

Sistema de Clasificación y Detección de Tumores con Redes Neuronales Basado en Imágenes de GammaGrama Ósea

Cahum D., Guerrero A., Ortiz I., Reyes P., Sanchez M.

May 2026

1 Introducción

La gammagrafía ósea es una técnica de imagen ampliamente utilizada en el área de medicina nuclear para la detección de alteraciones metabólicas en el tejido óseo, como metástasis, infecciones o fracturas. Éste tipo de estudios permite identificar regiones de alta captación, conocidas como hot-spots, las cuales pueden indicar la presencia de patologías. Sin embargo, la interpretación de estas imágenes depende en gran medida de la experiencia del especialista, lo que puede generar variabilidad diagnóstica y posibles inconsistencias en la evaluación clínica.

Sin embargo, el análisis de este tipo de imágenes depende principalmente de la evaluación realizada por especialistas en medicina nuclear, lo que puede representar un proceso complejo. De igual forma, la interpretación visual puede verse afectada por fatiga o diversos factores que alteren y retrasen el diagnóstico.

Ante esta problemática, surge la necesidad de desarrollar herramientas computacionales capaces de asistir al especialista en el análisis de gammagrafías óseas, facilitando la identificación temprana de anomalías y mostrando mejores diagnósticos.

En los últimos años el desarrollo de técnicas de inteligencia artificial, particularmente aquellas basadas en aprendizaje profundo, ha revolucionado el análisis de imágenes médicas. Las redes neuronales convolucionales (CNN) han demostrado una alta capacidad para extraer características relevantes y reconocer patrones complejos, siendo así una herramienta eficaz para la automatización de tareas de detección y clasificación en diversos contextos clínicos.

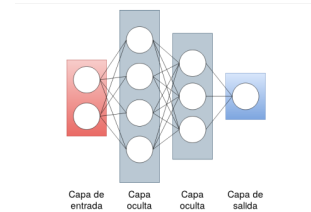


Figure 1: Estructura de una Red Neuronal Artificial.

El presente proyecto propone el desarrollo de un sistema computacional de apoyo al diagnóstico, orientado al análisis automatizado de gammagrafías óseas. La solución planteada se estructura en dos etapas principales: detección y clasificación. En la primera etapa, se emplea el modelo YOLO (You Only Look Once), el cual permite localizar regiones de interés dentro de las imágenes mediante la generación de bounding boxes y en una segunda etapa se trabajó con la red neuronal ResNet50.

2 Marco Teorico

En estudios de medicina nuclear, las redes neuronales convolucionales (CNN) han demostrado ser herramientas eficaces para la extracción de características complejas presentes en imágenes de gammagrafías, facilitando la identificación de lesiones metastásicas y otras alteraciones óseas.

Diversos estudios han implementado arquitecturas como ResNet34, U-Net, DenseNet, VGG16 y YOLO. Por ejemplo, trabajos basados en ResNet34 y U-Net reportaron precisiones de hasta 88.62 en

clasificación de metástasis (Lin et al., 2021), mientras que modelos basados en redes convolucionales profundas demostraron alta precisión diagnóstica en tareas de clasificación automática de gammagrafías óseas (Wang et al., 2024; Ibrahim et al., 2023). Asimismo, arquitecturas de detección aplicadas a gammagrafías óseas han mostrado resultados prometedores en la localización automática de lesiones y reducción de falsos positivos (Liu et al., 2021).

VGG16 es un red neuronal convolucional desarrollada por el Visual Geometry Group (VGG) de la universidad de Oxord. Esta arquitectura fue diseñada para tareas de clasificación de imágenes y se caracteriza por utilizar múltiples capas convolucionales pequeñas de 3x3, permitiendo extraer características visuales de manera progresiva.

Su principal función es extraer características relevantes de las imágenes como bordes, texturas, formas y patrones anatómicos presentes en estudios médicos.

Por otro lado SSD (Single Shot MultiBox Detector) es una arquitectura de detección de objetos basada en Deep Learnign que permite identificar y localizar múltiples objetos dentro de una imagen.

A diferencia de modelos tradicionales de clasificación, SSD no solo determina si existe una anomalía, también su ubicación exacta mediante bounding boxes. Esto resulta útil en aplicaciones médicas, donde es importante señalar visualmente las regiones sospechosas dentro de la gammagrafía.

El funcionamiento de SSD se basa en extraer características mediante una red backbone, en este caso VGG16. Genera predicciones sobre distintas regiones de la imagen, clasifica cada región y ajusta las coordenadas para delimitar las lesiones detectadas.

