

Determinación de la mejor Arquitectura de Redes Neuronales Convolucionales: VGG16, ResNet50 ó MobileNet para detección de la Neumonía 2023

The Determination of the Best Convolutional Neural Network Architecture: VGG16, ResNet50, or MobileNet for Pneumonia Detection in 2023

Enciso Ortiz Sergio Elías^A, Mamani Vilca Ecler^B y Ordoñez Ramos Erech^C

ORCID:0009-0005-0863-4889^A, ORCID:0000-0002-3058-9683^B y ORCID: 0000-0002-5205-3660^C

(Recepción: 8/08/2023 y aceptación 5/09/2023)

Resumen— Las Redes Neuronales Convolucionales (CNN) se utilizan para el reconocimiento de imágenes de radiografías y otras aplicaciones. Actualmente, existen trabajos que comparan la eficacia de las arquitecturas CNN, como VGG16, ResNet50 y MobileNet con parámetros de entradas diferentes en sus entrenamientos, creando incertidumbre entre los desarrolladores de aplicaciones de clasificación de imágenes, nosotros sí aplicamos entradas iguales para el entrenamiento de las CNN en estudio. Para abordar esta falta de información, se utilizó una base de datos de Kaggle que consta de 5856 imágenes. De esa base de datos, se seleccionó una muestra sistemática de 746 imágenes de radiografías de pulmones sanos y con Neumonía. Para asegurar la normalización de las imágenes, se utilizaron las herramientas *iloveimg* y *ReNamer*. Además, se utilizó Python con Google Colab y diversas librerías, como *tensorflow*, *matplotlib*, *numpy*, *os*, *cv2* y *random*, para ejecutar las diferentes arquitecturas. El diseño metodológico se basó en un enfoque cuantitativo, empleando tablas de comparación y las imágenes adquiridas de la base de datos de Kaggle. Los resultados obtenidos indicaron que el porcentaje de exactitud (Accuracy) fue de 80.83% para VGG16, 91.82% para ResNet50 y 78.28% para MobileNet, concluyéndose que ResNet50 es la arquitectura más precisa en este contexto.

Palabras claves: Arquitecturas de Redes Neuronales Convolucionales, MobileNet, ResNet50, VGG16.

Abstract— Convolutional Neural Networks (CNN) are used for the recognition of X-ray images and other applications. Currently, there are studies comparing the effectiveness of CNN architectures like VGG16, ResNet50, and MobileNet with different input parameters during their training, causing uncertainty among image classification application developers. In our study, we used consistent input parameters to address this lack of information. We employed a Kaggle database containing 5856 images. From this database, we systematically selected 746 images of X-rays depicting healthy and pneumonia-afflicted lungs. To ensure image normalization, we utilized tools such as *iloveimg* and *ReNamer*. Furthermore, we employed Python in conjunction with Google Colab and various libraries including *TensorFlow*, *Matplotlib*, *NumPy*, *os*, *cv2*, and *random* to execute the various CNN architectures. Our methodology was based on a quantitative approach, utilizing comparison tables and the images sourced from the Kaggle database. The results we obtained indicated an accuracy rate of 80.83% for VGG16, 91.82% for ResNet50, and 78.28% for MobileNet. We concluded that in this context, ResNet50 is the most accurate architecture.

Keywords: Convolutional Neural Network Architectures, MobileNet, ResNet50, VGG16.

- A. Enciso Ortiz Sergio Elías^A Escuela Académico Profesional de Ingeniería Informática y Sistemas de la Universidad Nacional Micaela Bastidas de Apurímac-Perú, sergioeliasencisoortiz@gmail.com
- B. Mamani Vilca Ecler^B Escuela Académico Profesional de Ingeniería Informática y Sistemas de la Universidad Nacional Micaela Bastidas de Apurímac- Perú, eclervirtual@unamba.edu.pe
- C. Ramos Ordoñez Erech^C Escuela Académico Profesional de Ingeniería Informática y Sistemas de la Universidad Nacional Micaela Bastidas de Apurímac - Perú. eordonhez@unamba.edu.pe

1 INTRODUCCIÓN

En el aprendizaje automático de la inteligencia artificial en adelante (IA) se basa en varios modelos computacionales y una de ellas es la Redes Neuronales Convolucionales (CNN, por sus siglas en inglés en adelante) se utilizan principalmente para tareas de procesamiento de imágenes y visión computacional, sin embargo dentro de ellas también existe diferentes arquitecturas como: Visual Geometry Group 16 (en adelante VGG16), Red Residual 50 (en adelante ResNet50) y Red Móvil (en adelante MobileNet), el problema es determinar **¿cuál de ellos tiene mejor accuracy?**, el objetivo es analizar y ejecutar las CNN VGG16, ResNet50 y MobileNet para determinar el que tiene mayor porcentaje de accuracy en el reconocimiento de Neumonía con imágenes RX mediante la métrica de evaluación (Matriz de confusión) y la hipótesis, es encontrar que exista una diferencia de accuracy entre las CNN VGG16, ResNet50 ó MobileNet, en la detección de la Neumonía.

En los trabajos de Maldonado Lezama, Lisset Fernanda y Moreano Rojas, Eder Omar [1], realizó el diagnóstico de enfermedades pulmonares cuyo resultado de accuracy fue la implementación propia de 96.19%, InceptionV3 de 86.46% y VGG16 de 96.32% siendo este el mejor, según Godoy Francisco, Vady Lee [2], el trabajo fue determinar el efecto del algoritmo de transferencia de aprendizaje con accuracy para resnet50 96.61%, VGG16 86.26% y RNN 95.89% donde ResNet50 resultó ser el mejor, también Caya Pérez, Jhan Carlos [3] realizó el requerimiento para el diagnóstico del análisis de RX de tórax, para evaluar el estado de los pulmones donde utilizó tres arquitecturas de CNN como: implementación propia, ResNet50 e InceptionV3, con accuracy de 96.59%, 97.63% y 98.48% respectivamente siendo InceptionV3 el mejor, así mismo según Cabrejos Yalán, Víctor Manuel [4] utilizaron CNN como: AlexNet, VGG y ResNet-50 para mejorar el diagnóstico de la Neumonía utilizando imágenes de RX, a fin de mejorar el tiempo de atención de pacientes con esta enfermedad y ayudar ante la demanda, obtuvieron resultados con accuracy del 95.22% para AlexNet que obtuvo el mejor resultado, 94.53% para VGG16 y 94.62% para ResNet50. Moura y otros [5] presenta la clasificación de imágenes RX de tórax de niños en dispositivos móviles utilizando CNN como MobileNetV1, MobileNetV2 e InceptionV3 con accuracy de 95.4%, 94.2% y 92.8% respectivamente siendo MobileNetV1 el mejor.

