



Universidad
Zaragoza

Trabajo Fin de Grado

Implementación de modelos de Machine Learning para
el análisis de ecografías de fascias plantares

Implementation of Machine Learning models for the
analysis of plantar fascia ultrasound images

Autor

Ismael Hernández Medrano

Director/es

Mario Morales Hernández

Miguel Malo Urriés

Grado en Ingeniería Mecánica

Universidad de Zaragoza

Escuela de Ingeniería y Arquitectura

2024 - 2025

Implementación de modelos de Machine Learning para el análisis de ecografías de fascias plantares

Resumen

Este proyecto se centra en el desarrollo de un modelo basado en técnicas de Machine Learning y Redes Neuronales para analizar imágenes ecográficas de las fascias plantares, con el objetivo de contribuir a la mejora en el diagnóstico médico. La investigación se divide en tres fases principales: la clasificación de la calidad de las ecografías, la segmentación de las partes más relevantes de las imágenes y la puntuación automática de dichas zonas según su estado de salud.

En la primera fase, se implementó un modelo para clasificar las ecografías en "buenas", "regulares" o "malas", lo que permitió filtrar aquellas imágenes que no eran aptas para un diagnóstico. Posteriormente, se llevó a cabo la segmentación automática de las partes críticas de las ecografías (hueso, fascia y límites de la fascia), empleando una arquitectura U-Net que fue optimizada para este propósito. Finalmente, en la tercera fase, se desarrolló un modelo capaz de puntuar automáticamente el estado de cada parte segmentada, basándose en evaluaciones realizadas previamente por un fisioterapeuta, asignando valores que representan el nivel de salud de las distintas partes de la ecografía.

A través del uso de técnicas avanzadas de inteligencia artificial, este trabajo propone una solución tecnológica que puede automatizar y optimizar procesos médicos, reduciendo la carga de trabajo de los especialistas y aumentando la precisión en los diagnósticos. Este enfoque interdisciplinar, que combina los fundamentos de la ingeniería mecánica y la innovación computacional, pone de manifiesto el papel crucial de la ingeniería en el desarrollo de herramientas que impactan directamente en la salud y el bienestar de las personas.

Índice

1	Introducción	4
1.1	Contexto	4
1.1.1	¿Qué es la fascitis plantar?.....	4
1.1.2	¿Qué es una ecografía y cómo funciona?.....	4
1.1.3	Limitaciones del diagnóstico actual y necesidad de automatización	5
1.2	Objetivo del proyecto	5
1.3	Procesamiento preliminar de datos.....	6
1.3.1	Extracción de imágenes.....	6
1.3.2	Clasificación manual de las imágenes	6
2	Clasificación de las ecografías.....	7
2.1	Modelo de clasificación.....	8
2.1.1	División del conjunto de datos y preprocesado	8
2.1.2	Arquitectura y entrenamiento del modelo.....	9
2.2	Modificación del modelo para clasificación multiclase.....	10
2.3	Análisis de los errores.....	12
2.4	Análisis de la reproducibilidad de las clasificaciones.....	12
3	Segmentación de las ecografías	15
3.1	Preparación del conjunto de datos.....	15
3.2	Tipos de ecografías segmentadas	16
3.3	Generación de las máscaras de segmentación.....	17
3.4	Características del modelo de segmentación.....	17
3.5	Compilación y entrenamiento del modelo	18
3.6	Modelo avanzado en el clúster Hermes	20
4	Puntuación de las zonas segmentadas.....	23
4.1	Modelo de Puntuación	24
4.2	Modificaciones del modelo de puntuación.....	26
4.3	Análisis de los resultados.....	28
5	Conclusiones y trabajo futuro.....	29
6	Bibliografía.....	30

Índice de Figuras

Figura 1 Ejemplo de imagen "buena" (izquierda), "regular" (centro) y "mala" (derecha)	8
Figura 2 Matriz de confusión 2x2	10
Figura 3 Matriz de confusión 3x3	11
Figura 4 Imágenes etiquetadas una vez como "malas" y otra como "buenas"	13
Figura 5 Imágenes etiquetadas una vez como "malas" y otra como "regulares"	14
Figura 6 Imágenes etiquetadas una vez como "regulares" y otra como "buenas"	14
Figura 7 Ecografía de talón segmentada.....	16
Figura 8 Ecografía del arco de la fascia segmentada.....	16
Figura 9 Resultado del primer modelo de segmentación zona de talón	19
Figura 10 Resultado del primer modelo de segmentación zona del arco.....	19
Figura 11 Historia de datos de entrenamiento (Precisión, Recall y Mean IoU) del modelo (Anexo 6) durante todas las etapas.....	21
Figura 12 Resultado del segundo modelo de segmentación zona de talón	21
Figura 13 Resultado del segundo modelo de segmentación zona del arco	21
Figura 14 Dos segmentaciones de una misma ecografía del arco realizadas por la misma persona en distinto tiempo	22
Figura 15 Dos segmentaciones de una misma ecografía del talón realizadas por la misma persona en distinto tiempo	22
Figura 16 Dos segmentaciones de una misma ecografía del talón realizadas por la misma persona en distinto tiempo	23
Figura 17 Historial del entrenamiento con la evolución de la pérdida (izquierda) y el MAE (derecha) a lo largo de todas las épocas del primero modelo de puntuación.	25
Figura 18 Historial del entrenamiento con la evolución de la pérdida (izquierda) y el MAE (derecha) a lo largo de todas las épocas del segundo modelo de puntuación.....	27

1 Introducción

1.1 Contexto

Las fascias plantares son estructuras fibrosas localizadas en la planta del pie que desempeñan un papel fundamental en la biomecánica del cuerpo humano. Estas estructuras actúan como soporte del arco plantar y absorben gran parte del impacto generado durante actividades como la marcha, la carrera o el salto, ayudando a distribuir correctamente las fuerzas en el pie. Sin embargo, diversos factores como la sobrecarga mecánica, el envejecimiento, el sobrepeso o incluso el uso de calzado inadecuado pueden desencadenar patologías en estas estructuras, siendo la fascitis plantar la más común. Esta condición inflamatoria puede causar un dolor crónico significativo, dificultando la movilidad y afectando gravemente la calidad de vida de quienes la padecen.

1.1.1 ¿Qué es la fascitis plantar?

La fascitis plantar [1] es una de las patologías más comunes que afectan al pie, caracterizada por dolor en la base del talón debido a la degeneración del tejido conectivo de la fascia plantar. Este tejido, que se extiende desde el hueso calcáneo en la zona del talón hasta los metatarsos, juega un papel fundamental en el mantenimiento del arco plantar y en la absorción de impactos durante la marcha. Aunque tradicionalmente se consideraba una condición inflamatoria, estudios recientes la describen como una fasciosis, debido a la evidencia de daño crónico y desorganización de las fibras, más que inflamación aguda. Entre los factores de riesgo se incluyen el sobrepeso, la marcha prolongada en superficies duras y alteraciones biomecánicas como el pie plano o cavo.

La detección temprana de estas patologías es crucial para aplicar tratamientos efectivos y prevenir complicaciones. En este sentido, las ecografías son una herramienta diagnóstica fundamental en el ámbito médico, ya que permiten obtener imágenes detalladas de los tejidos blandos, como las fascias, en tiempo real y de manera no invasiva.

1.1.2 ¿Qué es una ecografía y cómo funciona?

La ecografía es una técnica de diagnóstico por imagen que utiliza ondas de sonido de alta frecuencia para visualizar estructuras internas del cuerpo, como músculos, tendones, órganos y vasos sanguíneos. El dispositivo principal que se emplea es el transductor, que emite ondas sonoras y, al mismo tiempo, recoge los ecos reflejados por las distintas estructuras internas. Estas ondas reflejadas se convierten en señales eléctricas que el sistema procesa para generar una imagen en tiempo real.

Las principales ventajas de la ecografía son:

1. No invasiva: No requiere incisiones ni procedimientos agresivos.
2. Segura: Al no utilizar radiación ionizante, como los rayos X, es completamente inofensiva para el paciente.

3. Accesible y económica: Comparada con otras técnicas de imagen, como la resonancia magnética, es más asequible y rápida.
4. Versátil: Permite evaluar en tiempo real diferentes tipos de tejidos y estructuras anatómicas.

En el caso de las fascias plantares, la ecografía es especialmente útil para evaluar su grosor, su integridad estructural y la presencia de signos inflamatorios asociados a patologías como la fascitis plantar. Sin embargo, a pesar de sus múltiples ventajas, el análisis de las imágenes ecográficas presenta una limitación importante: la subjetividad del diagnóstico. La interpretación de las imágenes depende en gran medida de la experiencia y criterio del profesional, lo que puede dar lugar a resultados inconsistentes y diagnósticos poco precisos.

1.1.3 Limitaciones del diagnóstico actual y necesidad de automatización

El diagnóstico ecográfico de las fascias plantares requiere una evaluación visual detallada por parte de un especialista, quien debe identificar manualmente posibles signos de patología, como el aumento del grosor de la fascia o la aparición de irregularidades en su textura. Este proceso manual presenta varias limitaciones:

- Subjetividad: La precisión del diagnóstico puede variar según el nivel de experiencia del profesional que lo realiza y de las circunstancias en las que se encuentre.
- Tiempo y carga de trabajo: Evaluar manualmente un gran volumen de ecografías requiere un tiempo considerable, lo que puede retrasar los diagnósticos y congestionar el flujo de trabajo en centros médicos.
- Calidad variable de las imágenes: No todas las ecografías cumplen con los estándares de calidad necesarios para un diagnóstico preciso, lo que añade un desafío adicional al proceso.

En este contexto, la aplicación de técnicas avanzadas de Machine Learning (ML) y Redes Neuronales al análisis de imágenes médicas ha demostrado ser una solución prometedora [2] [3] [4]. Estas tecnologías permiten automatizar tareas complejas, como la clasificación, la segmentación y el análisis de imágenes, mejorando la objetividad del diagnóstico y acelerando el proceso.

1.2 Objetivo del proyecto

Este proyecto tiene como objetivo el desarrollo de un modelo en Python, basado en Redes Neuronales y Machine Learning, que sea capaz de analizar automáticamente imágenes de ecografías de fascias plantares y discernir entre fascias sanas y patológicas. Este enfoque pretende solucionar las limitaciones actuales mediante la automatización del diagnóstico, lo que contribuirá a:

- Aumentar la precisión de los diagnósticos.
- Reducir el tiempo necesario para el análisis de las ecografías.
- Descongestionar la carga de trabajo de los profesionales médicos.
- Mejorar la accesibilidad al diagnóstico para un mayor número de pacientes.

El presente trabajo se divide en tres fases principales:

1. Clasificación de las imágenes según su calidad: Se determina si una ecografía tiene la calidad necesaria para ser analizada.
2. Segmentación de las distintas partes de la ecografía: Se identifican automáticamente las áreas relevantes de la imagen.
3. Puntuación de las partes segmentadas: Se asigna una puntuación objetiva al estado de las fascias, permitiendo discernir entre fascias sanas y patológicas.

