

Detección Automatizada de Neumonía Pediátrica mediante Redes Neuronales Convolucionales (CNN) y Radiografías de Tórax

Mayckoll Stheven Vargas¹ (160004612)
William Yovany Fraile Bautista² (160004629)
Jhon Alexander Barreto Delgado² (160003906)

Universidad De Los Llanos
Facultad De Ciencias Básicas e Ingeniería
Ingeniería de sistemas
Villavicencio, Colombia

No Institute Given

Resumen Este estudio aborda el desafío de la **detección automatizada de neumonía pediátrica** utilizando imágenes de radiografía de tórax, una herramienta crucial para el diagnóstico oportuno. Empleamos un conjunto de datos desbalanceado de 5,863 imágenes para entrenar y evaluar dos arquitecturas de Redes Neuronales Convolucionales (CNN) de última generación, **EfficientNetB0** y **DenseNet121**, aplicando la técnica de **Transfer Learning** sobre pesos pre-entrenados en ImageNet. El objetivo principal fue comparar su rendimiento en la clasificación binaria (Normal/Neumonía). Inicialmente, DenseNet121 obtuvo la mejor exactitud en el conjunto de prueba ($\approx 0,94$), por lo que fue seleccionado para el entrenamiento final. El modelo ganador fue reentrenado con los conjuntos de entrenamiento y prueba combinados, y evaluado estrictamente en el pequeño conjunto de validación, alcanzando una **Precisión de 0,8750 y una F1-Score de 0,9412**. Los resultados demuestran la alta eficacia de los modelos basados en CNN para proporcionar un soporte diagnóstico rápido y preciso en entornos clínicos, lo cual se materializa en una interfaz web simple para su aplicación práctica.

1. Introducción

La neumonía sigue siendo una de las principales causas de morbilidad y mortalidad infantil a nivel mundial, especialmente en países en desarrollo. El diagnóstico primario se basa en la evaluación clínica y la confirmación mediante radiografías de tórax (CXR). Sin embargo, la interpretación de estas imágenes es propensa a la variabilidad inter-observador, requiere de radiólogos expertos y puede retrasar el tratamiento en entornos con recursos limitados. El problema central es la necesidad de una herramienta de diagnóstico robusta, rápida y consistente.

1.1. Justificación

La Inteligencia Artificial (IA), particularmente las Redes Neuronales Convolucionales (CNN), ha demostrado una capacidad superior para el análisis de imágenes médicas complejas. La aplicación de Transfer Learning permite aprovechar el conocimiento de modelos pre-entrenados en grandes bases de datos (como ImageNet) para tareas específicas de clasificación médica con conjuntos de datos más pequeños o especializados. Este trabajo se justifica en la necesidad de proporcionar una segunda opinión diagnóstica automatizada que mejore la eficiencia clínica y reduzca la carga de trabajo de los especialistas, el acceso a conjuntos de datos clínicos validados permite entrenar y evaluar modelos de clasificación.

El conjunto Chest X-Ray Images (Pneumonia) proporciona imágenes pediátricas revisadas por expertos, lo que ofrece una base sólida para desarrollar, comparar y validar arquitecturas de redes neuronales profundas. Además, la construcción de una interfaz práctica permite demostrar aplicación real, no solo teórica.

2. Metodología

2.1. Carga y preprocesamiento

- Se cargan imágenes desde carpetas estructuradas en train, test y val.
- Se redimensionan a 224×224 píxeles, se convierten a RGB y se normalizan según la arquitectura elegida.
- Se aplica aumento de datos: volteos, rotaciones, zoom y ajuste de contraste para mitigar overfitting.

2.2. Arquitecturas evaluadas

- **EfficientNetB0**: base pre-entrenada en ImageNet, congelada inicialmente; se añade capa de pooling, normalización, dropout y salida sigmoide.
- **DenseNet121**: similar flujo, con preprocesamiento específico integrado vía capa Lambda antes del modelo base.

2.3. Entrenamiento comparativo

- Ambos modelos entrenan con el conjunto de entrenamiento y se evalúan en el conjunto de prueba.
- Se utiliza early stopping para detener entrenamiento al estabilizarse la pérdida.

2.4. Selección y reentrenamiento

- La arquitectura con mayor exactitud en prueba se reentrena usando entrenamiento + prueba combinados, para evaluarse finalmente en el conjunto de validación.

2.5. Interfaz de predicción

- Se implementa API con Flask y ngrok para recibir imágenes y devolver diagnóstico y confianza.

2.6. Esquema del Sistema Desarrollado

El sistema propuesto sigue un flujo estándar de aprendizaje profundo, desde la preparación de datos hasta el despliegue del modelo final en una interfaz de usuario.