Se entiende por IA, Pino, Gómez y Martínez lo define como: “un campo de la ciencia y la ingeniería que se

ocupa de la comprensión, vista de un punto informático, que se denomina comportamiento inteligente” [6], en la medicina la IA es una rama de la informática capaz de analizar datos médicos complejos [7], también Recuenco y Reyes define como: La base sobre la que los sistemas se centran en el comportamiento para definirlos como inteligentes, así como la primera interpretación de tipo científico donde un sistema es inteligente cuando pasa con éxito el test de Turing [8] además Boden define como: IA no es una dimensión única, sino un espacio profusamente estructurado de capacidades diversas para procesar la información, de modo que esta utiliza muchas técnicas diferentes para resolver gran variedad de tareas [9]. El aprendizaje automático (Machine Learning) es una rama de IA. Gerón [10] lo define como: “El aprendizaje automático es la ciencia y el arte de programar máquinas para que puedan aprender de los datos”, En consecuencia, el aprendizaje automático puede ayudar a resolver problemas complejos, aprender de manera autónoma y descubrir patrones ocultos en grandes conjuntos de datos, lo que proporciona una valiosa herramienta tanto para la automatización como para el análisis de información. Este consta de tres partes: el algoritmo computacional, las variables y las funciones que conforman la decisión y el conocimiento base según el cual se sabe la respuesta que permite aprender al sistema (lo entrena). Así también se clasifican en: Aprendizaje supervisado, no supervisado y por refuerzo, el primero es automático se le proporciona una serie de datos de entrada cuya variable de salida es conocida, que tiene la finalidad de construir la regla general. El aprendizaje no supervisado no se utiliza la clase de salida esperada para los datos, el algoritmo no sabe nada. Por lo tanto, se tiene que estudiar el comportamiento de un conjunto de patrones de datos de entradas para determinar comportamientos parecidos que se agrupan y por último el consiste que algoritmo aprende por su cuenta propia [10], además el aprendizaje automático es combinar varios componentes del algoritmo, como una función de costo, una función de activación y una técnica de regularización, para formar un algoritmo de aprendizaje profundo completo [11].

Las Arquitecturas de Redes Neuronales según Martin del Brio y Serrano [12] lo denomina arquitectura a la topología, estructura o patrón de conexionado de una red neuronal. En general y de modo análogo a lo que sucede en el cerebro, las neuronas se suelen agrupar en unidades estructurales que denominaremos capas y el conjunto de una o más capas conforma la red neuronal y al inicio las arquitecturas tendrán muchas predicciones incorrectas [8]. Las CNN son un tipo especial de Redes Neuronales Artificiales, que esta réplica del funcionamiento del

cerebro humano es uno de los “modelos de moda” [13] y en este trabajo se definirán las antes mencionadas: VGG16 es un modelo de aprendizaje profundo presentado por Karen Simonyan y Andrew Zisserman en el artículo "Redes convolucionales muy profundas para el reconocimiento de imágenes a gran escala" [14], ResNet50 es según [15] ResNet50 es un modelo de Red Neuronal Convolucional propuesto en el 2016 por He et al. El cual, no posee una profundidad de capas fija y es dependiente del número de módulos que se utilicen, además permite un tamaño de entrada máximo de 224 x 224 y MobileNet es otro tipo de CNN que “reduce significativamente el número de parámetros en comparación con la red con convoluciones regulares con la misma profundidad en las redes” [16], más adelante cada arquitectura en estudio se definirán con más detalle. y la métrica a utilizar es la matriz de confusión para calcular así el accuracy de cada arquitectura analizada.

Con respecto al diagnóstico, es el arte de aproximarse al conocimiento de un ser humano, de su situación de salud física y emocional, mediante la comunicación verbal y no verbal y la exploración física de su organismo, en el marco de una relación respetuosa, cordial y la alerta, que tiene como fin la solución o alivio de los padecimientos o los riesgos descuidados descubiertos” [17] y se tiene las características del término diagnóstico como: clima adecuado en relación médico paciente, médico alerta con la vista aguda, conocimiento adecuado en el campo laboral y desarrollo de aptitudes, además, que se puede medir con la inferencia deductiva, descripciones teóricas, inferencia inductiva, inferencia análoga y inferencia estadística. Según la OMS [18] la neumonía es la principal causa de muerte infantil en todo el mundo. Se estima que esta mató a unos 920.136 niños menores de cinco años en 2015, lo que representa el 15% de todas las muertes entre niños en todo el mundo. La neumonía afecta a los niños y sus familias” y las características son: Streptococcus pneumoniae la causa más común, haemophilus influenzae de tipo b (Hib) es la segunda causa de neumonía bacteriana, el virus sincitial respiratorio es la causa común de neumonía viral, pneumocystis jiroveci es la principal causa de esta en niños menores de seis meses con VIH/SIDA causando una cuarta parte de las muertes en bebés VIH positivos, diagnóstico de streptococcus pneumoniae es la causa más común de neumonía bacteriana en niños, en consecuencia el diagnóstico de la neumonía es un componente fundamental de toda estrategia para reducir la mortalidad infantil y para el diagnóstico de neumonía suele basarse en la clínica y en la radiología [18] y se debe intentar una prueba microbiológica para orientar el tratamiento y el

diagnóstico nosológico” [19], luego el diagnóstico expone lo siguiente: Una radiografía de tórax analiza si hay inflamación en los pulmones. Se suele usar RX de tórax para detectar neumonía” [20], Los análisis de sangre, como un hemograma completo (CBC), se usan para saber si el sistema inmunitario está luchando contra una infección [20], La oximetría de pulso mide la cantidad de oxígeno en la sangre, este puede impedir que los pulmones dejen pasar suficiente oxígeno a la sangre [20] y un examen de RX para que el médico interprete dicho RX, el radiólogo buscará puntos blancos en los pulmones (llamados infiltrados) [20] y también las mediciones del diagnóstico es la presencia de taquipnea constituye uno de los indicadores más confiables de gravedad de la neumonía en los análisis univariados y multivariados, esto es sin embargo, en la práctica clínica, se recomienda que la frecuencia respiratoria mayor de 20 respiraciones/min sea considerada un signo de gravedad confiable en los pacientes con neumonía [21].

