	374.40	369.20	433.40	412.20	394.80
Promedio	139.87	125.70	124.80	123.07	144.47	137.40	131.60
Varianza	0.60	0.57	0.25	1.46	0.72	0.57	0.31

En la determinación de la resistencia a la compresión del mortero con diferentes tratamientos de óxido de calcio, se observó que el tratamiento T1 tuvo la mayor resistencia con un promedio de 139.877 kg/cm², seguido por T2 con 125.70 kg/cm², T3 con 124.80 kg/cm², y T4 con 123.07 kg/cm². En cuanto a los tratamientos con óxido de calcio hidratado, T5 mostró la mayor resistencia con 144.47 kg/cm² superando la muestra testigo, seguido por T6 con 137.40 kg/cm² y T7 con 131.60 kg/cm². El tratamiento T5 se destacó por su alta resistencia, lo que indica su potencial aplicación en morteros para mampostería.

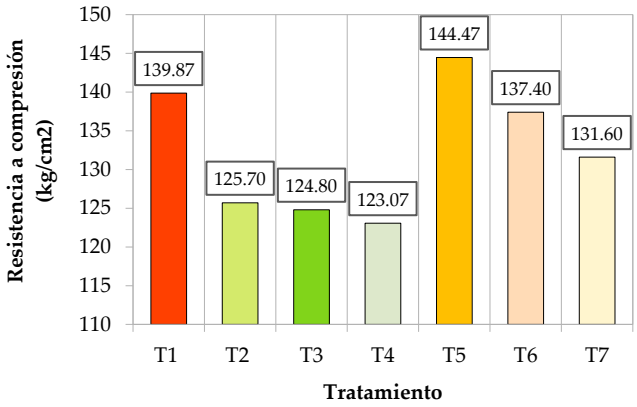


Fig. 03. Promedios de resistencia a compresión del mortero con diferentes dosis de óxido de calcio y sulfato de calcio hidratado para mampostería.

La figura 03 se observa que el tratamiento T5 presenta la mayor resistencia a la compresión, seguido por el tratamiento T1 (testigo), T6 y finalmente T7. La gráfica muestra una tendencia negativa, indicando que, al incrementar la dosis de óxido de calcio y sulfato de calcio hidratado, la resistencia disminuye, siendo inversamente proporcionales.

Tabla 03. Análisis de varianza de promedios de resistencia a la compresión.

Origen de variaciones	SS	df	MS	F	Valor-P	F crítico
Entre grupos	1 250.27	6.00	208.38	324.86	0.00	2.85
Dentro de los grupos	8.98	14.00	0.64			
Total	1 259.25	20.00				

La tabla 03 muestra los resultados del análisis de varianza para los promedios de la resistencia a la compresión del mortero con diferentes dosis de óxido de calcio y de sulfato de calcio hidratado. El p-valor es menor a 0.05, indicando que existe una diferencia significativa en la resistencia a la compresión entre las distintas dosificaciones. Para determinar cuál tratamiento tiene mayor resistencia, es necesario realizar un análisis de comparación múltiple de Tukey.

La prueba de Tukey demsotro una comparación múltiple al 95% de confiabilidad de los promedios de resistencia a la compresión del mortero con diferentes dosis de óxido de calcio y sulfato de calcio hidratado para mampostería. Se observa una diferencia significativa entre la mayoría de los tratamientos, excepto entre T6-T1, T3-T2, T4-T2 y T4-T3, que no presentan diferencias significativas, es decir, son estadísticamente iguales. Los resultados sugieren que la muestra patrón (T1, con dosificación 0.00:01:05) tiene la mayor resistencia, mientras que los tratamientos con óxido de calcio y sulfato de calcio hidratado disminuyen la resistencia a la compresión.

3.2 Fisuración - tipo de falla a compresión

Tabla 04. Tipo de falla del mortero con óxido de calcio y sulfato de calcio hidratado para mampostería.

Tratamiento	Tipo de falla a la compresión			Total
	Cónica (%)	Columnar (%)	Corte (%)	
T1	100.00	0.00	0.00	100
T2	100.00	0.00	0.00	100
T3	0.00	66.67	33.33	100
T4	0.00	66.67	33.33	100
T5	100.00	0.00	0.00	100
T6	0.00	33.33	66.67	100
T7	0.00	66.67	33.33	100

La tabla 04 muestra los datos observados en el laboratorio de concreto para determinar el tipo de falla del mortero con diferentes dosis de óxido de calcio (T1, T2, T3 y T4) y sulfato de calcio hidratado (T5, T6 y T7). Se observa que los tratamientos T1, T2 y T5 presentaron una falla cónica en el 100% de las muestras. Los tratamientos T3, T4 y T7 presentaron una falla columnar en el 66.67% de las muestras. El tratamiento T6 mostró una falla de tipo corte en el 66.67% de las muestras y una falla de tipo corte en el 33.3%.

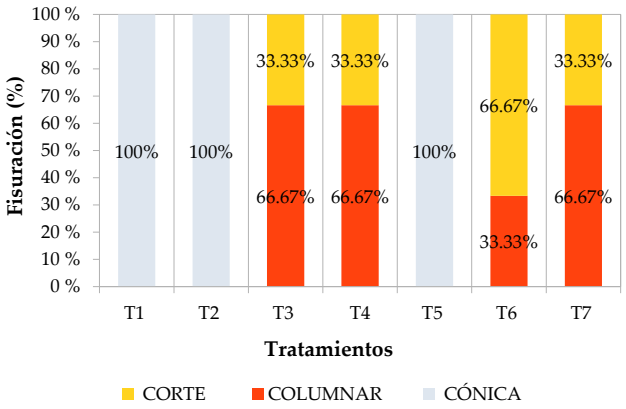


Fig. 04. Promedios del fisuración - tipo de falla del mortero con diferentes dosis de óxido de calcio y sulfato de calcio hidratado para uniones de mampostería.

La figura 04 presenta una representación gráfica de los promedios de fisuración y tipo de falla del mortero con diferentes dosis de óxido de calcio (T2, T3, T4) y sulfato de calcio hidratado (T5, T6, T7) para uniones de mampostería. Se observa que los tratamientos T1, T2 y T5 presentaron únicamente fallas de tipo cónica. Los tratamientos T3, T4 y T7 mostraron fallas de tipo columnar en el 66.67% de las muestras, mientras que el tratamiento T6 tuvo fallas de tipo corte en el 66.67% de las muestras.

3.3 Densidad

Tabla 05. Datos observados en el laboratorio de concreto para la determinación de la densidad del mortero con óxido de calcio y sulfato de calcio hidratado.

Densidad (kg/m³)	T1	T2	T3	T4	T5	T6	T7
R1	1 686.28	1 676.50	1 632.82	1 679.76	1 676.50	1 671.62	1 671.62
R2	1 661.87	1 636.03	1 671.62	1 743.93	1 676.50	1 644.09	1 657.01
R3	1 665.20	1 648.60	1 642.10	1 645.20	1 648.60	1 642.10	1 645.20
Total	5 013.36	4 961.13	4 946.53	5 068.89	5 001.60	4 957.80	4 973.83
Promedio	1671.12	1 653.71	1 648.85	1 689.63	1 667.20	1 652.60	1 657.94
Varianza	175.23	428.93	410.45	2 509.98	259.44	272.17	175.10

La tabla 05 muestra los datos de laboratorio para determinar la densidad del mortero con diferentes dosis de óxido de calcio (T2, T3 y T4) y sulfato de calcio hidratado (T5, T6 y T7) para uniones de mampostería. Los promedios de densidad fueron: T1 - 1671.12 kg/m³, T2 - 1653.71 kg/m³, T3 - 1648.85 kg/m³, T4 - 1689.63 kg/m³, T5 - 1667.20 kg/m³, T6 - 1652.60 kg/m³ y T7 - 1657.94 kg/m³. El tratamiento T4 tuvo la mayor densidad, seguido por T1, mientras que T3 tuvo la menor densidad.

