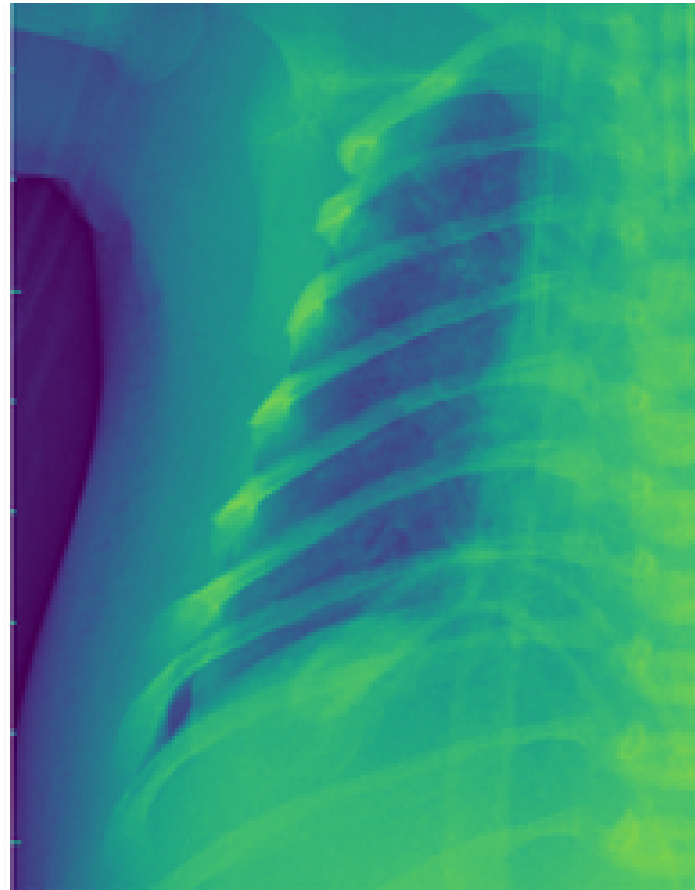


# Informe

## Clasificación en imágenes de rayos X

## Neumonía



- 
- ❖ Proyecto para el Instituto Superior Politécnico de Córdoba
  - ❖ Materias: Procesamiento de imágenes y Modelos de IA.
  - ❖ Profesores: Sol Figueroa y Carlos Charletti.
  - ❖ Equipo: Dante Javier Pagano, Eugenia Barozzi, Federico Gurrea,  
Juan Marcelo Molina, Julieta Battauz y Laura Peralta.

## Objetivo e introducción

El objetivo principal de este proyecto fue desarrollar un modelo de Deep Learning capaz de clasificar imágenes de radiografías de tórax para identificar la presencia de neumonía. Se buscaba crear una herramienta de apoyo al diagnóstico que mejorara la eficiencia y precisión en la detección de esta condición. Para ello, se trabajó con un dataset de imágenes de radiografías de tórax, divididas en conjuntos de entrenamiento, validación y prueba.

## Colaboratory del Proyecto

<https://drive.google.com/file/d/1QbAjdNhqPiYpznrtedOOAlu8j8QiUcLI/view?usp=sharing>

## 2. Importación de librerías

Se importan módulos de Tensor Flow y Keras, que son ampliamente utilizados en proyectos de Deep Learning. Las importaciones específicas indican que el proyecto se basa en un modelo de red neuronal para clasificación multiclase:

- layers y models: para construir la arquitectura de la red neuronal.
- Adam: optimizador eficiente que se adapta a los parámetros.
- CategoricalCrossentropy: función de pérdida adecuada para clasificación multiclase con etiquetas one-hot.
- AUC, Precision, Recall: métricas avanzadas que permiten evaluar el desempeño del modelo más allá de la simple exactitud.

## 2. Carga de datos e imágenes preprocesamiento

Se definen los parámetros de entrada:

- BATCH\_SIZE = 32: equilibrio entre velocidad de entrenamiento y uso de memoria.
- IMG\_SIZE = (224, 224): tamaño estándar compatible con redes como VGG, ResNet, etc.
- SEED = 42: asegura reproducibilidad.

Se cargan las imágenes con image\_dataset\_from\_directory, separadas en conjuntos de entrenamiento, validación y prueba. Se usa label\_mode='categorical' para codificación one-hot, adecuada para clasificación multiclase.