2 LAS ARQUITECTURAS DE REDES NEURONALES

2.1 VGG16

Según Pintos y Simonyan [22] tiene como entrada una imagen RGB de tamaño fijo de 224 x 224 píxeles [23]. "Redes convolucionales muy profundas para el reconocimiento de imágenes a gran escala" [14] la imagen se pasa a través de una serie de capas convolucionales (conv.) donde se utilizan filtros con un campo receptivo muy pequeño: 3 x 3 (que es el tamaño más pequeño para capturar la noción de izquierda/derecha, arriba/abajo, centro). VGG16 proviene del hecho de que el modelo tiene 16 capas en total. La Arquitectura se caracteriza por tener una estructura simple y uniforme, utilizando solo capas convolucionales de tamaño 3x3 con stride 1 y capas de pooling de tamaño 2x2 con stride 2. No se utilizan técnicas más avanzadas, como convoluciones dilatadas o convoluciones separables en profundidad. VGG16 es una Red Neuronal Convolucional profunda que utiliza convoluciones 3x3 y capas de pooling 2x2 para extraer características de una imagen. Consta de 13 capas convolucionales, 5 capas de pooling y 3 capas totalmente conectadas, seguidas de una capa de salida. Aunque esta arquitectura es más profunda y tiene más parámetros que algunas de las arquitecturas más modernas, su simplicidad y uniformidad han demostrado un buen rendimiento en muchas tareas de visión computacional.

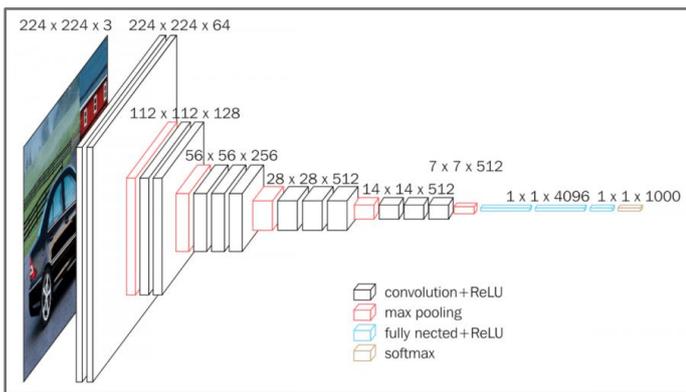


Fig. 1. Arquitectura de Red Neuronal Convolutiva VGG16 [14]

2.2 RESNET50

ResNet50 es una Red Neuronal Residual con 50 capas, su estructura es simple y efectiva, es beneficiosa por su amplia aplicación [24] como se observa en la figura 02. La red contiene un bloque de estructura residual, donde solo unos pocos dentro del mapeo de identidad, necesitan hacer una coincidencia de cotas y usar una capa convolutiva de 1x1 para agregar cotas. Con ResNet50 [25], conseguimos resolver el problema de desvanecimiento del gradiente, debido a que la red es demasiado profunda, además con esta red los gradientes pueden pasar directamente a través de las conexiones de salto hacia atrás desde las capas posteriores hasta los filtros iniciales, demostrando un rendimiento sobresaliente en tareas de clasificación de imágenes y detección de objetos. ResNet-50 ha influido en el diseño de numerosas arquitecturas posteriores y consta de más 23 millones de parámetros que claramente es menor a VGG16 y consiste en una serie de bloques residuales que aprenden características a diferentes niveles de abstracción. Al final, se utilizan capas de pooling, capas completamente conectadas y una capa de salida para realizar tareas específicas como clasificación de imágenes.

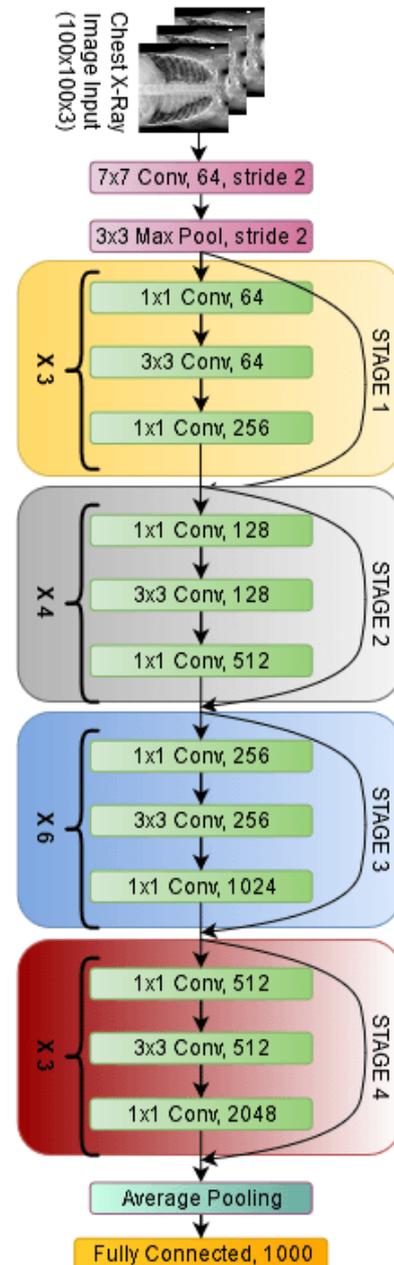


Fig. 2. Arquitectura de Red Neuronal Convolutiva ResNet50 [25]

2.3 MOBILENET

Para [26] MobileNet se basa en convoluciones separables en profundidad, que es una forma de convoluciones factorizadas que descomponen una convolución estándar en una convolución en profundidad y una convolución 1 x 1 llamada convolución de puntos. MobileNet "Reduce significativamente el número de parámetros en comparación con la red con convoluciones regulares con la misma profundidad en las redes" [27]. La convolución en profundidad aplica un único filtro a cada canal de entrada. Luego, la convolución de puntos aplica una convolución 1 x 1 para combinar las salidas de la convolución en profundidad y se basa en la convolución separable, que separa los filtros espaciales y de profundidad en dos etapas, así ayuda a reducir la carga computacional y el número de parámetros de la red, lo que la hace más eficiente en términos de recursos y tiempo de ejecución.

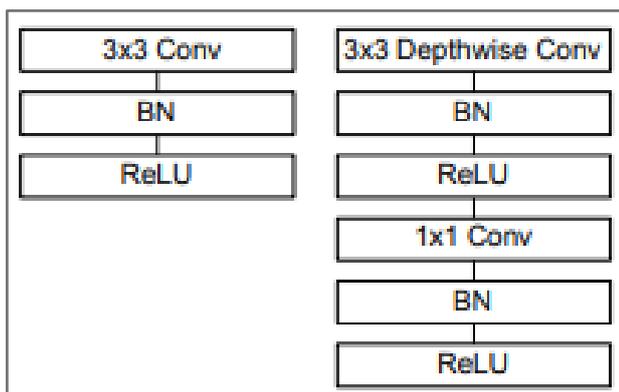


Fig. 3. Arquitectura de Red Neuronal Convolutiva MobileNet [28]

3 MÉTRICAS

3.1 Matriz

Para [29] la matriz de confusión es una de las métricas más intuitivas y sencillas para estimar la exactitud y otros, que es una tabla que describe el rendimiento de la arquitectura [30] y a partir de ella podemos calcular la métrica del accuracy [31].