1.3 Procesamiento preliminar de datos

1.3.1 Extracción de imágenes

El material médico con el que se inició este proyecto consistía en 84 vídeos anonimizados de aproximadamente 20 segundos de duración cada uno, correspondientes a ecografías de las fascias plantares de distintos pacientes. Estos vídeos eran grabaciones clínicas reales, cuyo objetivo principal era capturar imágenes en movimiento de las fascias plantares para su posterior análisis.

El primer paso fue extraer frames de cada uno de los vídeos a una tasa de 5 frames por segundo, lo que dio como resultado un total de 8.450 imágenes individuales de las fascias plantares. Este procedimiento fue fundamental para disponer de una gran cantidad de imágenes estáticas que pudieran ser empleadas en las distintas etapas del proyecto, como el entrenamiento, la validación y la prueba del modelo. La extracción de los frames se llevó a cabo mediante un script desarrollado en Python (ver Anexo 1 para la consulta completa del código), el cual permitió automatizar el proceso y generar de forma eficiente el conjunto completo de imágenes.

Sin embargo, debido a que estas imágenes fueron extraídas de vídeos, presentaban una gran variabilidad en la calidad visual. Muchos de los frames obtenidos no eran aptos para el análisis, ya que algunos estaban borrosos, otros mostraban movimiento excesivo, y algunos eran simplemente frames completamente en negro. Esta variabilidad en la calidad hacía necesario realizar una depuración previa de las imágenes, eliminando aquellas que no fueran útiles para el diagnóstico.

1.3.2 Clasificación manual de las imágenes

Con el conjunto de 8.450 imágenes extraídas, el siguiente paso fue realizar una clasificación manual de las mismas por el especialista con el fin de distinguir aquellas que eran útiles para el diagnóstico de las que no lo eran. Este proceso fue llevado a cabo por un fisioterapeuta especializado, quién utilizó una herramienta diseñada específicamente para este proyecto.

La herramienta mostraba de manera aleatoria una de las imágenes al fisioterapeuta, quién debía clasificarla en una de las siguientes categorías:

1. Buena: La imagen era lo suficientemente clara y nítida como para permitir un diagnóstico preciso.

2. Regular: La fascia se distinguía en la imagen, pero no con suficiente claridad como para asegurar un diagnóstico fiable.
3. Mala: La imagen no era apta para el diagnóstico, ya fuera porque estaba borrosa, en negro, o porque no se distinguían detalles anatómicos relevantes.

Este proceso de clasificación manual permitió filtrar las imágenes de baja calidad y preparar un conjunto de datos adecuado para el entrenamiento del modelo. Sin embargo, la clasificación no se limitaba únicamente a determinar la calidad de las imágenes, sino que, cuando una imagen era clasificada como “buena”, el fisioterapeuta debía además realizar una segmentación manual de las siguientes regiones anatómicas clave:

- Hueso → Marcado en color amarillo.
- Fascia → Marcado en color azul.
- Límites de la fascia plantar → Marcado en color magenta.

Finalmente, cada una de estas regiones segmentadas debía ser puntuada en una escala del 0 al 3, donde:

- 0 correspondía a una fascia sana.
- 3 indicaba una fascia con patología grave.

Una vez completada la clasificación y segmentación manual, se generó un archivo de metadatos que contenía información detallada sobre cada imagen, incluyendo:

- El nombre del frame original.
- El nombre de la imagen segmentada.
- La categoría de calidad asignada a la imagen (buena, regular o mala).
- La puntuación asignada a cada una de las regiones segmentadas (hueso, morfología, tejido, borde superior y borde inferior).

Este archivo de metadatos sirvió como base para estructurar el conjunto de datos que posteriormente sería utilizado en el desarrollo del modelo de machine learning.

2 Clasificación de las ecografías

Con el conjunto de imágenes clasificadas y segmentadas, se contaba ya con el material necesario para comenzar a entrenar el modelo. La primera tarea que se planteó fue entrenar un modelo capaz de replicar la clasificación manual realizada por el fisioterapeuta, es decir, diferenciar entre imágenes buenas, regulares y malas de manera automática. Este paso era crucial, ya que permitía depurar de manera automatizada el conjunto de datos, asegurando que sólo las imágenes de calidad adecuada fueran utilizadas en las etapas posteriores del proyecto.

Para esta clasificación, se disponía de un conjunto de 825 imágenes previamente clasificadas por un fisioterapeuta especializado, quién había etiquetado las imágenes en tres categorías como muestra la Figura 1.

- Buena: Imágenes con suficiente nitidez y detalle para permitir un diagnóstico fiable.
- Regular: Imágenes que muestran la fascia plantar, pero no de manera totalmente clara.
- Mala: Imágenes borrosas, movidas o completamente en negro, que no son útiles para el diagnóstico.



Figura 1 Ejemplo de imagen "buena" (izquierda), "regular" (centro) y "mala" (derecha)

El objetivo del modelo era realizar una clasificación binaria en la que las imágenes etiquetadas como “buenas” o “regulares” fueran agrupadas en una sola categoría denominada “buenas”, mientras que las imágenes etiquetadas como “malas” se mantuvieran en su categoría original. Esta decisión se tomó para asegurar que las imágenes útiles (buenas y regulares) se incluyeran en el conjunto de datos destinado a la segmentación.

2.1 Modelo de clasificación

2.1.1 División del conjunto de datos y preprocesado

Primero de todo, para el desarrollo y entrenamiento del modelo, las 825 imágenes se dividieron en tres subconjuntos:

- 70% para entrenamiento: Utilizado para ajustar los parámetros del modelo durante el proceso de aprendizaje.
- 20% para validación: Empleado para evaluar el desempeño del modelo durante el entrenamiento y evitar el sobreajuste.
- 10% para pruebas: Destinado a evaluar la capacidad del modelo una vez entrenado, con imágenes que no se habían utilizado en el proceso de aprendizaje.

El siguiente paso previo a entrenar el modelo fue someter a las imágenes a un proceso de preprocesado con los siguientes pasos:

1. Redimensionamiento:
Todas las imágenes fueron ajustadas a un tamaño uniforme de 128x128 píxeles. Este paso es fundamental para estandarizar el conjunto de datos y reducir la carga computacional del modelo, permitiendo que el entrenamiento sea más eficiente.
2. Normalización:
Los valores de los píxeles de cada imagen se dividieron entre 255, de modo que

quedaran en un rango entre 0 y 1. Este proceso mejora el rendimiento del modelo al facilitar la convergencia durante el entrenamiento.

2.1.2 Arquitectura y entrenamiento del modelo

La arquitectura que siguió la red neuronal del modelo de clasificación fue una red neuronal convolucional (CNN), una arquitectura ampliamente utilizada en problemas de procesamiento de imágenes debido a su capacidad para extraer características relevantes de las imágenes y realizar clasificaciones precisas.

La red neuronal consta de las siguientes capas:

1. Tres bloques de capas convolucionales y de pooling [11]:
 - Las capas convolucionales extraen características espaciales relevantes, como bordes y texturas.
 - Las capas de *pooling* reducen el tamaño de las características extraídas, disminuyendo la complejidad del modelo y evitando el sobreajuste.
2. Capa densa totalmente conectada:
 - Esta capa recibe las características extraídas por las capas convolucionales y realiza la clasificación final.
3. Capa de salida con activación *sigmoid* [12]:
 - La función de activación *sigmoid* se utiliza en la capa de salida para realizar una clasificación binaria, donde el modelo devuelve un valor entre 0 y 1, que se interpreta como la probabilidad de que una imagen sea “buena” o “mala”.

Finalmente, el modelo fue entrenado utilizando las siguientes configuraciones:

- Función de pérdida: Se utilizó la función de pérdida *binary crossentropy* [13], la cual es adecuada para problemas de clasificación binaria.
- Optimizador: El optimizador elegido fue *Adam*, ampliamente conocido por su eficiencia y capacidad de adaptación en problemas de clasificación.
- Técnica de *early stopping*: Para evitar el sobreentrenamiento (*overfitting*), se empleó la técnica de *early stopping*, que detiene automáticamente el entrenamiento cuando la pérdida en el conjunto de validación deja de mejorar durante varias épocas consecutivas. Esta técnica permite restaurar los mejores pesos obtenidos durante el entrenamiento.

Una vez completado el entrenamiento, el modelo (ver Anexo 2 para la consulta del código completo) fue evaluado utilizando el subconjunto de ecografías destinadas a comprobación, es decir, imágenes que no se habían empleado durante las fases de entrenamiento ni de validación. Esta evaluación se llevó a cabo con el fin de comprobar la capacidad del modelo para generalizar a nuevos datos y, para medir su desempeño, se utilizó una matriz de confusión. Esta herramienta es especialmente útil en problemas de clasificación, ya que permite visualizar de manera intuitiva los aciertos y errores del modelo en cada una de las categorías.

La matriz de confusión obtenida en esta primera prueba se observa en la Figura 2.

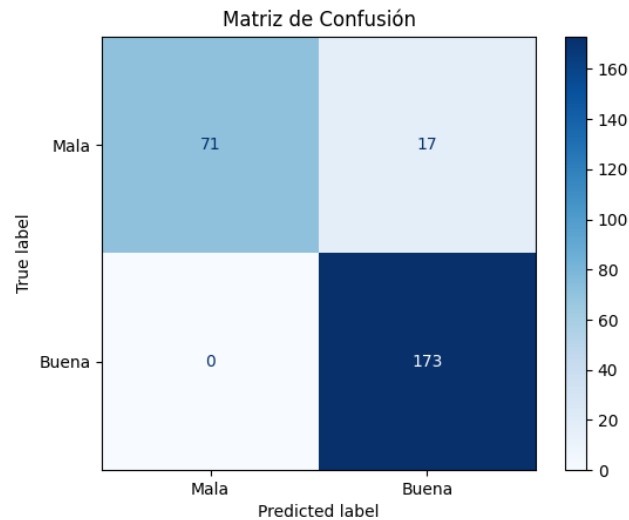


Figura 2 Matriz de confusión 2x2

Los resultados indicaron que el modelo alcanzó un alto grado de precisión al clasificar las imágenes en las categorías de “buenas” y “malas” con un 100% de las imágenes “buenas” clasificadas como tal. No obstante, se observó que 17 imágenes “malas” fueron clasificadas erróneamente como “buenas”. Este error podría explicarse por la similitud visual que, en ciertos casos, existe entre las imágenes etiquetadas como “malas” y aquellas clasificadas como “regulares”, lo que provoca que algunas de estas últimas sean confundidas y categorizadas como “buenas”.

2.2 Modificación del modelo para clasificación multiclase

Dado el comportamiento observado en la primera evaluación, se decidió modificar el modelo inicial para que pudiera diferenciar entre las tres categorías originales: buenas, malas y regulares.

Este cambio permitiría al modelo distinguir con mayor precisión las imágenes que, aunque no fueran ideales, podrían seguir siendo útiles para la etapa de segmentación. Para llevar a cabo esta mejora, se realizaron modificaciones clave en el código del modelo (ver Anexo 3 para la consulta del código completo) con el fin de transformar el problema de una clasificación binaria en una clasificación multiclase.

