



# Proyecto final

*Equipo 6*

---

Integrantes

**Sara Osorio Piñerez  
Darcy Daniela Ochoa Henao  
Juan Sebastian Fernández Montoya  
Juan Jose Giraldo Muñoz**

Medellín - Antioquia

Agosto 21 de 2025

# Índice

<b>1. Motivación</b>	<b>3</b>
<b>2. Introducción</b>	<b>3</b>
<b>3. Problemática</b>	<b>3</b>
<b>4. Objetivos</b>	<b>4</b>
4.1. Objetivo general: . . . . .	4
4.2. Objetivos específicos: . . . . .	4
<b>5. Metodología:</b>	<b>4</b>
<b>6. Resultados</b>	<b>5</b>
6.1. Modelo CNN sin ajuste de pesos . . . . .	5
6.2. Modelo CNN con ajuste de pesos . . . . .	5
6.3. Comparación de desempeño: CNN sin pesos vs. CNN con ajuste de pesos . . . .	6
6.4. Entrenamiento con modelo preentrenado EfficientNetB0 . . . . .	8
6.5. Visualización de atención mediante mapas de calor . . . . .	9
<b>7. Análisis y discusión</b>	<b>11</b>
<b>8. Conclusiones</b>	<b>11</b>
<b>9. Trabajo futuro / Recomendaciones</b>	<b>11</b>
<b>10. Referencias</b>	<b>12</b>
<b>Referencias</b>	<b>12</b>

## 1. Motivación

La motivación de este proyecto nace del compromiso con una atención médica más equitativa, eficiente y accesible. En un mundo donde millones de personas enfrentan barreras geográficas, económicas o estructurales para acceder a diagnósticos oportunos, la inteligencia artificial ofrece una vía para reducir estas desigualdades. Automatizar la detección de neumonía en radiografías de tórax no solo representa un avance técnico, sino también una respuesta concreta a la necesidad de mejorar la cobertura médica en regiones desatendidas. Este trabajo busca demostrar que, con herramientas adecuadas y modelos bien diseñados, es posible apoyar a los profesionales de la salud, acelerar el diagnóstico y contribuir a salvar vidas, especialmente en contextos donde cada minuto cuenta.

## 2. Introducción

La neumonía es una infección pulmonar grave que inflama los alvéolos, causando síntomas como tos, fiebre y dificultad para respirar. A nivel global, afecta aproximadamente a 450 millones de personas al año y provoca cerca de 4 millones de muertes [1]. Su detección temprana es fundamental para iniciar el tratamiento adecuado y mejorar las tasas de recuperación.

La radiografía de tórax es la herramienta de imagen más común para diagnosticar neumonía [2], pero su interpretación depende de la experiencia del radiólogo y está sujeta a variabilidad humana.

En la actualidad, la inteligencia artificial ha demostrado un impacto significativo en el ámbito médico, particularmente en el diagnóstico asistido por computadora [3]. El análisis de imágenes médicas mediante redes neuronales permite detectar patologías con alta precisión, agilizando los procesos diagnósticos y apoyando la labor de los especialistas. Entre estas aplicaciones, la detección de neumonía a través de radiografías de tórax es una de las más relevantes, ya que esta enfermedad representa una de las principales causas de morbilidad y mortalidad infantil en el mundo.

## 3. Problemática

El diagnóstico tradicional de neumonía requiere la interpretación de radiografías de tórax por parte de radiólogos experimentados. Sin embargo, la disponibilidad de estos especialistas es limitada, especialmente en áreas rurales o con infraestructura médica deficiente, lo que genera retrasos críticos en el tratamiento. Esta situación no solo aumenta el riesgo de complicaciones graves, como insuficiencia respiratoria o sepsis, sino que también contribuye a una de las principales causas de mortalidad infantil en el mundo: según la OMS, la neumonía es responsable de aproximadamente el 15 % de todas las muertes de niños menores de 5 años a nivel global [3].

A esto se suma el hecho de que la interpretación manual de radiografías puede estar sujeta a variabilidad entre observadores, lo que implica diagnósticos inconsistentes. En un contexto donde cada minuto cuenta, la Inteligencia Artificial emerge como una aliada estratégica, capaz de procesar y analizar imágenes médicas en segundos, identificando patrones invisibles al ojo

humano y reduciendo drásticamente los tiempos de respuesta. El desarrollo de estas herramientas no solo promete salvar vidas, sino también democratizar el acceso a diagnósticos de alta precisión, rompiendo barreras geográficas y económicas que históricamente han limitado la atención médica.