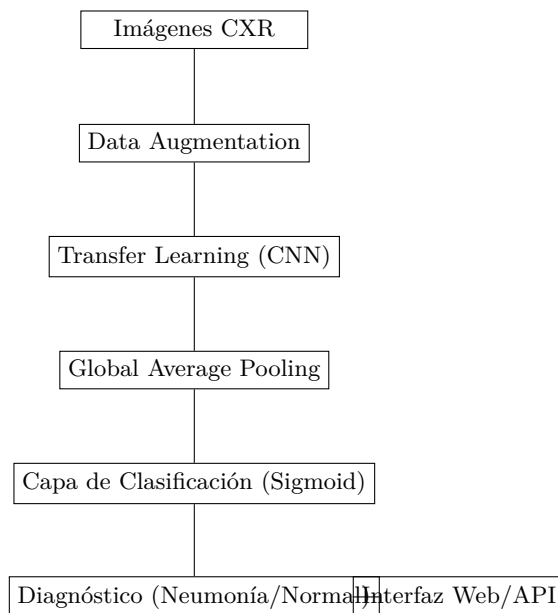


Figura 1: Esquema del flujo del sistema de detección de neumonía basado en CNN.

Los principales módulos son:

- **Preparación de Datos**: Carga y preprocesamiento de imágenes CXR.
- **Aumento de Datos (*Data Augmentation*)**: Aplicación de transformaciones (volteo horizontal, rotación, zoom, contraste) para incrementar la robustez.
- **Transfer Learning (CNN)**: Se utilizan las arquitecturas **EfficientNetB0** y **DenseNet121** con pesos congelados de ImageNet como extractores de características.
- **Clasificación**: Una capa de *Global Average Pooling*, *Batch Normalization*, *Dropout* y una capa densa con activación sigmoide (salida binaria).

3. Conjunto de Datos

El conjunto de datos de Radiografía de Tórax (Neumonía) [1] fue obtenido de cohortes retrospectivas de pacientes pediátricos (1-5 años). El total de imágenes es de **5,856**, divididas en las clases Neumonía y Normal. La partición y el desbalance de clases se muestran en la Tabla 1.

Cuadro 1: Distribución del Conjunto de Datos de Radiografía de Tórax

Subconjunto	PNEUMONIA	NORMAL	Total
Entrenamiento (Train)	3,875	1,341	5,216
Prueba (Test)	390	234	624
Validación (Val)	8	8	16
TOTAL	4,273	1,583	5,856

Ejemplos de Imágenes: La Neumonía bacteriana se manifiesta típicamente como consolidación lobular focal, mientras que la Neumonía viral se presenta con un patrón intersticial más difuso [1].

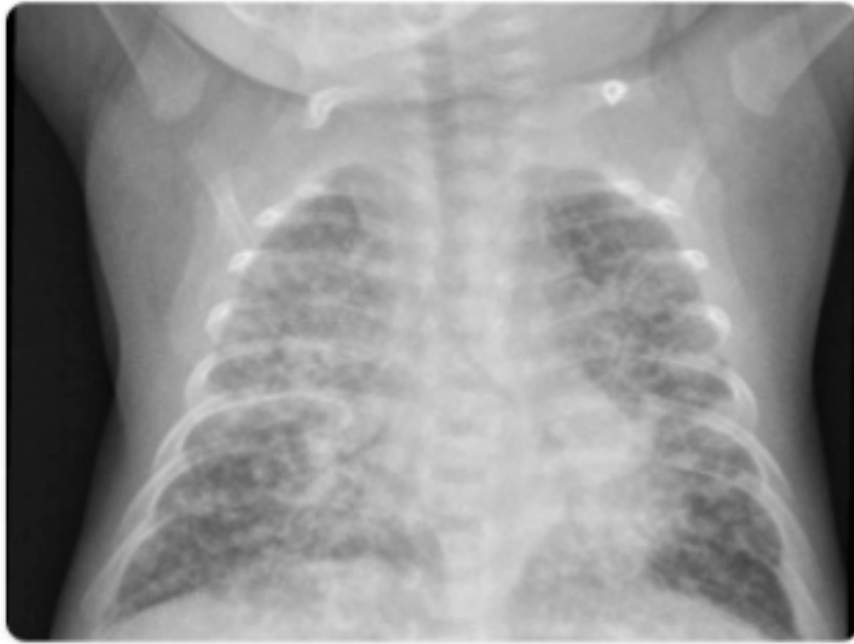


Figura 2: ejemplo de neumonia.