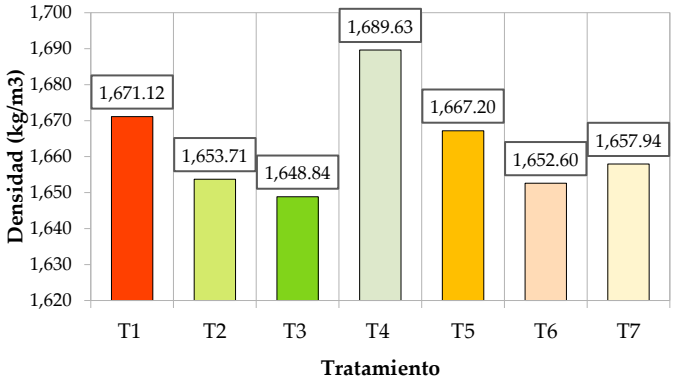


Fig. 05. Promedios de la densidad del mortero con diferentes dosis de óxido de calcio y sulfato de calcio hidratado para muros de mampostería.

La figura 05 muestra la representación gráfica de los promedios de densidad del mortero con diferentes dosis de óxido de calcio (T1, T2, T3, T4) y de sulfato de calcio hidratado (T5, T6, T7) para muros de mampostería. Se observa que el tratamiento T4 tiene la mayor densidad, seguido por el tratamiento T1 y T5. Los demás tratamientos tienen densidades inferiores a la muestra patrón.

Tabla 06. Análisis de varianza de promedios de densidad.

Origen de variaciones	SS	df	MS	F	Valor-P	F crítico
Entre grupos	3 639.02	6.00	606.50	1.00	0.46	2.85
Dentro de los grupos	8 462.58	14.00	604.47			
Total	12 101.60	20.00				

La tabla 06 muestra los resultados del análisis de varianza para los promedios de densidad del mortero con diferentes dosis de óxido de calcio (T2, T3, T4) y sulfato de calcio hidratado (T5, T6, T7) para muros de mampostería. El p-valor es mayor a 0.05, lo que indica que no hay una diferencia significativa en la densidad del mortero entre las diferentes dosis de óxido de calcio y sulfato de calcio hidratado, es decir, estadísticamente tienen igual densidad.

4 DISCUSIÓN DE RESULTADOS

4.1 Resistencia a la compresión

La tabla 01 presenta los resultados del análisis de varianza de los promedios de resistencia a la compresión del mortero con diferentes dosis de óxido de calcio (T1, T2, T3, T4) y sulfato de calcio hidratado (T5, T6, T7) para mampostería. El p-valor menor a 0.05 indica una diferencia significativa en la resistencia a la compresión entre las diferentes dosis de óxido de calcio y sulfato de calcio hidratado. Para determinar qué tratamiento tiene la mayor resistencia, se realizó un análisis de comparación múltiple de Tukey. Los resultados muestran que, a mayor concentración de óxido de calcio y sulfato de calcio en el mortero, disminuye la resistencia a la compresión. Sin embargo, el sulfato de calcio en una concentración de 0.25

mejora favorablemente la resistencia del mortero. Además, los resultados indican que la resistencia a la compresión del mortero se incrementó con el aumento de la cantidad de óxido de calcio, alcanzando su máximo con una adición del 5%. Sin embargo, cuando el porcentaje de óxido de calcio se incrementó a 6%, la resistencia a la compresión disminuyó[2]. De esto se concluye que un aumento del 1% en la resistencia a la compresión se logra con hasta un 5% de óxido de calcio, pero más allá de este punto, la resistencia disminuye en un 5% [8]. Del mismo modo se logró una resistencia máxima a la compresión de 21 kg/cm² con una dosis del 6% después de 28 días, demostrando que la resistencia a la compresión aumenta a medida que se incrementa la cantidad añadida [9]. Por otro lado, los resultados sobre la resistencia a la compresión del concreto, publicados en una relación a/c constante, mostraron que la mezcla con la mayor relación a/c tenía la mayor resistencia al envejecimiento prolongado debido al comportamiento puzolánico. Se concluyó que la resistencia a la compresión es mejor cuando la mezcla contiene una alta cantidad de escombros de mampostería en comparación con las mezclas con una gran cantidad de cal apagada [10].

4.2 Fisuración - tipo de falla del mortero

La figura 03 muestra una representación gráfica de los promedios de fisuración y tipo de falla del mortero con diferentes dosis de óxido de calcio (T1, T2, T3, T4) y sulfato de calcio hidratado (T5, T6, T7) para uniones de mampostería. Los tratamientos T1, T2 y T5 presentaron únicamente fallas de tipo cónica. Los tratamientos T3, T4 y T7 mostraron fallas de tipo columnar en el 66.67% de las muestras, mientras que el tratamiento T6 tuvo fallas de tipo corte en el 66.67% de las muestras. Sien embargo la ASTM C-39-86 [28], Las fallas típicas en el ensayo de compresión para briquetas incluyen la falla tipo cónica, que ocurre con una carga de compresión bien aplicada en un espécimen bien preparado. Sin embargo, también se observan fallas de columna debido a superficies de apoyo cóncavas, imperfecciones en los materiales o muescas en las placas de carga. La falla cilíndrica es causada por la falta de material, rugosidad de las placas de carga, mientras que la falla por cortante ocurre cuando la superficie de aplicación de carga está fuera de la tolerancia de paralelismo o el centro de la muestra está desalineado.

4.3 Densidad del mortero

La tabla 04 muestra los resultados del análisis de varianza para los promedios de densidad del mortero con diferentes dosis de óxido de calcio (T1, T2, T3, T4) y sulfato de calcio hidratado (T5, T6, T7) para muros de mampostería. El p-valor es mayor a 0.05, lo que indica que no hay una diferencia significativa en la densidad del mortero entre las diferentes dosis de óxido de calcio y sulfato de calcio hidratado, es decir, estadísticamente tienen la misma densidad. Según Fernández [29], La densidad real del hormigón depende principalmente de la densidad de los áridos, su granulación y el volumen involucrado en su composición. Una mayor densidad del hormigón puede indicar mejor resistencia mecánica y durabilidad, lo cual se logra con la máxima compacidad, es decir, con menos vacíos y mayor compactación.

5 CONCLUSIONES

El estudio sobre la evalaucion del óxido de calcio y sulfato de calcio hidratado en morteros para uniones de mampostería ha revelado que la adición de sulfato de calcio hidratado incrementa la resistencia a la compresión del mortero, alcanzando hasta 144.47 kg/cm², en comparación con 139.87 kg/cm² del testigo. Sin embargo, una dosificación excesiva de estos materiales reduce la resistencia estructural. Además, se observó que la fisuración cónica predomina con menores dosificaciones, mientras que la fisuración columnar y por corte son más frecuentes con mayores concentraciones de óxido de calcio y sulfato de calcio hidratado. No se encontraron variaciones significativas en la densidad entre los tratamientos, indicando que la densidad no se ve sustancialmente afectada por las diferentes dosificaciones. Además, el uso controlado de sulfato de calcio hidratado puede mejorar la resistencia a la compresión de los morteros, aunque se necesitan más estudios para evaluar los efectos a largo plazo y su compatibilidad con otros materiales de construcción.

6 REFERENCIAS BIBLIOGRÁFICAS

[1] Hidai Calleja García, Sulpicio Sánchez Tizapa, Al-fredo Cuevas Sandoval, Alma Delia Zacaria, and Vi-tal y Manuel Ramos Segura, " Influencia de la cal en la resistencia a compresión del mortero," Academia Journals, vol. 13, no. 1, pp. 136-141, Mar. 2021.

[2] S. Huaquisto Cáceres and G. Belizario Quispe, "Uti-lización de la ceniza volante en la dosificación del concreto como sustituto del cemento,"Revista de In-vestigaciones Altoandinas - Journal of High Andean Re-search, vol. 20, no. 2, pp. 255-234, Apr. 2018. https://doi.org/10.18271/ria.2018.366

[3] Y. F. Silva Urrego, J. E. Rojas Rojas, J. A. Gamboa, M. Gordillo, and S. Delvasto Arjona, "Optimización de la resistencia a compresión usando un diseño de mezcla de vértices extremos, en concretos ternarios basados en residuo de mampostería y cal hid-ratada," Revista EIA, vol. 16, no. 31, pp. 99-113, Jan. 2019. https://doi.org/10.24050/reia.v16i31.1177

[4] Arturo Tena Colunga, Artemio Juárez Ángeles, and Víctor Hugo Salinas Vallejo, "Resistencia y defor-mación de muros de mampostería combinada y confinada sujetos a cargas laterales," Ingeniería Sís-mica, vol. 76, no. 1, pp. 26-60, Mar. 2007. https://doi.org/10.18867/ris.76.58

[5] Ángel San Bartolomé and Daniel Quiun, "Propuesta normativa para el diseño sísmico de edificaciones de mampostería confinada," Scielo, vol. 42, no. 2, pp. 40-57, nov. 2004.