## 4. Objetivos

### 4.1. Objetivo general:

Desarrollar un modelo de inteligencia artificial basado en redes neuronales que, a partir de radiografías de tórax, clasifique si un paciente tiene neumonía o no.

### 4.2. Objetivos específicos:

1. Descargar y explorar la base de datos para entender su estructura y calidad.
2. Preprocesar las imágenes para optimizar su uso en el modelo, incluyendo redimensionado, normalización y división en conjuntos de entrenamiento, validación y prueba.
3. Entrenar un modelo de red neuronal convolucional empleando arquitecturas preentrenadas para mejorar el rendimiento en la clasificación.
4. Evaluar el desempeño del modelo utilizando métricas como exactitud, precisión, sensibilidad, y F1-score.
5. Incorporar técnicas de interpretabilidad para identificar las regiones de la imagen más relevantes en la predicción

## 5. Metodología:

1. **Recolección de datos:** Se empleará un conjunto de radiografías de tórax público **Base de datos de radiografías de tórax** con imágenes clasificadas en neumonía y normal.
2. **Preprocesamiento:** Aplicar redimensionado, normalización y técnicas de data augmentation (rotaciones, traslaciones, escalado) para mejorar la calidad y robustez del modelo.
3. **Diseño del modelo:** Implementación de una red neuronal convolucional (CNN) en TensorFlow utilizando transfer learning con arquitecturas como EfficientNetB0, ajustando la capa final para la clasificación binaria.
4. **Entrenamiento:** Dividir los datos en entrenamiento (76 %), validación (13 %) y prueba (11 %), optimizar hiperparámetros y aplicar regularización con Dropout y Early Stopping para prevenir sobreajuste.
5. **Evaluación:** Medir el rendimiento con métricas de clasificación y validar la robustez del modelo como precisión, sensibilidad y AUC, complementadas con visualizaciones Grad-CAM.

6. **Documentación:** Elaborar un informe final con los resultados, análisis, conclusiones y presentación de resultados.

## 6. Resultados

### 6.1. Modelo CNN sin ajuste de pesos

Para comenzar, se entrenó una red neuronal convolucional (CNN) sobre imágenes en escala de grises, sin aplicar ningún tipo de corrección por desbalance de clases, utilizando técnicas de regularización (L2 y Dropout), y una capa de salida con activación sigmoide para clasificación binaria.

El modelo fue compilado con:

- **Optimizador:** Adam (learning rate = 0.001).
- **Función de pérdida:** binary\_crossentropy.
- **Métricas:** accuracy, AUC, PR-AUC, precision, recall

Métrica	NORMAL	PNEUMONIA
Precisión	1.00	0.69
Recall (Sensibilidad)	0.25	1.00
F1-score	0.40	0.82
<b>AUC (global)</b>	0.91	

Cuadro 1: Métricas del modelo CNN sin ajuste de pesos de clase

- El modelo tiene alta precisión en ambas clases, pero el recall para NORMAL es muy bajo (solo 25 %), lo que indica que muchos casos normales fueron mal clasificados como neumonía.
- El F1-score para NORMAL es de apenas 0.40, lo que refleja un rendimiento desequilibrado
- El AUC de 0.91 indica buena capacidad para distinguir entre clases, pero el sesgo hacia la clase mayoritaria afecta la utilidad clínica.

### 6.2. Modelo CNN con ajuste de pesos

Se utilizó el mismo procedimiento de entrenamiento que en el modelo anterior, con imágenes en escala de grises, regularización L2, Dropout y activación sigmoide para clasificación binaria. Sin embargo, en este caso se aplicó la técnica de class\_weight para compensar el desbalance entre clases. Esta estrategia asigna un mayor peso a la clase minoritaria (NORMAL) y un menor peso a la clase mayoritaria (PNEUMONIA), obligando al modelo a prestar más atención a los casos menos representados.

El modelo fue compilado con el optimizador Adam (tasa de aprendizaje = 0.001), función de pérdida binary\_crossentropy y métricas como precisión, recall, AUC y PR-AUC. A continuación se presentan las métricas obtenidas en el conjunto de prueba:

Métrica	NORMAL	PNEUMONIA
Precisión	0.96	0.71
Recall (Sensibilidad)	0.32	0.99
F1-score	0.48	0.83
<b>AUC (global)</b>	0.93	

Cuadro 2: Métricas del modelo CNN con ajuste de pesos de clase

Los resultados muestran una mejora en la sensibilidad (recall) para la clase NORMAL, que pasó de 0.25 a 0.32. Aunque sigue siendo baja, representa un avance en la capacidad del modelo para detectar correctamente los casos normales. La precisión para esta clase también se mantiene alta (0.96), lo que indica que los pocos casos que logra identificar como normales son, en su mayoría, correctos.