### 3. Aumento de datos y preprocesamiento

Se aplican técnicas de aumento de datos (flip, rotación, zoom, contraste) sólo al entrenamiento para mejorar la generalización. Además, se normalizan los píxeles (0 a 1) con Rescaling. El uso de AUTOTUNE, cache, shuffle y prefetch permite acelerar el procesamiento y la alimentación del modelo.

### 4. Definición del modelo CNN

Se implementa una red CNN secuencial con 3 bloques convolucionales que extraen características, seguidos por capas densas que realizan la clasificación. Se utiliza Dropout como regularización. La capa final tiene 3 salidas softmax, correspondientes a las clases: NORMAL, VIRAL\_PNEUMONIA y BACTERIAL\_PNEUMONIA.

### 5. Compilación del modelo

Se usa el optimizador Adam con tasa de aprendizaje baja (0.0001) para entrenamiento estable. La función de pérdida es CategoricalCrossentropy y se evalúan métricas relevantes para medicina: precisión, recall y AUC además de la exactitud.

### 6. Callbacks

Se utilizan tres callbacks:

- EarlyStopping: detiene el entrenamiento si no mejora la validación por 5 épocas.
- ReduceLROnPlateau: reduce el learning rate si no hay mejoras.
- ModelCheckpoint: guarda el mejor modelo basado en val\_loss.

### 7. Entrenamiento del modelo

El modelo se entrena utilizando los datasets de entrenamiento y validación durante un máximo de 50 épocas. Gracias al uso de callbacks como EarlyStopping y ReduceLROnPlateau, el entrenamiento puede detenerse antes si no hay mejoras en la pérdida de validación. Esto previene el sobreajuste y mejora la eficiencia del entrenamiento.

### 8. Evaluación del modelo

Después del entrenamiento, el modelo se evalúa con el conjunto de prueba. Esto permite medir cómo generaliza el modelo a datos nuevos. Las métricas obtenidas incluyen:

- Accuracy: proporción de clasificaciones correctas.
- Precision: qué tan preciso es el modelo al predecir cada clase.
- Recall: qué tanto del total de positivos reales fue capturado.

- AUC: mide la capacidad del modelo para distinguir entre clases.

Estas métricas son especialmente importantes en medicina, donde los falsos negativos pueden tener consecuencias críticas.

## 9. Visualización de curvas de entrenamiento

Se grafican las curvas de pérdida y precisión para los conjuntos de entrenamiento y validación. Estas curvas permiten verificar si el modelo está aprendiendo correctamente y si hay signos de sobreajuste (por ejemplo, si la pérdida de validación aumenta mientras la de entrenamiento disminuye).