Las principales diferencias entre el modelo inicial y el modelo modificado fueron las siguientes:

1. Cambio de la salida del modelo:

En el primer modelo, la salida consistía en una única neurona con activación sigmoid, adecuada para clasificación binaria, donde se predice si una imagen es “buena” o “mala”.

Por el contrario, en el modelo modificado, la capa de salida fue sustituida por una capa con tres neuronas y activación softmax [14], lo que permite al modelo asignar una probabilidad a cada una de las tres categorías: buena, regular o mala.

2. Nueva categoría “regular”:

Se añadió la categoría “regular”, que corresponde a las imágenes etiquetadas como “fair” en el archivo de metadatos *.json*. Esta categoría intermedia representa imágenes que, aunque no son óptimas, aún pueden ser útiles para ciertos análisis y diagnósticos.

3. Función de pérdida:

La función de pérdida se cambió de *binary crossentropy* a *categorical crossentropy*, ya que esta última es la adecuada para problemas de clasificación multiclase, permitiendo al modelo aprender a diferenciar entre tres categorías en lugar de dos.

Tras entrenar el nuevo modelo multiclase con las nuevas configuraciones, se evaluó nuevamente su desempeño utilizando el subconjunto de ecografías destinado a comprobación. La nueva matriz de confusión obtenida para este nuevo modelo se observa en la Figura 3:

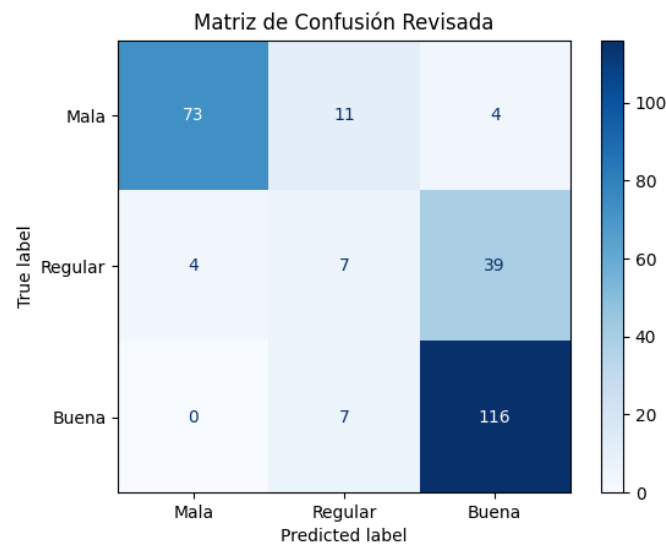


Figura 3 Matriz de confusión 3x3

El análisis de esta matriz de confusión arrojó los siguientes resultados:

- De las 88 imágenes malas, el modelo clasificó correctamente 73 como malas, pero 11 fueron clasificadas como regulares y 4 como buenas.
- De las 50 imágenes regulares, el modelo clasificó 39 como buenas, 7 como regulares y 4 como malas.

- De las 123 imágenes buenas, 116 fueron correctamente clasificadas como buenas, mientras que 7 fueron clasificadas como regulares y 0 como malas.

Estos resultados muestran que el modelo modificado mejoró notablemente su capacidad para distinguir entre imágenes buenas y malas, aunque se evidenció una tendencia a clasificar imágenes regulares como buenas, lo cual, en este contexto, no supone un problema importante, ya que ambas categorías son consideradas válidas para la etapa de segmentación.

2.3 Análisis de los errores

Uno de los principales aspectos que nos generaba preocupación era el hecho de que 4 imágenes malas habían sido clasificadas como buenas. Dado que la diferencia entre estas dos categorías es significativa para el diagnóstico, se decidió realizar un análisis más detallado de estas imágenes. Tras una reunión con el fisioterapeuta encargado de la clasificación manual, se identificó que el software que utilizaba para realizar la clasificación manual había presentado fallos técnicos en algunas ocasiones, lo que le obligaba a etiquetar ciertas imágenes como “malas” para poder continuar con el proceso. Por tanto, estas 4 imágenes no fueron mal clasificadas por el modelo, sino que fueron etiquetadas incorrectamente durante el proceso manual.

Por otro lado, se observó que la mayoría de las imágenes regulares fueron clasificadas como buenas. Este comportamiento no resultó problemático, ya que, para la etapa de segmentación, ambas categorías se consideran útiles. Por tanto, esta tendencia del modelo a agrupar las imágenes regulares con las buenas no afecta al objetivo final del proyecto.

La implementación del modelo multiclase permitió mejorar la precisión del sistema y resolver el problema inicial de confusión entre imágenes malas y buenas. Además, el modelo demostró ser robusto a la hora de identificar imágenes útiles para el diagnóstico, y su capacidad de distinguir entre tres categorías facilita una segmentación más precisa en etapas posteriores.

2.4 Análisis de la reproducibilidad de las clasificaciones

Finalmente, para dar cierre a esta parte, se decidió evaluar el grado de reproducibilidad de las clasificaciones realizadas por el fisioterapeuta especializado. El objetivo de esta evaluación era determinar si el experto, al ser expuesto a las mismas imágenes en diferentes momentos de tiempo, mantenía la consistencia en sus decisiones o si, por el contrario, se producían variaciones en la clasificación que pudieran deberse a factores subjetivos o al sesgo introducido por las imágenes previas analizadas.

Para dicho análisis, se llevó a cabo un análisis de las 1.115 imágenes clasificadas por el experto, identificando aquellas que habían sido presentadas repetidamente durante el proceso de clasificación. De las 1.115 imágenes, se detectó que 68 imágenes eran repetidas, es decir, habían sido evaluadas más de una vez. Al comparar las clasificaciones realizadas en diferentes momentos, se observó que, de estas 68 imágenes, 11 presentaron discrepancias en su clasificación inicial y posterior.

A continuación, se detallan las imágenes en las que se produjeron cambios en la clasificación y las categorías asignadas en cada ocasión:

- Imagen 11_T frame_0052.jpg: Clasificada primero como “regular” y posteriormente como “buena”.
- Imagen 11T_frame_0052.jpg: Clasificada primero como “regular” y posteriormente como “buena”.
- Imagen 49T_frame_0084.jpg: Clasificada primero como “mala” y posteriormente como “buena”.
- Imagen 19T_frame_0088.jpg: Clasificada primero como “regular” y posteriormente como “mala”.
- Imagen 67T_frame_0093.jpg: Clasificada primero como “mala” y posteriormente como “buena”.
- Imagen 3T_frame_0052.jpg: Clasificada primero como “regular”, luego como “buena” y finalmente como “regular”.
- Imagen 7T_frame_0061.jpg: Clasificada primero como “buena” y posteriormente como “regular”.
- Imagen 4T_frame_0002.jpg: Clasificada primero como “regular” y posteriormente como “mala”.
- Imagen 22T_frame_0060.jpg: Clasificada primero como “buena” y posteriormente como “regular”.
- Imagen 64T_frame_0069.jpg: Clasificada primero como “buena” y posteriormente como “regular”.
- Imagen 5T_frame_0059.jpg: Clasificada primero como “buena” y posteriormente como “regular”.

Para interpretar adecuadamente los resultados, se clasificaron las discrepancias observadas en tres grupos, según el tipo de cambio realizado por el experto:

1. Imágenes clasificadas una vez como “buenas” y otra como “malas”: Este grupo corresponde a las imágenes que mostraron la mayor discrepancia en su clasificación, ya que la diferencia entre estas dos categorías es significativa. Al revisar estas imágenes, se pudo constatar que ambas eran realmente “buenas”, lo que indica que la clasificación errónea como “malas” podría haberse debido a los problemas técnicos durante la clasificación manual comentados anteriormente (ver Figura 4).

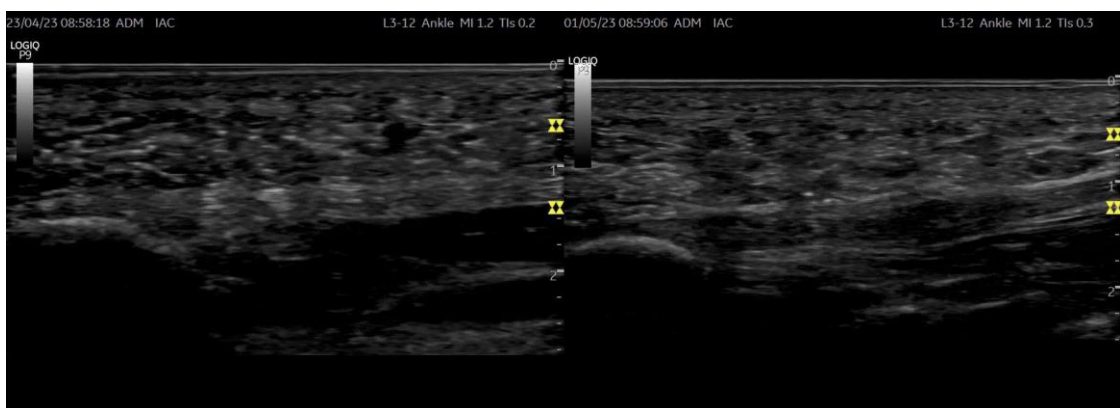


Figura 4 Imágenes etiquetadas una vez como “malas” y otra como “buenas”

2. Imágenes clasificadas una vez como “malas” y otra como “regulares”: Este grupo incluye imágenes en las que la clasificación fue más ambigua. Se observó que, aunque estas imágenes no estaban completamente en negro, presentaban baja calidad visual y, por tanto, resultaba difícil distinguir claramente las distintas partes anatómicas de la ecografía. Esta ambigüedad podría explicar la discrepancia en la clasificación realizada por el experto (ver Figura 5).

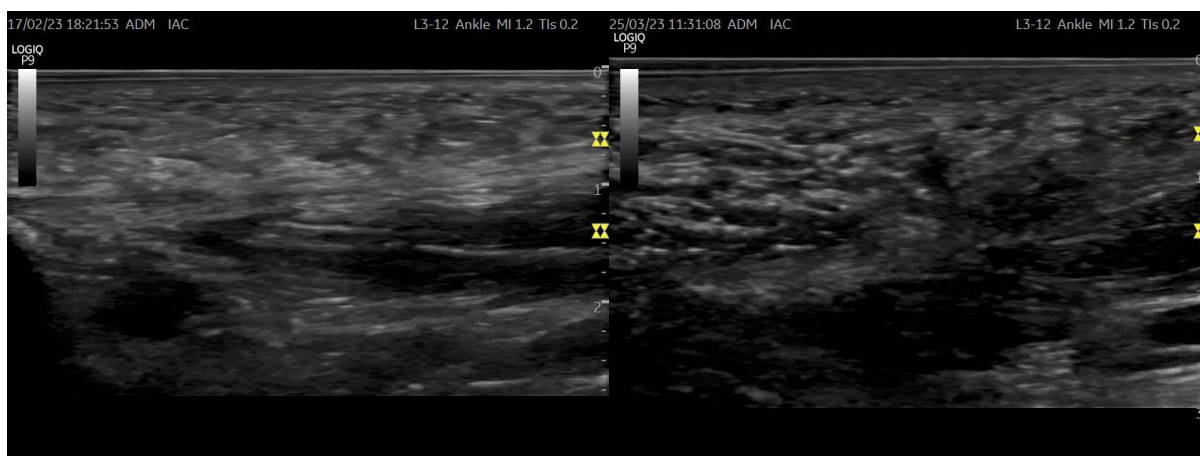


Figura 5 Imágenes etiquetadas una vez como "malas" y otra como "regulares"

3. Imágenes clasificadas una vez como “regulares” y otra como “buenas” 6: En este caso, las imágenes presentaban una calidad intermedia, lo que permitía que pudieran ser catalogadas en ambas categorías de manera razonable. La revisión de estas imágenes mostró que se trataba de ecografías en las que se distinguían claramente las diferentes zonas anatómicas, aunque no con la misma nitidez en todas ellas. Por este motivo, la variación en su clasificación es comprensible (ver Figura 6).