[6] Y. Zhou, S. Zheng, L. Chen, L. Long, and L. Dong, "Mechanical properties and seismic behavior of confined masonry walls in freeze-thaw environ-ment," Structures, vol. 31, pp. 647-659, jun. 2021. https://doi.org/10.1016/j.istruc.2020.12.051

[7] Hermawan, H. S. Djayaprabha, and H.-A. Nguyen, "Utilizing phosphogypsum waste to improve the mechanical and durability performances of cement-free structural mortar containing ground granulated blast furnace slag and calcium oxide," Journal of Building Engineering, vol. 72, p. 106557, Aug. 2023. https://doi.org/10.1016/j.jobe.2023.106557
[8] P. Vitti, "Lime and Gypsum Mortars in Historic Construction," in Encyclopedia of Archaeology (Second Edition), Elsevier, 2024, pp. 531-543. https://doi.org/10.1016/B978-0-323-90799-6.00064-1
[9] G. Baptista Junior et al., "Durability for coating mor-tars: Review of methodologies," Journal of Materials Research and Technology, vol. 30, pp. 657-671, May 2024. https://doi.org/10.1016/j.jmrt.2024.03.020
[10] D. V. Bompa and A. Y. Elghazouli, "Mechanical properties of hydraulic lime mortars and fired clay bricks subjected to dry-wet cycles," Constr Build Ma-ter, vol. 303, p. 124458, oct. 2021. https://doi.org/10.1016/j.conbuildmat.2021.124458
[11] "Guía para el Análisis de Estructuras de Mampostería."
[12] P. Maravelaki-Kalaitzaki, A. Bakolas, and A. Moropoulou, "Physico-chemical study of Cretan ancient mortars," Cem Concr Res, vol. 33, no. 5, pp. 651-661, May 2003. https://doi.org/10.1016/S0008-8846(02)01030-X
[13] "Guía práctica para los morteros con cal."
[14] "Evaluacion de resistencia a la compresion de mortero con proveta".
[15] James K. Wight and Basile G. Rabbat, "Requisitos de Reglamento para Concreto Estructural (ACI 318S-05) y Comentario (ACI 318SR-05)," American Con-crete Institute, vol. 1, no. 1, pp. 308-3018, Jan. 2005.
[16] Didier Combescure and Pierre Pegon, "Modelling of the in-plane behaviour of masonry infilled frames," European Seismic Design Practice, vol. 1, no. 1, pp. 621-629, oct. 1995. https://doi.org/10.1201/9780203756492-94
[17] Michael S. Mamlouk and John P. Zaniewski, Mate-riales Para Ingeniería Civil, 2nd ed., vol. 1. Madrid: Pearson Educación, 2009.
[18] P. Tian, W. Yang, J. Lu, X. Wu, and Z. Wang, "Study on the shear properties of masonry pagoda mortar and its influencing factors," Case Studies in Construc-tion Materials, vol. 20, p. e02981, jul. 2024. https://doi.org/10.1016/j.cscm.2024.e02981
[19] A. Soleymani, M. A. Najafgholipour, and A. Johari, "An experimental study on the mechanical proper-ties of solid clay brick masonry with traditional mortars,"Journal of Building Engineering, vol. 58, p. 105057, oct. 2022. https://doi.org/10.1016/j.jobe.2022.105057
[20] Y. Sébaïbi, R. M. Dheilly, and M. Quéneudec, "A study of the viscosity of lime-cement paste: influ-ence of the physico-chemical characteristics of lime," Constr Build Mater, vol. 18, no. 9, pp. 653-660, nov. 2004. https://doi.org/10.1016/j.conbuildmat.2004.04.028
[21] Z. Tang, H. Zhang, Y. Pan, L. Ke, Z. Xiang, and Z. Lai, "Experimental study on mechanical properties of basalt Fiber-Clay lime mortar and application in brick masonry," Constr Build Mater, vol. 398, p. 132479, Sep. 2023. https://doi.org/10.1016/j.conbuildmat.2023.132479
[22] D. Bednarska, M. Koniorczyk, N. Grzelak, and M. Czyż, "A contribution to preservation of salt-laden construction materials: An effective, eco-friendly and inexpensive system of renovation plasters," Journal of Building Engineering, vol. 83, p. 108444, Apr. 2024. https://doi.org/10.1016/j.jobe.2024.108444
[23] P. Tian, W. Yang, J. Lu, X. Wu, and Z. Wang, "Study on the shear properties of masonry pagoda mortar and its influencing factors," Case Studies in Construc-tion Materials, vol. 20, p. e02981, jul. 2024. https://doi.org/10.1016/j.cscm.2024.e02981
[24] Enrique Rivva López, Naturaleza y materiales del concreto, 1st ed., vol. 1. Lima: Aci Perú, 2000.

C&T Riqchary
Revista de Investigación en Ciencia y Tecnología
ISSN: 2810-8124 (en línea) / ISSN: 2706-543x

Vol. 6 Núm. 1 (2024)
Publicado: 25/07/2024
https://doi.org/10.57166/riqchary.v6.n1.2024.121
Licencia: CC BY-NC-ND

Determinación del mejor modelo de Machine Learning para la predicción del California Bearing Ratio de suelos en Abancay, 2024

Determination of the best Machine Learning model for the prediction of the California Bearing Ratio of soils in Abancay, 2024

Flor de Cantuta Tello-Sarmiento [A], Manuel J. Ibarra-Cabrera [B]
ORCID: 0009-0006-9567-5220 [A], 0000-0001-6711-4916 [B]
(Recepción: 16/07/2024 y aceptación 25/07/2024)

Resumen— El California Bearing Ratio (CBR) es un índice fundamental en la ingeniería geotécnica para evaluar la capacidad de soporte de los suelos, especialmente en el diseño y construcción de pavimentos y otras estructuras sobre terreno natural. Pero la determinación de este índice es una tarea costosa y laboriosa, por dicha razón en este estudio se propone la predicción del CBR mediante modelos de machine learning. Se desarrollaron 3 modelos de aprendizaje automático, redes neuronales profundas (DNN), árboles de decisión y máquinas de vectores de soporte. El trabajo consistió en recolectar 310 registros con características del suelo, de los cuales 217 registros fueron considerados para el entrenamiento, 62 para la validación y 31 para las pruebas; los datos fueron recolectados en 3 laboratorios de mecánica de suelos de la ciudad de Abancay, provincia de Abancay en la región Apurímac de Perú, donde se obtuvieron las siguientes características físicas del suelo: el porcentaje de grava, porcentaje de finos, el óptimo contenido de humedad (OCH), límite líquido, límite plástico, índice de plasticidad, máxima densidad seca (MDS) y para la característica a predecir el valor del CBR al 100%. Los modelos fueron evaluados con el coeficiente de determinación (R^2), el error absoluto medio (MAE), el error cuadrático medio (MSE) y la raíz del error cuadrático medio (RMSE). Los resultados muestran que el algoritmo o modelo de árboles de decisión es el más eficiente para predecir el CBR al 100% porque tiene el mejor coeficiente de determinación $R^2 = 0.9307$ y también los valores más bajos para el MSE = 9.199, MAE = 1.216 y RMSE = 3.033; estos valores son los mejores en relación con los encontrados para los modelos de redes neuronales profundas y el de máquina de vectores de soporte para regresión.

