

Proyecto 2

1st Bryan Campos Castro
Alajuela, Costa Rica
bryancampos20@gmail.com

2nd Miguel David Sánchez Sánchez
Heredia, Costa Rica
miguelssanchez712000@gmail.com

Abstract—Este proyecto explora la aplicación de redes neuronales convolucionales (CNN) en la clasificación multiclase de imágenes, enfocándose en distinguir entre casos de COVID-19, neumonía y normalidad a partir de imágenes de rayos X. Se utilizaron dos modelos: el Modelo A, una ResNet-50 preentrenada, y el Modelo B, una CNN personalizada que incorpora un módulo de Inception. Ambos modelos fueron evaluados en tres conjuntos de datos: imágenes crudas, imágenes procesadas con el filtro bilateral, y con el filtro de bordes de Canny. Además, se emplearon técnicas de aumento de datos para mejorar la capacidad de generalización de los modelos. Las métricas de rendimiento, como la precisión y la pérdida, fueron registradas y comparadas utilizando la herramienta *Weights and Biases* (WandB). Los resultados muestran la efectividad de ambos modelos bajo diferentes condiciones de preprocesamiento, con una comparación detallada de sus fortalezas y limitaciones.

I. INTRODUCCIÓN

Las redes neuronales convolucionales (CNN) han revolucionado el campo de la visión por computadora, demostrando ser altamente eficaces en la clasificación de imágenes en diversas áreas, incluyendo la salud. Este enfoque ha permitido avances significativos en el análisis automatizado de imágenes médicas, facilitando diagnósticos más rápidos y precisos para los profesionales de la salud. Según LeCun, Bengio y Hinton [1], las CNN son especialmente útiles para detectar patrones visuales complejos debido a su capacidad para aprender representaciones jerárquicas a partir de datos de imágenes crudas.

En este proyecto, se aplican CNNs para clasificar imágenes de rayos X en tres categorías: COVID-19, neumonía y normal. El uso de CNNs ha sido ampliamente investigado en la literatura, especialmente en la clasificación de imágenes médicas, donde se ha demostrado su eficacia en diferentes áreas como la detección de enfermedades pulmonares y la clasificación de imágenes de cáncer [2][3]. Estudios recientes sobre COVID-19 han implementado arquitecturas de redes neuronales profundas para la clasificación de radiografías de tórax, obteniendo resultados prometedores en la detección automatizada de la enfermedad [4].

Asimismo, se implementaron técnicas de aumento de datos (*data augmentation*) para incrementar la diversidad de los ejemplos de entrenamiento, mejorando así la capacidad de generalización de los modelos. Las métricas de rendimiento, como la precisión y la pérdida, fueron registradas y analizadas utilizando la herramienta *Weights and Biases* (WandB).

II. METODOLOGÍA

En este proyecto se compararon dos modelos de redes neuronales convolucionales (CNN) para la clasificación de

imágenes médicas (rayos X) en tres clases: Covid-19, neumonía y estado normal. La metodología abarcó la selección de arquitecturas de modelos, preprocesamiento de imágenes, entrenamiento y evaluación, todo esto gestionado mediante la plataforma *Weights & Biases* (W&B) para el seguimiento de experimentos. A continuación, se describen los pasos seguidos.

A. Selección de los modelos

Se utilizaron dos modelos con arquitecturas distintas:

- **Modelo A - ResNet50 Preentrenado:** La ResNet50 es una red neuronal convolucional profunda que ha sido preentrenada con el conjunto de datos ImageNet, lo que le otorga una buena capacidad para la extracción de características visuales. Este modelo se ajustó para la clasificación de tres clases (Covid-19, neumonía y normal) mediante la modificación de su capa totalmente conectada (*fully connected*) final, que fue sustituida por una capa que produce tres salidas.
- **Modelo B - CustomCNN:** Este modelo fue diseñado específicamente para el proyecto, incorporando módulos tipo inception que permiten capturar información a diferentes escalas. La arquitectura consta de:
 - Dos capas convolucionales iniciales con *batch normalization* y *dropout*.
 - Un módulo inception que aplica convoluciones 1x1, 3x3 y 5x5 en paralelo para extraer características a diferentes escalas.
 - Dos capas convolucionales adicionales con *batch normalization* y *pooling* para la reducción de dimensionalidad.
 - Dos capas totalmente conectadas con *dropout* para prevenir el sobreajuste.