En cuanto a la clase PNEUMONIA, el modelo conserva un rendimiento excelente, con precisión de 0.71, recall de 0.99 y F1-score de 0.83. Esto demuestra que el ajuste de pesos no afectó negativamente la capacidad del modelo para detectar casos positivos.

El AUC global aumentó ligeramente a 0.93, lo que confirma que el modelo mejora su capacidad para distinguir entre ambas clases de forma más equilibrada.

### 6.3. Comparación de desempeño: CNN sin pesos vs. CNN con ajuste de pesos

A continuación se presentan las visualizaciones comparativas que permiten evaluar el impacto del ajuste de pesos (class\_weight) sobre el rendimiento del modelo CNN. Se incluyen métricas por clase, matrices de confusión y curvas de aprendizaje.

#### 1. Métricas por clase

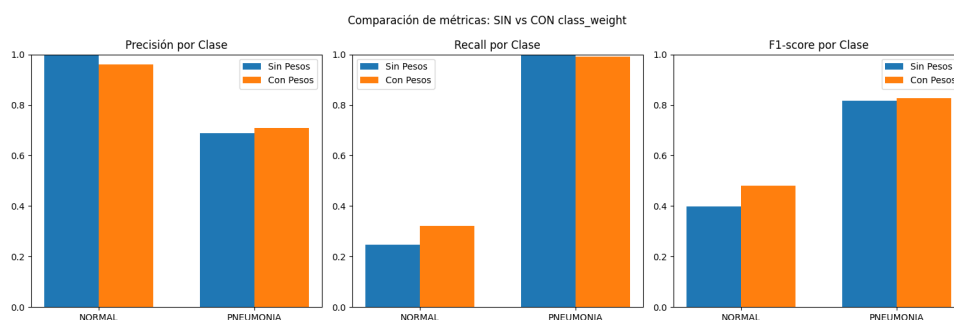


Figura 1: Comparación de precisión, recall y F1-score por clase entre el modelo CNN sin pesos y con ajuste de pesos.

Se confirma lo dicho anteriormente, para la clase NORMAL, se observa una mejora en el recall y el F1-score al aplicar pesos, lo que indica que el modelo logra detectar más casos normales sin comprometer demasiado la precisión. En cambio, para la clase PNEUMONIA, el rendimiento se mantiene alto en ambos modelos, con métricas cercanas a 1.0.

Esta visualización confirma que el uso de `class_weight` mejora el balance entre clases, especialmente en la detección de casos normales, sin afectar negativamente la capacidad del modelo para identificar neumonía.

## 2. Matrices de confusión

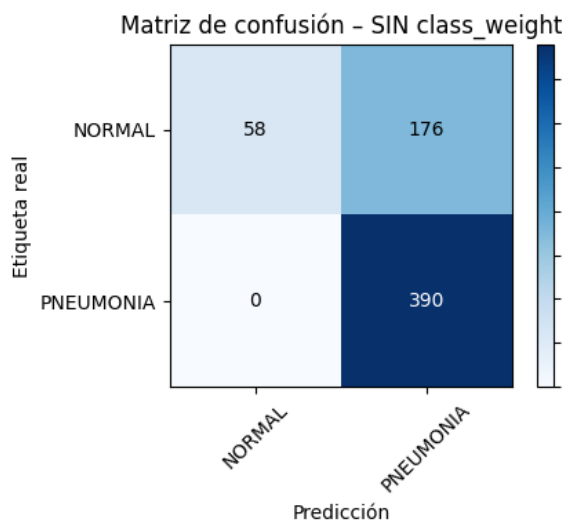


Figura 2: Matriz de confusión del modelo CNN sin ajuste de pesos.