Figura 6 Imágenes etiquetadas una vez como "regulares" y otra como "buenas"

Del análisis de las clasificaciones repetidas se concluye que el modelo desarrollado presenta una alta fiabilidad, especialmente en la tarea crítica de distinguir entre imágenes buenas y malas, donde se logró un 100% de acierto. Además, la tendencia del modelo a clasificar las imágenes regulares como buenas se considera aceptable, ya que, como se ha observado, esta misma tendencia se presentó en las clasificaciones realizadas por el fisioterapeuta, quien modificó su decisión en 1 de cada 6 imágenes repetidas.

Por otro lado, se identificó que ciertos errores en la clasificación manual inicial fueron causados por problemas técnicos del software utilizado por el experto, lo que explica la discrepancia observada en algunas imágenes. Estos errores no se debieron al desempeño del modelo, sino a incoherencias en la etiqueta original de las imágenes.

En general, los resultados obtenidos reflejan que el modelo es robusto y realiza una clasificación consistente y precisa, alineándose con el criterio del experto en la gran mayoría de los casos. Este comportamiento garantiza una adecuada depuración de las imágenes y sienta una base sólida para las siguientes etapas del proyecto, como la segmentación automática de las imágenes.

3 Segmentación de las ecografías

Con las ecografías ya clasificadas en buenas, regulares y malas, se procedió a la siguiente etapa del proyecto: la segmentación automática de las imágenes. La segmentación consiste en identificar y separar las distintas partes de interés presentes en una imagen. Es un tema ampliamente extendido debido a su utilidad no sólo en el ámbito médico sino en cualquier tarea de reconocimiento de imágenes [5] [6] [7] [8].

Lo que nos concierne en este proyecto, es realizar una segmentación médica [9] de acuerdo con los elementos anatómicos que se desean analizar o estudiar para obtener un diagnóstico de la fascia. Al tratarse de imágenes médicas de las fascias plantares, el objetivo era segmentar cuatro regiones específicas: el hueso del talón (en caso de que esté presente en la ecografía), los límites de la fascia plantar (tanto superior como inferior), el fascia y el fondo (correspondiente con las áreas que no pertenecen a ninguna de las clases anteriores).

3.1 Preparación del conjunto de datos

Para el entrenamiento del modelo de segmentación, se disponía de un conjunto de 1.115 imágenes segmentadas manualmente por un fisioterapeuta especializado (ver Figura 7 y 8). Estas imágenes ya contenían las anotaciones correspondientes a las diferentes regiones anatómicas de interés, marcadas con los siguientes colores:

- Hueso: Color amarillo.
- Límites de la fascia plantar: Color magenta.
- Fascia: Color azul.

- Fondo: No marcado, ya que corresponde a las áreas sin segmentar.

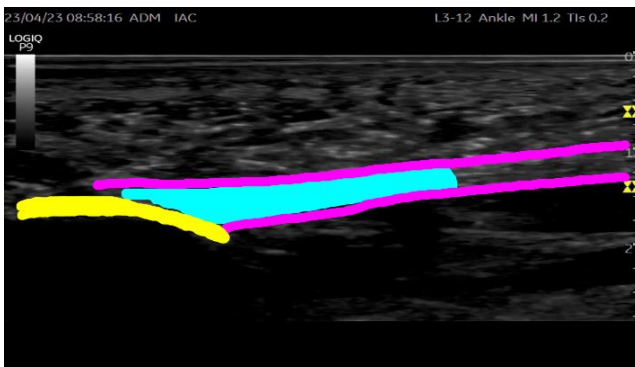


Figura 7 Ecografía de talón segmentada

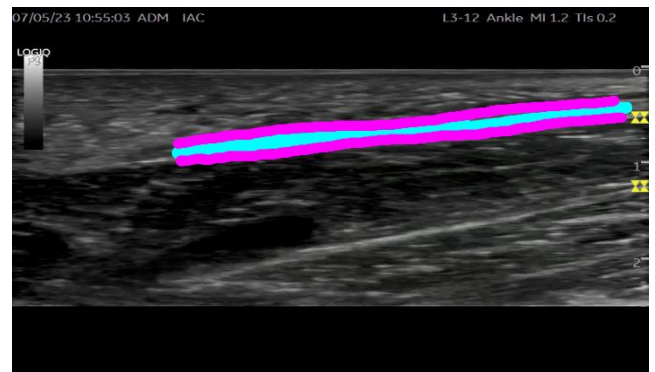


Figura 8 Ecografía del arco de la fascia segmentada

Siguiendo el mismo enfoque utilizado en la etapa de clasificación, el conjunto de imágenes fue dividido en tres subconjuntos:

- 70% para entrenamiento: Utilizado para ajustar los parámetros del modelo durante el aprendizaje.
- 20% para validación: Empleado para evaluar el desempeño del modelo durante el entrenamiento y prevenir el sobreajuste.
- 10% para pruebas: Utilizado para comprobar la capacidad del modelo en nuevas imágenes no vistas durante el entrenamiento ni la validación.

3.2 Tipos de ecografías segmentadas

Durante el proceso de segmentación, se observaron dos tipos principales de ecografías en función de la zona de la fascia plantar capturada en la imagen:

1. Ecografías de la zona cercana al talón (Ver figura 7): En este tipo de ecografías, se observan tres regiones claramente segmentadas:
 - Hueso del talón (amarillo).
 - Límites de la fascia (magenta).
 - Fascia (azul).
2. Ecografías de la zona más cercana a la planta del pie (Ver figura 8): En estas ecografías, las regiones segmentadas son las mismas, pero al tratarse de una zona más alejada del talón, el hueso del talón no está representado en la imagen.

Este detalle es importante, ya que las diferencias en el contenido anatómico de las imágenes influyen en el proceso de segmentación y en la capacidad del modelo para identificar correctamente las distintas regiones.

3.3 Generación de las máscaras de segmentación

El primer paso necesario para entrenar el modelo fue obtener las máscaras de segmentación a partir de las imágenes segmentadas manualmente por el fisioterapeuta. La máscara de segmentación asocia a cada píxel de la imagen una etiqueta correspondiente a una de las cuatro clases: hueso, límite de la fascia, fascia o fondo.

Para generar estas máscaras, se desarrolló un script en Python (ver Anexo 4 para la consulta completa del código) que automatizaba el proceso. Este script separó la segmentación realizada por el fisioterapeuta de la imagen original, generando así las máscaras correspondientes a las 1.115 imágenes disponibles. Las máscaras obtenidas fueron esenciales para entrenar el modelo de segmentación, ya que constituyen la referencia que el modelo debe aprender a replicar durante el proceso de entrenamiento.

Una vez generado el conjunto de datos con las imágenes ecográficas y sus respectivas máscaras de segmentación, se procedió al entrenamiento del modelo (ver Anexo 5 para la consulta completa del código). Para esta tarea, se implementó una red neuronal convolucional de tipo U-Net, ampliamente utilizada en problemas de segmentación de imágenes médicas debido a su capacidad para preservar tanto los detalles finos como el contexto general de la imagen.

3.4 Características del modelo de segmentación

El modelo utilizado sigue la arquitectura U-Net [15], diseñada específicamente para tareas de segmentación y muy utilizada en el ámbito de segmentaciones de imágenes médicas. Esta red se caracteriza por su estructura de encoder-decoder [16], donde:

El modelo cuenta con:

- Tres bloques de capas convolucionales y de pooling en el encoder.
- Una capa bottleneck que conecta el encoder con el decoder.
- Dos bloques de upsampling y concatenación en el decoder.
- Una capa de salida convolucional con función de activación *softmax*, que devuelve una probabilidad para cada uno de los cuatro tipos de regiones a segmentar: hueso, fascia, límites de la fascia y fondo. Similar a lo realizado en la tarea de clasificación pero ahora debía diferenciar entre las cuatro zonas de la ecografía en vez de la calidad de la misma.

Posteriormente, para entrenar el modelo, las imágenes y las máscaras se preprocesaron de manera similar al modelo de clasificación siguiendo los siguientes pasos:

1. Redimensionamiento: Tanto las imágenes como las máscaras se redimensionaron a un tamaño uniforme de 256x256 píxeles, con el fin de estandarizar el conjunto de datos y reducir la carga computacional del modelo.
2. Normalización de las imágenes: Las imágenes se normalizaron dividiendo los valores de los píxeles entre 255, de modo que sus valores quedaran en un rango entre 0 y 1. Este paso facilita el aprendizaje del modelo, ya que mejora la estabilidad numérica durante el entrenamiento.

3. Codificación de las máscaras: Las máscaras de segmentación, que inicialmente contenían valores de color correspondientes a las regiones segmentadas, se convirtieron en matrices de enteros, donde cada valor entero representa una clase (fondo, hueso, fascia o límites de fascia). Finalmente, se aplicó *One-Hot Encoding* para transformar las máscaras en un formato categórico, adecuado para el entrenamiento del modelo.

A la hora de entrenar el modelo, dado que las diferentes clases presentes en las imágenes (fondo, hueso, fascia y límite de la fascia) tienen una distribución desigual (por ejemplo el fondo es mucho más predominante a la fascia), se optó por utilizar una función de pérdida ponderada. Esta técnica asigna un mayor peso a las clases menos representadas, con el fin de que el modelo no tienda a ignorarlas durante el entrenamiento.

La función de pérdida elegida fue *Categorical Crossentropy* con pesos asignados a cada clase según su frecuencia relativa en el conjunto de datos. Los pesos utilizados fueron los siguientes:

- Fondo: 0.2
- Hueso: 22.98
- Fascia: 26.25
- Límite de fascia: 25.42

Estos pesos se ajustaron gracias a una función personalizada que calculó los pesos comparando el área que ocupaban cada una de las clases con respecto al área total en el conjunto de máscaras obtenidas para el entrenamiento del modelo, es decir, si una clase es muy predominante (como el fondo) tiene un peso menor a una clase poco predominante (la fascia).

