RIELAC
Revista de Ingeniería Electrónica, Automática y Comunicaciones

Optimización de la arquitectura U-NET en segmentación automática de úlceras de pie diabético

Dailys Pilar Oliver Borrero, María Matilde García Lorenzo, Rosalís Amador García y Yusely Ruiz González

RESUMEN / ***ABSTRACT***

Las úlceras de pie diabético representan una complicación médica crítica que requiere una monitorización exacta para prevenir amputaciones. Los enfoques previos basados en detección de objetos presentaron limitaciones al incluir ruido de fondo y tejido sano en el análisis, afectando la precisión en la delimitación de la lesión. El objetivo de este estudio fue desarrollar y optimizar un modelo de segmentación semántica automatizada para delimitar con exactitud la morfología irregular de estas lesiones. Metodológicamente, se implementó la arquitectura U-Net utilizando ResNet34 como codificador preentrenado. Se llevó a cabo una búsqueda sistemática de hiperparámetros, evaluando el impacto de los optimizadores Adam y AdamW, así como variaciones en la tasa de aprendizaje y el tamaño de lote sobre la convergencia del modelo. Los resultados experimentales demostraron que la configuración óptima, utilizando el optimizador AdamW con una tasa de aprendizaje de 0.0001, alcanzó un coeficiente Dice de 0.8107 y un IoU de 0.6863, además de una exactitud global a nivel de píxel del 0.9908. Se concluyó que el modelo propuesto permite una segmentación robusta a nivel de píxel, reduciendo la interferencia del fondo y proporcionando información morfológica más detallada de la lesión que la obtenida mediante cajas delimitadoras, con potencial para apoyar la cuantificación objetiva y el seguimiento de la evolución de la herida.

Palabras claves: Segmentación semántica, úlceras de pie diabético, U-Net, ResNet34, aprendizaje profundo.

Diabetic foot ulcers represent a critical medical complication requiring precise monitoring to prevent amputations. Previous approaches based on object detection exhibited limitations by including background noise and healthy tissue in the analysis, affecting the accuracy of lesion delineation. The aim of this study was to develop and optimize an automated semantic segmentation model to accurately delineate the irregular morphology of these lesions. Methodologically, the U-Net architecture was implemented using ResNet34 as a pre-trained encoder. A systematic hyperparameter search was conducted to evaluate the impact of the Adam and AdamW optimizers, as well as variations in learning rate and batch size, on model convergence. Experimental results demonstrated that the optimal configuration, using the AdamW optimizer with a learning rate of 0.0001, achieved a Dice coefficient of 0.8107 and an IoU of 0.6863, alongside an overall pixel-level accuracy of 0.9908. It was concluded that the proposed model enables robust pixel-level segmentation, reducing background interference and providing more detailed morphological information about the lesion than that obtained via bounding boxes, with the potential to support objective quantification and the monitoring of wound progression.

Keywords: Semantic segmentation, diabetic foot ulcers, U-Net, ResNet34, deep learning.

Optimization of the U-NET architecture in automatic segmentation of diabetic foot ulcers

1. -INTRODUCCIÓN

La diabetes mellitus se ha consolidado como una de las epidemias globales más críticas del siglo XXI, siendo las úlceras de pie diabético (UPD) una de sus complicaciones más debilitantes y costosas. Se estima que una proporción significativa de pacientes diabéticos desarrollará una úlcera a lo largo de su vida, condición que, ante un manejo inadecuado, es la causa principal de amputaciones no traumáticas de miembros inferiores. La evaluación clínica tradicional de estas lesiones depende en gran medida de la inspección visual subjetiva para determinar su extensión y gravedad (infección, isquemia o incluso

