

Aprendizaje Profundo

# Detección Automática de Neumonía en Radiografías de Tórax Utilizando una Red Neuronal Convolutiva Personalizada

Juan Sebastián Rodríguez Salazar

Profesor: Diego Gerardo Roldan

Universidad del Rosario

19 de septiembre de 2024

# 1. Introducción

La neumonía es una enfermedad respiratoria aguda que afecta a millones de personas en todo el mundo y es una de las principales causas de mortalidad infantil, especialmente en países en desarrollo. Según la Organización Mundial de la Salud (OMS), en 2019 la neumonía causó el 15 % de todas las muertes de niños menores de cinco años, lo que equivale a aproximadamente 800,000 fallecimientos [1].

El diagnóstico temprano y preciso de la neumonía es esencial para iniciar un tratamiento oportuno y reducir las tasas de mortalidad. La radiografía de tórax es una herramienta fundamental en el diagnóstico; sin embargo, su interpretación puede ser subjetiva y depende en gran medida de la experiencia del radiólogo. En áreas con recursos limitados, la falta de especialistas puede retrasar el diagnóstico y empeorar los resultados clínicos.

En los últimos años, las técnicas de aprendizaje profundo han demostrado un gran potencial en el campo de la visión por computadora, incluyendo aplicaciones médicas como la detección de enfermedades a partir de imágenes. Las redes neuronales convolucionales (CNN) son especialmente efectivas en tareas de clasificación de imágenes debido a su capacidad para extraer características relevantes de manera automática.

Este trabajo tiene como objetivo desarrollar y evaluar un modelo basado en CNN para la detección automática de neumonía en imágenes de radiografías de tórax. Se utiliza un conjunto de datos público [2] que contiene imágenes clasificadas como 'PNEUMONIA' y 'NORMAL'. A diferencia de otros enfoques que emplean modelos preentrenados y transferencia de aprendizaje, este proyecto se enfoca en la construcción de un modelo personalizado desde cero, optimizado para esta tarea específica.

El informe se organiza de la siguiente manera: En la Sección 2, se describe la metodología, incluyendo el preprocesamiento de datos, la construcción del modelo CNN, y las técnicas de entrenamiento empleadas. La Sección 3 presenta los resultados obtenidos, como la precisión, la curva ROC, y la matriz de confusión normalizada, analizando el desempeño del modelo. Finalmente, en la Sección 4, se exponen las conclusiones del estudio, destacando las contribuciones, las limitaciones del modelo, y posibles direcciones para trabajos futuros.

## 2. Procesamiento de Datos e Imágenes

Para entrenar un modelo de red neuronal convolucional (CNN) que pueda clasificar imágenes de radiografías de tórax, se prepararon los datos mediante un proceso de preprocesamiento de imágenes. Este paso es esencial para garantizar que todas las imágenes tengan un formato y tamaño uniformes, lo que permite una entrada consistente al modelo y mejora su rendimiento.

Las imágenes del conjunto de datos original estaban clasificadas en dos categorías: 'PNEUMONIA' y 'NORMAL'. Cada imagen fue cargada en escala de grises para reducir la complejidad computacional, ya que la información relevante para la detección de neumonía está contenida en las características estructurales de la radiografía, que son más pronunciadas en una sola escala de color.

Una vez cargadas, las imágenes fueron redimensionadas a un tamaño fijo de  $150 \times 150$  píxeles. Este tamaño se seleccionó para mantener un equilibrio entre la retención de

detalles importantes de la imagen y la eficiencia computacional. Luego, se normalizaron dividiendo los valores de los píxeles por 255, escalándolos a un rango entre 0 y 1.

Además, debido a la naturaleza del conjunto de datos, se verificó la distribución de las clases en el conjunto de entrenamiento. Para abordar un posible desbalance de clases, se calcularon pesos de clase para asignar una mayor importancia a la clase minoritaria durante el entrenamiento. Esto garantiza que el modelo no se incline hacia la clase mayoritaria, proporcionando un equilibrio en su capacidad predictiva.

El preprocesamiento se completó al dividir los datos en tres conjuntos: entrenamiento, validación y prueba. Esta división es fundamental para evaluar el desempeño del modelo y ajustar sus hiperparámetros adecuadamente. En la siguiente sección, se describe la construcción y entrenamiento del modelo propuesto.

## 2.1. Ajuste de Pesos de Clase y Aumentación de Datos

Dado que los conjuntos de datos de imágenes médicas suelen presentar un desbalanceo de clases, se calculó el peso de las clases para este problema de clasificación binaria ('PNEUMONIA' y 'NORMAL'). Este desbalance se da cuando hay una cantidad significativamente mayor de imágenes de una clase en comparación con la otra. En este caso particular, la clase mayoritaria fue 'PNEUMONIA' con 3875 imágenes, mientras que la clase 'NORMAL' contaba con solo 1341 imágenes. Para abordar este desbalance, se ajustaron los pesos de las clases durante el entrenamiento del modelo, resultando en los siguientes pesos: 0.673 para la clase 'PNEUMONIA' y 1.945 para la clase 'NORMAL'. Estos pesos otorgan una mayor importancia a la clase minoritaria durante la fase de aprendizaje.