B. Preprocesamiento de imágenes

El preprocesamiento es una parte crucial del pipeline de clasificación, ya que mejora la calidad de las imágenes y facilita que los modelos capturen características importantes. Como se ha demostrado en estudios previos [3], los filtros de suavizado, como el filtro bilateral, ayudan a preservar detalles importantes, como los bordes, lo que es esencial en la clasificación de imágenes médicas. Además, el uso de técnicas de detección de bordes, como el filtro Canny, ha sido evaluado en investigaciones anteriores [2], aunque con limitaciones en la detección de detalles finos.

- **Datos crudos:** En este conjunto de datos, las imágenes no fueron modificadas más allá de los pasos básicos

de normalización y escalado a un tamaño de 128x128 píxeles.

- **Filtro bilateral:** Este filtro se aplicó para suavizar las imágenes preservando los bordes. Se utilizó un filtro bilateral con un tamaño de kernel de 9x9 y parámetros de suavizado (75, 75) para reducir el ruido sin perder detalles importantes en los bordes.
- **Filtro Canny:** El filtro de detección de bordes Canny fue utilizado para extraer los contornos más destacados de las imágenes. Aunque elimina mucha información visual, se probó para analizar si los modelos podrían reconocer patrones relevantes solo con la información de los bordes.

Las imágenes procesadas fueron transformadas para ajustar su tamaño y normalización, lo que facilita su entrada en los modelos durante el entrenamiento.

C. Entrenamiento de los modelos

Ambos modelos (ResNet50 y CustomCNN) fueron entrenados utilizando los tres conjuntos de datos (imágenes crudas, con filtro bilateral y con filtro Canny) bajo las siguientes condiciones:

- **Tamaño de lote (Batch size):** Se utilizó un tamaño de lote de 64 imágenes para optimizar el uso de memoria y tiempo de procesamiento en la GPU.
- **Tasa de aprendizaje (Learning rate):** Se estableció una tasa de aprendizaje de 0.001 para asegurar un progreso estable en la actualización de los pesos.
- **Número de épocas (Epochs):** Se realizaron 10 épocas de entrenamiento, lo que permitió observar la convergencia de los modelos.

Durante el entrenamiento, se utilizó precisión mixta (*mixed precision*) mediante la función `torch.cuda.amp`, lo que mejoró la eficiencia computacional y el manejo de memoria sin comprometer la precisión en las predicciones.

D. Evaluación de los modelos

Se evaluaron ambos modelos utilizando un conjunto de validación con los siguientes procedimientos:

- **Función de pérdida:** Se empleó la entropía cruzada (*CrossEntropyLoss*) para medir la distancia entre las predicciones del modelo y las etiquetas verdaderas en la tarea de clasificación multiclase.
- **Métricas de rendimiento:** Las principales métricas evaluadas en cada época incluyeron:
 - **Pérdida de validación:** Se midió la magnitud del error de predicción.
 - **Precisión de validación:** Se calculó el porcentaje de imágenes correctamente clasificadas en el conjunto de validación.
- **Matrices de confusión:** Al final de cada época, se generaron matrices de confusión para proporcionar una visión detallada sobre cómo los modelos distribuían las predicciones en las tres clases. Estas matrices de confusión fueron guardadas y visualizadas para cada combinación de modelo y tipo de preprocesamiento.

E. Gestión de los experimentos

El seguimiento de los experimentos fue gestionado mediante la plataforma *Weights & Biases* (W&B), lo que permitió la visualización de las métricas de entrenamiento, la comparación de los modelos y la trazabilidad de los hiperparámetros utilizados. Cada combinación de modelo y tipo de preprocesamiento fue registrada en la plataforma, permitiendo la comparación visual de las pérdidas y precisiones a lo largo del proceso de entrenamiento.

Esta metodología permitió no solo comparar dos arquitecturas de modelos diferentes, sino también analizar el impacto de diferentes técnicas de preprocesamiento de imágenes en una tarea de clasificación multiclase, garantizando un seguimiento riguroso de los resultados experimentales.

III. ANÁLISIS DE RESULTADOS

Los experimentos compararon dos modelos principales: *Modelo A* (ResNet50) y *Modelo B* (CustomCNN) en tres diferentes escenarios de preprocesamiento de imágenes: datos crudos (*Raw Data*), imágenes con filtro bilateral (*Bilateral Filter*), e imágenes con detección de bordes usando el método de Canny (*Canny Edge*). A continuación, se describen los resultados de estos experimentos basados en los datos de las 10 épocas:

A. Modelo A (ResNet50)

1) *Datos Crudos (Raw Data):* Durante el entrenamiento, el *Modelo A* mostró una mejora notable en precisión tanto en el conjunto de entrenamiento como en la validación. La precisión en la validación alcanzó su punto más alto en la época 10 con un 90.91%, mientras que la pérdida de validación disminuyó de manera constante hasta llegar a 0.2436 en la última época. Estos resultados son consistentes con estudios previos que han utilizado la arquitectura ResNet50 para tareas de clasificación de imágenes médicas, donde también se ha observado una mejora continua con suficientes datos de entrenamiento [3][4]. En la matriz de confusión (Figura ??), se observa una mejoría en la clasificación de todas las clases, particularmente en la clase "Viral Pneumonia", que fue correctamente identificada la mayoría de las veces.