TP	FN
FP	TN

Fig. 4. Matriz de confusión

3.2 Accuracy

Accuracy o exactitud es una de las métricas más utilizadas y confiables para conocer el rendimiento de las CNN [30], para Mora [29] es el cálculo en el porcentaje de muestras que se clasifican correctamente y se muestra la ecuación respectiva:

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

4 IMÁGENES

4.1 Imágenes de RX

Según [20] "Una radiografía de tórax analiza si hay inflamación en los pulmones. Se suele usar una radiografía de tórax para detectar neumonía" para interpretar los RX, el radiólogo buscará puntos blancos en los pulmones (llamados infiltrados) que son característicos de una infección, así como se muestra en la figura 5.



Fig. 5. Radiografía de un pulmón con neumonía [32]

5 METODOLOGÍA

La investigación es aplicada con nivel experimental y diseño factorial por bloques, para Hernández Sampieri y otros [33], los diseños factoriales manipulan dos o más variables independientes e incluyen dos o más niveles o modalidades de presencia en cada una de las variables independientes.

5.1 Población y Muestra

La población en este trabajo de investigación ha sido adquirida de la base de datos de Kaggle [32], entre datos de entrenamiento, validación y prueba todos suman la cantidad de la población total es 5856 y se trabajó con la carpeta de prueba (test) correspondiente a 746 datos que son imágenes de RX de pulmones sanos y de imágenes de RX de pulmones con la enfermedad de la Neumonía. Mediante el **Muestreo Sistemático** $K = 4$ se obtuvo $n = 187$.

5.2 Procedimiento

Primero se realizó búsquedas de CNN que tengan características de entradas semejantes, con la finalidad de que los resultados en el procesamiento de imágenes equivalente al tamaño de píxeles de 224×224 . Segundo se evaluó y se ejecutó las arquitecturas con los parámetros originales y también variando otros parámetros para poder comparar los diferentes resultados para luego determinar el mejor entre las tres, como se detalla en la sección 6.2. Tercera aplicación de instrumentos de recolección de datos que se obtuvo de la data de la base de datos de kaggle [32] y se normalizaron las imágenes recolectadas para tener el mismo tamaño y renombrar, con el propósito que no exista error a la hora del entrenamiento y testeo. Cuarto Análisis de los datos, el cual se utilizaron los datos para la prueba de las CNN.

5.3 Técnicas e instrumentos

Registro de Base datos, CNN como: VGG16, ResNet50 y MobileNet, herramientas software iloveimg [34] para redimensionar a un solo tamaño y ReNamer [35] para renombrar en bloque las imágenes, el uso de Python con Google Colab con diferentes librerías como tensorflow, matplotlib, numpy, os, cv2, random y otros.

6 RESULTADOS

6.1 Distribución de las imágenes

En la tabla 1 presenta la distribución de las imágenes donde se estableció la división de datos basándonos en trabajos similares sugiere la proporción [36] para los datos del entrenamiento un 80% del total, y para los datos de validación el 20%, la normalización de las imágenes de RX utilizando un software iloveimg [34] para redimensionar las imágenes a un solo tamaño de 224×224 y para renombrar se utilizó el software ReNamer [35].

TABLA 1

Distribución de las imágenes en directorios

Directorio Principal	Directorio secundario	Tipo	Cantidad
TRAIN	NORMAL	JPG	1104
	PNEUMONIA	JPG	2984
TEST	NORMAL	JPG	203
	PNEUMONIA	JPG	543
VAL	NORMAL	JPG	276
	PNEUMONIA	JPG	746
TOTAL		JPG	5856

6.2 Análisis y ejecución de las Arquitecturas de Redes Neuronales Convolucionales

Se emplean funciones de pérdida específicas en el análisis de las tres CNN. Estas funciones son "categorical_crossentropy" (loss = "categorical_crossentropy", optimizer = "adam", métrica = "accuracy") y "binary_crossentropy" (loss = "binary_crossentropy", optimizer = "adam", métrica = "accuracy"). La elección de estas funciones se basa en la naturaleza de este trabajo, que implica la clasificación en dos categorías: radiografías de pulmones normales y radiografías con neumonía. Como resultado, se llevarán a cabo dos conjuntos de experimentos para cada CNN. Uno de ellos utilizará los valores predeterminados a 50 épocas, mientras que el otro variará la función de pérdida a "binary_crossentropy" y se entrenará a 25 épocas, manteniendo constante el optimizador Adam [37] y la métrica Accuracy [30].

6.2.1 Experimento con VGG16

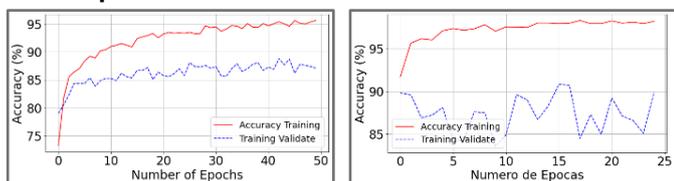


Fig. 6. Validación de la Arquitectura VGG16 de accuracy. **Izquierda:** Training accuracy con los parámetros predeterminados con 50 épocas de entrenamiento ambas curvas son similares dando un resultado aceptable. **Derecha:** Training accuracy con los parámetros modificados con 25 épocas de entrenamiento, las curvas son muy diferentes dando un resultado no aceptable.

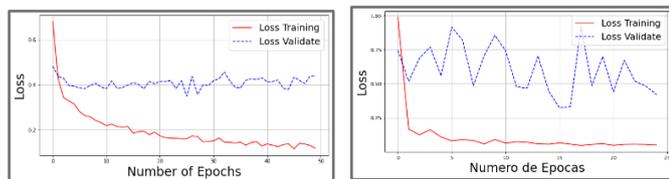


Fig. 7. Validación de la Arquitectura VGG16 de la función de pérdida. **Izquierda:** Training loss con los parámetros predeterminados con 50 épocas de entrenamiento ambas curvas con comportamiento similar y aceptable. **Derecha:** Training loss con los parámetros modificados con 25 épocas de entrenamiento, el comportamiento de la curva de validación es diferente cuyo resultado no es aceptable.

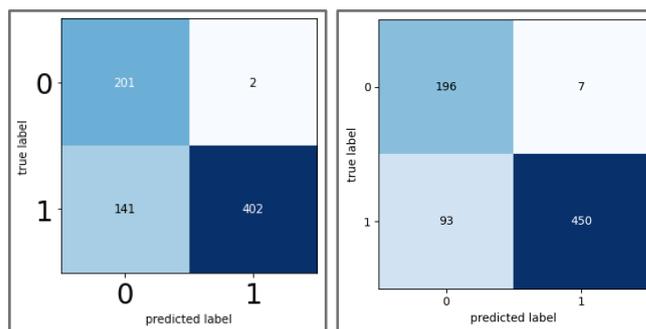


Fig. 8. Matriz de confusión de la Arquitectura VGG16 con 50 y 25 épocas de entrenamiento respectivo. **Izquierda:** es el resultado de una matriz de confusión con los parámetros predeterminados. **Derecha:** es el resultado de una matriz de confusión con los parámetros modificados.