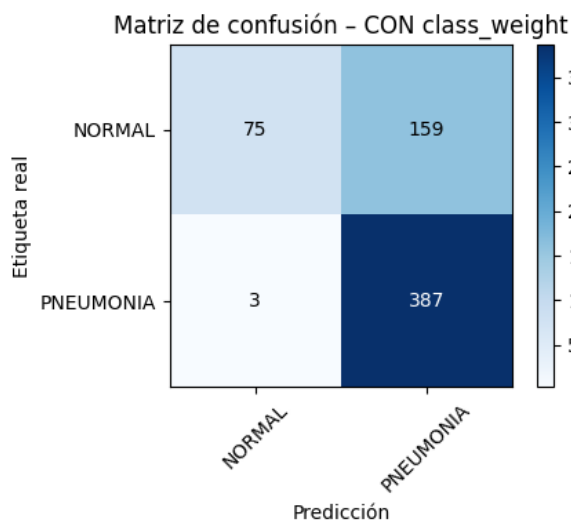


Figura 3: Matriz de confusión del modelo CNN con ajuste de pesos.

- **Sin ajuste de pesos:** El modelo tiene alta sensibilidad para detectar neumonía (390 aciertos), pero comete muchos falsos negativos en la clase NORMAL (176 casos mal clasificados).
- **Con ajuste de pesos:** Mejora la detección de casos normales (75 aciertos vs. 58), reduciendo los falsos negativos a 159. A cambio, introduce 3 falsos positivos en la clase PNEUMONIA.
- En conclusión el ajuste de pesos mejora el balance entre clases, especialmente en la detección de NORMAL, sin comprometer significativamente la precisión en PNEUMONIA.

## 3. Curvas de aprendizaje

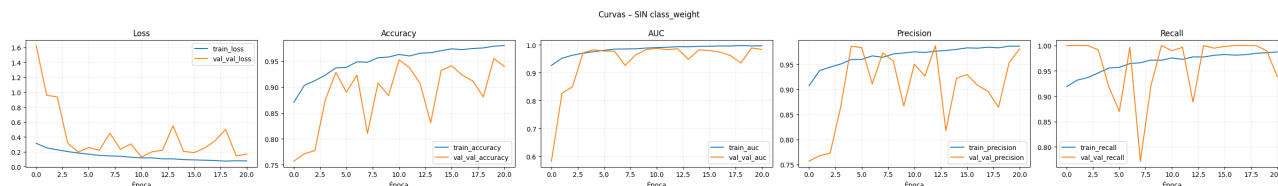


Figura 4: Curvas de aprendizaje sin ajuste de pesos.

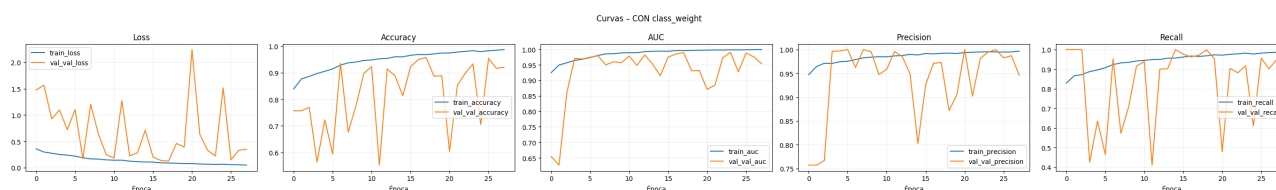


Figura 5: Curvas de aprendizaje con ajuste de pesos.

- **Pérdida (Loss):** Ambos modelos muestran una disminución progresiva en la pérdida de entrenamiento. Sin embargo, el modelo con `class_weight` presenta una curva de validación más estable, lo que sugiere mejor generalización y menor sobreajuste.
- **Precisión (Accuracy):** Aunque ambos modelos alcanzan niveles similares de precisión, el modelo con `class_weight` muestra menor variabilidad en la validación, lo que indica mayor consistencia en la clasificación.
- **AUC:** Las curvas AUC se mantienen altas en ambos casos, pero el modelo con `class_weight` logra una validación más estable, lo que refuerza su capacidad para discriminar entre clases, incluso en presencia de desbalance.
- **Precisión y Recall:** El modelo con `class_weight` mejora ligeramente la recuperación (recall) en validación, lo cual es clave en contextos clínicos donde los falsos negativos deben minimizarse. La precisión también se mantiene más estable, lo que sugiere un mejor equilibrio entre sensibilidad y especificidad.

El uso de `class_weight` contribuye a una mayor estabilidad en las métricas de validación, especialmente en recall y AUC, lo que lo convierte en una estrategia útil para abordar el desbalance de clases en imágenes médicas.

## 6.4. Entrenamiento con modelo preentrenado EfficientNetB0

Se implementó un modelo basado en EfficientNetB0, una arquitectura preentrenada sobre ImageNet que ha demostrado alto rendimiento en tareas de clasificación con bajo costo computacional. A diferencia del modelo CNN anterior, esta red trabaja con imágenes en RGB (3 canales), por lo que se adaptó el preprocesamiento para convertir las radiografías en escala de grises a formato compatible.