Palabras clave: CBR, machine learning, arboles de decisión, red neuronal, regresión lineal múltiple

Abstract— The California Bearing Ratio (CBR) is a fundamental index in geotechnical engineering to evaluate the bearing capacity of soils, especially in the design and construction of pavements and other structures on natural ground. However, the determination of this index is a costly and laborious task, for that reason in this study, the prediction of CBR using machine learning models is proposed. Three machine learning models were developed, deep neural networks (DNN), decision trees, and support vector machines. The work consisted of collecting 310 records with soil characteristics, of which 217 records were considered for training, 62 for validation and 31 for testing; the data were collected in 3 soil laboratories in the city of Abancay, province of Abancay in the Apurimac region of Peru, where the following physical soil characteristics were obtained: gravel percentage, percentage of fines, optimum moisture content (OCH), liquid limit, plastic limit, plasticity index and maximum dry density (MDS) and for the characteristic to be predicted the CBR value at 100%. The models were evaluated with the coefficient of determination (R^2), the mean absolute error (MAE), the mean square error (MSE), and the root mean square error (RMSE). The results show that the decision tree algorithm or model is the most efficient for predicting the CBR at 100% because it has the best coefficient of determination $R^2 = 0.9307$ and also the lowest values for the MSE = 9.199, MAE = 1.216 and RMSE = 3.033; these values are the best in relation to those found for the deep neural network and support vector regression machine models.

Keywords: CBR, machine learning, decision tree, neural network, multiple linear regression.

A. *Flor de Cantuta Tello-Sarmiento, Escuela Profesional de Ingenieria Informática y Sistemas de la Universidad Nacional Micaela Bastidas de Apurímac-Perú, 111181@unamba.edu.pe.*
B. *Manuel J. Ibarra-Cabrera, Departamento Académico de Informática y Sistemas de la Universidad Nacional Micaela Bastidas de Apurímac-Perú, mibarra@unamba.edu.pe.*

1 Introducción

El empleo de técnicas de aprendizaje automático han sido fundamentales para resolver problemas en la ingeniería, especialmente cuando las variables implicadas tienen relaciones no lineales y resulta difícil representar el problema mediante funciones matemáticas convencionales [1].

El índice de California (CBR) es una medida de la resistencia del suelo, el CBR se mide en laboratorio aplicando una carga a una muestra de suelo compactado y midiendo la resistencia a la penetración de un pistón estándar, el CBR viene siendo considerado como la única forma para evaluar la capacidad de soporte de los suelos, especialmente en el diseño y construcción de pavimentos y otras estructuras sobre terreno natural. Un CBR alto indica que el suelo tiene una buena capacidad de soporte y puede reducir los costos de construcción

al requerir menos materiales de refuerzo [2].

Sin embargo, debido a que el ensayo de CBR es un proceso laborioso y que consume mucho tiempo aproximadamente 4 días y en ocasiones, los resultados pueden carecer de precisión debido a errores humanos y a la posible falta de habilidades técnicas suficientes del personal de laboratorio [2]. La aplicación de algoritmos de machine learning para calcular el CBR basado en ciertas características del suelo es una excelente opción, principalmente en lo que respecta a una mayor precisión y reducción de tiempo.

Una de las alternativas para mejorar el proceso de diseño y construcción de pavimentos es la aplicación de modelos de machine learning para predecir el CBR de los suelos, ya que teniendo valores aproximados se pueden estimar presupuestos preliminares en obras viales.

Asimismo, la ingeniería ha sido transformada por la revolución digital, donde el crecimiento masivo de datos y el avance de la inteligencia artificial, en particular el aprendizaje automático (ML), han sido cruciales. Machine learning se ha convertido en una herramienta esencial en diversas áreas de ingeniería, desde mejorar procesos industriales hasta diseñar sistemas autónomos y prever comportamientos complejos. Adicionalmente, el aprendizaje automatizado, es un área de estudio de la inteligencia artificial y de las ciencias de la computación que aporta bastante a la ingeniería geotécnica [3], el adecuado uso de técnicas de machine learning trae consigo beneficios económicos, técnicos y contribuye a la optimización de procesos, así como, por ejemplo, se usa para la estimación de la densidad del suelo [4], también para la evaluación de algunas propiedades del suelo para la caracterización de macizos rocosos [5].

En el Perú, existe una variedad de suelos y la construcción de carreteras y pavimentos es esencial para el desarrollo socioeconómico del país. Estas infraestructuras mejoran la conectividad, impulsan el comercio, promueven la integración regional, facilitan el acceso a servicios básicos y aumentan la seguridad vial. Estos proyectos viales son cruciales para fomentar el progreso y elevar la calidad de vida de la población peruana. En Abancay, como en todo el Perú, los pavimentos suelen ser descuidados tras su construcción, lo que resulta en su deterioro visible en el centro de la ciudad. Las autoridades a menudo no toman medidas de mantenimiento adecuadas, lo que lleva a reparaciones mal ejecutadas o incompletas, acelerando el deterioro de las infraestructuras. Asimismo, la falta de un plan de mantenimiento para las infraestructuras viales se refleja en los daños no reparados, empeorando las condiciones viales. La congestión de tráfico se agrava por la gran cantidad de vehículos pesados, afectando tanto a vehículos particulares como al transporte público. Además, el sistema de drenaje deficiente causa obstrucciones durante las lluvias, afectando el sistema de agua y desagüe y provocando daños en las pistas debido a filtraciones y colapsos de tuberías subterráneas [6].

En cuanto al espesor de la capa de asfalto, es esencial realizar estudios adecuados para la construcción de carreteras, ya que estas vías unen diferentes regiones y facilitan el desarrollo económico, ignorar estos aspectos puede conllevar a altos costos logísticos para la economía, por lo que es vital considerar estos factores durante la planificación y construcción de infraestructuras viales [7]. Al diseñar un pavimento para una carretera, es esencial determinar la capacidad de soporte del suelo subyacente, ya que esto influye en la selección de materiales y el espesor del pavimento necesario para garantizar una estructura segura, duradera y eficiente. El índice California Bearing Ratio (CBR) se usa para evaluar la capacidad de soporte del suelo, siendo crucial para decisiones informadas sobre el diseño de pavimentos en vías y campos de aterrizaje [8]. Sin embargo, la escasez de datos y la ausencia de análisis de suelos son problemas significativos en la red vial de carreteras en el Perú [9].

Este estudio busca comparar distintas técnicas de machine learning para determinar cuál es la más eficiente, para calcular el indicador de la calidad de suelo basado en su resistencia (California Bearing Ratio), considerando varias características del suelo como entrada. Después de un análisis exploratorio de datos (EDA) [10] donde tenemos múltiples variables predictoras para predecir la variable de respuesta (en este caso el CBR al 100%) se opta por hacer la comparación de 3 modelos de regresión múltiple para determinar el mejor modelo de machine learning, utilizando varias métricas de evaluación como el coeficiente de determinación(R^2), error cuadrático medio (MSE), error absoluto medio (MAE) y raíz del error cuadrático medio (RMSE) [11].

2 Trabajos relacionados

En este contexto, existen algunos autores que han realizado estudios para predecir características del suelo, por ejemplo, Johan y José realizaron una investigación cuyo objetivo fue construir bibliotecas espectrales con técnicas de machine learning para los suelos tropicales de Costa Rica y determinar las bandas hiper-espectrales óptimas en el rango espectral visible infrarrojo cercano e infrarrojo de onda corta para caracterizar propiedades de suelo. El método propuesto logró una estimación precisa del contenido de diferentes componentes (Ca, Mg, Fe, C, N y CICE) con un R^2 superior a 0,8 y un error cuadrático medio (RMSE) inferior al 10% [12].

Por otro lado, Frank y Heber realizaron una investigación cuyo objetivo fue desarrollar un modelo predictivo de las propiedades del suelo usando RNA del tipo perceptrón multicapa, para predecir individualmente la máxima densidad seca (MDD), el óptimo contenido de humedad (OMC), el valor de la relación de soporte de California (CBR) al 95% y el CBR al 100% usando como variables de entrada otras propiedades del suelo. La investigación contó con un dataset de 285 ejemplos. Teniendo como resultados la predicción de la MDD con R^2=0,90, OMC con R^2=0,87, CBR al 95% con R^2=0,92 y CBR al 100% con R^2=0,89, respectivamente, demostrando que los

modelos son eficientes para predecir las propiedades del suelo [13].