A continuación, con los resultados de ambas matrices, calculamos el accuracy del experimento realizado con los valores originales de la CNN VGG16 con un resultado de 80.83% y también con los valores modificados de la función de pérdida con un resultado de 86.60%, como se muestran con la fórmula en ambos resultados.

$$Accuracy_{Predeterminado} = \frac{201 + 402}{201 + 402 + 2 + 141} = 80.83\%$$

$$Accuracy_{Modificado} = \frac{196 + 450}{196 + 450 + 7 + 93} = 86.60\%$$

6.2.2 Experimento con ResNet50

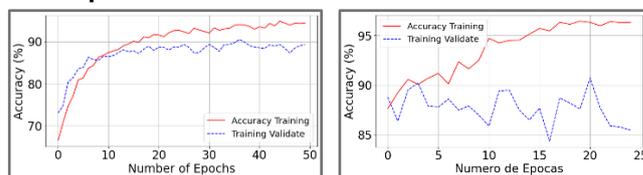


Fig. 9. Validación de la Arquitectura ResNet50 de accuracy. **Izquierda:** Training accuracy con los parámetros predeterminados con 50 épocas de entrenamiento cuyas curvas son idénticas y aceptable. **Derecha:** Training accuracy con los parámetros modificados con 25 épocas de entrenamiento cuyas curvas son diferentes y no es aceptable.

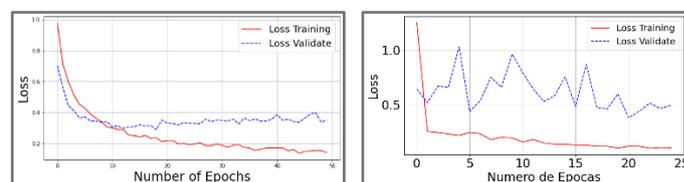


Fig. 10. Validación de la Arquitectura ResNet50 de la función de pérdida. **Izquierda:** Training loss con los parámetros predeterminados con 50 épocas de entrenamiento cuyas curvas con comportamiento similar y aceptable. **Derecha:** Training loss con los parámetros modificados con 25 épocas de entrenamiento cuya curva de validación es diferente a la del entrenamiento y no es aceptable.

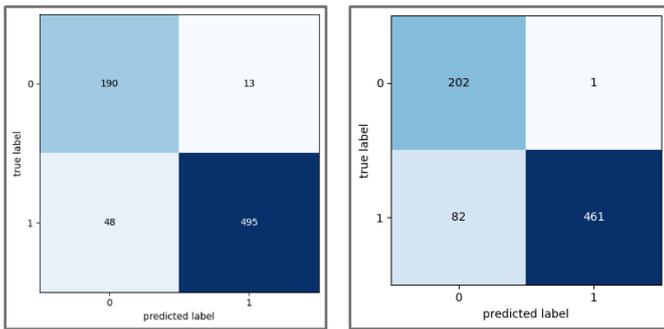


Fig. 11. Matriz de confusión de la Arquitectura ResNet50 con 50 y 25 épocas de entrenamiento respectivo. **Izquierda:** es el resultado de una matriz de confusión con los parámetros predeterminados. **Derecha:** es el resultado de una matriz de confusión con los parámetros modificados.

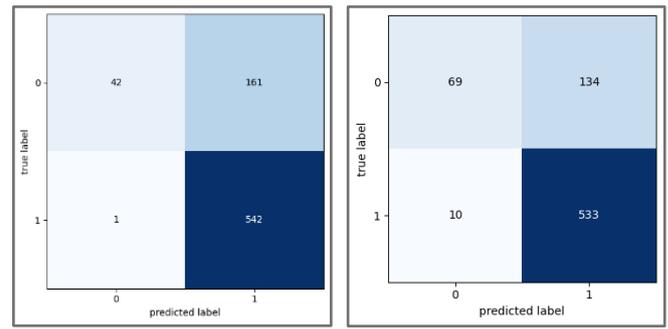


Fig. 14. Matriz de confusión de la Arquitectura MobileNet con 50 y 25 épocas de entrenamiento respectivo. **Izquierda:** es el resultado de una matriz de confusión con los parámetros predeterminados. **Derecha:** es el resultado de una matriz de confusión con los parámetros modificados.

A continuación, con los resultados de ambas matrices, calculamos el accuracy tanto del experimento realizado con los valores originales de la CNN ResNet50 con un resultado de 91.82% de accuracy y también con los valores modificados de la función de pérdida con resultado de 88.87% de accuracy, tal como se muestran con la fórmula para ambos resultados.

$$Accuracy_{Predeterminado} = \frac{190 + 495}{190 + 495 + 13 + 48} = 91.82\%$$

$$Accuracy_{Modificado} = \frac{202 + 461}{202 + 461 + 1 + 82} = 88.87\%$$

6.2.3 Experimento con MobileNet

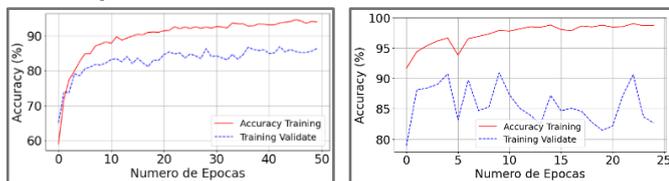


Fig. 12. Validación de la Arquitectura MobileNet de accuracy. **Izquierda:** Training accuracy con los parámetros predeterminados con 50 épocas de entrenamiento cuyas curvas con comportamiento similar y aceptable. **Derecha:** Training accuracy con los parámetros modificados con 25 épocas de entrenamiento y la curva de validación es diferente con respecto al de entrenamiento y no es aceptable.

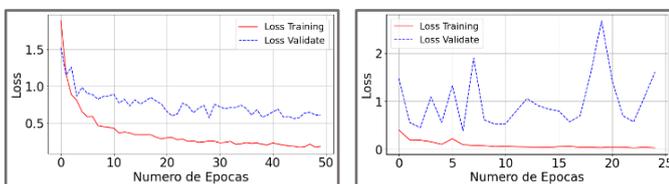


Fig. 13. Validación de la Arquitectura MobileNet de la función de pérdida. **Izquierda:** Training loss con los parámetros predeterminados con 50 épocas de entrenamiento con curvas con comportamiento semejante y aceptable. **Derecha:** Training loss con los parámetros modificados con 25 épocas de entrenamiento con la curva de validación diferente al de entrenamiento y no es aceptable.