## 10. Matriz de confusión y reporte de clasificación

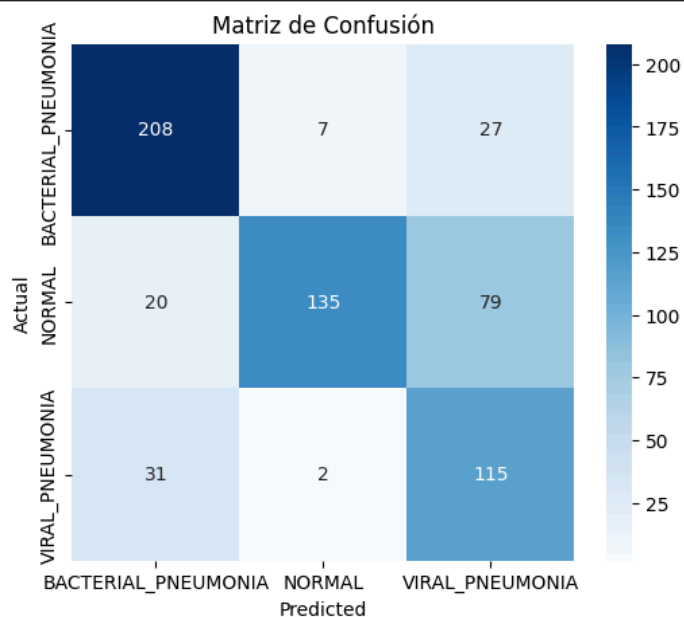
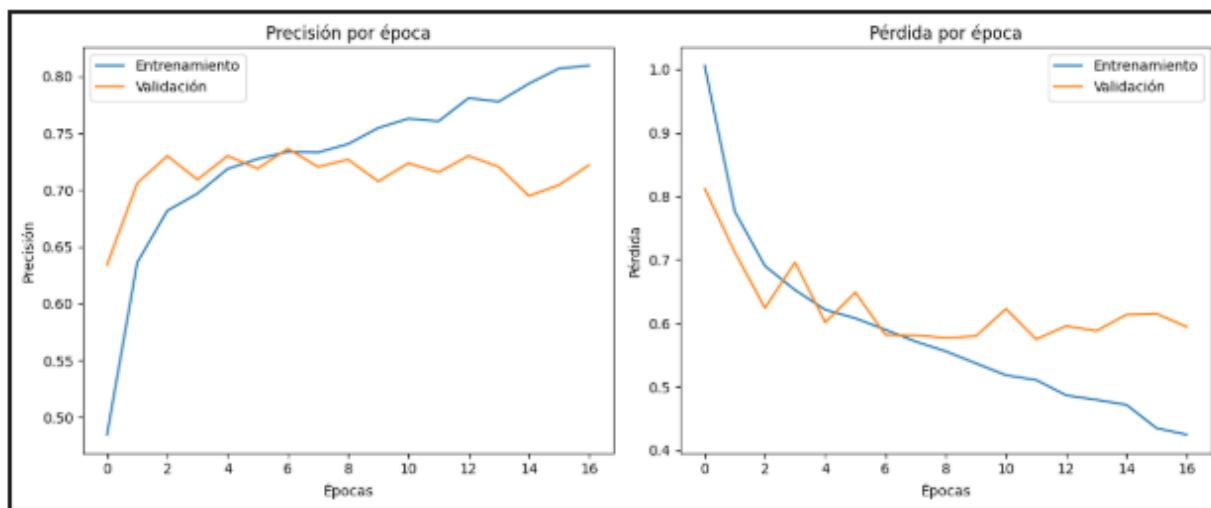
La matriz de confusión permite visualizar los aciertos y errores del modelo en cada clase. Se acompaña de un informe de clasificación que detalla precisión, recall y F1-score por clase. Esto es esencial para entender el comportamiento del modelo con cada tipo de neumonía y con imágenes normales.

### Ejecución N° 1

Esta primera prueba se inició con el objetivo de ver los primeros resultados de la clasificación para posteriormente poder continuar con los posteriores cambios. Se utilizó la configuración base del modelo y los parámetros iniciales para el entrenamiento.

#### 1. Gráfico de Precisión y Pérdida por Época:

La precisión de entrenamiento ascendió de manera constante y la precisión de validación subió al principio luego tendió a estabilizarse para descender en las últimas épocas sugiriendo posible sobreajuste. En la pérdida de entrenamiento se vio disminuyendo progresivamente y la de validación fue similar aunque con fluctuaciones y leve tendencia a subir al final del entrenamiento, reforzando el sobreajuste.



## 2. Matriz de Confusión:

El modelo clasificó correctamente 208 casos de Neumonía Bacteriana, 135 casos Normales y 115 casos de Neumonía Viral. Los principales errores se observaron en la confusión de casos "NORMALES" predichos como "VIRAL\_PNEUMONIA" (79 casos) y casos de "BACTERIAL\_PNEUMONIA" predichos como "VIRAL\_PNEUMONIA" (27 casos).

Accuracy global: El modelo alcanzó una precisión general del 73%.

Classification Report:				
	precision	recall	f1-score	support
BACTERIAL_PNEUMONIA	0.80	0.86	0.83	242
NORMAL	0.94	0.58	0.71	234
VIRAL_PNEUMONIA	0.52	0.78	0.62	148
accuracy			0.73	624
macro avg	0.75	0.74	0.72	624
weighted avg	0.79	0.73	0.74	624

Por clase:

**BACTERIAL\_PNEUMONIA:** Presentó una buena precisión (0.80) y un excelente recall (0.86), indicando que el modelo es eficaz detectando la mayoría de los casos reales.

**NORMAL:** Tuvo una alta precisión (0.94) al predecir casos normales, pero un recall bajo (0.58), lo que significa que un número significativo de casos normales reales fueron clasificados erróneamente en otras categorías (principalmente como neumonía viral, como se vio en la matriz de confusión).