La combinación de VGG16 y SSD resulta efectiva para detectar y localizar anomalías en imágenes médicas. VGG16 actúa como extractor de características, identifica patrones relevantes dentro de las gammagrafías óseas, mientras SSD utiliza esas características para localizar regiones sospechosas mediante bounding boxes (Liu et al., 2021; Ibrahim et al., 2023). Debido a que esta arquitectura combina extracción de características y detección en un mismo modelo, permite identificar posibles anomalías de forma más rápida y facilita el análisis de las gammagrafías óseas.

Posteriormente se utilizo YOLO (*You Only Look Once*) es un sistema del estado del arte, que utiliza la Red Neuronal para la detección de objetos en tiempo real. Esta red divide la imagen en regiones y predice múltiples cajas delimitadoras (*bounding box*), y la probabilidad de detección de las clases de entrenamiento para cada caja delimitadora. Como resultado, el sistema imprime los objetos que detectó, su confianza y el tiempo de ejecución (Joseph Redmon Ali Farhadi. 2018).

La arquitectura ResNet, particularmente su variante ResNet-50, se ha consolidado como una de las redes neuronales convolucionales más utilizadas en el análisis de imágenes médicas debido a su capacidad para entrenar modelos profundos mediante conexiones residuales, las cuales permiten reducir el problema del desvanecimiento del gradiente y mejorar la extracción de características complejas. De acuerdo con Liu et al. (2021), ResNet50 puede emplearse eficazmente en la identificación automática de lesiones metastásicas sospechosas en gammagrafías óseas de cuerpo completo, alcanzando una precisión promedio de 81.23 por ciento, además de mostrar un desempeño competitivo frente a arquitecturas como InceptionV3, VGG16 y DenseNet169.

Asimismo, Papandrianos et al. (2020) señalan que las redes convolucionales profundas como ResNet50 representan herramientas altamente prometedoras para sistemas de diagnóstico asistido por computadora (CAD) en medicina nuclear, debido a su capacidad para aprender patrones visuales complejos presentes en imágenes de gammagrafía ósea utilizadas en la detección de metástasis derivadas del cáncer de mama.

3 Metodología

La metodología desarrollada en este proyecto se dividió en cuatro fases principales, enfocadas en la preparación del dataset, preprocesamiento de imágenes e implementación de arquitectura de inteligencia artificial para la detección y por último la clasificación de anomalías en gammagrafías óseas.

Etapa 1

En esta primera fase se utilizó el dataset "BS-

80K”, compuesto por más de 3 mil imágenes. Inicialmente, se desarrolló un código para verificar la cantidad total de imágenes y sus respectivos archivos XML de anotaciones. Durante esta revisión se identificó que no todas las imágenes contaban con un archivo XML asociado, por lo que únicamente se seleccionaron aquellas que tenían información válida para el entrenamiento. Como resultado, se obtuvo un conjunto de 2,924 imágenes, de cuales 1,741 correspondían a estudios normales y 1,183 estudios anormales.

Debido al desbalance, se realizó un balanceo del conjunto de datos reduciendo la cantidad de imágenes normales a 1,183, con el objetivo de trabajar con la misma cantidad de muestras de ambas categorías y evitar sesgos durante el entrenamiento.

Etapas 2

Se desarrolló un código en Google Colab para generar un archivo en formato Excel a partir de las anotaciones XML incluidas en el BS-80K. Este procedimiento permitió organizar la información correspondiente a las bounding box, considerando que una misma imagen podía contener múltiples regiones de interés. Cada fila del archivo representaba una bounding box distinta, facilitando posteriormente el procesamiento de las anotaciones durante el entrenamiento.

Se aplicaron técnicas de preprocesamiento sobre las imágenes médicas, incluyendo normalización y reescalamiento. Estas técnicas permitieron mejorar la calidad de las imágenes, sus dimensiones y optimizar la extracción de características necesarias para las arquitecturas implementadas.

Etapas 3 y 4

En esta etapa se desarrollaron 3 arquitecturas diferentes orientadas a tareas de detección y clasificación: YOLO (You Only Look Once), SSD combinado con VGG16 y ResNet50. Las arquitecturas YOLO y SSD + VGG16 fueron utilizadas para tareas de detección de anomalías mediante bounding boxes, mientras que ResNet50 se utilizó para tareas de clasificación automática de las gammagrafías óseas. Se inicializó el modelo YOLOv8 con imágenes redimensionadas a 640x640 píxeles para no sobrecargar el CPU. Se utilizó un tamaño (batch size) de 4 y se programaron 100 épocas. Se implementó un

mecanismo de parámetro (early stopping), para asegurar que si la red no mejoraba la precisión en 10 épocas, el entrenamiento se detendría para evitar un sobreajuste (overfitting).