Recibido: 08/2026 Aceptado: 10/2026

ambas categorías) [1]. En este escenario, la Inteligencia Artificial (IA) y la Visión por Computadora emergen como herramientas auxiliares vitales para estandarizar el diagnóstico y cuantificar la progresión de la herida de manera objetiva. El uso de Redes Neuronales Convolucionales (CNN) [2] y, más recientemente, arquitecturas basadas en Transformers (como ViT) [3], ha dominado la literatura sobre UPD. En este contexto,[4] exploraron el uso de técnicas de aprendizaje profundo para la detección y clasificación automática de UPD. Como resultado, se implementó un sistema inteligente para la clasificación de lesiones de UPD, registrado como "Cuidado Inteligente del Pie Diabético versión 1.0". La versión implementada incluye un módulo de captación y pre-procesamiento de la imagen del pie del paciente, en la detección de las lesiones utiliza el modelo de detección de objetos YOLOv8l [5] con un mAP50 [6] de 0.799 y en el módulo de clasificación de las distintas lesiones utiliza la red convolucional DenseNet201 [7] que reporta una eficacia en entrenamiento de 0.97 con una pérdida del 0.23, 0.94 de eficacia en validación con una pérdida del 0.45 y 0.4545 de eficacia en el conjunto de prueba. Ambos modelos fueron entrenados con las bases de datos liberadas durante los DFU Challenge 2020 (detección) [8] y 2021 (clasificación) [9]. Sin embargo, estos enfoques presentan limitaciones intrínsecas, la clasificación global de la imagen es susceptible al ruido del entorno y no localiza la lesión, mientras que las cajas delimitadoras incluyen invariablemente piel sana y fondo irrelevante, introduciendo sesgos en el análisis de características finas. Existe una brecha significativa en la literatura respecto a la transición de la mera detección a la segmentación semántica precisa a nivel de píxel, capaz de delinear la morfología irregular de la úlcera y excluir tejidos circundantes que confunden a los clasificadores tradicionales, como el tejido adiposo o callosidades.

La presente investigación aborda esta problemática mediante la implementación de técnicas de segmentación semántica [10], las cuales permiten una comprensión granular de la escena médica. Al aislar píxel a píxel la región de interés, se elimina la interferencia visual que afecta a los modelos de "caja negra" o de detección simple. Desde un punto de vista práctico, la segmentación precisa es un prerrequisito indispensable para el cálculo automatizado del área de la herida, un biomarcador clave para evaluar la eficacia del tratamiento a lo largo del tiempo. Metodológicamente, este trabajo justifica la integración de backbones robustos, y pre-entrenados, específicamente ResNet34 [11], dentro de arquitecturas encoder-decoder (U-Net) [12], para maximizar la extracción de características en conjuntos de datos médicos limitados. En la arquitectura U-Net-ResNet34, el backbone ResNet34 actúa como el extractor de características principal (el encoder). Su función es analizar la imagen de la úlcera y producir una representación compacta y rica en información. La elección de ResNet34 se debe a su probada eficacia, su capacidad para manejar redes profundas gracias a las conexiones residuales y, sobre todo, a la ventaja del Transfer Learning [13], que permite aprovechar un conocimiento visual ya adquirido para acelerar el entrenamiento y mejorar la precisión en la tarea específica de segmentación de úlceras.

El objetivo general de este estudio es desarrollar y optimizar un modelo de segmentación semántica basado en la arquitectura U-Net con codificador ResNet34 para la delimitación automática de UPD. Para ello, se evalúa el impacto de diferentes estrategias de optimización y tasas de aprendizaje en la convergencia del modelo, se cuantifica el rendimiento de la segmentación mediante métricas como el Coeficiente Dice, la Intersección sobre Unión (IoU) y la Precisión, y se determina la configuración de hiperparámetros que maximiza la fidelidad de la segmentación y minimiza los falsos positivos y negativos.

En la sección 3, se detalla la metodología empleada, describiendo la arquitectura de red propuesta, el preprocesamiento de datos y el diseño experimental de la búsqueda de hiperparámetros. Posteriormente, se presentan y discuten los resultados cuantitativos obtenidos, contrastando las diferentes configuraciones evaluadas para finalmente ofrecer conclusiones sobre el desempeño y las posibilidades de aplicación del modelo desarrollado.

