

Detección de Fracturas en Radiografías Mediante Transfer Learning

Álvaro Salgado López

I. INTRODUCCIÓN

LA detección de fracturas en radiografías es una tarea crucial en el diagnóstico médico, ya que permite una identificación temprana y precisa de lesiones óseas, facilitando el tratamiento adecuado para los pacientes. Sin embargo, la evaluación manual de estas imágenes requiere de un alto grado de experiencia por parte de los especialistas y está sujeta a errores humanos. En este contexto, la aplicación de algoritmos de aprendizaje profundo, en particular las redes neuronales convolucionales (CNNs), ha demostrado ser una solución eficiente para la automatización del proceso de clasificación y detección de anomalías en imágenes médicas.

El presente estudio tiene como objetivo el diseño e implementación de un modelo basado en la arquitectura VGG16, utilizando técnicas de *transferlearning* y *fine – tuning*, para clasificar radiografías en dos categorías: con fractura y sin fractura. Esta metodología se justifica por el excelente desempeño reportado por modelos preentrenados en tareas de reconocimiento de patrones en imágenes médicas.

II. METODOLOGÍA

El conjunto de datos empleado en este estudio fue obtenido de *Kaggle* y está compuesto por radiografías clasificadas en dos grupos: imágenes de huesos fracturados y no fracturados, tanto para entrenamiento como para validación se aplicó un *ImageDataGenerator* para normalizar y aumentar artificialmente la cantidad de imágenes, generando variaciones en iluminación, rotaciones y escalados.

A. Arquitectura del Modelo

Se seleccionó la arquitectura VGG16 preentrenada con pesos de *ImageNet* debido a su eficacia en tareas de clasificación de imágenes. Se llevaron a cabo las siguientes modificaciones:

Se eliminaron las capas superiores del modelo original y se añadió una nueva capa completamente conectada con activación sigmoide para la clasificación binaria.

Se congelaron las primeras capas convolucionales para emplear *transferlearning*, permitiendo que el modelo aproveche las representaciones aprendidas previamente.

Se realizó *fine – tuning* sobre las capas superiores del modelo con el objetivo de adaptar las representaciones a las características específicas de las radiografías.

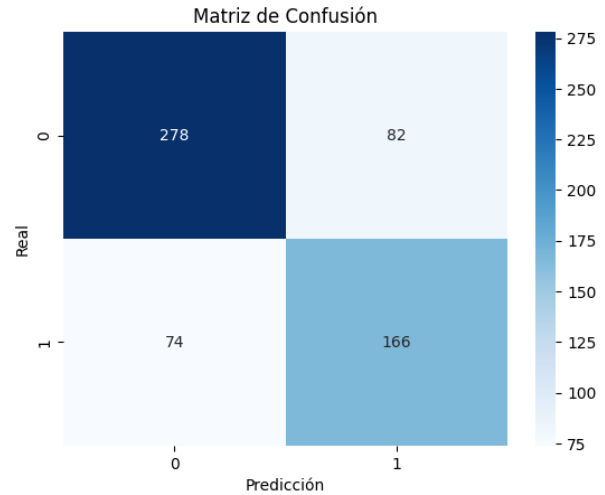


Fig. 1. Matriz de confusión en el conjunto de validación

B. Entrenamiento

El modelo se entrenó empleando la siguiente configuración:

- Optimizador: Adam con una tasa de aprendizaje inicial de $1e-4$, ajustada de manera dinámica mediante reducción de la tasa de aprendizaje en caso de estancamiento en la validación.
- Función de pérdida: *BinaryCrossentropy*, adecuada para problemas de clasificación binaria.
- Estrategias de regularización: Se utilizó *Dropout* para reducir el riesgo de sobreajuste y *EarlyStopping* para detener el entrenamiento cuando no se observaba mejora significativa en la validación.

El modelo fue entrenado durante múltiples épocas y se seleccionaron los pesos óptimos en función del mejor desempeño en el conjunto de validación.

III. RESULTADOS

Para evaluar el desempeño del modelo de clasificación basado en VGG16, se analizaron diversas métricas sobre el conjunto de datos de validación. En esta sección, se presentan los resultados obtenidos en términos de exactitud, junto con una interpretación detallada de la matriz de confusión. Estos análisis permiten identificar las fortalezas y limitaciones del modelo en la detección de fracturas en imágenes de rayos X, así como posibles estrategias para mejorar su desempeño.