El ajuste de los pesos se realizó utilizando la función `class_weight` de la librería `scikit-learn`. Esto ayuda al modelo a aprender de manera más equitativa las características de ambas clases, mejorando la sensibilidad hacia la clase menos representada.

Para mejorar aún más la capacidad del modelo de generalizar a nuevas imágenes, se aplicaron técnicas de aumentación de datos. La aumentación de datos es un proceso que genera nuevas muestras a partir de las existentes mediante la aplicación de transformaciones aleatorias, tales como:

- **Rotación:** Se aplicó una rotación aleatoria de hasta 15 grados para hacer al modelo más robusto frente a la variabilidad en la orientación de las imágenes.
- **Zoom:** Se introdujo un ligero zoom aleatorio para simular cambios en el tamaño de las imágenes.
- **Desplazamiento:** Se permitieron desplazamientos horizontales y verticales de hasta un 10 % de la imagen, representando variaciones en la colocación del paciente durante la toma de radiografías.
- **Volteo horizontal:** Se incluyó el volteo horizontal para aumentar la diversidad de las imágenes sin cambiar su naturaleza clínica.

Estas técnicas permiten que el modelo vea una variedad más amplia de ejemplos durante el entrenamiento, mejorando su capacidad de generalizar y reducir el sobreajuste (*overfitting*). Las imágenes aumentadas se generaron en tiempo real durante el

entrenamiento utilizando la clase `ImageDataGenerator` de Keras, lo que permite un uso eficiente de la memoria y proporciona nuevas muestras en cada época.

En la próxima sección, se describe la arquitectura de la red neuronal convolucional personalizada y el proceso de entrenamiento.

## 2.2. Arquitectura del Modelo y Entrenamiento

Para abordar el problema de la detección de neumonía en radiografías, se diseñó una red neuronal convolucional (CNN) personalizada. Las CNN son particularmente efectivas para tareas de clasificación de imágenes debido a su capacidad para extraer características relevantes mediante capas de convolución y pooling. La arquitectura de la red se construyó desde cero, adaptándola específicamente a este problema en lugar de utilizar un modelo preentrenado.

La arquitectura del modelo consta de las siguientes capas:

- **Capas Convolucionales:** El modelo contiene cinco capas convolucionales. La primera capa tiene 32 filtros, mientras que las capas subsiguientes contienen 64, 64, 128, y 256 filtros respectivamente, todos con un tamaño de kernel de  $3 \times 3$ . Cada capa utiliza la función de activación ReLU, que introduce no linealidad al modelo y permite aprender características complejas. Las capas convolucionales están acompañadas por capas de normalización por lotes (*batch normalization*) para estabilizar y acelerar el proceso de entrenamiento.
- **Capas de Pooling:** Después de cada capa convolucional, se incluye una capa de pooling (*MaxPooling2D*) con un tamaño de  $2 \times 2$ . Estas capas reducen la dimensionalidad de las características, ayudando a disminuir la complejidad computacional y a evitar el sobreajuste.
- **Capas de Dropout:** Para evitar el sobreajuste, se aplicó regularización mediante la técnica de dropout después de algunas capas convolucionales y densas. Las tasas de dropout varían entre 0.1 y 0.2, lo que significa que se desactivan aleatoriamente un porcentaje de neuronas durante el entrenamiento, forzando al modelo a aprender características más robustas.
- **Capa de Aplanamiento:** Después de las capas convolucionales, se incluye una capa de aplanamiento (*Flatten*) para transformar las características extraídas en un vector unidimensional que puede ser procesado por las capas densas.
- **Capas Densas:** El modelo tiene dos capas densas. La primera capa contiene 128 neuronas con función de activación ReLU, mientras que la capa de salida tiene una única neurona con función de activación sigmoide para realizar la clasificación binaria ('PNEUMONIA' o 'NORMAL').

La elección de esta arquitectura se basa en la necesidad de capturar patrones complejos en las imágenes de rayos X, como las infiltraciones y opacidades características de la neumonía. Se prefirió construir un modelo personalizado en lugar de usar un modelo preentrenado debido a la especificidad de las características en imágenes médicas y para evitar transferencias de conocimiento que no sean relevantes para este dominio específico.

El modelo se compiló utilizando el optimizador *RMSprop* con una tasa de aprendizaje ajustable durante el entrenamiento. La función de pérdida seleccionada fue la *binary cross-entropy*, adecuada para problemas de clasificación binaria. La métrica principal de evaluación fue la precisión (*accuracy*).

Durante el entrenamiento, se utilizaron callbacks para mejorar el rendimiento y prevenir el sobreajuste:

- **Reducción de la Tasa de Aprendizaje:** Mediante el callback `ReduceLROnPlateau`, se redujo la tasa de aprendizaje si la precisión en el conjunto de validación no mejoraba durante 2 épocas consecutivas.
- **Detención Temprana:** Se aplicó detención temprana (*EarlyStopping*) para detener el entrenamiento si la pérdida en el conjunto de validación no mejoraba durante 10 épocas consecutivas, restaurando los pesos del modelo al mejor estado observado.