Las contribuciones científicas fundamentales que se hacen en este trabajo son:

1. Implementación de un modelo de segmentación semántica basado en U-Net con backbone ResNet34 para UPD.
2. Optimización sistemática de hiperparámetros para analizar el efecto de Adam y AdamW sobre el rendimiento del modelo.
3. Cuantificación del rendimiento mediante Dice, IoU y precisión.
4. Comparación referencial del rendimiento obtenido con resultados publicados para el desafío DFUC2022.

2.- ENFOQUES DE SEGMENTACIÓN TRADICIONALES Y DE APRENDIZAJE AUTOMÁTICO

El seguimiento preciso del proceso de cicatrización de las UPD es un pilar fundamental para la evaluación de la eficacia de los tratamientos y la prevención de complicaciones graves como la amputación. La medición del área de la lesión constituye el biomarcador cuantitativo más relevante para este fin. Históricamente, esta evaluación se ha realizado mediante métodos manuales, los cuales son subjetivos, propensos a errores y conllevan riesgos de infección. La irrupción de las técnicas de análisis de imágenes y la visión por computadora ha abierto la vía hacia la automatización de este proceso, prometiendo mediciones objetivas, no invasivas y repetibles. A continuación, se presenta una revisión crítica de los enfoques más relevantes

en la literatura, agrupados por su paradigma tecnológico, para contextualizar la novedad y el aporte de la presente investigación.

Las primeras aproximaciones a la segmentación automática de UPD se basaron en algoritmos de procesamiento de imágenes clásicos y modelos de aprendizaje automático que no eran profundos. Estos métodos dependen en gran medida de la ingeniería de características manual y del preprocesamiento de la imagen para resaltar la región de interés. Un ejemplo representativo de esta línea de investigación es el estudio presentado en [14]. En este trabajo, los autores compararon cualitativamente métodos de segmentación como Chan-Vese [15], el modelo de mezclas gaussianas (GMM) [16] y GrabCut [17]. Para mejorar los resultados, se aplicó un preprocesamiento mediante la Transformada Wavelet Discreta Logarítmica en un modelo S-LIP, buscando incrementar el contraste entre la piel sana y el tejido ulcerado.

Los resultados de este y otros estudios similares demuestran que, si bien los métodos tradicionales pueden ser funcionales en escenarios controlados, presentan limitaciones significativas. Su rendimiento es altamente sensible a condiciones variables de iluminación, al color de la piel del paciente y a la baja definición de los bordes de la úlcera. La dependencia de características predefinidas limita su capacidad de generalización a conjuntos de datos diversos y complejos, un escenario común en la práctica clínica. Aunque el trabajo concluye que el GMM fue el método más factible, su precisión no alcanza los niveles de robustez requeridos para una aplicación clínica fiable, dejando abierta la necesidad de enfoques más potentes y adaptables.

La revisión de la literatura revela una clara trayectoria evolutiva desde métodos tradicionales con un rendimiento limitado hacia arquitecturas de aprendizaje profundo cada vez más complejas y especializadas. Si bien trabajos como el de [18] sentaron las bases demostrando la idoneidad de U-Net, y propuestas como DFUNet [19] o los ganadores de competiciones como DFUC2022 [20] han establecido puntos de referencia con arquitecturas de vanguardia, persiste una brecha de conocimiento crucial.

El paradigma dominante parece ser que para mejorar el rendimiento es necesario crear arquitecturas nuevas y cada vez más intrincadas. Sin embargo, esta especialización conlleva desventajas significativas: aumenta la complejidad computacional, dificulta la reproducibilidad de los resultados y aleja los modelos de los estándares robustos y bien comprendidos por la comunidad científica. La presente investigación se sitúa precisamente en esta brecha, proponiendo una alternativa a esta tendencia.