El entrenamiento se realizó en dos fases:

- **Fase 1 – Base congelada:** Se mantuvieron congeladas las capas convolucionales del modelo preentrenado, entrenando únicamente las capas superiores personalizadas. Esta etapa permitió adaptar el modelo a la tarea específica sin alterar los pesos base.
- **Fase 2 – Fine-tuning:** Se descongelaron las últimas 50 capas del modelo para realizar un ajuste fino, permitiendo que el modelo refine sus representaciones internas en función del nuevo dominio (radiografías de tórax).

Durante ambas fases se aplicaron técnicas como EarlyStopping, ReduceLROnPlateau y ModelCheckpoint, además del uso de `class_weight` para compensar el desbalance de clases.



A continuación se presenta la comparación de curvas de aprendizaje entre EfficientNetB0 y el modelo CNN:

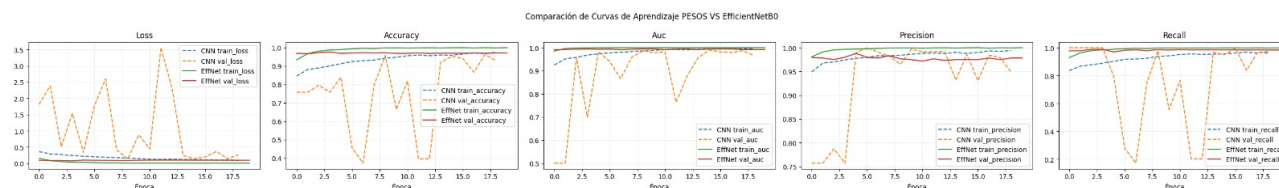


Figura 6: Comparación de curvas de aprendizaje entre EfficientNetB0 y CNN. Se muestran las métricas de pérdida, precisión, AUC, precisión y recall en entrenamiento y validación.

La figura 6 muestra que EfficientNetB0 supera al modelo CNN en todas las métricas clave:

- **Loss:** EfficientNetB0 presenta menor pérdida en validación, con curvas más estables y sin señales de sobreajuste.
- **Accuracy:** La precisión en validación es consistentemente más alta en EfficientNetB0, con menor variabilidad.
- **AUC:** El modelo preentrenado alcanza valores más altos de AUC, lo que indica mejor capacidad para distinguir entre clases.
- **Precision y Recall:** EfficientNetB0 mantiene valores elevados y estables en ambas métricas, lo que refleja un mejor equilibrio entre sensibilidad y especificidad.

Estos resultados confirman que el uso de un modelo preentrenado como EfficientNetB0 mejora significativamente el rendimiento en la tarea de detección de neumonía, especialmente en escenarios clínicos donde la precisión y el recall son críticos.

## 6.5. Visualización de atención mediante mapas de calor

Como parte del esfuerzo por mejorar la interpretabilidad del sistema, se incorporó la técnica Grad-CAM para visualizar las regiones de las radiografías que los modelos consideran relevantes al emitir una predicción. Esta herramienta permite generar mapas de calor sobre la imagen original, facilitando la trazabilidad del modelo y ofreciendo una guía visual que puede apoyar la interpretación clínica.

A continuación se presentan ejemplos de activación Grad-CAM para dos modelos entrenados: CNN con ajuste de pesos (class\_weight) y EfficientNetB0 fine-tuned. En cada caso se muestra la radiografía original, el mapa de calor generado y la predicción del modelo.

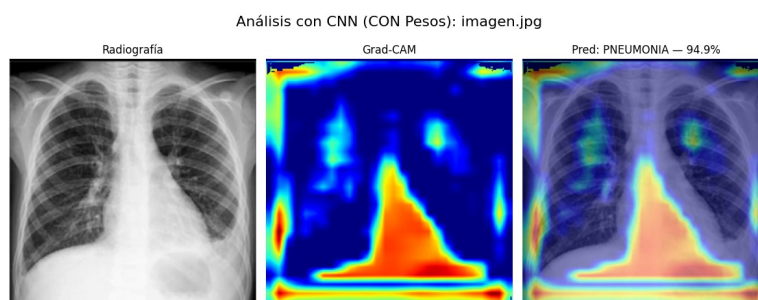


Figura 7: CNN — Activación Grad-CAM en caso clasificado como PNEUMONIA (94.9 %). Se observa atención en regiones pulmonares.