El modelo se entrenó durante un máximo de 30 épocas, con los datos de entrenamiento y validación generados a partir de las imágenes preprocesadas y aumentadas. En la siguiente sección, se presentan y analizan los resultados obtenidos.

### 3. Evaluación del Modelo y Resultados

Para evaluar el rendimiento del modelo, se utilizó una matriz de confusión, que proporciona una representación visual de las predicciones del modelo en relación con las etiquetas reales de las imágenes de prueba. La Figura 1 muestra la matriz de confusión obtenida en el conjunto de prueba.

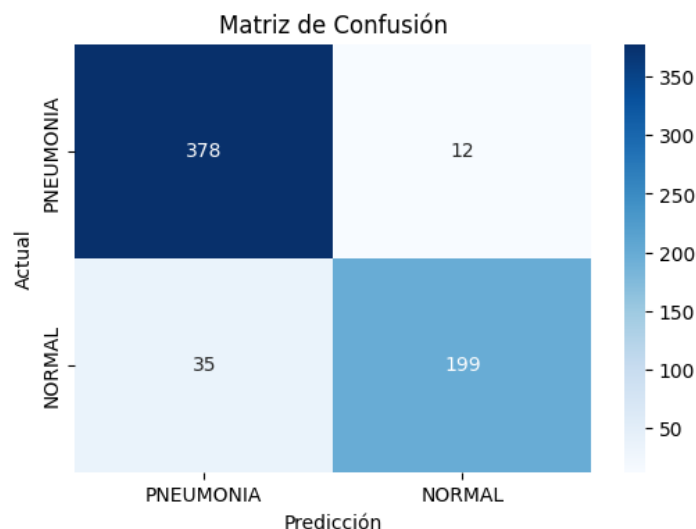


Figura 1: Matriz de confusión del modelo en el conjunto de prueba.

La matriz de confusión se interpreta de la siguiente manera:

- **Verdaderos Positivos (VP):** 378 imágenes de neumonía fueron correctamente clasificadas como 'PNEUMONIA'.

- **Falsos Negativos (FN):** 12 imágenes de neumonía fueron incorrectamente clasificadas como 'NORMAL'.
- **Verdaderos Negativos (VN):** 199 imágenes normales fueron correctamente clasificadas como 'NORMAL'.
- **Falsos Positivos (FP):** 35 imágenes normales fueron incorrectamente clasificadas como 'PNEUMONIA'.

Estos resultados indican que el modelo tiene una buena capacidad para detectar correctamente la neumonía, con un alto número de verdaderos positivos (378) y verdaderos negativos (199). Sin embargo, se observan algunos errores en la clasificación:

- El número de falsos negativos (12) es relativamente bajo, lo que es importante en aplicaciones médicas, ya que minimizar los casos donde la enfermedad no se detecta es fundamental para evitar consecuencias graves.
- El modelo también muestra un cierto número de falsos positivos (35), lo que implica que algunas imágenes normales fueron clasificadas como neumonía. Aunque este error es menos crítico que los falsos negativos, aún es relevante, ya que puede llevar a diagnósticos incorrectos y tratamientos innecesarios.

Para complementar este análisis, se presenta la Figura 2, que muestra la matriz de confusión normalizada. Esta representación porcentual ofrece una comprensión más clara del rendimiento del modelo en términos relativos para cada clase.

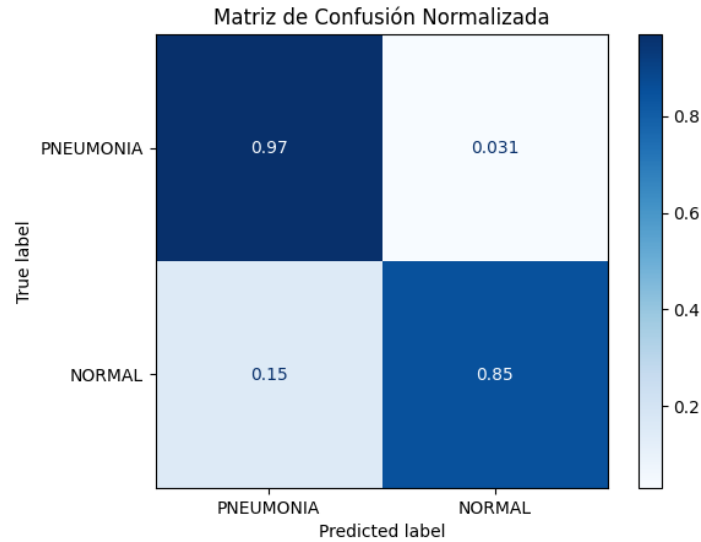


Figura 2: Matriz de confusión normalizada del modelo en el conjunto de prueba.

La matriz de confusión normalizada se interpreta de la siguiente manera:

- **Verdaderos Positivos (PNEUMONIA):** El modelo clasificó correctamente el 97 % de las imágenes de neumonía.
- **Falsos Negativos (PNEUMONIA):** El 3.1 % de las imágenes de neumonía fueron incorrectamente clasificadas como 'NORMAL'.