3.5 Compilación y entrenamiento del modelo

El modelo fue compilado utilizando el optimizador *Adam* con una tasa de aprendizaje (*learning rate*) de $1e-4$, debido a su eficacia en problemas de clasificación y segmentación. La compilación incluyó también las siguientes métricas de evaluación:

- Precisión: Mide la proporción de predicciones correctas realizadas por el modelo.
- Recall [17]: Evalúa la capacidad del modelo para identificar correctamente las regiones positivas de cada clase.
- Mean Intersection over Union (Mean IoU) [18]: Métrica estándar en problemas de segmentación, que mide la superposición entre las regiones predichas por el modelo y las regiones reales.

El entrenamiento se llevó a cabo durante 50 épocas con un tamaño de *batch* de 4 imágenes. Además, se configuraron dos *callbacks* importantes:

1. ModelCheckpoint: Permite guardar el modelo con los mejores pesos en función de la pérdida en el conjunto de validación.

2. EarlyStopping: Detiene el entrenamiento de manera anticipada si la pérdida en el conjunto de validación deja de mejorar durante 10 épocas consecutivas, evitando así el sobreentrenamiento.

Con este primer modelo obtuvimos estos resultados de segmentación (ver Figuras 9 y 10):

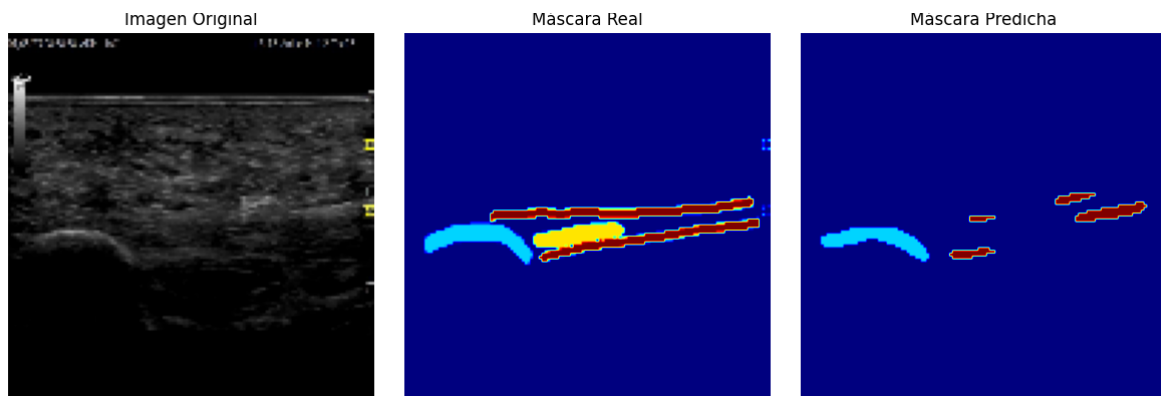


Figura 9 Resultado del primer modelo de segmentación zona de talón

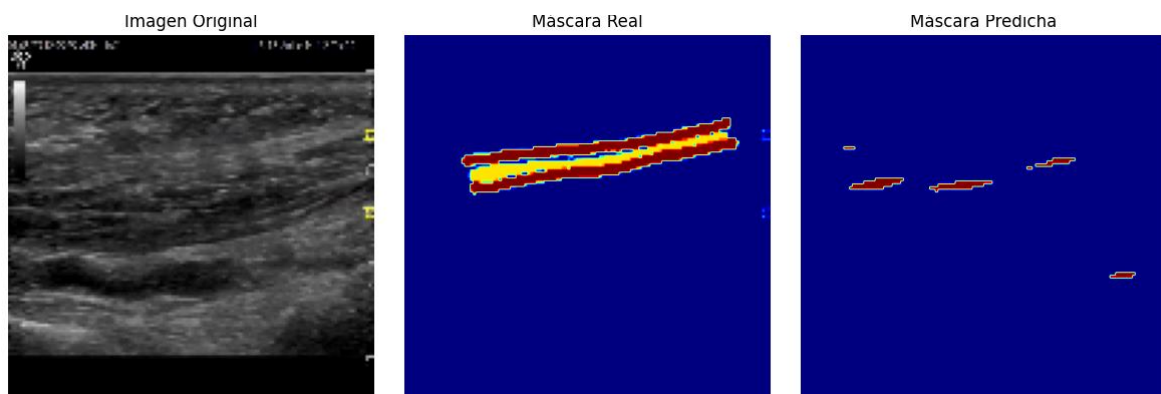


Figura 10 Resultado del primer modelo de segmentación zona del arco

Como observamos, son resultados bastante pobres. Resultados que no mejoraron aún realizando diversas pruebas cambiando valores como el *learning rate*, la cantidad de filtros de cada capa, el tamaño del redimensionamiento de las imágenes e incluso la propia arquitectura de la red. Dado que estos cambios no surgieron efecto y que la GPU con la que estaba corriendo el modelo era limitada, ya que era la de mi ordenador personal, tomamos la decisión de conectarnos remotamente al Clúster Hermes del I3A de la Universidad de Zaragoza para tener acceso a GPUs (tarjetas gráficas) con más memoria y así poder entrenar un modelo más potente y en menos tiempo, lo que nos daría la oportunidad de realizar una mayor cantidad de pruebas para optimizar nuestro modelo de segmentación.

3.6 Modelo avanzado en el clúster Hermes

Una vez conectados al Clúster Hermes, se procedió a implementar una serie de mejoras en el modelo inicial, con el objetivo de optimizar el proceso de segmentación y obtener mejores resultados. Estas mejoras fueron posibles gracias a la mayor capacidad de procesamiento y memoria proporcionada por las GPUs del clúster, lo que permitió entrenar modelos más complejos y realizar un mayor número de pruebas en un tiempo razonable. Éstos fueron los cambios que mejores resultados arrojaron y los que fueron finalmente implementados en el modelo:

1. Reducción de la resolución de entrada: En el modelo inicial, las imágenes eran redimensionadas a 256x256 píxeles, lo que requería un elevado uso de memoria durante el entrenamiento. En el modelo mejorado, se optó por redimensionar las imágenes a 128x128 píxeles, reduciendo significativamente la carga computacional.
2. Ajuste del número de filtros por capa: Dado que el modelo inicial no lograba un desempeño adecuado, se decidió reducir el número de filtros en cada capa convolucional y aumentar el número de bloques. Mientras que en el primer modelo se utilizaban 64, 128 y 256 filtros, en el modelo mejorado se emplearon 16, 32, 64, 128 y 256 filtros, distribuidos en cinco bloques del *encoder* y del *decoder*.
3. Ajuste del tamaño del *batch*: Gracias al acceso a GPUs con mayor capacidad, se pudo aumentar el tamaño del *batch* de 4 imágenes a 20 imágenes. Esto permitió estabilizar el entrenamiento y obtener mejores resultados en menos épocas.
4. Función de pérdida ponderada ajustada manualmente: Si bien en el primer modelo se utilizaba una función de pérdida ponderada, en el segundo modelo se ajustaron manualmente los pesos de cada clase, aumentando los valores manualmente en aquellas clases que le resultaba más complicado realizar la segmentación. Este cambio permitió al modelo prestar mayor atención a las clases menos representadas, como los límites de la fascia o la propia fascia, mejorando su capacidad para segmentar estas regiones críticas.
5. Mayor número de bloques en el *encoder* y el *decoder*: En el modelo mejorado se añadió un bloque adicional tanto en el *encoder* como en el *decoder*, incrementando la capacidad del modelo para extraer y reconstruir características a diferentes escalas. Esta ampliación de la red fue posible gracias al uso de GPUs con más memoria, que permitieron entrenar un modelo más profundo sin comprometer el tiempo de entrenamiento.

Gracias a estas mejoras, el modelo mejorado (ver Anexo 6 para la consulta del código completo) no sólo fue más eficiente en términos computacionales, sino que también logró una mayor precisión en la segmentación de las ecografías. Además, el uso de GPUs más potentes y el ajustar el modelo para que fuese más liviano computacionalmente nos permitía realizar

un mayor número de pruebas para ajustar parámetros. Los resultados (ver Figuras 11, 12 y 13) que obtuvimos con las mejoras presentadas anteriormente fueron los que se muestran:

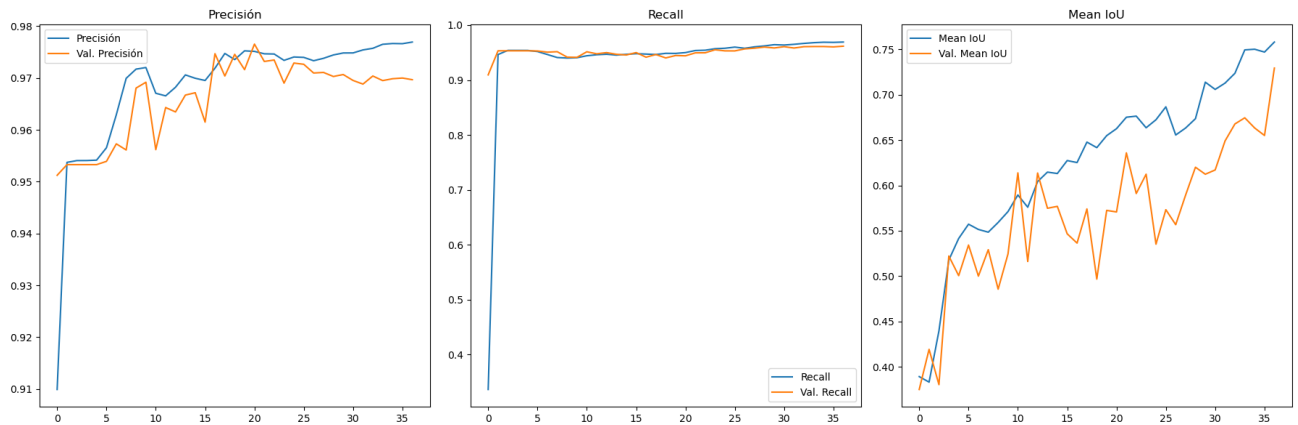


Figura 11 Historia de datos de entrenamiento (Precisión, Recall y Mean IoU) del modelo (Anexo 6) durante todas las etapas