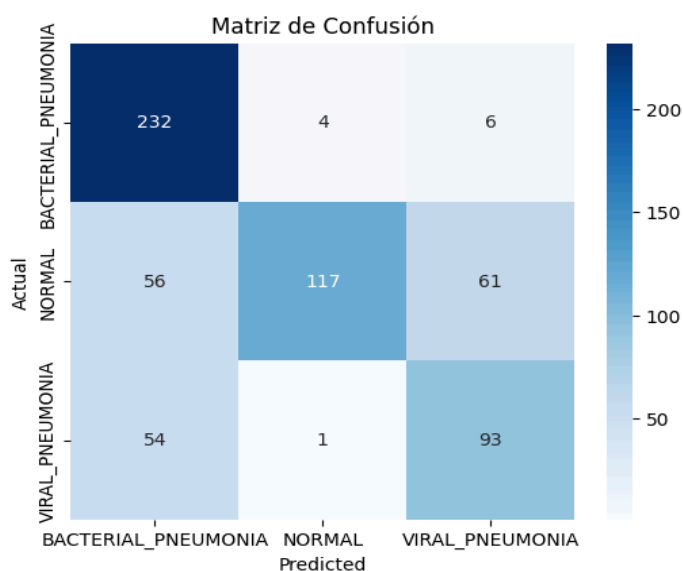
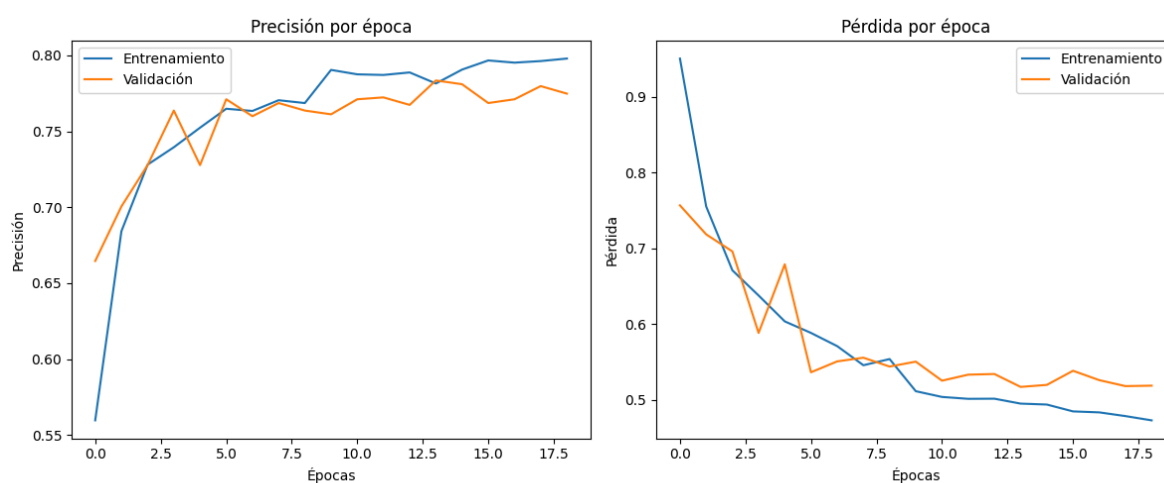
**VIRAL\_PNEUMONIA:** Tuvo la precisión más baja (0.52), indicando que muchas de sus predicciones de neumonía viral eran incorrectas, pero su recall (0.78) fue decente, capturando una buena parte de los casos reales.

## Ejecucion 2

Se balancean los datos de forma manual para evitar sesgos, ya que el modelo tendía a inclinarse por la categoría de Neumonía Bacteriana la cual poseía un aproximado de 1500 datos extra por sobre las otras 2 categorías y se realizó una distribución de 70, 15, 15 de los datos para train, test y val.

1. Gráficos de Precisión y Pérdida por Época:

La precisión de entrenamiento siguió una tendencia creciente, y la precisión de validación (línea naranja) también mostró una mejora general, manteniéndose más estable en las últimas épocas en comparación con la Ejecución 1. Sugiriendo que el balanceo de datos ayudó a mejorar la generalización, pero existiendo aún una brecha con la precisión de entrenamiento. En la pérdida de entrenamiento se vio disminuyendo y la de validación fue más estable con valores más bajos hacia el final en comparación que la ejecución 1, siendo positivo por que muestra una mejor capacidad de generalización.



## 2. Matriz de confusión

En comparación con la Ejecución 1, observamos algunos cambios en la distribución de las predicciones.

**Aciertos:** El modelo clasificó correctamente 232 casos de Neumonía Bacteriana (un aumento significativo respecto a los 208 anteriores), 117 casos Normales (una

ligera disminución de 135) y 93 casos de Neumonía Viral (una disminución respecto a los 115 anteriores).