- **Verdaderos Positivos (NORMAL):** El modelo clasificó correctamente el 85 % de las imágenes normales.
- **Falsos Positivos (NORMAL):** El 15 % de las imágenes normales fueron incorrectamente clasificadas como 'PNEUMONIA'.

Esta representación normalizada confirma que el modelo tiene un alto desempeño en la clasificación de neumonía, priorizando la sensibilidad (97 % de imágenes de neumonía clasificadas correctamente), lo cual es crucial en el contexto clínico. Sin embargo, hay un mayor porcentaje de falsos positivos en la clase 'NORMAL' (15 %), indicando que el modelo tiende a ser conservador, prefiriendo clasificar erróneamente algunas imágenes normales como neumonía para evitar falsos negativos.

Para evaluar de manera integral el desempeño del modelo, se generaron tanto un informe de clasificación como gráficas de precisión y F1-score para cada clase. Esto permite analizar cómo se comporta el modelo en la detección de las imágenes de neumonía y las normales.

La Tabla 1 muestra el informe de clasificación que incluye métricas como precisión, recall y F1-score para cada clase ('PNEUMONIA' y 'NORMAL'). Las Figuras 3, 4, y 5 presentan gráficas que ilustran el porcentaje de predicciones correctas e incorrectas, así como el F1-score para cada clase.

Informe de clasificación:				
	precision	recall	f1-score	support
PNEUMONIA	0.92	0.97	0.94	390
NORMAL	0.94	0.85	0.89	234
accuracy			0.92	624
macro avg	0.93	0.91	0.92	624
weighted avg	0.93	0.92	0.92	624

Cuadro 1: Informe de clasificación del modelo en el conjunto de prueba.

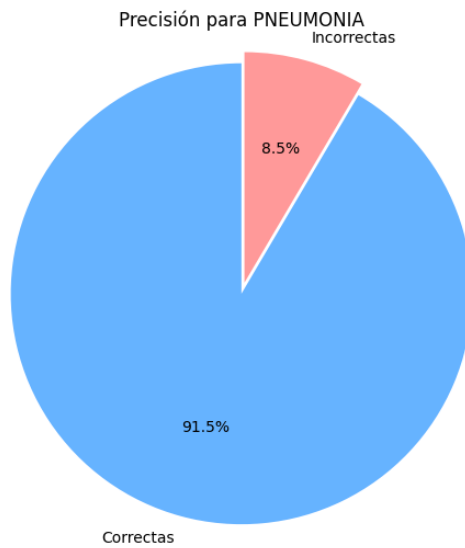


Figura 3: Precisión para la clase 'PNEUMONIA'. El 91.5 % de las predicciones fueron correctas, mientras que el 8.5 % fueron incorrectas.

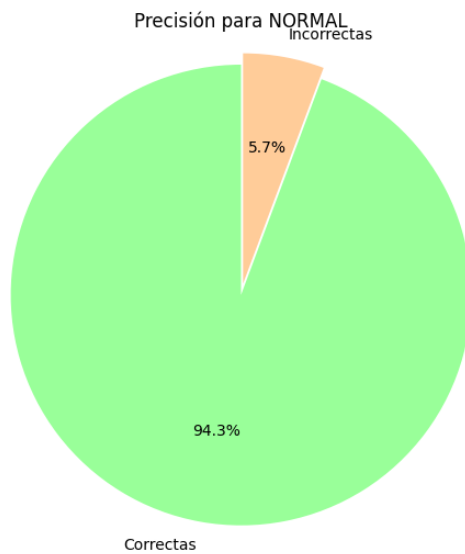


Figura 4: Precisión para la clase 'NORMAL'. El 94.3 % de las predicciones fueron correctas, mientras que el 5.7 % fueron incorrectas.



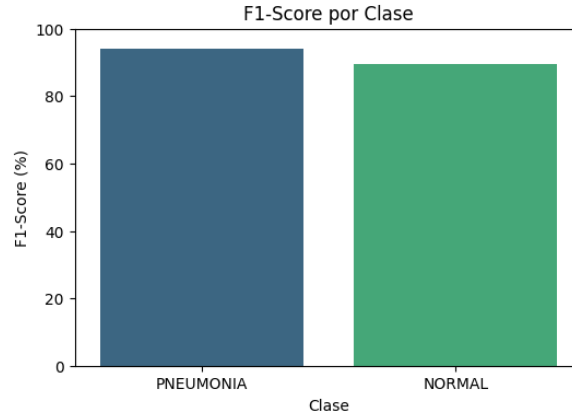


Figura 5: F1-Score para cada clase.

El informe de clasificación se interpreta de la siguiente manera:

- **Precisión:** La precisión para la clase 'PNEUMONIA' es de 0.92, mientras que para la clase 'NORMAL' es de 0.94. Esto indica que, cuando el modelo predice una clase determinada, es correcto aproximadamente el 92-94 % de las veces.
- **Recall (Sensibilidad):** El recall para la clase 'PNEUMONIA' es de 0.97, lo que significa que el modelo es capaz de identificar el 97 % de los casos de neumonía. Sin embargo, el recall para la clase 'NORMAL' es de 0.85, lo que indica que el modelo detecta el 85 % de los casos normales.
- **F1-Score:** El F1-score proporciona una medida equilibrada entre la precisión y el recall. Los valores de F1-score son 0.94 para 'PNEUMONIA' y 0.89 para 'NORMAL', indicando un buen equilibrio en el rendimiento del modelo, como se muestra en la Figura 5.
- **Exactitud Global (Accuracy):** La exactitud del modelo es del 92 %, lo que refleja el porcentaje de todas las predicciones correctas en el conjunto de prueba.