También Saúl y Alaín realizaron un proyecto de investigación donde el objetivo principal fue determinar el modelo de redes bayesianas para predecir los parámetros de resistencia al corte del suelo (ángulo de fricción y cohesión) en función de sus propiedades físicas (granulometría, límites de Atterberg, contenido de humedad y peso específico). La investigación concluyó con un modelo de redes bayesianas con un coeficiente de determinación R^2 de 0.89790 para el ángulo de fricción y de 0.92819 para la cohesión y reafirmó el potencial de las técnicas avanzadas de aprendizaje automático, como las redes bayesianas, como herramientas innovadoras y valiosas para la predicción y la resolución de problemas complejos en el campo de la ingeniería [14].

Asimismo, Marisabel y Rodrigo en su investigación donde su objetivo fue evaluar el desempeño de un modelo de inteligencia artificial basado en redes neuronales artificiales (RNA) para predecir los parámetros de resistencia al corte de suelos, específicamente el ángulo de fricción interna y la cohesión, a partir de sus propiedades físicas como los límites de Atterberg, granulometría, humedad y peso específico. El modelo de RNA fue entrenado utilizando la metodología de propagación inversa (Feed-Forward Backprop) en Matlab, con un 72% de los datos para el entrenamiento y el 28% restante para la validación. Los resultados, evaluados mediante el análisis estadístico de error medio cuadrático (MSE), mostraron un coeficiente de determinación R-cuadrado de 0.93927 en el entrenamiento, 0.99746 en la validación, y 0.96465 en la prueba, obteniendo un modelo final con R^2=0.95507. Esto demuestra que el modelo de RNA es eficaz con un error menor al 5%, proponiéndose, así como una alternativa viable para estudios geotécnicos en la planificación, diseño y ejecución de proyectos de construcción [15].

Como se observa, existen varios estudios donde se aplicaron técnicas de machine learning en el área de la ingeniería geotécnica como una alternativa en la predicción de características del suelo.

El objetivo de este proyecto de investigación fue determinar el mejor algoritmo de machine learning para determinar el valor del CBR al 100% en base a ciertas características del suelo ya que la predicción del CBR es esencial para determinar el espesor y los materiales adecuados para la construcción de pavimentos y carreteras, para lo cual se recolectó registros de ensayos mecánicos llevados a cabo en los diferentes laboratorios especializados en suelos y concretos y algunos expedientes técnicos de obras de carreteras en el departamento de Apurímac.

El entrenamiento se realizó con tres modelos o algoritmos de machine learning: árboles de decisión, máquinas de vectores de soporte y redes neuronales profundas.

3 Método

3.1 Ámbito de estudio

Este estudio se realizó en la Universidad Nacional Micaela Bastidas de Apurímac (UNAMBA) en la ciudad de Abancay. Los datos fueron recolectados de informes de estudios de mecánica de suelos de los laboratorios de suelos de la ciudad de Abancay y expedientes técnicos de obras. Los datos proporcionados pertenecen a diferentes proyectos de desarrollo y mejoramiento de pavimentos en el departamento de Apurímac entre los años 2022 y 2024.

3.2 Tipo y nivel de investigación.

Este trabajo es una investigación de tipo aplicada y de nivel descriptivo.

3.3 Procedimiento

El procedimiento fue el siguiente:

a) Recopilación y adquisición de datos con las características de los suelos y el indicador de calidad de suelo basado en su resistencia (CBR).
b) Preparación de datos, tener los datos limpios y preparados para en el entrenamiento de los modelos, esto incluye la normalización de datos, manejo de valores faltantes y la separación de datos de entrenamiento, validación y pruebas.
c) Entrenamiento de los modelos, en esta etapa se entrena cada modelo con el 70% de datos.
d) Validación de los modelos, en esta etapa de utiliza el 20% de datos para evaluar el rendimiento de cada modelo con las métricas relevantes.
e) Ajuste de los modelos, en esta etapa, si corresponde se ajusta la arquitectura del modelo o los parámetros de entrenamiento y se vuelve a entrenar los modelos para mejorar su rendimiento.
f) Proceso pruebas de los modelos, en esta etapa se utilizan los modelos entrenados para realizar predicciones con los datos de pruebas donde se usa el 10% de los datos.
g) Comparación y análisis de los resultados obtenidos con cada uno de los modelos de machine learning y se selecciona el más eficiente y con menor margen de error.

Para la implementación de los modelos de machine learning se utilizó el lenguaje de programación pytnon v3.10.12 junto a las librerías sklearn v1.2.2, tensorflow v2.15.0 y keras v2.15.0. Asimismo, para el análisis y visualización de datos pandas v2.0.3 numpy v1.25.2, seaborn v0.13.1 y matplotlib v3.7.1.

3.4 Población y muestra

Los datos han sido recolectados en 3 laboratorios de suelos: Obregon S.C.R.L, Geolef E.I.R.L y Ecx ingenieros, ubicados en la ciudad de Abancay y también de expedientes técnicos de obras de pavimentos y carreteras del departamento de Apurímac. Se obtuvieron registros históricos de estudios de mecánica de suelos, alcanzando a recopilar un total de 310 registros esto debido a la disponibilidad limitada de datos.

Donde 6 columnas son características del suelo que intervienen en la capacidad de carga del suelo asimismo como variable dependiente el valor del soporte de california (CBR).

Las variables independientes se obtuvieron de los ensayos granulométricos (porcentaje de grava, porcentaje de finos), de los límites de Atterberg (límite líquido, límite plástico e índice de plasticidad) y del ensayo del Proctor Modificado (óptimo contenido de humedad y máxima densidad seca) y la columna target con el valor del CBR al 100%. Del total de datos, el 70% se usa para el entrenamiento y el 20% se usa para evaluar la eficiencia de los modelos y 10% para las pruebas. Para el proceso de entrenamiento se obtuvo una muestra de 217 registros, 62 registros para la validación y 31 para el testing. Este muestreo fue de forma aleatoria.



Fig. 1. Imágenes tomadas de los ensayos de laboratorio

La tabla 1 muestra las variables con los 10 primeros registros de la base de datos considerando las primeras 6 características del suelo como los datos de entrada y el CBR al 100% como dato de salida. Donde se observan que todos los valores de las variables independientes son datos de tipo cuantitativos continuos, todos se miden en porcentajes a excepción de la densidad máxima seca (MDS) que se mide en gramos por centímetro cúbico. Mientras que la variable dependiente (CBR) se mide en porcentaje y también los valores son de tipo cuantitativos continuos, mientras más alto el valor del CBR indica mayor capacidad de soporte.

TABLA 1
Registro de las 10 primeras observaciones del dataset

Grava (%)	Finos (%)	OCH (%)	LL (%)	LP (%)	IP (%)	MDS (gr/cm³)	CBR al 100%
57.5	42.5	7.3	30.5	22.7	7.8	2.117	38.1
22.6	77.4	10.2	29.8	17.7	12.1	1.928	17.4
52.6	47.4	7.1	25.5	19.1	6.4	2.165	34.3
34.1	65.9	12.2	24.1	16.7	7.4	1.975	14.1
40.9	59.1	7.7	23.1	18.2	4.9	2.113	38.4
8	92	5.7	31	22.4	8.6	1.681	8.5
67.7	32.3	5.2	32.8	24.5	8.3	2.234	45.1
36.9	63.1	7	34	26.5	7.5	2.095	39.7
0	100	16.3	39.1	31.7	7.4	1.696	6.8
43.2	56.8	5.4	34.2	24.4	9.7	2.143	32.4

3.5 Materiales e Instrumentos

Dentro de los materiales utilizados tenemos:

a) Hojas de calculo Excel.
b) Google Colab.

Dentro de los instrumentos utilizados están:

a) Registro de los resultados de informes de ensayos de laboratorios de estudios de mecánica de suelos en formato .xlsx y .csv.
b) Cuaderno de notebooks con extensión .ipynb.