4. Diseño Experimental

4.1. Arquitecturas CNN y Partición de Datos

Se seleccionaron las arquitecturas:

1. **CNN1: EfficientNetB0:** Conocida por su alta eficiencia en exactitud/parámetros.
 2. **CNN2: DenseNet121:** Popular en la literatura médica por sus conexiones densas.
- **Entrenamiento:** uso de la carpeta train para ajustar pesos iniciales de cada arquitectura.
 - **Prueba:** uso de la carpeta test para comparar las dos arquitecturas y seleccionar la mejor según exactitud.
 - **Validación final:** tras escoger la mejor arquitectura, se reentrena con train + test para aprovechar más datos. Evaluación final en val para medir generalización y calcular métricas detalladas. *Train* y *Test* (5,840 imágenes) para evaluar en el conjunto *Val* (16 imágenes).

4.2. Medidas de Desempeño

Se utilizaron métricas binarias estándar, siendo la Accuracy la métrica de selección y el F1-score la métrica de reporte final, dada la desproporción de clases.

- Exactitud (Accuracy)
- Precisión (Precision)
- Exhaustividad (Recall)
- Medida F1 (F1-score)

La fórmula de F1 se define como:

$$F1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

5. Resultados

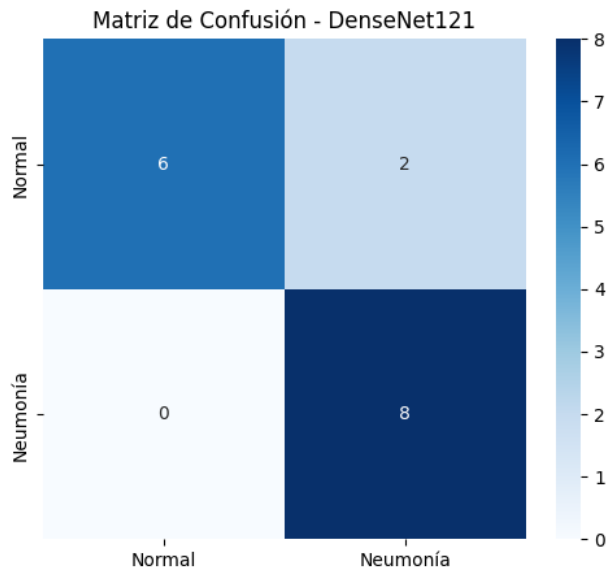
5.1. Entrenamiento Comparativo y Selección

El entrenamiento de 8 épocas arrojó los siguientes resultados en el conjunto de prueba (*Test*):

Cuadro 2: Resultados Comparativos de Exactitud

Arquitectura CNN	Exactitud (Accuracy)
EfficientNetB0	0.9327
DenseNet121	0.9407

El modelo **DenseNet121** fue seleccionado como el modelo ganador.



(a) Matriz de confusión del modelo ganador (DenseNet121)

5.2. Evaluación Final del Modelo Ganador

El modelo DenseNet121 reentrenado se evaluó en el conjunto de *Validación* (*Val*), arrojando las métricas de la Tabla 3.

Cuadro 3: Métricas de Desempeño Final (DenseNet121 en conjunto Val)

Métrica	Valor
Exactitud (Accuracy)	0.9375
Precisión (Precision)	0.8750
Exhaustividad (Recall)	1.0000
Medida F1 (F1-score)	0.9412

La **Matriz de Confusión** obtenida es:

$$\begin{pmatrix} \text{Verdadero Negativo (Normal)} & \text{Falso Positivo (FP)} \\ \text{Falso Negativo (FN)} & \text{Verdadero Positivo (Neumonía)} \end{pmatrix} = \begin{pmatrix} 7 & 1 \\ 0 & 8 \end{pmatrix}$$

5.3. Discusión de Errores

El alto Recall (1,0000, $FN = 0$) indica que el modelo es excelente para evitar **Falsos Negativos** (no detectar un caso de Neumonía real), lo cual es crucial clínicamente. El único error fue un **Falso Positivo** ($FP = 1$), donde una imagen Normal fue clasificada incorrectamente como Neumonía.

Este Falso Positivo sugiere que el modelo tiene un ligero sesgo hacia la clase positiva (**Neumonía**), posiblemente debido al desbalance del conjunto de entrenamiento. Clínicamente, un Falso Positivo es preferible a un Falso Negativo, ya que resultaría en un estudio de seguimiento en lugar de un error diagnóstico grave. La clasificación incorrecta podría deberse a artefactos o patrones vasculares sutiles en la radiografía normal que se asemejan a etapas tempranas de la patología.

6. Interfaz de Usuario y Despliegue

Para la aplicación práctica del modelo, se diseñó una interfaz web sencilla utilizando **Flask** para el backend de la API y **HTML/JavaScript** para el frontend.