Las Figuras 3 y 4 refuerzan la evaluación cuantitativa, mostrando visualmente que el modelo clasifica correctamente el 91.5 % de las imágenes de neumonía y el 94.3 % de las imágenes normales. Los errores cometidos se representan por las secciones más pequeñas de las gráficas, correspondiendo al 8.5 % y 5.7 % de las imágenes mal clasificadas, respectivamente. La Figura 5 complementa este análisis al mostrar que el F1-score es relativamente alto para ambas clases, lo que indica un buen equilibrio entre precisión y recall.

En conjunto, estas métricas y visualizaciones demuestran que el modelo tiene un buen desempeño en la clasificación de radiografías, aunque existen algunas limitaciones, especialmente en la detección de casos normales. Este análisis destaca la necesidad de utilizar el modelo como una herramienta de apoyo en el diagnóstico clínico.

### 3.1. Radiografías Correctamente Clasificadas

Para entender mejor el rendimiento del modelo, se seleccionaron algunas imágenes de radiografías que fueron correctamente clasificadas. La Figura 6 muestra cinco ejemplos de radiografías clasificadas correctamente por el modelo como 'PNEUMONIA'.

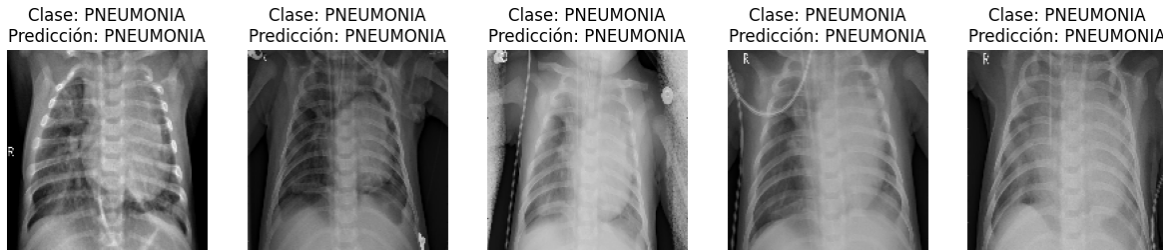


Figura 6: Ejemplos de imágenes correctamente clasificadas por el modelo como 'PNEUMONIA'.

Estas imágenes ilustran cómo el modelo es capaz de identificar correctamente las características visuales asociadas con la neumonía en las radiografías de tórax. Las áreas opacas y las infiltraciones en los pulmones, que son signos comunes de neumonía, parecen haber sido detectadas de manera efectiva por la red neuronal convolucional.

La capacidad del modelo para clasificar correctamente estas imágenes demuestra su eficacia para aprender y reconocer los patrones característicos de la neumonía en las radiografías. Aunque estas imágenes representan casos exitosos, es importante tener en cuenta que el modelo también comete errores, como se discutió en la sección anterior con la matriz de confusión. Por lo tanto, aunque el modelo muestra un buen desempeño, siempre debe ser utilizado como una herramienta de apoyo al diagnóstico clínico y no como un reemplazo de la evaluación médica profesional.

### 3.2. Radiografías Incorrectamente Clasificadas

Para una evaluación más completa del modelo, también es importante analizar los casos en los que el modelo cometió errores. La Figura 7 muestra cinco ejemplos de radiografías que fueron incorrectamente clasificadas por el modelo. En estos ejemplos, todas las imágenes muestran casos de neumonía que fueron erróneamente predichos como 'NORMAL'.

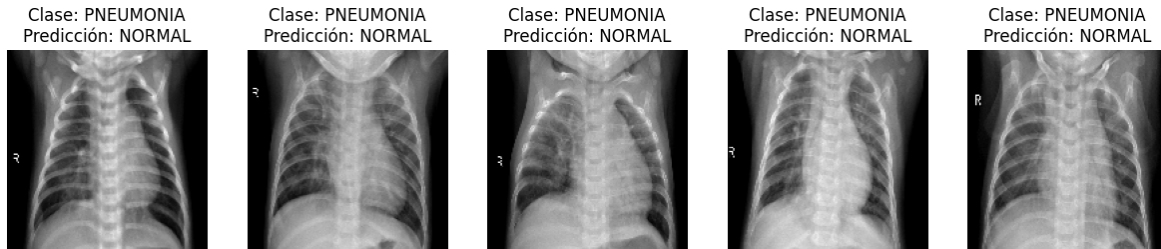


Figura 7: Ejemplos de imágenes incorrectamente clasificadas por el modelo. Estas imágenes muestran casos de neumonía que fueron clasificados erróneamente como 'NORMAL'.