A continuación, con los resultados de ambas matrices, calculamos el accuracy del experimento realizado con los valores originales de la CNN VGG16 con un resultado de 78.28% y también con los valores modificados de la función de pérdida con resultado de 80.70%, tal como se muestran con la fórmula para ambos casos.

$$Accuracy_{Predeterminado} = \frac{42 + 542}{42 + 542 + 161 + 1} = 78.28\%$$

$$Accuracy_{Modificado} = \frac{69 + 533}{69 + 533 + 134 + 10} = 80.70\%$$

6.3 Resultados de las Arquitecturas VGG16, RESNET50 Y MOBILENET

En esta sección se presenta la matriz de confusión de las Arquitecturas de Redes Neuronales Convolucionales VGG16, ResNet50 y MobileNet, Matriz de observación y Métricas de evaluación, las cuales fueron calculadas a partir de la ecuación explicada en la sección 3.2

En la Figura 15 se observa la matriz de confusión obtenida al finalizar el proceso de validación de las 2 clases. En esta matriz se muestra el cálculo de los verdaderos positivos y negativos, falso positivos y negativos, para cada una de las dos clases donde se tiene que "0" representa a los datos de las imágenes de RX de personas sanas y "1" representa a los datos de las imágenes de RX de personas con Neumonía.

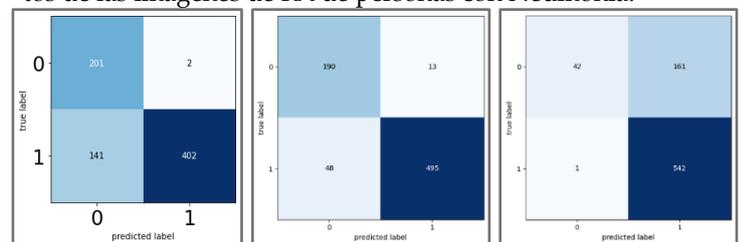


Fig. 15. Matriz de confusión de la Arquitectura con 50 épocas de entrenamiento. **Izquierda:** es el resultado de una matriz de confusión de VGG16 con los parámetros predeterminados. **Medio:** es el resultado de una matriz de confusión de ResNet50 con los parámetros predeterminados. **Derecha:** es el resultado de una matriz de confusión de MobileNet con los parámetros predeterminados.

A continuación, empleando los resultados obtenidos de las tres matrices, aplicamos la fórmula para calcular la precisión (accuracy) de las redes neuronales convolucionales VGG16, ResNet50 y MobileNet en los experimentos realizados utilizando sus valores originales o por defecto.

$$Accuracy_{VGG16} = \frac{201 + 402}{201 + 402 + 2 + 141} = 80.83\%$$

$$Accuracy_{ResNet50} = \frac{190 + 495}{190 + 495 + 13 + 48} = 91.82\%$$

$$Accuracy_{MobileNet} = \frac{42 + 542}{42 + 533 + 161 + 1} = 78.28\%$$

En el marco de esta investigación, se han obtenido resultados específicos que se detallan en la Tabla 2. En dicha tabla, se puede observar que la elección óptima para la Arquitectura de Redes Neuronales Convolucionales es la RESNET50. Por lo tanto, se sugiere emplear RESNET50 en la aplicación de detección de neumonía. Esto se recomienda con el propósito de proporcionar a los profesionales de la salud una herramienta esencial que les permita ofrecer diagnósticos más precisos en un tiempo más breve, evitando así largos períodos de espera para los resultados de los exámenes por parte de los pacientes.

TABLA 2

Resultado de las arquitecturas analizadas con accuracy

Nº	Parámetros de entranamiento	Arquitectura	Accuracy
1	(loss='categorical_crossentropy',	VGG16	80.83%
	optimizer='adam',	ResNet50	91.82%
	metrics=['accuracy']	MobileNet	78.28%
2	(loss='binary_crossentropy',	VGG16	86.60%
	optimizer='adam',	ResNet50	88.87%
	metrics=['accuracy']	MobileNet	80.70%

Entonces, se determina que la Arquitectura de Red Neuronal Convolutiva RESNET50 corresponde a la mayor puntuación de Accuracy, siendo la Arquitectura de Red Neuronal Convolutiva RESNET50 con una predicción del 91.82%, como se muestra en la siguiente Figura 16.

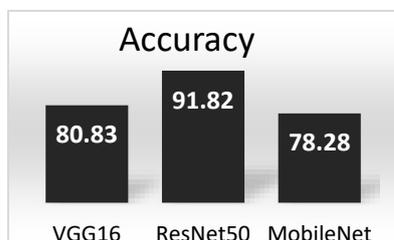


Fig. 16. Tasa de Reconocimiento de imágenes de RX de las Arquitecturas VGG16, ResNet50 y MobileNet

Estos resultados señalan claramente que, en lo que respecta a la predicción en el reconocimiento de imágenes de radiografías de neumonía, la Arquitectura ResNet50 es indiscutiblemente la que presenta el mayor porcentaje de acierto. En segunda posición se encuentra la Arquitectura VGG16, aunque

con una diferencia considerable respecto a la primera. Por último, la Arquitectura MobileNet ocupa el tercer lugar y se caracteriza por tener un porcentaje muy bajo de acierto, lo que la hace no recomendable para este tipo de investigación, tal como se ilustra en la Figura 16.

6.4 Análisis estadístico

Se utilizó la prueba de Friedman al no poderse normalizar los datos, el cual consiste en una prueba no paramétrica desarrollado por el economista Milton Friedman. Para varias muestras relacionadas. Equivalente a la prueba ANOVA para medidas repetidas es la no paramétrica, el método consiste en ordenar los datos por filas o bloques, reemplazándolos por su respectivo orden [37] que se explica en la sección 6.5 de contrastación de hipótesis.

6.5 Contrastación de hipótesis

Resultado de las CNN VGG16, ResNet50 y MobileNet para determinar el que tiene mayor porcentaje de accuracy en el reconocimiento de Neumonía con imágenes RX.

Planteamiento de la Hipótesis

Ho: $\mu_1 = \mu_2 = \mu_3$ [No hay diferencia entre las arquitecturas VGG16, ResNet50 o MobileNet en la precisión de la detección de la Neumonía]

Ha: $\mu_1 \neq \mu_2 \neq \mu_3$ [Alguna de las arquitecturas VGG16, ResNet50 o MobileNet es diferente, en la precisión de la detección de la Neumonía]

Nivel de significancia: $\alpha = 0.05$ y la confiabilidad del 95%

Estadístico

Tras intentar normalizar los datos y no lograrlo, se recurrió a la prueba no paramétrica de Friedman debido a que se obtuvo un valor de $p < 0.01$ y en la Figura 17, se puede observar claramente que la curva de color azul presenta una diferencia significativa en comparación con la pendiente de color rojo. Esto justifica la elección de la prueba de Friedman en estas circunstancias particulares.