4 Resultados y Discusión

4.1 Analisis exploratorio de datos(EDA)

La Fig. 2 muestra el mapa de calor, en el cual se puede observar la correlación entre variables, donde podemos observar que la característica de Grava tiene una alta correlación positiva con el CBR al 100%, por lo que se puede inferir que a medida que el porcentaje de Grava incrementa también se incrementa el valor del CBR al 100%; asimismo, con la densidad máxima seca (MDS) que también tiene una alta correlación positiva con el CBR al 100%. Por otro lado, el porcentaje de Finos, el óptimo contenido de humedad (OCH) y el Límite Líquido (LL) tienen una correlación negativa con el CBR al 100% con lo que se puede inferir que a medida que estos valores bajan también baja el CBR al 100%. En cuanto al límite plástico (LP) y el índice de plasticidad (IP) tienen una correlación baja, por lo que podemos inferir que tienen menor influencia en el CBR al 100%.

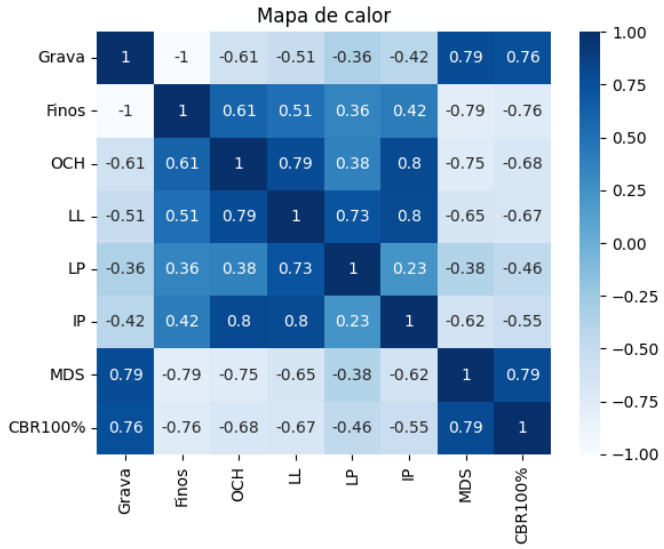


Fig. 2. Mapa de Calor (Heatmap)

La Fig. 3 muestra la distribución de la variable dependiente(CBR al 100%) en el cual se puede observar que tenemos valores atípicos o outliers a partir del valor 50, en total se presentaron 4 registros del total de las observaciones, y revisando los informes de los ensayos de laboratorio concluimos que estos datos corresponden a muestras con algún adherido o añadidos para el mejoramiento del CBR por lo que se tomó la decisión de remover estos registros ya que son irrelevantes para este estudio porque estamos tomando muestras de suelos naturales.

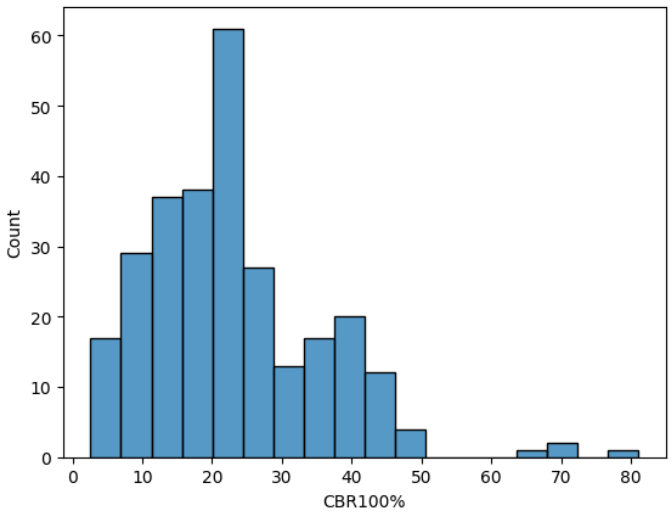


Fig. 3. Distribución del CBR al 100%

Como la variable dependiente(CBR) es un valor continuo el problema es de regresión por lo que no se aplico técnicas de balanceo de datos solo eliminación de valores atipicos. Para el caso de las variables independientes no se contaban con valores atipicos relevantes pero si se aplicó normalización optando el método estandar escaler para uniformizar la escala de datos.

4.2 Pruebas y resultados para el modelo SVR

La tabla 2 muestra las métricas para el modelo de Máquinas de Vectores de Soporte, en la cual se puede observar que el modelo SVR tiene un buen rendimiento en la predicción del CBR al 100%, con un R^2 superior a 0.7, lo que sugiere que el modelo es capaz de capturar una parte significativa de la variabilidad en los datos.

El R^2 de 0.701 sugiere que el modelo explica aproximadamente el 70.15% de la variabilidad en los valores del CBR basados en las características del suelo. Este es un indicador positivo del rendimiento del modelo, aunque también muestra que hay un 29.85% de la variabilidad que no está siendo capturada. El MSE de 36.83 indica el promedio de los cuadrados de los errores. El RMSE de 6.07 indica que en promedio las predicciones del modelo tienen un error de aproximadamente 6.07 unidades del CBR al 100%. El MAE de 4.41 indica que, en promedio, las predicciones del modelo están desviadas en 4.41 unidades del CBR al 100%.

TABLA 2
Métricas para el modelo SVR

Algoritmo o modelo de machine learning	R^2	MSE	RMSE	MAE
SVM - máquinas de vectores de soporte de regresión	0.701	36.834	6.069	4.409

La Fig. 4 muestra el gráfico de dispersión de modelo para la data de training de las predicciones versus los valores reales del modelo.

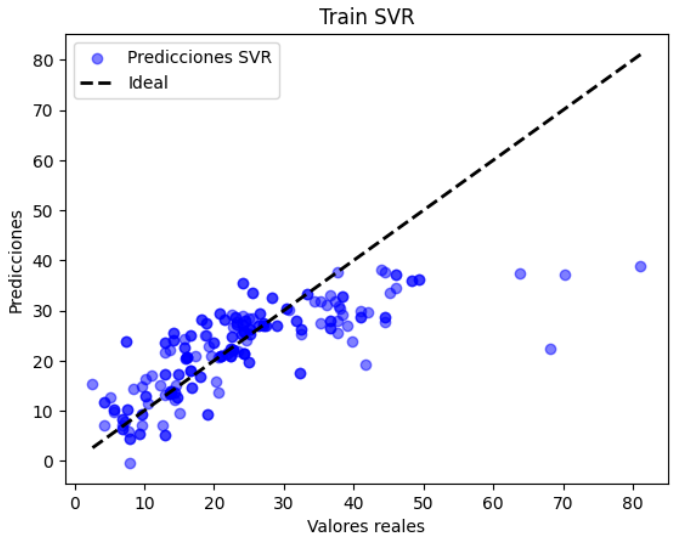


Fig. 4. Gráficos de dispersión del modelo de SVR para la data de training

La Fig. 5 muestra el gráfico de dispersión de modelo para la data de pruebas de las predicciones versus los valores reales del modelo.

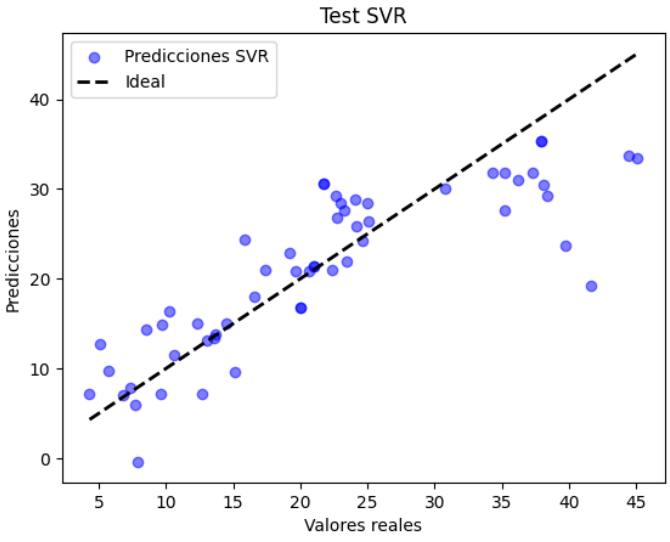


Fig. 5. Gráficos de dispersión del modelo de SVR para la data de testing

3.3 Pruebas y resultados para el modelo con redes neuronales profundas(DNN)

La tabla 3 muestra las métricas para el modelo de redes neuronales profundas, en la cual se puede observar que el modelo tiene un buen rendimiento en la predicción del CBR al 100%, con un R^2 superior a 0.784, lo que sugiere que el modelo es capaz de capturar una parte significativa de la variabilidad en los datos. Estas métricas sugieren que el modelo DNN tiene un rendimiento bastante bueno para predecir el CBR basado en las características del suelo. El valor de R^2 de 0.7875 indica que el modelo explica bien la variabilidad en los datos. Tanto el RMSE de 5.12 como el MAE de 3.76 son razonablemente bajos, lo que sugiere que el modelo tiene un error de predicción aceptable. Adicionalmente, en las pruebas se pudo observar que para el entrenamiento del modelo no afecta significativamente tener los datos normalizados, asimismo, el excluir la característica del índice de plasticidad (IP) tuvo un impacto positivo en el rendimiento del modelo.