Estos errores, conocidos como falsos negativos, son críticos en un contexto médico, ya que representan casos donde la neumonía no fue detectada. La revisión visual de estas imágenes muestra que las características distintivas de la neumonía pueden ser menos evidentes o estar parcialmente ocultas, lo que puede explicar por qué el modelo no logró clasificarlas correctamente.

Las posibles razones para estos errores podrían incluir:

- **Características sutiles:** Las imágenes pueden contener características de neumonía que son sutiles y difíciles de distinguir, incluso para un modelo bien entrenado.
- **Ruido en las imágenes:** La presencia de artefactos, variaciones en la calidad de la imagen o diferencias en la exposición pueden haber afectado el rendimiento del modelo.
- **Limitaciones del modelo:** La arquitectura del modelo puede no ser lo suficientemente compleja para captar todos los patrones posibles de neumonía, especialmente en casos atípicos.

Estos errores resaltan la necesidad de utilizar este modelo como una herramienta complementaria en el proceso de diagnóstico médico, proporcionando una segunda opinión que debe ser validada por profesionales de la salud. La detección de neumonía en radiografías es una tarea compleja, y aunque el modelo ofrece un buen desempeño general, no es infalible.

### 3.3. Precisión Durante el Entrenamiento

La Figura 8 muestra la evolución de la precisión del modelo durante el entrenamiento y la validación a lo largo de las épocas. La curva azul representa la precisión en el conjunto de entrenamiento, mientras que la curva naranja representa la precisión en el conjunto de validación.

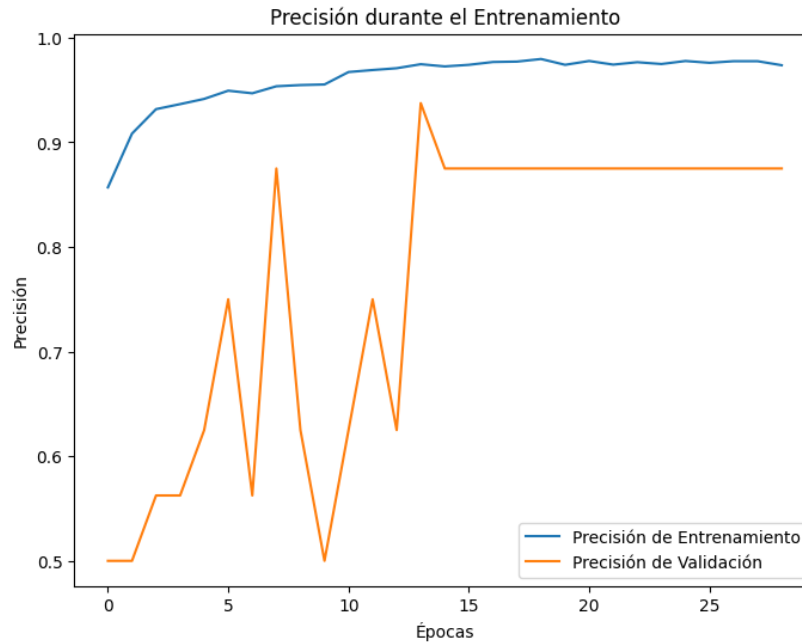


Figura 8: Precisión del modelo durante el entrenamiento y la validación a lo largo de las épocas.

El gráfico revela varios aspectos importantes sobre el proceso de entrenamiento del modelo:

- **Precisión en el Entrenamiento:** La precisión en el conjunto de entrenamiento aumenta de manera constante y se estabiliza alrededor del 95 % después de unas pocas épocas. Esto indica que el modelo está aprendiendo las características presentes en los datos de entrenamiento de manera efectiva.
- **Precisión en la Validación:** La precisión en el conjunto de validación muestra una tendencia más fluctuante, especialmente en las primeras épocas. Sin embargo, hacia el final del entrenamiento, la precisión se estabiliza alrededor del 92 %. Las fluctuaciones iniciales pueden deberse a la variabilidad en el conjunto de validación y al ajuste de los pesos del modelo durante el aprendizaje.
- **Estabilización:** La estabilización de la precisión en el conjunto de validación sin una caída significativa hacia el final del entrenamiento sugiere que el modelo no está sobreajustándose gravemente a los datos de entrenamiento. Esto indica una buena generalización a nuevos datos, lo cual es crucial para el desempeño en un contexto clínico.

La estabilización de las curvas de precisión tanto en entrenamiento como en validación demuestra que el modelo ha alcanzado un equilibrio en su capacidad de aprender de los datos de entrenamiento y de generalizar a los datos de validación. Esto refuerza la validez del modelo para la tarea de clasificación de neumonía en radiografías de tórax.

En resumen, la Figura 8 proporciona una visión general del proceso de entrenamiento, mostrando que el modelo ha aprendido de manera efectiva las características necesarias para clasificar las imágenes con un alto grado de precisión.

### 3.4. Pérdida Durante el Entrenamiento

La Figura 9 muestra la evolución de la pérdida del modelo durante el entrenamiento y la validación a lo largo de las épocas. La curva azul representa la pérdida en el conjunto de entrenamiento, mientras que la curva naranja representa la pérdida en el conjunto de validación.