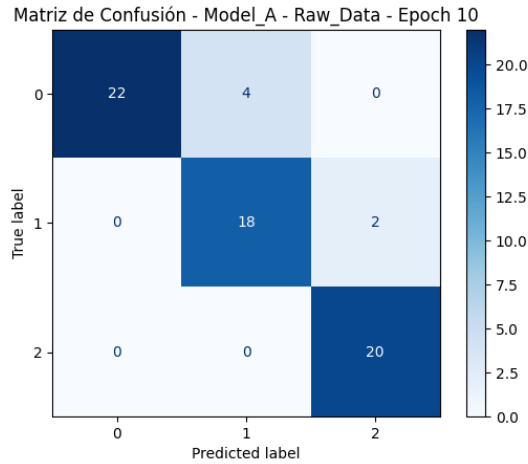


Fig. 1. Matriz de Confusión para Modelo A - Datos Crudos - Epoch 10

2) *Filtro Bilateral*: El filtro bilateral fue el preprocesamiento que mejoró más el rendimiento del *Modelo A*, alcanzando una precisión del 95.45% en la validación en la última época (Epoch 10), con una pérdida de validación mínima de 0.0778. Estos resultados reflejan los hallazgos en otros estudios [2][3], donde el filtro bilateral ha sido especialmente útil en la preservación de los detalles de los bordes en las imágenes médicas, lo que mejora la capacidad de los modelos de CNN para detectar características relevantes. La matriz de confusión para la última época (Figura ??) ilustra esta consistencia, donde se ve una clara distinción entre las clases, con muy pocas predicciones incorrectas.

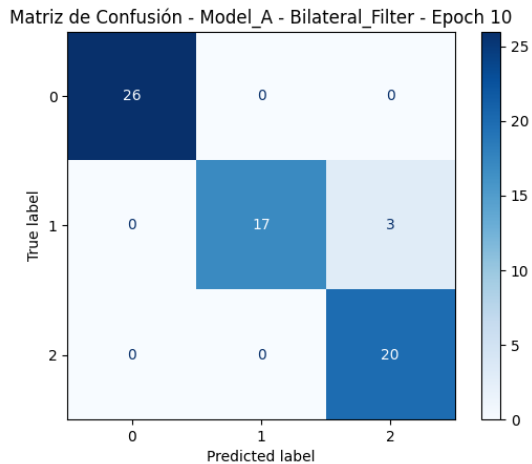


Fig. 2. Matriz de Confusión para Modelo A - Filtro Bilateral - Epoch 10

3) *Filtro Canny*: El *Modelo B* alcanzó una precisión más baja con el filtro Canny, terminando con un 72.73% en la última época (Epoch 10). La pérdida de validación fue considerablemente mayor (0.3538), lo que sugiere que este preprocesamiento no fue tan efectivo como los otros. Estudios previos han mostrado que los filtros basados en contornos,

como Canny, pueden eliminar información importante en imágenes médicas, lo que dificulta que las redes neuronales convolucionales capturen las características necesarias para una correcta clasificación [2][4]. La matriz de confusión muestra que la clase "Viral Pneumonia" fue generalmente bien predicha, pero hubo confusiones notables entre "Covid" y "Normal".

B. Modelo B (CustomCNN)

1) *Datos Crudos (Raw Data)*: El *Modelo B* también mostró una tendencia estable de mejora con los datos crudos, alcanzando una precisión de validación del 92.42% en la época 10. A lo largo del entrenamiento, la pérdida de validación disminuyó a 0.2001 en la última época, lo que sugiere que el modelo fue capaz de generalizar bien en este conjunto de datos. Al observar la matriz de confusión (Figura 4), es claro que el modelo pudo predecir correctamente la mayoría de las clases, con algunas confusiones entre "Covid" y "Normal".