Prueba de normalidad

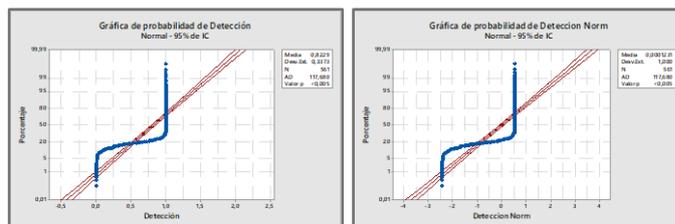


Fig. 17. La primera figura corresponde de accuracy de las tres arquitecturas sin normalizar y lo segundo corresponde datos que no cumplen con la normalización como $p < 0.01$, para los datos no normalizados corresponde a datos no paramétricos que utiliza la prueba de Friedman

Fórmula de la prueba de Friedman

$$X^2 = \frac{12}{Nk(k+1)} \sum_{j=1}^k R_j^2 - 3N(k+1)$$

Donde: **k** el número de tratamientos, **N** es el número total de observaciones y **R_j** = es la suma de los rangos de los tratamientos en el j-ésimo bloque

TABLA 3
Resultado de estadístico

Estadísticos de prueba	Valores
N	187
Chi-cuadrado	11,797
gl	2
Sig. asintótica	,003

Región crítica

Se rechaza la hipótesis nula **H₀** sí: **FrC > X²**

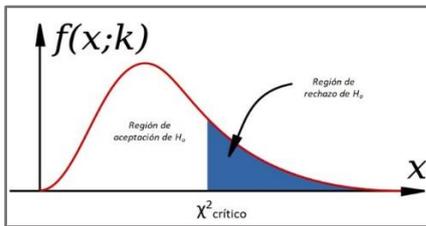


Fig. 18. Región de rechazo de **H₀**

Decisión:

FrC = 11.797 > X² = 5.99, por tanto, rechazamos la Ho y aceptamos la Ha donde una de las arquitecturas VGG16, ResNet50 ó MobileNet es diferente, en la precisión de la detección de la Neumonía.

TABLA 4

Resultados del promedio de cada Arquitectura

Arquitecturas	Rango promedio
RESNET50	2,11
VGG16	1,80
MOBILENET	2,09

La Arquitectura ResNet50 lidera en términos de promedio, lo que confirma que es la superior entre las tres Arquitecturas de Redes Neuronales Convolucionales analizadas en este estudio de investigación, destacándose especialmente en la precisión del diagnóstico de la neumonía.

6.6 Discusiones

Tras realizar los experimentos en nuestro trabajo y determinar la mejor arquitectura con accuracy más alta en porcentaje siendo la CNN ResNet50 la que obtuvo este reconocimiento, antes de eso los datos no pudieron ser normalizados motivo, por el cual se utilizó la prueba no paramétrica de Friedman para así llegar al resultado esperado por tanto para trabajos similares corresponde realizar con esta prueba mencionada y con los resultados obtenidos respecto al accuracy de VGG16 se obtuvo resultados con parámetro original/modifi-

cado 80.83%/86.60% respectivamente, ResNet50 con resultados con parámetro original/modificado 91.82%/88.87% respectivamente y MobileNet de resultados con parámetro original/modificado 78.28%/80.70% y para realizar la comparación con los trabajos de Maldonado y otros [1] que obtuvieron resultados de accuracy de la implementación propia de 96.19%, InceptionV3 de 86.46% y VGG16 de 96.32%, Godoy y otros [2] llegaron a resultados de accuracy para resnet50 96.61%, VGG16 86.26%, Caya y otros [3] llegaron a los resultados de accuracy de la implementación propia, ResNet50 e InceptionV3, con 96.59%, 97.63% y 98.48%, Cabrejos y otros [4] utilizaron CNN como: AlexNet, VGG y ResNet-50 llegando al resultado de accuracy de 95.22%, 94.53 y 94.62% y finalmente Moura y otros [5] presenta CNN como MobileNetV1, MobileNetV2 e InceptionV3 con accuracy de 95.4%, 94.2% y 92.8%. Nuestros resultados, aunque menores en términos de porcentaje en comparación con algunos de estos estudios, presentan una ventaja al obtener una mayor precisión en dos de los resultados en comparación con dos trabajos previos. Es importante destacar que, en nuestro enfoque, realizamos dos experimentos para cada CNN, el primero con parámetros originales con 50 épocas, y el segundo modificando la función de pérdida de "categorical_crossentropy" a "binary_crossentropy", dado que nuestro análisis involucra dos clases, manteniendo la métrica de accuracy y el optimizador adam a 25 épocas.

7 CONCLUSIONES Y RECOMENDACIONES

7.1 Conclusiones

Aplicando los mismo parámetros de entrada para el preprocesamiento y conjunto de datos en las tres arquitecturas de CNN es posible realizar la comparación con la métrica de evaluación denominado "Accuracy" [13], obteniendo los siguientes resultados según el orden de análisis de la investigación: 80.83% de Accuracy para VGG16, un 91.82% de Accuracy para ResNet50 y por último con un 78.28% de Accuracy para MobileNet y según los resultados obtenidos muestran que la Arquitectura de CNN ResNet50 es más efectiva que sobresale con el resto de sus similares, por lo que se determina que es la mejor Arquitectura de Red Neuronal Convolutiva.

7.2 Recomendaciones

Después de culminar el trabajo de investigación se recomienda utilizar una BD de la región de origen junto con la arquitectura de CNN ResNet50 para el desarrollo de aplicativos móviles o web como herramienta de apoyo a los profesionales en la salud, es por los altos valores obtenidos en la experimentación.

REFERENCIAS

[1] L. F. Maldonado y E. O. Moreano, «Diseño y aplicación de redes neuronales convolucionales para el diagnóstico de enfermedades pulmonares, a partir de radiografías de tórax frontal.» 2021.
[2] V. L. Godoy, «Algoritmo de diagnóstico preliminar de neumonía a partir de

imágenes radiográficas del tórax.» Lima, 2020.