3.- DISEÑO GENERAL DE LA SOLUCIÓN

El sistema propuesto se basa en un paradigma de aprendizaje profundo supervisado para tareas de visión por computadora. A diferencia de los enfoques de clasificación que asignan una etiqueta global a la imagen, o de detección que generan cajas delimitadoras, este diseño busca generar una máscara binaria donde cada píxel es clasificado como "lesión" o "fondo".

El flujo de procesamiento, ilustrado en la Fig. 1, consta de cuatro etapas secuenciales. La ingesta y preprocesamiento, donde se realiza el redimensionamiento y la normalización de las imágenes y máscaras para adecuarlas a la entrada del modelo. La codificación de características (Encoder) [21] para la extracción de mapas de características mediante una red vertebral pre-entrenada. La decodificación y reconstrucción (Decoder) [21] que consiste en la recuperación de la resolución espacial mediante capas de muestreo ascendente y conexiones residuales y por último la generación de máscara [22] aplicando un umbral de decisión para obtener la segmentación final.

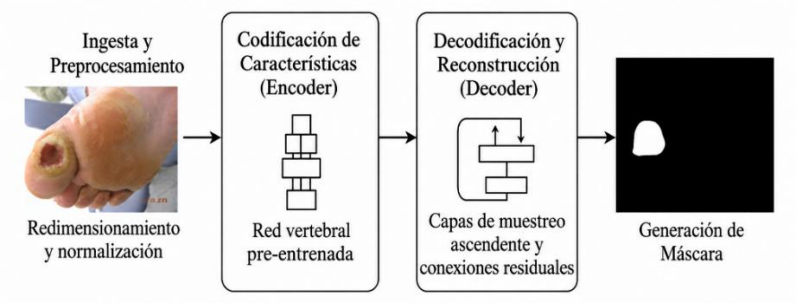


Figura 1

Diagrama de flujo general del sistema de segmentación semántica propuesto, desde la entrada de la imagen original hasta la generación de la máscara binaria.

3.1.- CONJUNTO DE DATOS Y PREPROCESAMIENTO

Para el entrenamiento y validación del modelo se utilizaron 2000 imágenes provenientes del conjunto de datos DFUC2022 (Diabetic Foot Ulcer Challenge) [20], junto con sus respectivas máscaras de segmentación. Los pares imagen-máscara fueron mezclados aleatoriamente utilizando una semilla fija de 42 y posteriormente divididos en dos subconjuntos: el 90% de las imágenes se destinó al entrenamiento y el 10% restante a la validación.

Como parte del preprocesamiento, las imágenes fueron cargadas en escala de grises y redimensionadas a una resolución uniforme de 256 × 256 píxeles. Los valores de intensidad se normalizaron al intervalo [0,1] mediante su división por 255. Las máscaras de segmentación también fueron convertidas a escala de grises y redimensionadas a 256 × 256 píxeles; posteriormente, se binarizaron mediante un umbral de intensidad de 127 para representar las clases lesión y fondo.

En los experimentos reportados no se aplicaron transformaciones de aumento de datos. El codificador ResNet34 fue inicializado con pesos preentrenados en ImageNet como parte de la estrategia de Transfer Learning.

3.2.- ARQUITECTURA DEL MODELO: U-NET CON BACKBONE, RESNET34

El núcleo de la solución es una arquitectura U-Net [23], seleccionada por su eficacia probada en la segmentación biomédica. U-Net se caracteriza por su estructura en forma de "U", compuesta por una ruta de contracción (encoder) para capturar el contexto y una ruta de expansión (decoder) para permitir una localización precisa.

En lugar de entrenar el codificador desde cero, se sustituyó la ruta de contracción estándar por una red ResNet34 pre-entrenada en ImageNet. Esta estrategia, conocida como Transfer Learning, permite aprovechar extractores de características robustos y jerárquicos (desde bordes simples hasta texturas complejas de tejido) aprendidos previamente en grandes volúmenes de datos.