### Errores:

Los falsos positivos de "NORMAL" clasificados como "VIRAL\_PNEUMONIA" (61 casos) se redujeron (antes eran 79).

Sin embargo, se observa un aumento considerable de casos de "NORMAL" clasificados erróneamente como "BACTERIAL\_PNEUMONIA" (56 casos, antes solo 20).

Además, los casos de "VIRAL\_PNEUMONIA" predichos como "BACTERIAL\_PNEUMONIA" (54 casos) aumentaron sustancialmente (antes eran 31).

#### Classification Report:

	precision	recall	f1-score	support
BACTERIAL_PNEUMONIA	0.68	0.96	0.79	242
NORMAL	0.96	0.50	0.66	234
VIRAL_PNEUMONIA	0.58	0.63	0.60	148
accuracy			0.71	624
macro avg	0.74	0.70	0.69	624
weighted avg	0.76	0.71	0.70	624

El accuracy global disminuyó ligeramente a **0.71%**

Por clase:

BACTERIAL\_PNEUMONIA: El recall mejoró notablemente a 0.96 (antes 0.86), lo que significa que el modelo ahora detecta casi todos los casos reales de neumonía bacteriana. Sin embargo, su precisión disminuyó a 0.68 (antes 0.80), indicando que, cuando predice neumonía bacteriana, es menos fiable porque incluye más falsos positivos.



NORMAL: La precisión se mantuvo alta en 0.96 (antes 0.94), pero el recall disminuyó aún más a 0.50 (antes 0.58), lo que confirma que el modelo sigue teniendo dificultades para identificar correctamente todos los casos normales, confundándose con neumonía.

VIRAL\_PNEUMONIA: Tanto la precisión (0.58) como el recall (0.63) experimentaron una ligera mejora comparado con la Ejecución 1 (0.52 y 0.78 respectivamente), aunque sigue siendo la clase con el rendimiento más bajo.

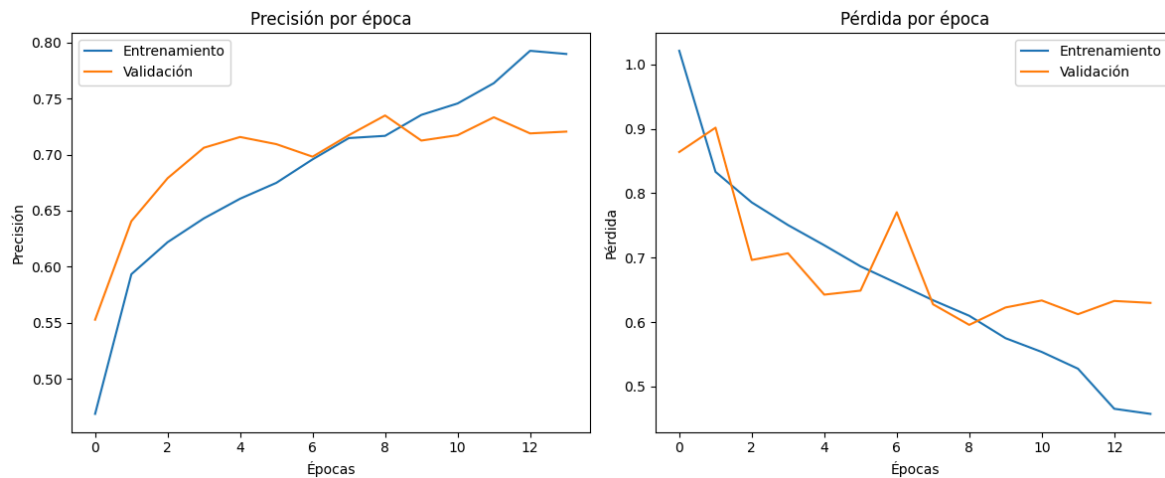
### Ejecución 3

En esta ejecución se tocaron los valores de `BATCH_SIZE`, se redujeron de 32 a 16 y también se tocaron los hiperparametros de rotación, zoom y contraste para darle más variedad a los datos generados.