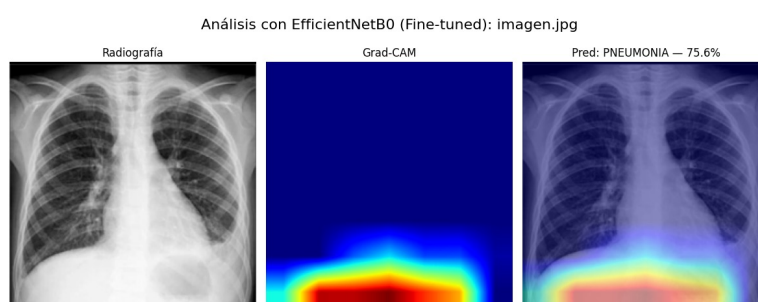


Figura 8: EfficientNetB0 — Activación Grad-CAM en caso clasificado como PNEUMONIA (75.6 %). Atención más difusa, con foco en campos medios.

Durante las pruebas, Grad-CAM permitió observar que los modelos tienden a enfocar su atención en regiones pulmonares relevantes, como bases y campos medios, especialmente en casos positivos. Sin embargo, también se detectaron comportamientos no deseados:

- Activaciones en zonas no patológicas, como huesos (costillas, clavículas), mediastino, dispositivos médicos (marcapasos), rótulos o textos fuera del tórax.
- Generación de puntos fijos “clavados” en los extremos de las imágenes, sin relación anatómica con la patología.
- En algunos casos, el mapa de calor se extendía hacia áreas periféricas o fuera del campo pulmonar, lo que puede inducir interpretaciones erróneas.

Estos patrones, aunque ocasionales, representan un riesgo en escenarios clínicos reales, especialmente en casos limítrofes donde la confianza del usuario es crítica. Por esta razón, el equipo decidió deshabilitar temporalmente el componente de mapas de calor en el despliegue final, priorizando la seguridad y claridad del sistema. No obstante, se mantiene como línea de trabajo activa para futuras versiones del proyecto, dado su potencial para mejorar la transparencia y la aceptación clínica del modelo.

## 7. Análisis y discusión

- **Modelos comparados:** CNN básica, CNN con `class_weight`, y EfficientNetB0 con fine-tuning.
- **Desempeño:** EfficientNetB0 fue el más balanceado y preciso. El uso de `class_weight` mejoró el recall para la clase NORMAL.
- **Generalización:** Las curvas de validación mostraron menor sobreajuste en los modelos ajustados.
- **Interpretabilidad:** Grad-CAM mostró activaciones poco confiables; se deshabilitó para evitar riesgos clínicos.
- **Aplicabilidad:** Los ajustes mejoran la sensibilidad y reducen sesgos, lo que favorece su uso en entornos reales.

## 8. Conclusiones

La comparación entre modelos demostró que el uso de técnicas como `class_weight` y transfer learning con EfficientNetB0 mejora significativamente la detección de neumonía en radiografías, especialmente en escenarios con desbalance de clases. EfficientNetB0 ofreció el mejor rendimiento global, con mayor estabilidad y sensibilidad. Aunque Grad-CAM aportó valor interpretativo, su uso requiere ajustes para garantizar confianza clínica. En conjunto, los resultados respaldan el uso de modelos robustos y balanceados como base para sistemas de apoyo diagnóstico confiables.

## 9. Trabajo futuro / Recomendaciones

Para que este sistema no solo detecte neumonía, sino transforme el diagnóstico médico:

- Integrar interpretabilidad confiable que no solo explique, sino que inspire confianza clínica.
- Validar en múltiples centros médicos para garantizar robustez y equidad.
- Diseñar un flujo de despliegue que combine precisión algorítmica con supervisión humana.
- Incorporar metadatos clínicos (edad, síntomas, historial) para enriquecer el contexto diagnóstico.
- Establecer métricas de impacto real: ¿cuántos diagnósticos se corrigen, cuántas vidas se salvan?

## 10. Referencias

- [1] Wikipedia, la enciclopedia libre, “Neumonía.” <https://es.wikipedia.org/wiki/Neumon%C3%ADa>.
- [2] American Journal of Roentgenology, “Radiología.” <https://ajronline.org/doi/10.2214/AJR.19.21521>.
- [3] National library of medicine, “Artificial intelligence in healthcare.” <https://pubmed.ncbi.nlm.nih.gov/39583770/>.