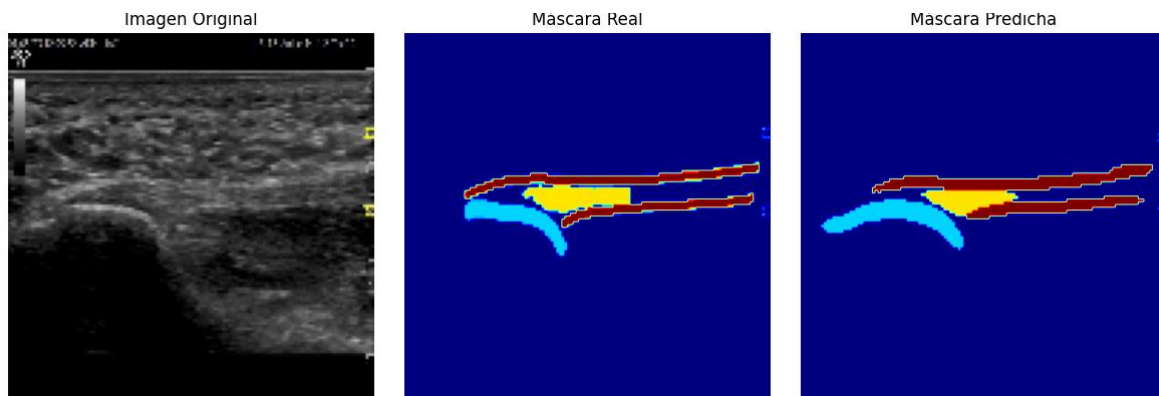


Figura 12 Resultado del segundo modelo de segmentación zona de talón

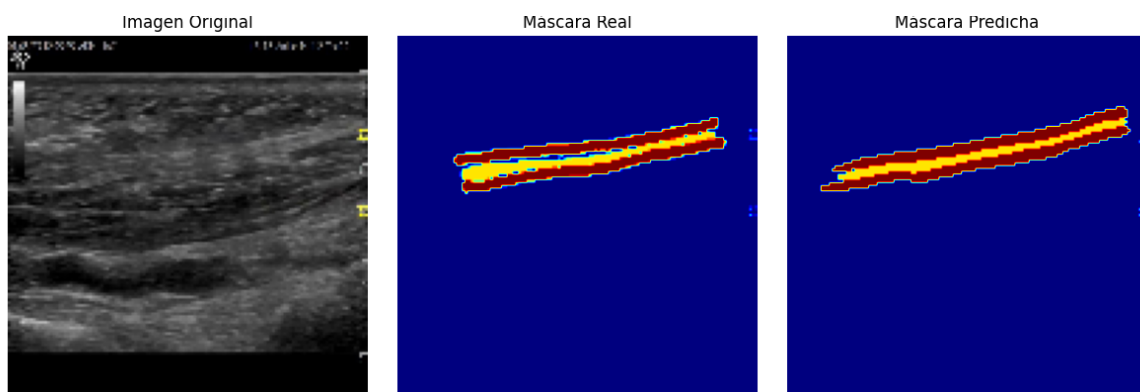


Figura 13 Resultado del segundo modelo de segmentación zona del arco

Como observamos, el modelo muestra un buen desempeño general con métricas altas de precisión, *recall* y *Mean IoU* tanto en entrenamiento como en validación. Además, la estabilidad y convergencia de las métricas sugieren que el modelo es capaz de aprender patrones relevantes para la tarea de segmentación. Por otro lado se observa que el valor del *Mean IoU* fluctúa bastante y continúa aumentando, tendencia que no se mantiene

aumentando las épocas de entrenamiento por lo que fue este modelo el que mejores resultados se obtuvieron.

Cabe destacar que a la hora de realizar valoraciones observando las imágenes y comparando la “máscara real” con la “máscara predicha”, hay que tener en cuenta que no se puede esperar un resultado perfecto. Esto se debe a que, principalmente, si solicitaras a dos fisioterapeutas distintos la segmentación de una misma ecografía o incluso al mismo fisioterapeuta en momentos diferentes, los resultados van a diferir en mayor o menor medida, pero nunca serían idénticos. Por todo esto, esperar que el modelo replique a la perfección la máscara dibujada por el fisioterapeuta es una expectativa poco realista.

Para ilustrar este hecho, se han analizado las ecografías analizadas por el fisioterapeuta y hemos comparado las máscaras dibujadas sobre ellas para corroborar este comportamiento (ver Figuras 14, 15 y 16):

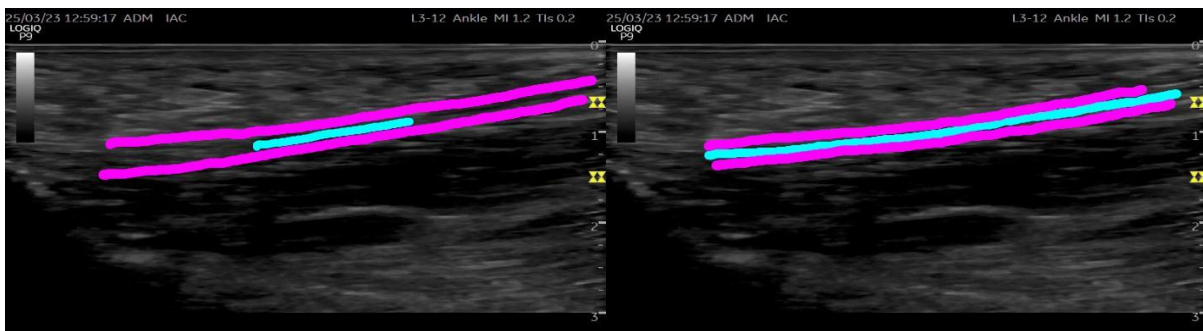


Figura 14 Dos segmentaciones de una misma ecografía del arco realizadas por la misma persona en distinto tiempo

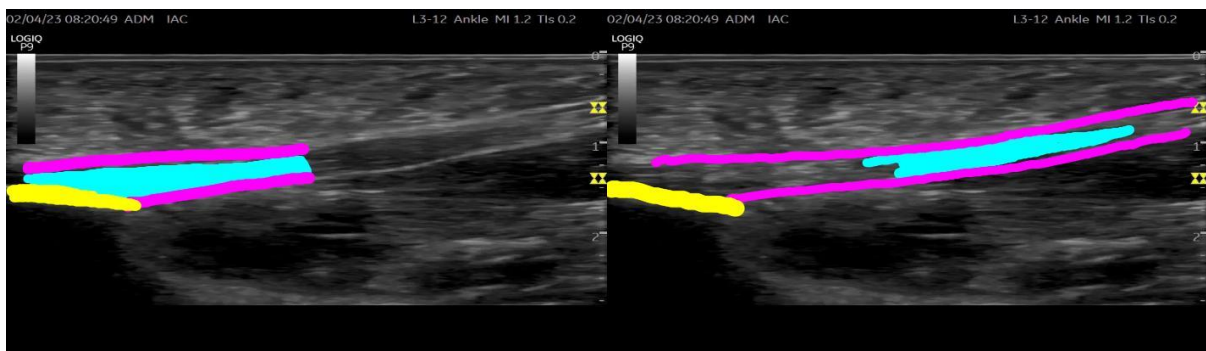


Figura 15 Dos segmentaciones de una misma ecografía del talón realizadas por la misma persona en distinto tiempo

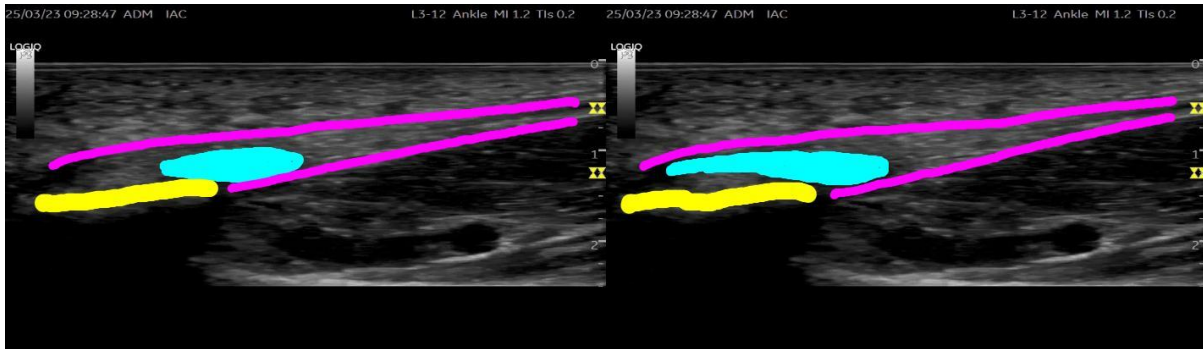


Figura 16 Dos segmentaciones de una misma ecografía del talón realizadas por la misma persona en distinto tiempo

Como se observa, cada par de imágenes corresponde con exactamente la misma ecografía, pero la segmentación realizada por el fisioterapeuta difiere bastante entre una y otra por lo que podemos concluir que la segmentación de nuestro modelo es lo suficientemente buena como para realizar una aproximación bastante fiel a como lo haría un experto especializado.

4 Puntuación de las zonas segmentadas

Una vez que las distintas zonas críticas de las ecografías han sido segmentadas, el siguiente paso del proyecto consiste en desarrollar un modelo capaz de puntuar automáticamente el nivel de salud de cada una de estas zonas segmentadas. El objetivo de este proceso es determinar cuán sana o patológica es cada región de interés, basándose en las evaluaciones previas realizadas por un fisioterapeuta especializado.

Para este propósito, se disponía de los datos anotados manualmente en la etapa inicial del proyecto. De todas las imágenes evaluadas, 539 ecografías fueron clasificadas como “buenas” y por tanto puntuadas en cinco aspectos clave:

1. Tejido.
2. Morfología.
3. Borde superior.
4. Borde inferior.
5. Hueso.

La puntuación asignada a cada uno de estos aspectos oscila entre 0 y 3, donde 0 indica un estado completamente sano y 3 un estado altamente patológico. Partiendo de estas evaluaciones manuales, se diseñó y entrenó un modelo basado en redes neuronales que, a partir de una imagen ecográfica y su correspondiente máscara de segmentación, pudiera predecir automáticamente estas cinco puntuaciones.

Partiendo de todas estas evaluaciones pudimos empezar a construir el modelo de puntuación que se estructuró de la siguiente manera:

4.1 Modelo de Puntuación

La última parte del proyecto consistía en que las partes segmentadas anteriormente fueran evaluadas por el modelo siguiendo la tendencia con la que un experto lo haría.

Para ello, la primera tarea del modelo consistió en cargar las imágenes ecográficas, las máscaras de segmentación y las puntuaciones manuales realizadas por el fisioterapeuta. Esto se logra mediante la función `load_data`, que:

- Recorre el directorio de imágenes y localiza sus respectivas máscaras de segmentación.
- Redimensiona las imágenes y máscaras a un tamaño uniforme de 256x256 píxeles.
- Normaliza los valores de las imágenes dividiendo los píxeles entre 255 para que sus valores se encuentren en el rango [0, 1].
- Extrae las puntuaciones manuales desde el archivo JSON correspondiente. Estas puntuaciones se asocian a cada imagen en forma de una lista de cinco valores, cada uno correspondiente a uno de los aspectos evaluados: tejido, morfología, borde superior, borde inferior y hueso.

Finalmente, la función devuelve tres conjuntos de datos: las imágenes normalizadas, las máscaras de segmentación normalizadas de dichas imágenes y las puntuaciones asociadas a cada imagen.

Una vez cargados todos estos datos, se realiza una división aleatoria del conjunto en:

- 80% para entrenamiento.
- 20% para validación.

Esta división permite evaluar el desempeño del modelo durante el entrenamiento y evitar el sobreajuste. He de aclarar que no se guardan imágenes para comprobación ya que se separaron ya anteriormente en otra carpeta distinta.