ResNet50

Para la red neuronal elaborada con ResNet50 se trabajó en conjunto con una red ya pre-entrenada llamada “imagenet” la cual está compuesta por imágenes naturales, las primeras capas convolucionales de estas redes aprenden características generales y de bajo nivel como bordes, texturas, contrastes y patrones espaciales, que pueden reutilizarse eficazmente en dominios médicos. Esto permite reducir el tiempo de entrenamiento, mejorar la convergencia del modelo y obtener un desempeño adecuado incluso cuando se dispone de conjuntos de datos médicos limitados, situación común en aplicaciones clínicas. Además, diversos estudios han demostrado que los modelos preentrenados en ImageNet logran un rendimiento competitivo en tareas de clasificación médica al adaptarse mediante fine-tuning a imágenes especializadas (Shin et al., 2016). Durante una primera fase de entrenamiento, el modelo fue ajustado a lo largo de 15 épocas utilizando el optimizador Adam con una tasa de aprendizaje de 1×10^{-5} , la función de pérdida categorical crossentropy y la métrica de evaluación accuracy. Posteriormente, se aplicó una etapa de fine-tuning mediante entrenamiento adicional con un máximo de 20 épocas, incorporando un mecanismo de Early Stopping monitoreado a partir de la pérdida de validación con una paciencia de 5 épocas y restauración automática de los mejores pesos obtenidos. Esta estrategia permitió reducir el riesgo de sobreajuste y conservar el punto óptimo de generalización del modelo durante el entrenamiento.

SSD + VGG16

Para el desarrollo e implementación del sistema de detección de anomalías, se llevó a cabo una etapa de preparación y procesamiento de conjunto de datos. Como la finalidad de simplificar el procesamiento de datos, las imágenes normales y anormales fueron unificadas en un único directorio utilizando las bibliotecas os y shutil. Asimismo, se cargó un archivo Excel que contenía las coordenadas de las bounding boxes necesarias para identificar las regiones de in-

terés dentro de cada imagen.

Posteriormente, mediante la biblioteca Pandas, se estandarizaron los nombres de los archivos y se verificó la existencia de las imágenes dentro del directorio de trabajo, conservando únicamente aquellas válidas para el entrenamiento.

Una vez organizado el conjunto de datos, se implementó una clase personalizada denominada BoneDataset de PyTorch, encargada de cargar las imágenes y sus respectivas etiquetas. Las imágenes fueron redimensionadas a 300*300 píxeles para adaptarse al modelo SSD300, mientras que las coordenadas de las bounding boxes fueron escaladas proporcionalmente para conservar la correctalocalización de las anomalías.

Adicionalmente, el conjunto de datos fue dividido en entrenamiento y validación utilizando Scikit-learn, destiando 70 por ciento para el entrenamiento y el 30 por ciento en validación. Asimismo, se configuraron los objetos DataLoader para procesar las imágenes en lotes durante el entrenamiento.

El sistema de detección implementado estuvo basado en la arquitectura SSD300 con VGG16 preentrenado, obtenida mediante Torchvision. La arquitectura original fue modificada para trabajar en dos clases: Normal y anaormal. El modelo fue entrenado utilizando el optimizador Adam con una tasa de aprendizaje de 0.0001 y un máximo de 100 épocas. Además, se implementó la técnica de Early Stopping para prevenir el sobreajuste y conservar automáticamente el modelo con mejor desempeño.

Finalmente, el rendimiento del sistema fue evaluado mediante curvas de pérdida, matriz de confusión y la métrica Intersection over Union (IoU), utilizada para medir la coincidencia entre las cajas de predichas y las cajas reales. Dicha métrica permitió evaluar la precisión especial de las detecciones realizadas por la red neuronal, considerando como detecciones correctas aquellas cuyo valor de IoU fue mayor o igual a 0.5.

4 Resultados

El modelo YOLOv8 exhibió un rendimiento de detección solido. La imagen de distribucion demues-

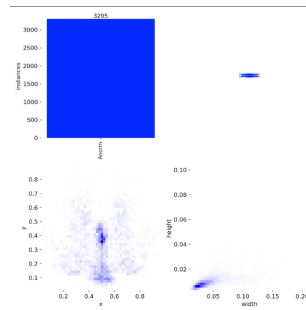


Figure 2: Distribucion de Etiquetas(Graficas de Dispersion).