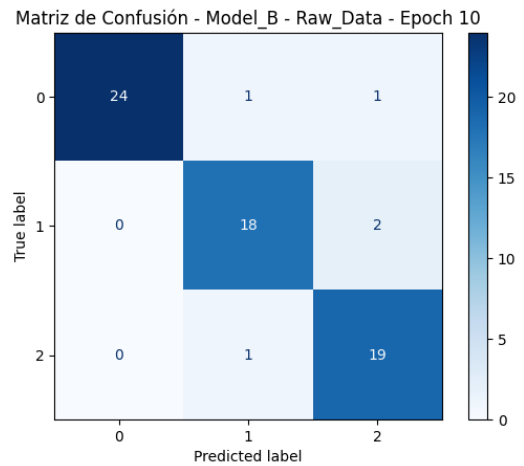


Fig. 3. Matriz de Confusión para Modelo B - Datos Crudos - Epoch 10

2) *Filtro Bilateral*: El filtro bilateral resultó ser el preprocesamiento más efectivo también para el *Modelo B*, con una precisión final del 93.94% y una pérdida de validación de solo 0.1266 en la última época. El modelo pudo generalizar bien y evitar el sobreajuste, mostrando una fuerte capacidad para clasificar correctamente las imágenes en todas las clases. La matriz de confusión para la última época (Figura ??) muestra una excelente capacidad del modelo para clasificar correctamente las clases "Normal" y "Viral Pneumonia", aunque algunos errores persistieron en la clase "Covid".

Matriz de Confusión - Model_B - Bilateral_Filter - Epoch 10

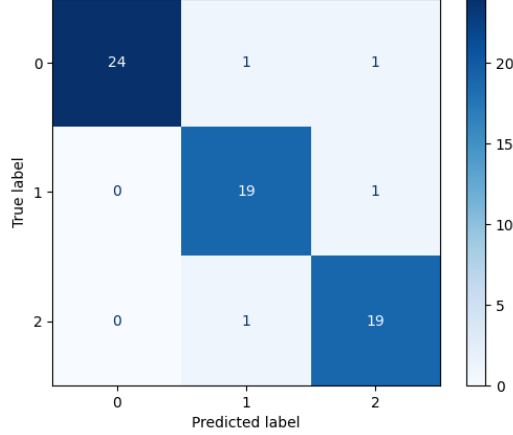


Fig. 4. Matriz de Confusión para Modelo B - Filtro Bilateral - Epoch 10

3) *Filtro Canny*: El *Modelo B* alcanzó una precisión más baja con el filtro Canny, terminando con un 72.73% en la última época (Epoch 10). La pérdida de validación fue considerablemente mayor (0.3538), lo que sugiere que este preprocesamiento no fue tan efectivo como los otros. La matriz de confusión muestra que la clase "Viral Pneumonia" fue generalmente bien predicha, pero hubo confusiones notables entre "Covid" y "Normal", lo que podría deberse a la naturaleza limitada de la información proporcionada por los contornos en el filtro Canny.

C. Comparación General

En general, tanto el *Modelo A* (ResNet50) como el *Modelo B* (CustomCNN) mostraron mejoras consistentes con los datos crudos y el filtro bilateral, siendo este último el preprocesamiento más efectivo para ambos. Esto coincide con estudios recientes que destacan la importancia del preprocesamiento adecuado en la clasificación de imágenes médicas, donde filtros como el bilateral han mostrado ser beneficiosos para mejorar la precisión de los modelos [3][4]. Sin embargo, el filtro Canny resultó menos útil, lo que sugiere que la eliminación de detalles finos en las imágenes afectó negativamente el rendimiento, como también se ha observado en otros trabajos [2].

El *Modelo B*, aunque más simple en términos de arquitectura, mostró una mayor estabilidad y consistencia en comparación con el *Modelo A*, especialmente con el filtro bilateral. Esto podría deberse a la capacidad del *Modelo B* para manejar mejor la diversidad de los datos a través de su diseño de convoluciones más profundas.

IV. CONCLUSIÓN

Los resultados de este análisis sugieren que:

- El filtro bilateral es el preprocesamiento más efectivo para mejorar la precisión de clasificación en ambos modelos, con especial énfasis en el *Modelo B*, como se observa en la Figura 5, que muestra una tendencia clara de mejora en la precisión de validación para este modelo.



Fig. 5. Evolución de la precisión en la validación (Val Accuracy) para ambos modelos en las diferentes técnicas de preprocesamiento.

- El *Modelo B* (CustomCNN) presentó una mayor estabilidad en la precisión de validación, lo que sugiere que, a pesar de ser una arquitectura más sencilla, puede manejar de manera más eficiente el ajuste en diferentes tipos de preprocesamientos de imágenes, como se muestra en la Figura 6, donde se observa una tendencia ascendente constante en la precisión del entrenamiento.