ResNet34 introduce bloques residuales que mitigan el problema del desvanecimiento del gradiente en redes profundas, permitiendo un entrenamiento más efectivo. El encoder reduce progresivamente la dimensión espacial de los mapas de características mientras aumenta su profundidad.

El decodificador reconstruye la resolución espacial de la imagen mediante operaciones de convolución transpuesta. Un componente crítico son las conexiones de salto, mostradas en la Fig. 2. Estas conexiones concatenan los mapas de características de alta resolución del encoder (ResNet34) con las características procesadas en el decoder. Esto permite al modelo recuperar información espacial fina perdida durante la codificación, resultando en bordes de lesión mejor definidos.

La selección de ResNet34 como codificador respondió a la búsqueda de un equilibrio entre capacidad de extracción de características y complejidad computacional. Su utilización como encoder de U-Net ha mostrado resultados satisfactorios en tareas de segmentación de imágenes médicas, aprovechando las conexiones residuales para facilitar el entrenamiento y la extracción de características [24]. Asimismo, estudios comparativos de arquitecturas residuales en segmentación médica muestran que el incremento de la profundidad de la red y del costo computacional no necesariamente se traduce en una mejora proporcional del rendimiento. En este sentido, ResNet34 ofrece una alternativa de menor complejidad frente a variantes más profundas como ResNet50 y ResNet101, manteniendo las ventajas del aprendizaje residual y del Transfer Learning. Considerando además el tamaño limitado del conjunto de datos empleado en este estudio, se seleccionó ResNet34 como backbone de la arquitectura U-Net.

3.3.- CONFIGURACIÓN EXPERIMENTAL Y ENTRENAMIENTO

Para identificar la configuración óptima del modelo, se diseñó un experimento factorial variando hiperparámetros críticos que influyen en la dinámica del descenso del gradiente. Se evaluaron dos optimizadores para actualizar los pesos de la red, uno es Adam (Adaptive Moment Estimation) [25] que es estándar en aprendizaje profundo por su eficiencia en el manejo de tasas de aprendizaje adaptativas y el otro es AdamW [26], que es una variante de Adam que desacopla la caída de pesos (weight decay) de la optimización del gradiente, lo cual teóricamente mejora la generalización en modelos complejos.

Se realizaron pruebas sistemáticas combinando los siguientes valores, tal como se refleja en los registros experimentales:

Tasa de Aprendizaje (Learning Rate - LR) [27]: Se probaron valores de orden de magnitud 0.0001 y 0.001 para encontrar el equilibrio entre velocidad de convergencia y estabilidad en el mínimo global.

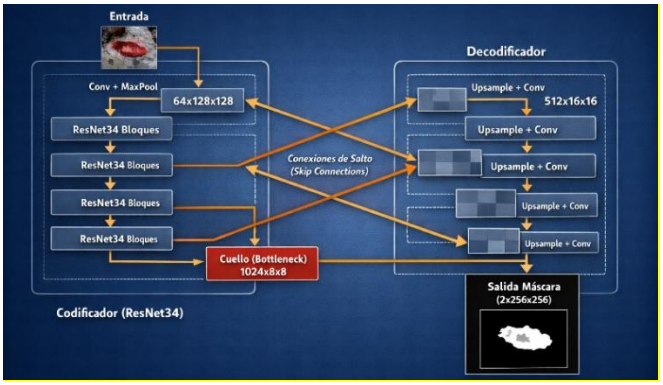


Figura 2

Arquitectura detallada del modelo U-Net implementado, destacando la integración del backbone ResNet34 en la etapa de codificación y las conexiones de salto hacia el decodificador.

Tamaño del Lote (Batch Size) [28]: Se evaluaron tamaños de 8 y 16 imágenes por lote, condicionados por la memoria de la GPU y la necesidad de estimar correctamente el gradiente.

Número de épocas [29]: Se monitoreó el entrenamiento hasta 10 épocas para observar la estabilización de las métricas de pérdida (loss).