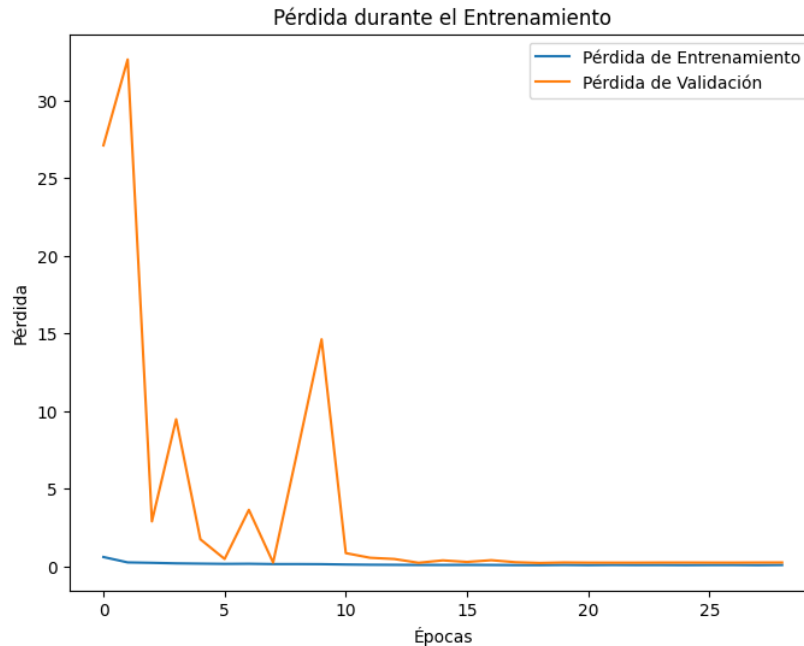


Figura 9: Pérdida del modelo durante el entrenamiento y la validación a lo largo de las épocas.

La gráfica revela varios aspectos importantes sobre el proceso de entrenamiento del modelo:

- **Pérdida en el Entrenamiento:** La pérdida en el conjunto de entrenamiento disminuye rápidamente al principio y luego se estabiliza en un valor bajo. Esto indica que el modelo está aprendiendo a ajustar sus parámetros para minimizar el error durante el entrenamiento.
- **Pérdida en la Validación:** La pérdida en el conjunto de validación muestra una tendencia a la baja con fluctuaciones significativas en las primeras épocas. Después de un número considerable de épocas, la pérdida en el conjunto de validación se reduce a niveles bajos y se estabiliza, lo que sugiere que el modelo ha alcanzado una buena generalización.
- **Fluctuaciones Iniciales:** Las fuertes fluctuaciones en la pérdida de validación durante las primeras épocas podrían deberse a la variabilidad en el conjunto de datos de validación y a la calibración de los pesos del modelo. A medida que el modelo aprende, estas fluctuaciones se reducen, lo que indica que el modelo se está volviendo más estable.

La estabilización de la pérdida en ambas curvas sugiere que el modelo ha alcanzado un punto de equilibrio, donde ha aprendido a capturar las características importantes de las imágenes sin sobreajustarse a los datos de entrenamiento. Esto es esencial para asegurar que el modelo tenga un buen rendimiento en datos no vistos.

En resumen, la Figura 9 muestra que el modelo ha sido entrenado de manera efectiva, logrando una reducción en la pérdida tanto en el conjunto de entrenamiento como en el de validación. La disminución de la pérdida indica que el modelo ha mejorado su capacidad para realizar predicciones precisas durante el proceso de entrenamiento.

### 3.5. Curva ROC

La Figura 10 muestra la Curva Característica Operativa del Receptor (ROC) del modelo, que es una herramienta importante para evaluar el rendimiento de un clasificador binario. La curva ROC representa la relación entre la tasa de verdaderos positivos (sensibilidad) y la tasa de falsos positivos a diferentes umbrales de decisión.

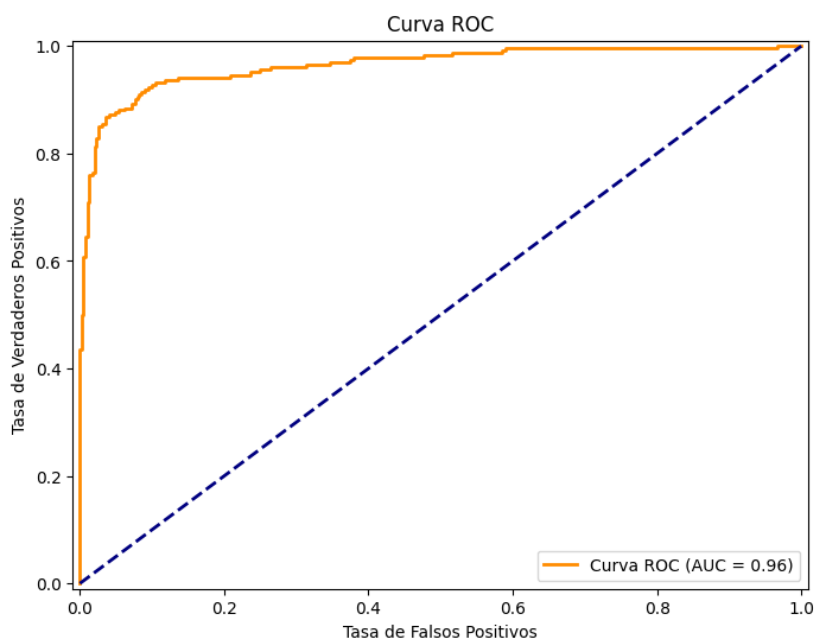


Figura 10: Curva ROC del modelo con un AUC de 0.96.