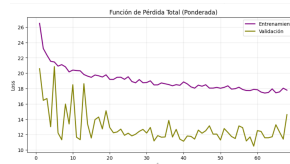


Figure 3: Funcion de Perdida Total (Ponderada).

tra que se extrajeron 3,200 instancias marcadas como "Anormal".

La grafica de "Funcion de Perdida Total" refleja un comportamiento bueno de aprendizaje. Se muestra un descenso continuo y estable, indicando que la red esta aprendiendo a extraer las características. Aunque presenta una leve varianza, se mantiene por debajo de la perdida de entrenamiento.

En la grafica se muestra que termina abruptamente alrededor de la epoca 68. Esto demuestra que nuestro algoritmo de *Early Stopping* funciono de manera exitosa, protegiendo al modelo de *overfitting*.

En la matriz de confusion, se obtiene 147 detecciones correctas, y 68 casos donde la red confundio tejido sano o ruido de fondo con una anomalia.

La red se conserva, cuando predice que hay una lesion. Presenta una buena tasa de acierto; sin embargo suele mostrar "*Falsos Negativos*".Omitio el 621. Tamaño minuscuro de las *bounding box* 2. Contraste escaso entre el tejido sano y una lesion real.

Como resultado final, se presenta el rendimiento final del aprendizaje de la Red Neuronal YOLOv8 como exitosa, detectando correctamente las anoma-

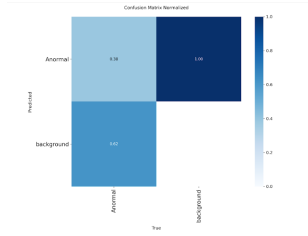


Figure 4: Matriz de Confusión

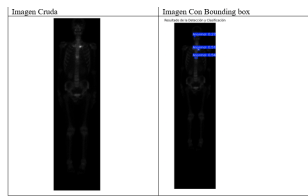


Figure 5: Red Neuronal ya entrenada

lias en las imagenes de test. A continuacion se muestra la tabla de la imagen en crudo limpia y la otra con las bounding box detectadas como anomalias.

SSD + VGG16

La gráfica muestra la evolución de la pérdida (loss) durante el entrenamiento y la validación del modelo a lo largo de las épocas. En las primeras épocas se observa una variación considerable en la pérdida de entrenamiento, alcanzando un valor elevado en la época 2, lo que indica inestabilidad inicial en el proceso de aprendizaje. Sin embargo, conforme avanzan las épocas, tanto la pérdida de entrenamiento como la de validación disminuyen y tienden a estabilizarse alrededor de valores cercanos a 4 (Figura 6).

Además, ambas curvas permanecen relativamente próximas entre sí después de las primeras iteraciones, lo cual sugiere que el modelo logró un aprendizaje estable sin presentar un sobreajuste significativo (overfitting). Esto indica que el modelo es capaz de generalizar adecuadamente sobre los datos de validación.

La imagen presenta una comparación entre las predicciones realizadas por el modelo y las anotaciones reales en una gammagrafía ósea.

En la parte izquierda se muestran las predicciones

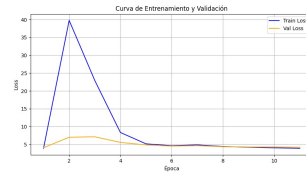


Figure 6: Curva de Entrenamiento y Validación

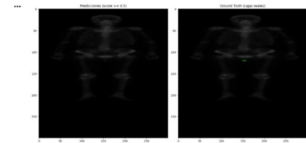


Figure 7: Comparación de predicción

generadas por la red neuronal utilizando un umbral de confianza (score 0.5), donde el modelo logró identificar una región sospechosa dentro de la imagen médica (Figura 7).

En la parte derecha se observa la ubicación real de la lesión marcada manualmente mediante una caja delimitadora verde, correspondiente a la región de interés establecida en el conjunto de datos.

La similitud entre la región detectada por el modelo y la anotación real indica que el sistema fue capaz de localizar adecuadamente la zona sospechosa, demostrando la capacidad del modelo para apoyar en la detección automática de posibles anomalías en gammagrafías óseas.

ResNet50

Los resultados obtenidos a partir de la matriz de confusión muestran que el modelo alcanzó un accuracy de 55.9 por ciento, lo que indica que aproximadamente poco más de la mitad de las imágenes fueron clasificadas correctamente entre las categorías normal y anormal. Este valor sugiere que el desempeño general del modelo aún es limitado para una aplicación clínica robusta, ya que existe una proporción considerable de errores de clasificación.