#### 1. Gráficos de Precisión y Pérdida por Época:

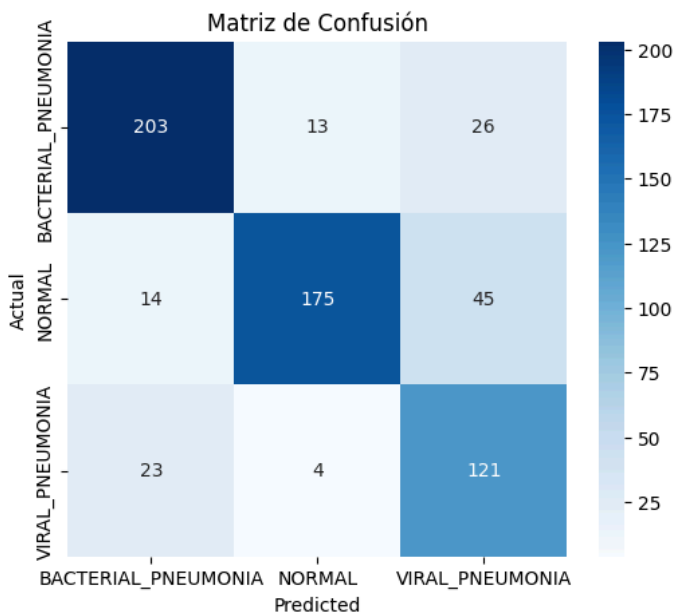
La precisión de entrenamiento continuó su ascenso, mientras que la precisión de validación (línea naranja) mostró una mejora significativa y una mayor estabilidad en comparación con las ejecuciones anteriores. La brecha entre entrenamiento y validación se mantuvo relativamente pequeña, lo que es un indicio de menor sobreajuste. Alrededor de la época 8-9, ambas curvas convergen muy bien, y la validación mantiene una buena línea hasta el final. La pérdida de entrenamiento siguió disminuyendo. La pérdida de validación (línea naranja) también mostró una tendencia general a la baja y, aunque con alguna fluctuación, se mantuvo en valores bajos y siguió más de cerca la curva de entrenamiento, lo que sugiere que el modelo está

aprendiendo de manera más eficiente y generalizando mejor a los datos no vistos.



## 2. Matriz de confusión

Esta matriz de confusión muestra una distribución de aciertos y errores notablemente mejorada en comparación con la Ejecución 2.



Hubo un aumento significativo en los aciertos para casos NORMALES (175) y VIRALES (121), mientras que los de Neumonía Bacteriana se mantuvieron fuertes (203). Lo más importante es que se logró una reducción sustancial en los errores donde casos NORMALES eran confundidos con neumonías (bajando de 56 a 14 para bacteriana y de 61 a 45 para viral), y la confusión de neumonía viral con bacteriana también disminuyó (de 54 a 23)

Classification Report:				
	precision	recall	f1-score	support
BACTERIAL_PNEUMONIA	0.85	0.84	0.84	242
NORMAL	0.91	0.75	0.82	234
VIRAL_PNEUMONIA	0.63	0.82	0.71	148
accuracy			0.80	624
macro avg	0.80	0.80	0.79	624
weighted avg	0.82	0.80	0.80	624

El accuracy global **mejoró significativamente a 0.80%** (un aumento de 0.71% en la Ejecución 2 y 0.73% en la Ejecución 1).

Por clase:

BACTERIAL\_PNEUMONIA: Mantiene un excelente recall de 0.84 (ligeramente inferior a 0.96 de Ejecución 2,) y una precisión mejorada a 0.85 (antes 0.68),.

NORMAL: Experimenta una mejora notable en el recall a 0.75 (antes 0.50), y su precisión se mantiene muy alta (0.91). Esto significa que el modelo es mucho mejor identificando los casos normales y también es muy confiable cuando los predice.