TABLA 3
Métricas para el modelo DNN

Algoritmo o modelo de machine learning	R^2	MSE	RMSE	MAE
Redes neuronales profundas (DNN)	0.784	26.606	5.158	3.743

La Fig. 6 muestra el gráfico de dispersión de modelo para la

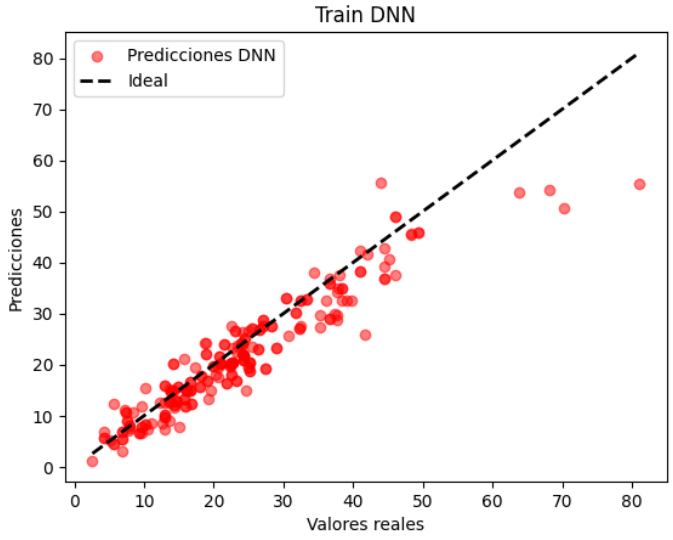


Fig. 6. Gráficos de dispersión del modelo de DNN para la data de training

La Fig. 7 muestra el gráfico de dispersión de modelo para la data de testing de las predicciones versus los valores reales del modelo.

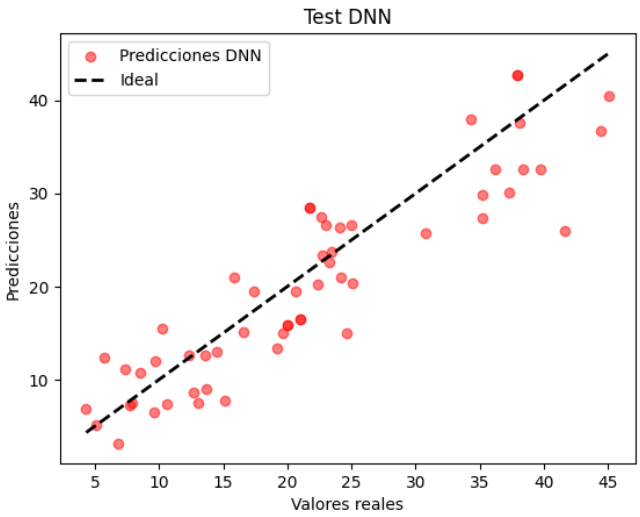


Fig. 7. Gráficos de dispersión del modelo de DNN para la data de test

3.3 Pruebas y resultados para el modelo con Arboles de decisión(Decision tree)

La tabla 4 muestra las métricas para el modelo de árboles de decisión, en la cual se puede observar que el modelo tiene un buen rendimiento en la predicción del CBR al 100%. Este valor indica que el modelo explica aproximadamente el 93.07% de la variabilidad en los valores de CBR al 100%. El valor de R^2 cercano a 1 sugiere que el modelo tiene un muy buen ajuste, capturando la mayor parte de la variabilidad en los datos. En cuanto a la importancia de las características, el orden es el siguiente: la densidad máxima seca (MDS) contribuye aproximadamente en 53.5.% a las decisiones del modelo, el porcentaje de finos es la segunda más importante, con una contribución del 23.1%, el límite plástico (LP) tiene una importancia del 10.6%, El óptimo contenido de humedad (OCH) tiene una importancia del 9.0%, aunque menos influyente que las tres primeras características, OCH sigue siendo relevante en la predicción del CBR al 100% y la característica Grava tiene la menor importancia, con una contribución del 3.8%. Asimismo, se excluyó el límite líquido (LL) y el índice de plasticidad (IP) porque influyen en la mejora del rendimiento del modelo. El número de nodos del árbol es 393, que nos indica la complejidad del modelo que está capturando detalles en los datos. Las métricas de error (MAE, RMSE y MSE) son relativamente bajas, sugiriendo que el modelo tiene una buena precisión.

TABLA 4
Métricas para el modelo Arboles de decisión

Algoritmo o modelo de machine learning	R^2	MSE	RMSE	MAE
árboles de decisión	0.9307	9.199	3.033	1.216

La Fig. 8 muestra el gráfico de dispersión de modelo para la data de training de las predicciones versus los valores reales del modelo.

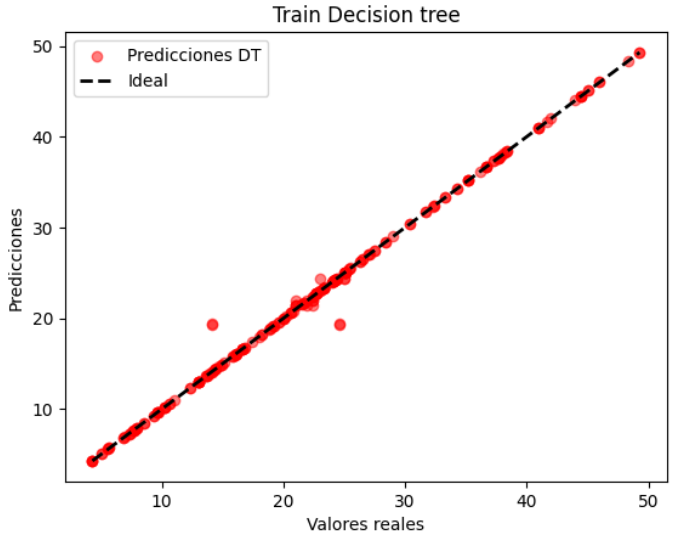


Fig. 8. Gráficos de dispersión del modelo de Decision tree para la data de training

La Fig. 9 muestra el gráfico de dispersión de modelo para la data de testing de las predicciones versus los valores reales del modelo.

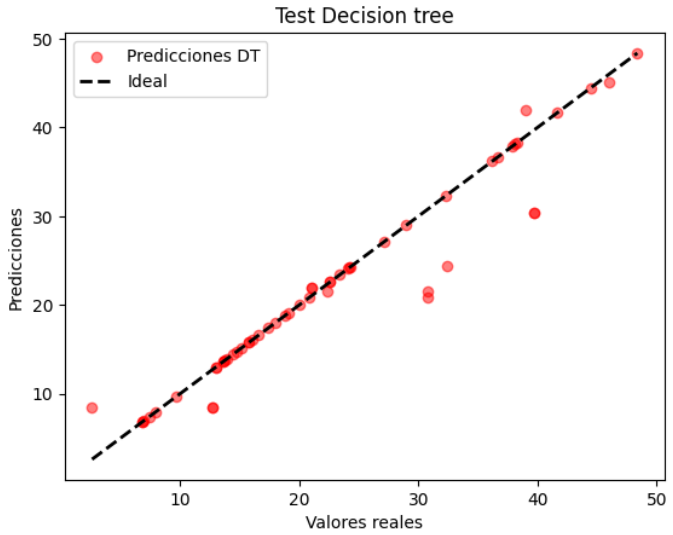


Fig. 9. Gráficos de dispersión del modelo de Decision tree para la data de testing

3.2 Discusión

En este estudio, se han entrenado y evaluado tres modelos de machine learning para predecir el CBR (california bearing ratio). Los modelos evaluados fueron máquinas de vectores de soporte de regresión (SVM), redes neuronales profundas (DNN) y árboles de decisión (DT), la tabla 6 muestra la comparación de métricas de los modelos donde podemos observar que el modelo de árboles de decisión (DT) ha demostrado ser el más eficaz para predecir el CBR basado en las características del suelo, superando a los modelos de SVR y DNN en todas las métricas de desempeño. El modelo de árboles de decisión no solo ofrece el mejor coeficiente de determinación (R^2), sino que también tiene los valores más bajos de MSE, RMSE y MAE, lo que indica una menor variabilidad del error y una mayor precisión en las predicciones.