La arquitectura utilizada en el modelo para la puntuación automática se compone de una red neuronal convolucional que toma como entrada:

- La imagen ecográfica redimensionada.
- La máscara de segmentación correspondiente.

Ambas entradas se concatenan en una única representación, lo que permite al modelo combinar la información visual de la imagen y su segmentación para realizar la predicción. La arquitectura del modelo es la siguiente:

- Entrada de imagen: Tensor de tamaño (256, 256, 3).
- Entrada de máscara: Tensor de tamaño (256, 256, 1).
- Concatenación de ambas entradas.
- Tres bloques convolucionales, cada uno compuesto por:
 - Una capa *Conv2D* con 32, 64 y 128 filtros respectivamente, activación *ReLU* y *padding 'same'*.
 - Una capa *MaxPooling2D* para reducir la resolución y extraer características relevantes.

- Capa de *Flatten* para aplanar las características extraídas.
- Capa densa de salida con 5 neuronas (una para cada puntuación), con activación lineal.

La activación lineal permite al modelo generar valores continuos que representan las puntuaciones estimadas para cada aspecto evaluado.

El modelo se compila utilizando el optimizador *Adam*, una función de pérdida *MSE (Mean Squared Error)* y la métrica de evaluación *MAE (Mean Absolute Error)* [19], que mide el error absoluto promedio entre las puntuaciones reales y las predichas.

Con el código ya completado, el modelo se entrenó durante 50 épocas con un tamaño de *batch* de 16 imágenes. Durante el entrenamiento, el modelo optimiza sus parámetros para minimizar el error cuadrático medio entre las puntuaciones reales y las predichas.

Además, se utiliza el conjunto de validación para monitorear el desempeño del modelo y evitar el sobreentrenamiento. El resultado final es un modelo capaz de predecir automáticamente las puntuaciones de las diferentes zonas segmentadas de las ecografías.

Una vez completado el entrenamiento, el modelo entrenado se guarda, lo que permite reutilizarlo posteriormente para realizar predicciones sin necesidad de volver a entrenarlo.

Con el modelo (ver Anexo 7 para la comprobación del código completo) ya entrenado y guardado se realizó una prueba para comprobar su eficiencia a la hora de puntuar las distintas zonas segmentadas. Para poder cuantificar el error cometido por el modelo, se calculó el error medio de las distintas zonas de la ecografía sumando el valor absoluto de la diferencia entre la puntuación real y la puntuación predicha por el modelo y dividiendo ese valor por el número de ecografías. Estos fueron los resultados que obtuvimos con las 164 imágenes de comprobación (ver Figura 16):

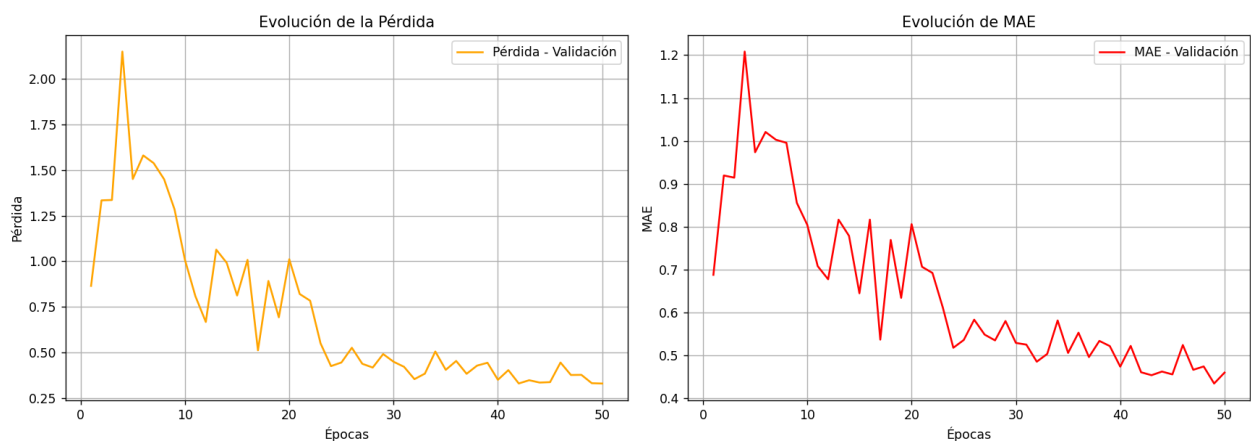


Figura 17 Historial del entrenamiento con la evolución de la pérdida (izquierda) y el MAE (derecha) a lo largo de todas las épocas del primero modelo de puntuación.

- En la zona de tejido el error medio fue de 0,56.
- La morfología tuvo un error medio de 0,45.
- En la zona de borde superior el error medio fue de 0,47.
- En la zona de borde inferior el error medio fue de 0,48.

- En la zona de hueso el error medio fue de 0,34.

Con todos estos errores medios realizamos de nuevo una media para obtener el valor de error medio del modelo, el cuál fue de 0,46.

Tras este análisis de los resultados de este primer modelo de puntuación, concluimos que era una muy buena primera aproximación, pero realizamos algunas modificaciones y ajustes en el modelo para que llevase a cabo una mejor puntuación de las ecografías.

4.2 Modificaciones del modelo de puntuación

Tras varias modificaciones y pruebas del modelo inicial, pudimos obtener un modelo que realizaba una puntuación más fiel a la del fisioterapeuta que el presentado anteriormente.

Éstas fueron las modificaciones que obtuvieron los mejores resultados:

En el primer modelo, se utilizó una resolución de entrada de 256x256 píxeles, lo que implicaba un mayor uso de memoria y un tiempo de entrenamiento más prolongado. Viendo que en el modelo de segmentación obtuvimos mejores resultados disminuyendo el costo computacional, en el segundo modelo se optó por reducir la resolución de entrada a 128x128 píxeles, manteniendo un buen equilibrio entre el detalle capturado y la carga computacional. Esta reducción del costo computacional permitió además aumentar el tamaño del *batch* del modelo lo que mejoró la estabilidad del entrenamiento y redujo el tiempo necesario para completar cada época.

Además, el primer modelo contaba con tres bloques convolucionales, lo que limitaba la capacidad del modelo para extraer características complejas de las imágenes ecográficas. En el segundo modelo, se añadió un bloque convolucional adicional, resultando en una red con cuatro bloques convolucionales, cada uno con un mayor número de filtros. Así, el primer modelo contaba con 32, 64 y 128 filtros en tres bloques mientras que el segundo modelo contaba con un bloque extra con 256 filtros. Este incremento en la profundidad permitió al modelo capturar de manera más efectiva las diferencias sutiles presentes en las distintas regiones anatómicas.

Por otra parte, una de las limitaciones del primer modelo era la ausencia de *Batch Normalization* [20], una técnica que estabiliza y acelera el proceso de entrenamiento al normalizar la salida de cada capa convolucional. Por tanto, en el segundo modelo, se añadió *Batch Normalization* después de cada capa convolucional, lo que permitió reducir la variación interna de las activaciones y mejorar la velocidad de convergencia del modelo.

Otra diferencia del segundo modelo fue que, si bien en el primer modelo se utilizó *Dropout* para evitar el sobreentrenamiento, en el segundo modelo se aplicó una estrategia de regularización más intensiva, aumentando la frecuencia de *Dropout* al final de cada bloque convolucional con una tasa del 25% lo que redujo el riesgo de sobreajuste, especialmente debido al incremento en la profundidad del modelo.

Finalmente, en el primer modelo se utilizó una tasa de aprendizaje fija durante todo el proceso de entrenamiento. Sin embargo, en el segundo modelo se incorporó el *callback ReduceLROnPlateau*, que reduce automáticamente la tasa de aprendizaje a la mitad si la pérdida en validación no mejora durante 5 épocas consecutivas lo que permitió que el modelo realizara ajustes más finos en las etapas finales del entrenamiento, logrando así una mejor precisión en la predicción de las puntuaciones.

Con todas estas mejoras, obtuvimos los siguientes resultados (ver Figura 17 y Tabla 1) para este nuevo modelo (ver Anexo 8 para la consulta completa del código) con las mismas 164 imágenes de comprobación que usamos para el modelo anterior:

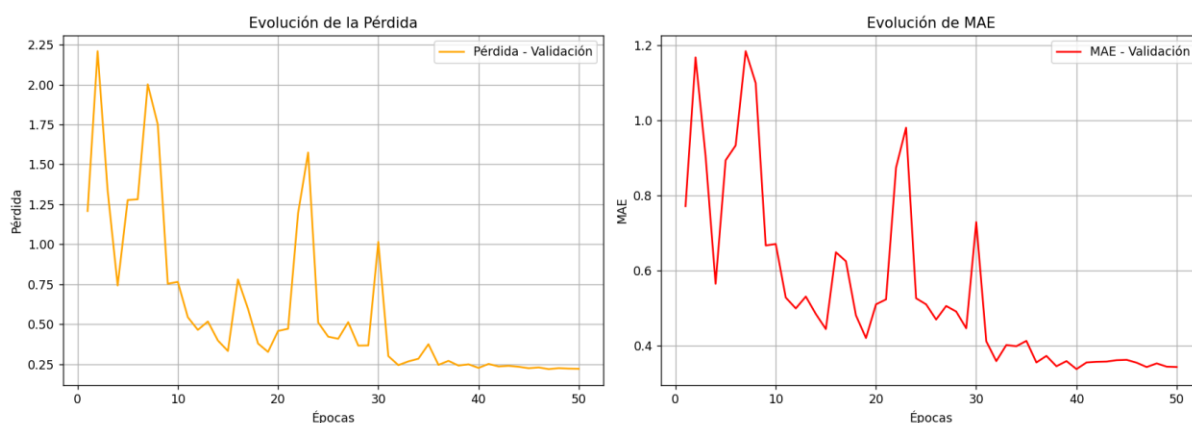


Figura 18 Historial del entrenamiento con la evolución de la pérdida (izquierda) y el MAE (derecha) a lo largo de todas las épocas del segundo modelo de puntuación.

- En la zona de tejido el error medio fue de 0,53.
- La morfología tuvo un error medio de 0,3.
- En la zona de límite superior el error medio fue de 0,3.
- En la zona de límite inferior el error medio fue de 0,36.
- En la zona de hueso el error medio fue de 0,27.

Todos estos resultados nos dieron como resultado que el error medio en este nuevo modelo había disminuido hasta 0,35.

	Modelo 1	Modelo 2
Característica	Error medio	Error medio
Tejido	0,56	0,53
Morfología	0,45	0,3
Límite Superior	0,47	0,3
Límite Inferior	0,48	0,36
Hueso	0,34	0,27
Error medio total	0,46	0,35

Tabla 1: Tabla comparativa con los resultados del error medio de ambos modelos

4.3 Análisis de los resultados

Observando y analizando los resultados durante las distintas pruebas para mejorar el modelo de puntuación se realizó la siguiente observación:

De las 1115 ecografías que habían sido analizadas por el fisioterapeuta, 539 habían sido clasificadas como “buenas” y por tanto posteriormente segmentadas y puntuadas. En cada una de las 539 imágenes se puntúan 5 valores diferentes (tejido, morfología, borde superior, borde inferior y hueso) lo que da lugar a 2695 puntuaciones totales entre todas las imágenes y zonas. De las 2695 puntuaciones entre 0 y 3, solamente 19 puntuaciones fueron valoradas con un 3.