En la Fig 1 se aprecia que el modelo tiene un mejor desempeño en la detección de imágenes sin fractura que en

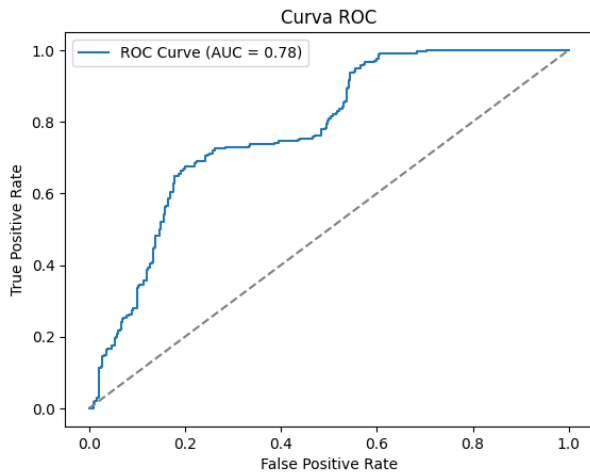


Fig. 2. Curva ROC

la identificación de fracturas. Esto se evidencia en la mayor cantidad de falsos negativos (74), lo que implica que hay fracturas que el modelo no está detectando.

La precisión general del modelo alcanzó un 74%, lo que indica un desempeño moderado. Aunque esta cifra es aceptable en un problema de clasificación binaria en imágenes médicas, un rendimiento superior es deseable debido a la importancia clínica del diagnóstico.

En la Fig 2 se observa la curva ROC obtenida, ésta muestra una capacidad moderada del modelo para distinguir entre imágenes con y sin fractura, con un área bajo la curva (AUC) de 0.78. Este valor indica que el modelo es significativamente mejor que un clasificador aleatorio, pero aún tiene margen de mejora.

En el contexto de la detección de fracturas, donde la precisión del diagnóstico es crítica, un AUC de 0.78 sugiere que el modelo puede diferenciar correctamente en la mayoría de los casos, aunque todavía presenta limitaciones. Si bien logra identificar una buena proporción de fracturas, el hecho de que no se aproxime a valores superiores a 0.85 indica que existen casos en los que puede generar falsos positivos o falsos negativos de manera significativa.

Dado que los falsos negativos pueden implicar riesgos clínicos considerables al no detectar fracturas reales, es importante considerar estrategias para mejorar la sensibilidad del modelo sin comprometer en exceso la especificidad. Algunas opciones incluyen un ajuste más profundo de los hiperparámetros, la aplicación de técnicas de preprocesamiento enfocadas en resaltar las estructuras óseas o la incorporación de arquitecturas más avanzadas. En general, si bien el desempeño del modelo es adecuado para un primer enfoque, es recomendable seguir optimizando su precisión antes de su posible implementación en un entorno clínico real.

IV. CONCLUSIÓN

El modelo basado en VGG16 mostró un desempeño aceptable en la clasificación de fracturas en radiografías, con una precisión global del 74% y un AUC de 0.78 en la curva ROC. Estos resultados indican que el modelo es capaz de diferenciar

entre imágenes con y sin fractura en la mayoría de los casos, pero aún presenta limitaciones que podrían comprometer su aplicabilidad en un entorno clínico.

Si bien la detección de fracturas es adecuada en términos generales, el modelo tiene un desempeño desigual entre clases, mostrando mejor precisión para la categoría “No Fractura” en comparación con “Fractura”. Esto sugiere que el modelo puede estar sesgado hacia la detección de imágenes sin fracturas, lo que podría derivar en falsos negativos con consecuencias potencialmente graves en un contexto médico.

Para mejorar el rendimiento, sería recomendable aplicar técnicas de aumento de datos enfocadas en mejorar la representación de fracturas, ajustar aún más las capas intermedias del modelo mediante *fine-tuning*. También se podría considerar el uso de enfoques híbridos con métodos de interpretación de imágenes médicas, como *Grad-CAM*, para mejorar la comprensión de las decisiones del modelo.

En conclusión, aunque el modelo representa un avance prometedor en la clasificación automatizada de fracturas en radiografías, se requieren mejoras adicionales para garantizar una precisión diagnóstica confiable antes de su implementación en entornos clínicos reales.