Para el entrenamiento del modelo se utilizó la función de pérdida CrossEntropyLoss de PyTorch con dos clases (lesión y fondo). Esta función combina una capa LogSoftmax y la pérdida Negative Log Likelihood (NLL), y cuantifica la discrepancia entre las probabilidades predichas por el modelo y las etiquetas reales de cada píxel. La misma función de pérdida se empleó en todas las configuraciones experimentales reportadas en la Tabla 1. Para evaluar el desempeño, se monitorearon las siguientes métricas cuantitativas:

Coeficiente Dice [30]: Métrica principal que mide la superposición armónica entre la máscara predicha y la real.

IoU (Intersección sobre Unión) [31]: Mide el porcentaje de superposición entre la predicción y la verdad fundamental.

Precisión (Precision) y Sensibilidad (Recall) [32]: Para evaluar la exactitud en la clasificación de píxeles positivos (lesión).

3.4.- HERRAMIENTAS Y ENTORNO DE IMPLEMENTACIÓN

La implementación de los modelos se realizó utilizando el lenguaje de programación Python. Se empleó la biblioteca Segmentation Models [33] basada en el framework de aprendizaje profundo PyTorch [34], lo que facilitó la integración de backbones pre-entrenados.

Los experimentos se ejecutaron en la plataforma Kaggle [35], utilizando aceleración mediante GPU NVIDIA Tesla T4 x2. Durante la ejecución, PyTorch detectó y utilizó el dispositivo CUDA para realizar las operaciones de entrenamiento. El tiempo total de entrenamiento de cada configuración y la convergencia de la función de pérdida fueron registrados con el objetivo de comparar el rendimiento y la eficiencia computacional de las configuraciones evaluadas. Los tiempos de entrenamiento obtenidos se presentan en la Tabla 1.

4.- RESULTADOS Y ANÁLISIS

Los experimentos realizados con el modelo de segmentación U-Net con backbone ResNet34 se diseñaron para analizar el impacto de tres hiperparámetros: el optimizador (Adam y AdamW), la tasa de aprendizaje (Learning Rate - LR) y el tamaño del lote (Batch Size), evaluando el rendimiento a las 5 y 10 épocas.

Se evaluaron 8 configuraciones de hiperparámetros, resultantes de la combinación de dos optimizadores (Adam y AdamW), dos tasas de aprendizaje (0.001 y 0.0001) y dos tamaños de lote (8 y 16). Para cada configuración se registró el rendimiento

en las épocas 5 y 10, obteniéndose un total de 16 registros experimentales. La Tabla 1 resume las métricas de rendimiento obtenidas para las configuraciones evaluadas, agrupadas según el optimizador y el número de épocas. El Coeficiente Dice se estableció como la métrica principal para la selección del modelo debido a su robustez ante el desbalance de clases inherente a la segmentación de úlceras.

El análisis de los resultados muestra que, entre las configuraciones evaluadas, AdamW combinado con una tasa de aprendizaje de 0.0001 alcanzó el mayor rendimiento al finalizar las 10 épocas. La combinación de AdamW con LR de 1e−4 y Batch Size de 16 obtuvo un Coeficiente Dice de 0.8107 y un índice IoU de 0.6863. Bajo las mismas condiciones de tasa de aprendizaje, tamaño de lote y número de épocas, el optimizador Adam obtuvo un Dice de 0.7955 y un IoU de 0.6689. Por otra parte, con AdamW, una tasa de aprendizaje de 1e−3 y Batch Size de 16, se obtuvo un Dice de 0.7019 a las 10 épocas, inferior al alcanzado con LR de 1e−4. Estos resultados indican que, dentro de las configuraciones evaluadas, la tasa de aprendizaje de 0.0001 proporcionó el mejor desempeño para AdamW.