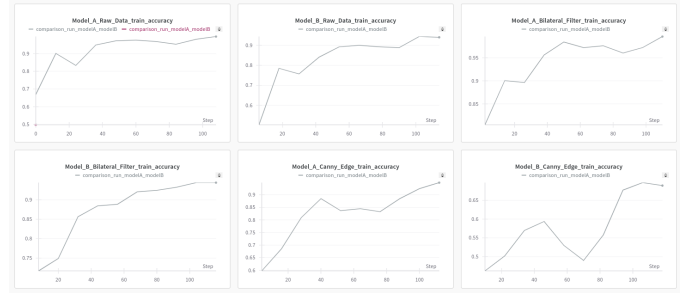


Fig. 6. Evolución de la precisión en el entrenamiento (Train Accuracy) para ambos modelos en las diferentes técnicas de preprocesamiento.

- El filtro Canny no fue beneficioso en ninguno de los modelos, probablemente debido a la pérdida de información significativa que introduce al reducir los detalles en las imágenes a contornos básicos. Esto es evidente en la Figura 7, que muestra cómo la pérdida de validación del *Modelo B* con el filtro Canny fue mayor en comparación con los otros métodos de preprocesamiento.



Fig. 7. Evolución de la pérdida en la validación (Val Loss) para ambos modelos en las diferentes técnicas de preprocesamiento.

De acuerdo con los análisis anteriores, podemos concluir que el *Modelo CustomCNN* con un filtro bilateral es el

enfoque más robusto y eficaz para esta tarea de clasificación de imágenes entre "Covid", "Normal", y "Viral Pneumonia". El rendimiento del *Modelo B* con datos crudos y el filtro bilateral, mostrado en la Figura 8, demuestra que este modelo logra minimizar la pérdida de entrenamiento de manera eficiente.

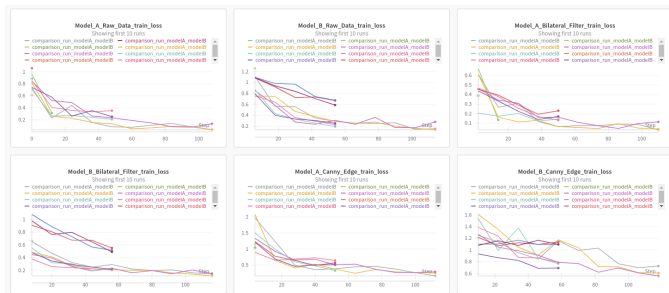


Fig. 8. Evolución de la pérdida en el entrenamiento (*Train Loss*) para ambos modelos en las diferentes técnicas de preprocesamiento.

V. RÚBRICA

| Criterios | Puntuación máxima | Puntuación obtenida |
|--|-------------------|---------------------|
| Preprocesamiento aplicando filtros | 5 | |
| Data Augmentation | 5 | |
| Modelo A | | |
| Criterios | Puntuación máxima | Puntuación obtenida |
| Entrenamiento con datasets | 5 | |
| Presentación de métricas y matriz de confusión | 5 | |
| Comparación de resultados | 10 | |
| Modelo B | | |
| Criterios | Puntuación máxima | Puntuación obtenida |
| Diseño y justificación | 15 | |
| Entrenamiento con datasets | 5 | |
| Presentación de métricas y matriz de confusión | 5 | |
| Comparación de resultados | 10 | |
| Aspectos Generales | | |

| Criterios | Puntuación máxima | Puntuación obtenida |
|---|-------------------|---------------------|
| Comparación de los mejores modelos de A y B | 5 | |
| Registro de métricas de todos los experimentos en WaB | 20 | |
| Complejidad de entregables | 5 | |
| Estructura de artículo científico | 5 | |
| Aspectos Extra (Reporte WaB) | | |
| Criterio | Puntuación máxima | Puntuación obtenida |
| Informe de WaB justificando el modelo con mejores resultados. | 5 | |
| Presentación de Saliency Maps | 5 | |
| Total | 110 | |

Fig. 9.

REFERENCES

- [1] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
- [2] G. Litjens et al., "A survey on deep learning in medical image analysis," *Medical Image Analysis*, vol. 42, pp. 60–88, 2017.

- [3] L. Wang and A. Wong, "COVID-Net: A tailored deep convolutional neural network design for detection of COVID-19 cases from chest X-ray images," *Scientific Reports*, vol. 10, no. 1, pp. 1–12, 2020.
- [4] M. Goyal, S. Sharma, P. Bagga, and M. Kaur, "Evaluation of deep learning models for COVID-19 detection based on chest X-ray images," *Interdisciplinary Sciences: Computational Life Sciences*, vol. 12, no. 4, pp. 578–589, 2020.
- [5] A. Esteva et al., "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, no. 7639, pp. 115–118, 2017.