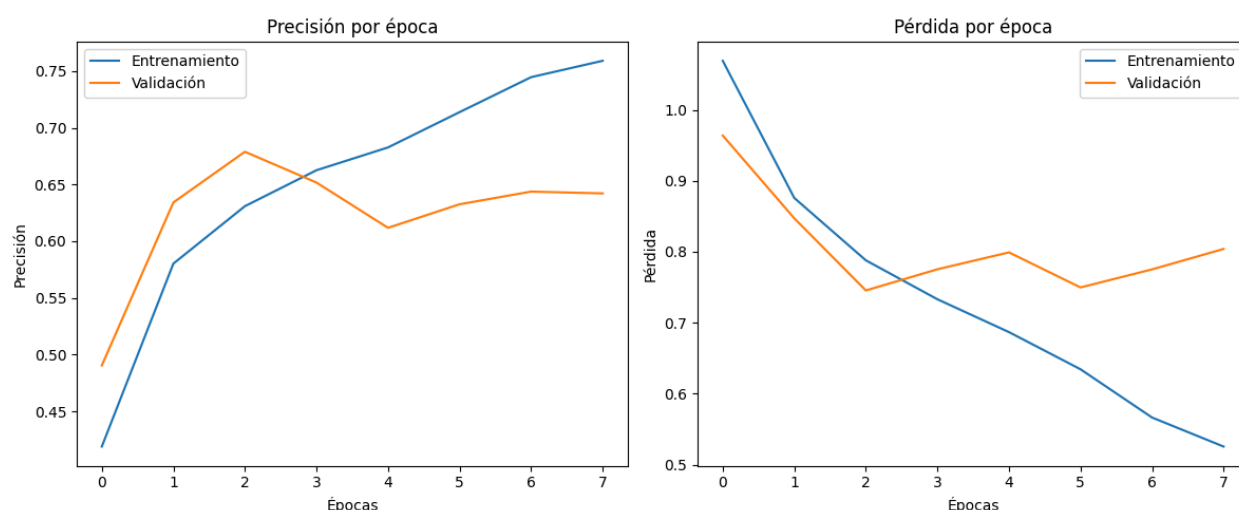
VIRAL\_PNEUMONIA: Muestra una mejora en la precisión a 0.63 (antes 0.58) y un recall de 0.82 (antes 0.63), lo que representa una mejora.

#### Ejecución 4

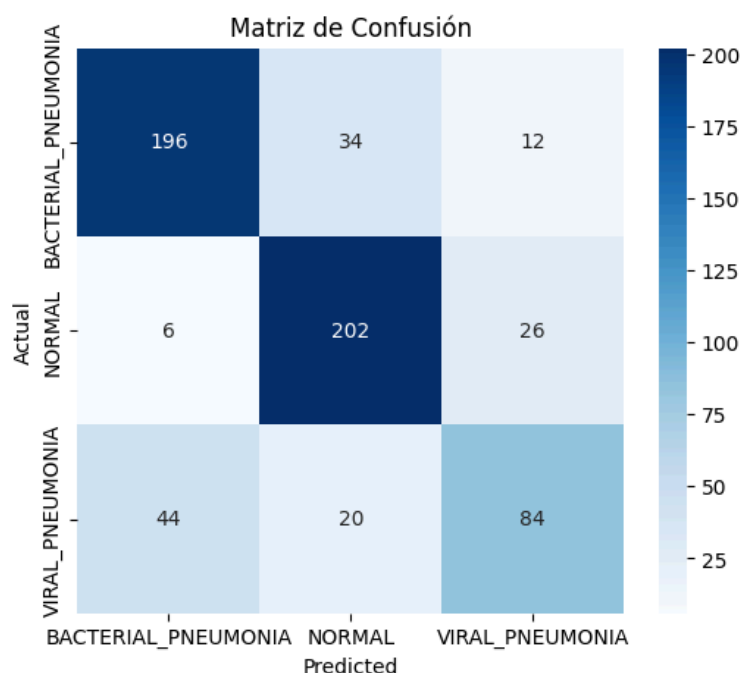
Se decidió incrementar la cantidad de filtros en las capas convolucionales con el objetivo de aumentar la capacidad del modelo para extraer y representar características más complejas de las imágenes. Al duplicar la cantidad de filtros, cada capa puede aprender una mayor variedad de patrones espaciales (bordes, texturas, formas), lo que resulta especialmente útil en tareas de clasificación multiclase con imágenes de alta resolución

1. Gráficos de Precisión y Pérdida por Época: Precisión:

La precisión de entrenamiento continuó aumentando, mientras que la precisión de validación (línea naranja) mostró un comportamiento mixto mostrando una brecha creciente con la precisión de entrenamiento. Esto indica que el modelo, al tener más capacidad de aprendizaje, podría estar sobre ajustándose a los datos de entrenamiento con los filtros adicionales, perdiendo un poco de su capacidad de generalización en esta configuración. La pérdida de entrenamiento disminuyó constantemente. Sin embargo, la pérdida de validación (línea naranja) se mantuvo por encima de la de entrenamiento y mostró fluctuaciones, con una tendencia general a no disminuir de manera tan consistente como en la Ejecución 3.