La curva ROC se interpreta de la siguiente manera:

- **Área Bajo la Curva (AUC):** El Área Bajo la Curva (AUC) es una medida de la capacidad del modelo para distinguir entre las clases 'PNEUMONIA' y 'NORMAL'. En este caso, el AUC es de 0.96, lo que indica un rendimiento excelente del modelo. Un AUC de 0.5 indica un clasificador aleatorio, mientras que un AUC cercano a 1 indica un clasificador perfecto.
- **Curva ROC:** La curva ROC (línea naranja) se encuentra muy por encima de la diagonal (línea punteada), lo que indica que el modelo tiene una alta sensibilidad y una baja tasa de falsos positivos en la mayoría de los umbrales. Esto

refleja la capacidad del modelo para discriminar eficazmente entre las imágenes de neumonía y las normales.

La alta AUC de 0.96 sugiere que el modelo tiene un excelente equilibrio entre la sensibilidad y la especificidad, lo que es fundamental en aplicaciones médicas donde es crucial minimizar tanto los falsos negativos como los falsos positivos.

En resumen, la Figura 10 demuestra que el modelo es altamente efectivo en la detección de neumonía en radiografías de tórax, con un rendimiento general fuerte en términos de su capacidad para diferenciar entre las clases. Este alto valor de AUC refuerza la validez del modelo como una herramienta de apoyo en el diagnóstico clínico.

## 4. Conclusiones

En este trabajo, se desarrolló y evaluó un modelo de red neuronal convolucional (CNN) para la detección automática de neumonía en radiografías de tórax. A través de un enfoque personalizado de construcción del modelo, se logró una alta precisión en la clasificación de imágenes, con una exactitud global del 92 %, y un área bajo la curva (AUC) de 0.96 en la curva ROC. Estos resultados demuestran que las CNN pueden ser una herramienta efectiva para apoyar el diagnóstico clínico de la neumonía.

La matriz de confusión normalizada reveló que el modelo es capaz de identificar correctamente el 97 % de las imágenes de neumonía, lo que es crucial para minimizar los falsos negativos en un contexto clínico. Sin embargo, se observó una tasa más alta de falsos positivos en la clase 'NORMAL', lo que sugiere que el modelo tiende a ser conservador en su enfoque, prefiriendo errar en la clasificación hacia la detección de la enfermedad. Aunque esto puede conducir a diagnósticos falsos positivos, es preferible en el contexto de la salud, donde la falta de diagnóstico (falsos negativos) puede tener consecuencias más graves.

A lo largo del entrenamiento, se evidenció una mejora constante en la capacidad del modelo para aprender características relevantes de las imágenes, logrando una reducción significativa de la pérdida tanto en el conjunto de entrenamiento como en el de validación. Las técnicas de preprocesamiento y aumentación de datos jugaron un papel fundamental en la mejora de la generalización del modelo, permitiendo manejar la variabilidad inherente a las imágenes médicas.

Pese a los resultados prometedores, este modelo tiene ciertas limitaciones. Su desempeño depende de la calidad y representatividad del conjunto de datos utilizado. Además, aunque el modelo demostró una alta precisión, no está exento de errores. Por lo tanto, se debe utilizar como una herramienta complementaria y no como un reemplazo del diagnóstico médico profesional. En la práctica clínica, los resultados de este modelo deben ser interpretados por expertos en conjunto con otros métodos diagnósticos.

Para trabajos futuros, se podría explorar la integración de datos adicionales, como la historia clínica del paciente, para mejorar aún más la precisión del modelo. Además, se podrían considerar modelos más avanzados, como redes neuronales profundas preentrenadas, para mejorar la capacidad del modelo para detectar características sutiles en las imágenes de radiografías.

En conclusión, este estudio demuestra la viabilidad de las redes neuronales convolucionales como herramienta de apoyo en la detección de neumonía, destacando la importancia de la IA en el avance de las aplicaciones médicas.

## Referencias

- [1] Organización Mundial de la Salud. *Neumonía*. 2019. Disponible en: <https://www.who.int/es/news-room/fact-sheets/detail/pneumonia>.
- [2] Paul Mooney. *Chest X-Ray Images (Pneumonia)*. Kaggle, 2018. Disponible en: <https://www.kaggle.com/datasets/paultimothymooney/chest-xray-pneumonia>.
- [3] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner. *Gradient-based learning applied to document recognition*. Proceedings of the IEEE, 86(11):2278-2324, 1998.
- [4] I. Goodfellow, Y. Bengio, and A. Courville. *Deep Learning*. MIT Press, 2016. Disponible en: <http://www.deeplearningbook.org>.
- [5] F. Chollet. *Keras*. GitHub, 2015. Disponible en: <https://github.com/fchollet/keras>.