Por otro lado, el modelo obtuvo un F1-Score de 55.0 por ciento, métrica que evalúa el equilibrio entre la precisión y la sensibilidad del modelo. Este resultado refleja que el sistema presenta dificultades tanto para identificar correctamente los casos anor-

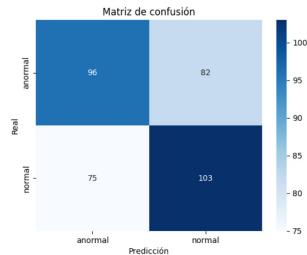


Figure 8: Matriz de confusión

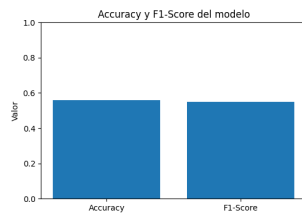


Figure 9: F1- score y precision

males como para evitar clasificaciones erróneas. En aplicaciones médicas, un F1-Score moderado puede indicar que el modelo todavía requiere optimización, especialmente porque los falsos negativos representan un riesgo importante al no detectar posibles patologías presentes en las imágenes.

En conjunto, las métricas obtenidas evidencian que el modelo posee una capacidad inicial de clasificación, pero aún necesita mejoras en entrenamiento y ajuste de parámetros.

5 Conclusion

El desarrollo de las redes neuronales presentados en este artículo demostraron la viabilidad de integrar de forma estructurada el procesamiento de datos médicos y el modelado predictivo para el análisis de gammagrafías óseas. La automatización en la extracción y organización de anotaciones XML a formato estructurado permitió gestionar eficientemente múltiples regiones de interés (ROIs) por imagen, optimizando la etapa de entrenamiento mediante un canal de datos ordenado. Asimismo, la aplicación de técnicas de preprocesamiento (normal-

ización, reescalamiento y filtrado) fue crucial para homogeneizar las dimensiones de entrada y mejorar la calidad de las imágenes médicas, facilitando una extracción de características más robusta.

6 Referencias

- Lin, Q., Li, T., Cao, C., et al. (2021). Deep learning based automated diagnosis of bone metastases with SPECT thoracic bone images. *Scientific Reports*, 11(1), 1–11. <https://doi.org/10.1038/s41598-021-83083-6>
- Wang, Y., Lin, Q., Zhao, S., et al. (2024). Automated Diagnosis of Bone Metastasis by Classifying Bone Scintigrams Using a Self-defined Deep Learning Model. *Current Medical Imaging*, 20(4), 1–10. <https://www.eurekaselect.com/225829/article>
- Ibrahim, A., Vaidyanathan, A., Primakov, S., et al. (2023). Deep learning based identification of bone scintigraphies containing metastatic bone disease foci. *Cancer Imaging*, 23(1), 1–12. <https://doi.org/10.1186/s40644-023-00524-3>
- Liu, Y., Yang, P., Pi, Y., et al. (2021). Automatic identification of suspicious bone metastatic lesions in bone scintigraphy using convolutional neural network. *BMC Medical Imaging*, 21(1), 1–10. <https://doi.org/10.1186/s12880-021-00662-9>
- Simonyan, K., Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv Preprint arXiv:1409.1556*. <https://doi.org/10.48550/arXiv.1409.1556>
- Liu, W., Anguelov, D., Erhan, D., et al. (2016). SSD: Single Shot MultiBox Detector. In *Lecture Notes in Computer Science* (Vol. 9905, pp. 21–37). Springer. <https://doi.org/10.1007/978-3-319-46448-0>
- Kim, D. H., Seo, J., Lee, Ji. H., Jeon, E.-T., Jeong, D., Chae, H. D., Lee, E., Kang, J. H., Choi, Y.-H., Kim, H. J., Chai, J. W. (2024). Automated detection and segmentation of bone metastases on spine MRI using U-Net: A multicenter study. *Korean Journal of Radiology*, 25(4), 363. <https://doi.org/10.3348/kjr.2023.0671>
- Chen H, Dou Q, Yu L, Qin J, Heng PA. VoxResNet: deep voxelwise residual networks for brain

segmentation from 3D MR images. *Neuroimage* 2018;170:446-455

-Papandrianos, N., Papageorgiou, E., Anagnostis, A., Feleki, A. (2020). A deep-learning approach for diagnosis of metastatic breast cancer in bones from whole-body scans. *Applied Sciences*, 10(3), 997. <https://doi.org/10.3390/app10030997>