## 2. Matriz de confusión



Comparada con la Ejecución 3, esta matriz muestra un rendimiento mixto. Si bien hubo una mejora notable en la identificación de casos NORMALES (202 aciertos) y una reducción de sus errores como neumonía bacteriana (de 14 a 6), se observó un deterioro significativo en la detección de VIRAL\_PNEUMONIA (solo 84 aciertos). Esto se debe a un aumento considerable de falsos negativos para la

clase viral, donde el modelo ahora confunde más la neumonía viral con la bacteriana o incluso con casos normales. También aumentaron los falsos positivos para NORMAL desde BACTERIAL\_PNEUMONIA.

Classification Report:				
	precision	recall	f1-score	support
BACTERIAL_PNEUMONIA	0.80	0.81	0.80	242
NORMAL	0.79	0.86	0.82	234
VIRAL_PNEUMONIA	0.69	0.57	0.62	148
accuracy			0.77	624
macro avg	0.76	0.75	0.75	624
weighted avg	0.77	0.77	0.77	624

Accuracy global: El accuracy global disminuyó a 0.77% (antes 0.80% en Ejecución 3). Esto sugiere que el aumento de filtros, en esta configuración, no mejoró el rendimiento global y podría haber introducido sobreajuste o complejidades innecesarias.

Por clase:

BACTERIAL\_PNEUMONIA: La precisión (0.80) se mantuvo similar.

NORMAL: Tuvo una mejora significativa en el recall a 0.86 (antes 0.75), lo cual es muy positivo, lo que significa que el modelo es mucho mejor identificando la mayoría de los casos normales. Su precisión se mantuvo alta (0.79).

VIRAL\_PNEUMONIA: Esta clase fue la más afectada. Su precisión aumentó a 0.69 (antes 0.63), pero su recall disminuyó drásticamente a 0.57 (antes 0.82). Esto significa que el modelo es más preciso cuando predice neumonía viral, pero pasa por alto una gran cantidad de casos reales de neumonía viral.

## 11. Conclusiones de los cambios

El proyecto demuestra que un modelo de Deep Learning puede ser una herramienta prometedora para la clasificación de neumonías. La iteración clave fue la Ejecución 3, donde el balanceo de datos (para tratar el desequilibrio en la cantidad de imágenes por tipo de

neumonía) en conjunto con una cuidadosa calibración del **BATCH\_SIZE** (el número de imágenes procesadas a la vez durante el entrenamiento, impactando la estabilidad y la exploración del modelo) y los parámetros de aumento de datos (como la rotación o el zoom para crear más variedad de imágenes), condujo al modelo con el mejor rendimiento general y el balance más deseable entre precisión y recall para todas las clases. Esto hace resaltar que la efectividad no siempre reside en aumentar la complejidad del modelo (por ejemplo, con más filtros para buscar más detalles, como se hizo en la Ejecución 4), sino en optimizarlo de forma más global considerando tanto la arquitectura, los parámetros de entrenamiento como la calidad y la diversidad de los datos de entrada. Futuras mejoras podrían centrarse en explorar arquitecturas de modelos pre-entrenados o técnicas más avanzadas de manejo de clases desbalanceadas o sobreajuste para refinar aún más la detección de la neumonía viral sin sacrificar el rendimiento en las otras clases.

## 12. Consideraciones claves para uso Clínico y Ético

Para que una herramienta como esta sea útil en la práctica médica, es vital ir más allá de los resultados técnicos.

Para la aplicación práctica, el modelo necesitaría validación estricta de expertos médicos y ajustar los criterios de decisión para necesidades clínicas específicas (por ejemplo, si se prefiere un mayor **Recall** a pesar de una disminución en la **Precision** para no perder casos críticos). Es una herramienta de apoyo, no un diagnóstico definitivo; su buen aprovechamiento radicaría en complementar el juicio profesional clínico, no en reemplazarlo.

Factores Éticos Cruciales:

- **Sesgo y Equidad:** Hay que asegurar que los datos de entrenamiento sean diversos y representativos para evitar sesgos que puedan llevar a diagnósticos erróneos o desigualdades en la atención de pacientes.
- **Transparencia y Explicabilidad:** Los médicos deben poder entender cómo el modelo llega a sus conclusiones. Un diagnóstico de IA sin una justificación clara puede generar desconfianza y dificultad en el manejo o aceptación del mismo.
- **Responsabilidad:** La IA funciona como un asistente inteligente, proporcionando una segunda opinión valiosa. La decisión final sobre el diagnóstico y el tratamiento siempre será del profesional de la salud.

Este proyecto destaca el potencial de la IA en la medicina, pero también la necesidad crítica de un desarrollo responsable y un uso consciente de estas tecnologías.