Este hecho da lugar a que la cantidad de ecografías de entrenamiento que son muy patológicas es muy baja como para que el modelo pueda llegar a aprender a puntuar de una manera correcta las zonas que deberían tener mayor puntuación.

A continuación, se muestran algunos resultados de algunas pruebas que se llevaron a cabo para observar el distinto comportamiento del modelo entre ecografías sanas y patológicas.

Si llevamos a cabo la misma comprobación realizada anteriormente, pero en vez de usando todas las imágenes de comprobación (168) la realizamos solamente con las 18 imágenes más patológicas (las que tienen al menos un valor de 3 en alguna de las zonas), los errores medios aumentaban sensiblemente de la siguiente manera:

- En la zona de tejido el error medio aumentó hasta 0,82.
- La morfología tuvo un aumento del error medio hasta 0,53.
- En la zona de límite superior el error medio aumentó hasta 0,53.
- En la zona de límite inferior el error medio aumentó hasta 1,05.
- En la zona de hueso el error medio aumentó hasta 0,47.

Finalmente, el error medio absoluto prácticamente se vio duplicado llegando a un valor de 0,68.

También se quiso comprobar si el modelo realizaba correctamente unas puntuaciones más elevadas hacia las ecografías más patológicas, por lo que se compararon las puntuaciones medias que obtuvieron cada una de las clases en la primera comprobación con las 168 imágenes de entrenamiento con las que se habían obtenido con la comprobación solamente con las 18 imágenes más patológicas. La tabla 2 refleja estos resultados.

Modelo					
	Tejido	Morfología	Borde Superior	Borde Inferior	Hueso
Conjunto de 168 imágenes	0,49	0,2	0,3	0,4	0,28
Conjunto de las 18 más patológicas	1,47	1,29	1	0,94	1,18

Fisioterapeuta					
	Tejido	Morfología	Borde Superior	Borde Inferior	Hueso
Conjunto de 168 imágenes	0,5	0,4	0,28	0,3	0,24
Conjunto de las 18 más patológicas	2,29	1,82	1,29	1,76	1,53

Tabla 2: Tabla de comparación entre las distintas puntuaciones del modelo y del fisioterapeuta según la gravedad patológica

Se puede observar que el modelo propone una puntuación más elevada para las zonas de las fascias que son más patológicas, lo que es correcto. El problema es que al haber un número tan bajo de imágenes patológicas con las que el modelo ha sido entrenado, el modelo no es capaz de dar puntuaciones tan elevadas como el evaluador en estos casos.

Como conclusión podemos decir que el modelo obtiene resultados buenos a la hora de puntuar ecografías de fascias mayoritariamente sanas, pero sufre a la hora de dar valores altos de puntuación en los casos de alta patología. Sin embargo, la solución a este problema no radica expresamente en el modelo en sí, sino en el conjunto de datos de entrenamiento que se le ha suministrado ya que, observamos que la tendencia del modelo es puntuar de una manera más elevada las fascias patológicas frente a las sanas. Ésto corroboraría que si alimentamos el modelo con más casos de fascias patológicas en la fase del entrenamiento, podría llegar a puntuar este tipo de ecografías de una forma tan fiable como lo hace con las fascias más sanas.

5 Conclusiones y trabajo futuro

El presente proyecto ha demostrado como las técnicas de Machine Learning y Redes Neuronales pueden contribuir significativamente al análisis y diagnóstico de patologías en las fascias plantares, automatizando procesos que tradicionalmente dependen del criterio subjetivo del especialista. Las tres fases del trabajo (clasificación de la calidad de las ecografías, segmentación de las zonas críticas y puntuación automática) han permitido desarrollar un sistema robusto y eficiente, capaz de asistir a los profesionales de la salud en la toma de decisiones clínicas.

En cuanto a los resultados obtenidos, los modelos desarrollados han mostrado un desempeño satisfactorio en cada fase. Destaca la capacidad del modelo de segmentación para identificar con alta fidelidad las partes más relevantes de las ecografías y la precisión del modelo de clasificación para determinar si una ecografía es apta o no para un diagnóstico. Por otro lado, el modelo destinado a la puntuación automática de las zonas segmentadas demostró su potencial, aunque se identificó la necesidad de contar con un mayor número de imágenes

patológicas para mejorar su rendimiento. Además, las limitaciones técnicas de los modelos desarrollados también incluyen la dependencia de la calidad de las imágenes ecográficas utilizadas, lo que podría influir en su desempeño en situaciones reales donde las imágenes pueden ser generadas por diferentes equipos de ecografía o con configuraciones no estandarizadas. Este aspecto plantea posibles problemas de generalización del modelo a nuevos entornos clínicos.

Este proyecto ha supuesto un desafío significativo al explorar un ámbito innovador y en constante evolución, como lo son las Redes Neuronales y el Machine Learning, especialmente en un contexto donde no se ofrece formación específica durante el grado. El desarrollo de este trabajo ha requerido una intensa autoformación, tanto en los fundamentos teóricos como en la implementación práctica de estas técnicas, así como la adaptación a metodologías de trabajo propias de proyectos interdisciplinarios. Adicionalmente, la creación y optimización de scripts complejos en Python (tanto para los modelos principales como para los auxiliares) desde cero han representado un reto significativo que ha permitido un aprendizaje profundo durante todo el proyecto.

De cara al trabajo futuro, sería crucial abordar las limitaciones identificadas, particularmente la obtención de un mayor número de datos patológicos para mejorar la capacidad del modelo de puntuación. Esto podría lograrse mediante colaboraciones con hospitales, el acceso a bases de datos públicas o la creación de un protocolo conjunto con profesionales del ámbito médico para capturar imágenes representativas de casos reales. Además, sería interesante explorar la posibilidad de extender el modelo a otras patologías musculoesqueléticas relacionadas con las fascias plantares o incluso a otras regiones anatómicas, ampliando así su aplicabilidad clínica. Por último, una línea de investigación prometedora sería la integración de los modelos en sistemas de diagnóstico asistido en tiempo real, lo que podría acelerar la toma de decisiones médicas y aumentar el impacto positivo de este proyecto en la práctica clínica.

En compromiso con la ciencia abierta, los códigos desarrollados en este trabajo han sido depositados en un repositorio público de GitHub, fomentando la transparencia y la reproducibilidad científica e incentivando al posible trabajo futuro sobre temas similares a los tratados en este proyecto. El enlace de dicho repositorio se muestra a continuación.

<https://github.com/xlismaa/TFG.git>

6 Bibliografía

[1] Muhammad Mannan, Faisal Karim, Usman Hafeez, Sarmad Khalil, Efficacy of Platelet-Rich Plasma (PRP) in Treating Plantar Fasciitis, Cureus, 10.7759/cureus.72454, (2024).

[2] Song K, Feng J and Chen D (2024), A survey on deep learning in medical ultrasound imaging. Front. Phys. 12:1398393. doi: 10.3389/fphy.2024.1398393

[3] Department of Ultrasound, Xijing Hospital, Fourth Military Medical University, Xi'an 710032, China

Department of Ultrasound, Yuncheng Central Hospital, Shanxi Medical University, Yuncheng 044000, China

State Key Laboratory for Manufacturing Systems Engineering, Mechanics Institute, Xi'an Jiaotong University, Xi'an 710049, China

[4] Micucci, M.; Iula, A. Recent Advances in Machine Learning Applied to Ultrasound Imaging. *Electronics* 2022, 11, 1800. <https://doi.org/10.3390/electronics11111800>

[5] Nikhila Ravi, Valentin Gabeur, Yuan-Ting Hu, Ronghang Hu, Chaitanya Ryali, Tengyu Ma, Haitham Khedr, Roman Rädle, Chloe Rolland, Laura Gustafson, et al. Sam 2: Segment anything in images and videos. arXiv preprint arXiv:2408.00714, 2024.

[6] Liang-Chieh Chen, Yukun Zhu, George Papandreou, Florian Schroff, and Hartwig Adam. Encoder-decoder with atrous separable convolution for semantic image segmentation. In *Proceedings of the European conference on computer vision (ECCV)*, pages 801–818, 2018.

[7] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L Yuille. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE transactions on pattern analysis and machine intelligence*, 40(4):834–848, 2017.

[8] Enze Xie, Wenhai Wang, Zhiding Yu, Anima Anandkumar, Jose M Alvarez, and Ping Luo. Segformer: Simple and efficient design for semantic segmentation with transformers. *Advances in neural information processing systems*, 34:12077–12090, 2021.

[9] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical image computing and computer-assisted intervention—MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18*, pages 234–241. Springer, 2015.

[10] Cifuentes, A., Mendoza, E., Lizcano, M., Santrich, A., & Moreno-Trillos, S. (2019). Desarrollo de una red neuronal convolucional para reconocer patrones en imágenes. *Investigación y desarrollo en TIC*, 10(2), 7-17.

[11] Morocho Jiménez, J. I. (2019). Detección de tumores cutáneos malignos y benignos usando una red neuronal convolucional (Bachelor's thesis, Quito, 2019.).

[12] MESRAN, M., et al. Investigating the Impact of ReLU and Sigmoid Activation Functions on Animal Classification Using CNN Models. *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, 2024, vol. 8, no 1, p. 111-118.

[13] ZHANG, Zhilu; SABUNCU, Mert. Generalized cross entropy loss for training deep neural networks with noisy labels. *Advances in neural information processing systems*, 2018, vol. 31.

- [14] Memisevic, R., Zach, C., Pollefeys, M., & Hinton, G. E. (2010). Gated softmax classification. *Advances in neural information processing systems*, 23.
- [15] Krithika Alias AnbuDevi, M., & Suganthi, K. (2022). Review of semantic segmentation of medical images using modified architectures of UNET. *Diagnostics*, 12(12), 3064.
- [16] Nguyen, N. Q., & Lee, S. W. (2019). Robust boundary segmentation in medical images using a consecutive deep encoder-decoder network. *IEEE Access*, 7, 33795-33808.
- [17] Müller, D., Soto-Rey, I., & Kramer, F. (2022). Towards a guideline for evaluation metrics in medical image segmentation. *BMC Research Notes*, 15(1), 210.
- [18] Furtado, P. (2021). Testing segmentation popular loss and variations in three multiclass medical imaging problems. *Journal of Imaging*, 7(2), 16.
- [19] Wang, W., & Lu, Y. (2018, March). Analysis of the mean absolute error (MAE) and the root mean square error (RMSE) in assessing rounding model. In *IOP conference series: materials science and engineering* (Vol. 324, p. 012049). IOP Publishing.
- [20] Bjorck, N., Gomes, C. P., Selman, B., & Weinberger, K. Q. (2018). Understanding batch normalization. *Advances in neural information processing systems*, 31.