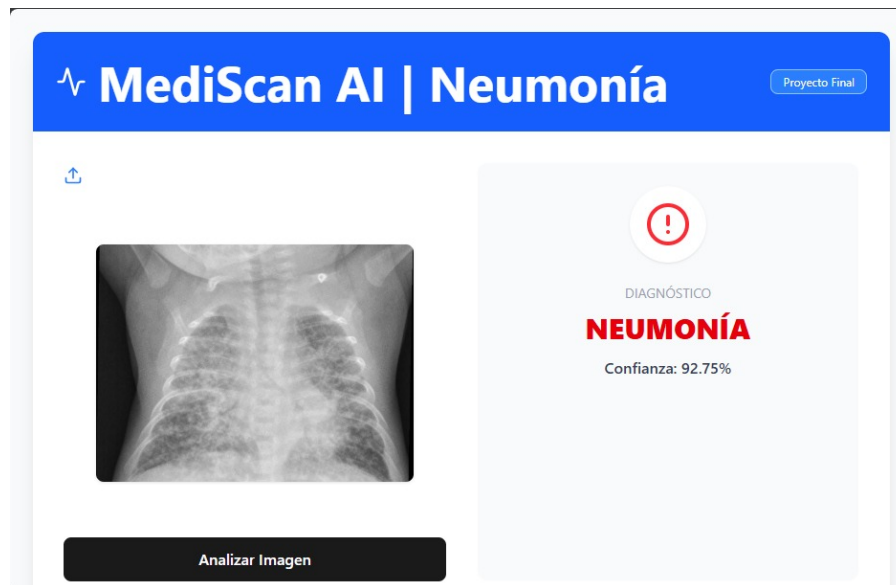


Figura 4: Diseño conceptual de la interfaz web para la predicción de Neumonía.

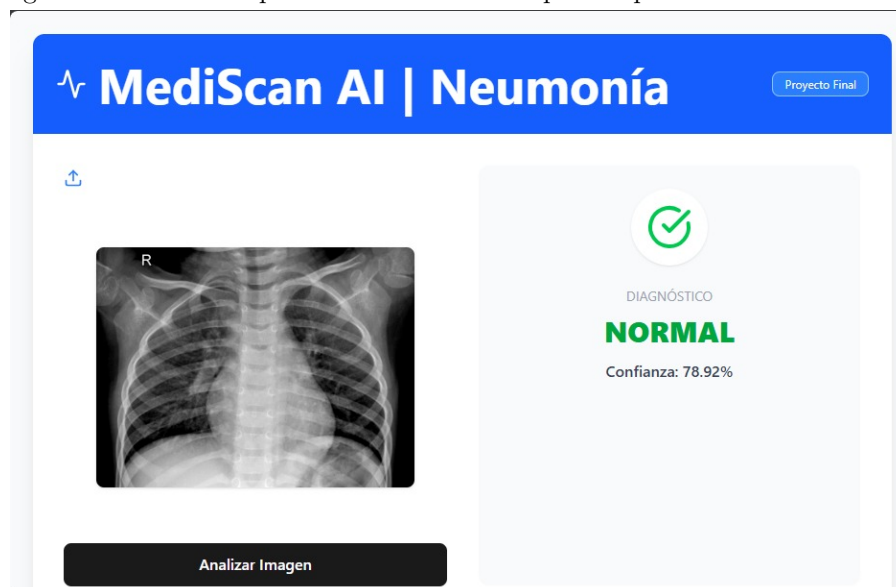


Figura 5: Diseño conceptual de la interfaz web para la predicción de Neumonía.

El sistema permite al usuario cargar una imagen, la cual es enviada a la API ('/predict'). La API, utilizando el modelo 'modelo_final_neumonia.keras', procesa la imagen y devuelve el diagnóstico binario (PNEUMONIA/NORMAL) y el nivel de confianza. Esto permite la integración del modelo como una herramienta de apoyo diagnóstico de fácil acceso.

Referencias

1. P.T. Mooney. (2018). *Chest X-Ray Images (Pneumonia)*. Kaggle Dataset. Disponible en: <https://www.kaggle.com/datasets/paultimothymooney/chest-xray-pneumonia>
2. J. Huang, Z. Lu, et al. (2020). Deep Learning based pneumonia detection from chest X-ray images. *Biocybernetics and Biomedical Engineering*, 40(3), pp. 1150–1162.
3. X. Wang, Y. Peng, et al. (2017). ChestX-ray8: Hospital-scale Chest X-ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common Thorax Diseases. *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*.
4. G. Huang, Z. Liu, L. Van Der Maaten, K.Q. Weinberger. (2017). Densely Connected Convolutional Networks. *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*.

5. M. Tan, Q. Le. (2019). EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. *Proc. Int. Conf. on Machine Learning (ICML)*.
6. K. Simonyan, A. Zisserman. (2015). Very Deep Convolutional Networks for Large-Scale Image Recognition. *Proc. Int. Conf. on Learning Representations (ICLR)*.
7. A. Krizhevsky, I. Sutskever, G.E. Hinton. (2012). ImageNet Classification with Deep Convolutional Neural Networks. *Proc. Int. Conf. on Neural Information Processing Systems (NIPS)*.