- [3] J. C. Caya, «Evaluación de Modelos de Redes Neuronales Convolucionales aplicado a Radiografías de Tórax, para apoyar al Proceso de Diagnóstico de Neumonía asociada al COVID-19,» Lima, 2020.
- [4] V. M. Cabrejos, «Las Redes Neuronales Convolucionales y la mejora en el diagnóstico de Neumonía – área de Radiología,» Lima, 2022.
- [5] J. V. Moura Sousa, W. Rosa de Almedia, A. Andrade Saraiva, D. B. Sousa Santos, P. M. Cunha Pimentel, L. P. Lopez de Sousa y L. Lopez de Sousa, «Clasificación de imágenes de Neumonía en dispositivos móviles con Red Neuronal Cuantizada,» Brasil, 2020.
- [6] P. R. y G. A., «Introducción a la inteligencia artificial: sistemas expertos, redes neuronales artificiales y computación evolutiva,» 2001.
- [7] G. Dalvinder Singh, «A Critical Conceptual Analysis of Definitions of Artificial Intelligence as Applicable to Computer Engineering,» IOSR Journal of Computer Engineering (IOSR-JCE), vol. 16, nº 2, pp. 9-13, 2014. <https://doi.org/10.9790/0661-16210913>
- [8] A. D. Recuenco Cabrera y W. A. Reyes Alva, «Inteligencia artificial: Camino a un nuevo esquema del mundo,» Trujillo, 2020.
- [9] M. A. Boden, «Inteligencia artificial,» España, 2017.
- [10] A. Gerón, Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow, vol. Segunda, O'Reilly, 2019.
- [11] I. Goodfellow, Y. Bengio y A. Courville, «Deep Learning.»
- [12] B. Martin del Brio y C. Serrano Cinca, «Fundamentos de las redes neuronales artificiales: hardware y software,» Zaragoza, 1995. <https://doi.org/10.54886/scire.v1i1.1036>
- [13] L. J. Sandoval, «Algoritmos de aprendizaje automático para análisis y predicción de datos,» 2018.
- [14] Neuro, «VGG16 - Convolutional Network for Classification and Detection,» 28 Agosto 2022. [En línea]. Available: <https://neurohive.io/en/popular-networks/vgg16/>.
- [15] A. Gómez, S. Tabik, J. Luengo y F. Herrera, «Redes Neuronales Convolucionales para Una Clasificación Precisa de Imágenes de Corales. In XVIII Conferencia de la Asociación Española para la Inteligencia Artificial,» 2019.
- [16] Abhijeet Pujara, «Image Classification With MobileNet,» medium, 10 12 2022. [En línea]. Available: <https://builtin.com/machine-learning/mobilenet>.
- [17] Á. M. Botero y J. A. Chacon, El proceso diagnóstico, una aproximación al raciocinio clínico, Caldas: Editorial Universidad de Caldas, 2007.
- [18] OMS-Neumonía, «OMS-Neumonía,» 11 Noviembre 2021. [En línea]. Available: <https://www.who.int/es/news-room/fact-sheets/detail/pneumonia>.
- [19] C. J. Álvarez, Neumonías: concepto, clasificación y diagnóstico diferencial.
- [20] NIH, «National heart, lung, and blood institute,» 04 Agosto 2022. [En línea]. Available: <https://www.nhlbi.nih.gov/>.
- [21] F. Saldías y J. Pavié, «Evaluación de la gravedad en la neumonía del adulto adquirida en la comunidad,» Revista chilena de enfermedades respiratorias, Santiago, 2005. <https://doi.org/10.4067/S0716-10182005000400006>
- [22] N. Urbano Pintos, H. A. Lacomí y M. B. Lovarato, «B-VGG16: Red Neuronal de Convolución cuantizada binariamente para la clasificación de imágenes,» elektron ambar, Argentina, 2022. <https://doi.org/10.37537/rev.elektron.6.2.169.2022>
- [23] K. Simonyan y A. Zisserman, «Very deep convolutional networks for large-scale image recognition,» 2015.
- [24] É. Moreno Puig, «Reconocimiento de enfermedades a largo plazo,» 2022.
- [25] K. He, X. Zhang, S. Ren y J. Sun, «Deep Residual Learning for Image Recognition,» arXiv, 10 Diciembre 2015. <https://doi.org/10.1109/CVPR.2016.90>
- [26] A. G. Howard, Z. Menglong, C. Bo, K. Dmitry, W. Weijun, W. Tobias, A. Marco y A. Hartwig, «MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications,» 2017.
- [27] Abhijeet Pujara, «medium,» 10 12 2022. [En línea]. Available: <https://medium.com/analytics-vidhya/image-classification-with-mobilenet-cc6fb2cd470>.
- [28] P. Sirawan y O. Surinta, «Food Image Classification with Improved MobileNet Architecture and Data Augmentation,» 2020.
- [29] E. A. Huerta Mora, V. Gonzales Huitrón, H. Rodriguez Range y L. E. Amabilis Sosa, «Detección de enfermedades foliares con arquitecturas de redes neuronales convolucionales,» RINDERESU, Mexico, 2020.
- [30] W. A. Lozada Portilla, M. J. Suarez Barón y E. Avendaño Fernandez, «Aplicación de redes neuronales convolucionales para la detección del tizón tardío Phytophthora infestans en papa Solanum tuberosum,» 2021.
- [31] J. C. Valero Gómez, A. P. Zuñiga Incalla y J. C. Clares Perca, «Detección de la tuberculosis con algoritmos de Deep Learning en imágenes de radiografías del tórax,» VIVE, Moquegua, 2021. <https://doi.org/10.33996/revista-vive.v4i12.119>
- [32] Kaggle, «Kaggle,» 2023. [En línea]. Available: <https://www.kaggle.com/datasets/paultimothymooney/chest-xray-pneumonia>. [Último acceso: 22 Agosto 2022].
- [33] R. Hernández, C. Fernández y M. d. P. Baptista, Metodología de la investigación, 6ta Edición ed., 2000.
- [34] Iloveimg, «Iloveimg,» [En línea]. Available: <https://www.iloveimg.com/es/redimensionar-imagen#resize-options,pixels>. [Último acceso: 07 Julio 2023].
- [35] softonic, «softonic,» 23 Mayo 2023. [En línea]. Available: <https://renamer-free.softonic.com/>. [Último acceso: 07 Julio 2023].
- [36] A. Rác, D. Bajusz y K. Héberger, «Effect of Dataset Size and Train/Test Split Ratios in QSAR/QSPR Multiclass Classification,» 2021. <https://doi.org/10.3390/molecules26041111>
- [37] D. P. Kingma y B. Jimmy, «Adam: A method for stochastic optimization,» arXiv, 22 Diciembre 2014.
- [38] A. Quispe A, K. M. Calla V, J. S. Yangali V, J. L. Rodríguez L y I. I. Pumacayo P, «Estadística no paramétrica aplicada a la investigación científica con software SPSS, MINITAB Y EXCEL,» EIDEC, Colombia, 2019.

BIOGRAFÍAS

Sergio Elías Enciso Ortiz, Técnico Profesional en Electrónica en el Instituto Superior Tecnológico IDAT y bachiller en Ingeniería Informática y Sistemas de la Universidad Nacional Micaela Bastidas de Apurímac.

Ecler Mamani Vilca, Universidad Nacional Micaela Bastidas de Apurímac - Perú, Dr. en Ciencias de la Computación, desarrollador de aplicaciones multimedia.

Erech Ordoñez Ramos, catedrático Asociado en la Universidad Nacional Micaela Bastidas de Apurímac - Perú, Director de la Carrera Profesional de Ingeniería Informática y Sistemas.