TABLA 5
Comparación de los modelos

Algoritmo o modelo de machine learning	R^2	MSE	RMSE	MAE
SVM - Máquinas de Vectores de Soporte de Regresión	0.701	36.834	6.069	4.409
Redes neuronales profundas (DNN)	0.784	26.606	5.158	3.743
árboles de decisión (DT)	0.9307	9.199	3.033	1.216

4 Conclusiones

En esta investigación se corrobora que las propiedades físicas del suelo pueden ser utilizadas para predecir otras características como el CBR, puesto que, utilizando variables como el porcentaje de grava, el porcentaje de finos, el óptimo contenido de humedad, la densidad máxima seca, el límite líquido, el límite plástico, el índice de plasticidad del suelo y aplicando modelos de machine learning se puede predecir el índice de CBR al 100%. Asimismo, a pesar del tamaño limitado de muestras, los hallazgos son significativos para proporcionar una base sólida para futuras investigaciones.

Adicionalmente, este trabajo de investigación muestra que, el algoritmo más eficiente para la predicción del CBR al 100% basadonos en el coeficiente de determinación y las métricas de error es el modelo de árboles de decisión, seguido de redes neuronales profundas (DNN) y finalmente el modelo de máquinas de vectores de soporte de regresión (SVR).

5 Trabajos futuros

A partir de los resultados de esta investigación se puede ver el potencial que tiene el desarrollo de modelos de machine learning para la predicción de propiedades del suelo, por lo cual se espera que en trabajos futuros se puedan mejorar los modelos propuestos, incorporando características adicionales del suelo, como propiedades de compactación y características de deformación. Asimismo, se podrían aplicar y experimentar diferentes algoritmos como random forest. También se puede optar por el desarrollo de aplicaciones web que permitan a los profesionales predecir el CBR en tiempo real utilizando modelos de machine learning entrenados. Adicionalmente, se puede incorporar la automatización con el uso de sensores y dispositivos IoT para la recolección automática de datos del suelo. También sería genial fomentar la

colaboración entre ingenieros civiles, geotécnicos, científicos de datos para desarrollar modelos más integrales y precisos, asimismo desarrollar programas de educación y capacitación para ingenieros sobre el uso de técnicas avanzadas de machine learning aplicadas a la geotecnia.

AGRADECIMIENTOS

Se agradece la colaboración de los ingenieros del grupo corporativo Obregon S.C.R.L por sus conocimientos compartidos durante el desarrollo de la investigación. Agradecer también a los laboratorios de suelos y concreto Geolef E.I.R.L y Ecx ingenieros por el apoyo en la recolección de datos.

REFERENCIAS

[1] J. R. Lópe, "Herramientas digitales y uso de inteligencia artificial en geotecnia: Un enfoque en la evaluación de taludes con Matlab," no. 29 de enero de 2024, 2024. https://doi.org/10.56712/latam.v5i1.1640

[2] D. K. Talukdar, "A Study of Correlation Between California Bearing Ratio (CBR) Value With Other Properties of Soil," vol. 4, India, 2014.

[3] J. V.-L. J. E. A.-B. K. F.-F. C. L. C.-E. K. M. M.-A. E. N. L.-M. M. T. &. D. D. PERRET, Desarrollo de métodos de análisis de espectroscopia y algoritmos de aprendizaje automático para la evaluación de algunas propiedades del suelo en Costa Rica. Agronomía Costarricense, Universidad de Costa Rica. Colegio de Ingenieros y Agrónomos. Ministerio de Agricultura y Ganadería, 2022.

[4] D. L. COBA, M. HERRERA SUAREZ, M. M. GARCIA LORENZO and R. y BELTRAN, Modelo computacional para la estimación de la densidad del suelo a través del sensoramiento continuo. Revista Ciencia y Técnica Agrícola, vol. 27, 2018.

[5] V. C. LOME, "Análisis fotogramétrico de nube de puntos y aprendizaje automático como herramientas útiles en la caracterización de macizos rocosos," Puebla, Benemérita Universidad Autónoma de Puebla, 2023.

[6] M. J. CCASANI and Y. I. FERRO, "Evaluación y análisis de pavimentos en la ciudad de Abancay, para proponer una mejor alternativa estructural en el diseño de pavimentos," Abancay, Universidad Tecnológica de los Andes, 2017.

[7] M.T.C, "MANUAL DE CARRETERAS: DISEÑO GEOMÉTRICO DG – 2018," 2018.

[8] M.T.C, "MANUAL DE ENSAYO DE MATERIALES," Lima, Ministerio de Transportes y Comunicaciones, 2016.

[9] H. O. BECERRA, "ANÁLISIS Y ESTUDIOS DE SUELOS Y SU APLICACIÓN PARA EL MEJORAMIENTO DEL TRAMO 19 DE UNA CARRETERA EN LA PROVINCIA DE CORONEL PORTILLO UCAYALI 2018," Lima, 2020.

[10] J. W. Tukey, Exploratory Data Analysis, Massachusetts: Addison-Wesley, 1977.

[11] G. James, D. Witten, T. Hastie and R. Tibshirani, An Introduction to Statistical Learning, Springer, 2013. https://doi.org/10.1007/978-1-4614-7138-7

[12] J. E. V.-L. K. A.-B. Johan Perret, Desarrollo de métodos de análisis de espectroscopia y algoritmos de aprendizaje automático para la evaluación de algunas propiedades del suelo en Costa Rica, 2019.

[13] F. J. Valderrama, H. I. Mejía, S. P. Muñoz Pérez and V. Tuesta, "“Desarrollo de un modelo predictivo de las propiedades mecánicas del suelo usando redes neuronales artificiales," 2021.

[14] A. R. Paucar and S. Esteban, "Parámetros de resistencia al corte del suelo en función a sus propiedades físicas, empleando redes bayesianas y ensayo triaxial-Callqui Grande," 2023.

[15] M. M. O. R. Boza Capani, PARÁMETROS DE RESISTENCIA AL CORTE DE SUELOS A PARTIR DE SUS PROPIEDADES FÍSICAS, UTILIZANDO REDES NEURONALES ARTIFICIALES Y EQUIPO TRIAXIAL, UNH, 2018.

[16] K. TERZAGHI, R. B. PECK and G. MESRI, Soil mechanics in engineering practice., John wiley & sons, 1996.

BIOGRAFÍA

Flor de Cantuta Tello Sarmiento, bachiller en Ingeniería Informática y Sistemas de la Universidad Nacional Micaela Bastidas de Apurímac, actualmente es software engineer en la empresa Rappi S.A.C.

Manuel Jesús Ibarra Cabrera, doctor en ciencias de la computación e investigador en las áreas de ingeniería de software, serious game, informática educativa, computación móvil, IoT e industria y sociedad. Docente universitario de pre y pos grado en la Universidad Nacional Micaela Bastidas de Apurimac, Universidad Tecnológica de los Andes, Universidad Nacional del Altiplano y la Universidad Nacional Mayor de San Marcos. Actualmente, es miembro activo de conferencias importantes en latinoamerica: LACLO, CLEI, CONTIE, HCI, SABTIC, CISTI, Decisioning y otros.