Tabla 1

Comparativa de rendimiento de las configuraciones experimentales evaluadas para la arquitectura U-Net-ResNet34

Optimizador	LR	Batch	Épocas	Dice	IoU	Train Loss	Val Loss	Tiempo (s)
Adam	0.001	8	5	0.7292	0.5923	0.0564	0.0348	212.88
Adam	0.001	16	5	0.7567	0.6161	0.0531	0.0334	178.12
Adam	0.0001	8	5	0.7580	0.6221	0.1528	0.0369	187.04
Adam	0.0001	16	5	0.7475	0.6061	0.1332	0.0460	173.05
AdamW	0.001	8	5	0.7203	0.5823	0.0479	0.0330	184.76
AdamW	0.001	16	5	0.6295	0.4652	0.0493	0.0568	169.85
AdamW	0.0001	8	5	0.7648	0.6394	0.0845	0.0362	184.35
AdamW	0.0001	16	5	0.8073	0.6804	0.0791	0.0360	173.10
Adam	0.001	8	10	0.7371	0.5997	0.0386	0.0348	368.97
Adam	0.001	16	10	0.7353	0.5940	0.0438	0.0379	349.17
Adam	0.0001	8	10	0.7874	0.6652	0.0551	0.0331	370.67
Adam	0.0001	16	10	0.7955	0.6689	0.0788	0.0355	356.71
AdamW	0.001	8	10	0.5729	0.4350	0.0401	0.0434	385.19
AdamW	0.001	16	10	0.7019	0.5499	0.0425	0.0466	355.13
AdamW	0.0001	8	10	0.7990	0.6784	0.0327	0.0339	372.65
AdamW	**0.0001**	**16**	**10**	**0.8107**	**0.6863**	**0.0746**	**0.0335**	**342.69**

El aumento del tamaño del lote de 8 a 16 demostró beneficios en la estabilidad del gradiente y la precisión final en la mayoría de los casos exitosos. En el mejor escenario (AdamW, LR de 0.0001, 10 épocas), pasar de un Batch Size de 8 a 16 incrementó el coeficiente Dice de 0.7990 a 0.8107. Simultáneamente, este cambio aumentó la pérdida de entrenamiento (Train Loss) de 0.0339 a 0.0746.

4.1.- RESULTADOS EXPERIMENTALES

A partir de los resultados obtenidos en la búsqueda de hiperparámetros, se seleccionaron las cinco configuraciones con mejor desempeño para comparar sus métricas de Dice e IoU, como se muestra en la Fig. 3.

La Figura 4 muestra ejemplos cualitativos de la segmentación obtenida con la configuración óptima. Se observa que el modelo logra delimitar correctamente la región de la lesión en la mayoría de los casos, con bordes bien definidos que se aproximan a la máscara real. Sin embargo, en lesiones con bordes difusos o de pequeño tamaño se aprecian ligeras diferencias respecto a la máscara real, lo cual es consistente con la tasa de falsos negativos reportada (FNE = 0.1787). Estos resultados confirman que el modelo es capaz de capturar la morfología irregular de las úlceras y de discriminar la lesión del tejido sano circundante.

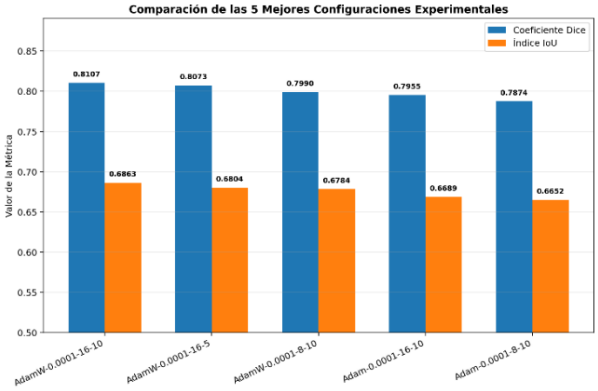


Figura 3

Gráfico de barras comparativo de las métricas Dice e IoU para las 5 mejores configuraciones experimentales.

Basado en la maximización del Coeficiente Dice y la minimización de la pérdida, se seleccionó el modelo entrenado con AdamW, LR 0.0001, Batch Size 16 durante 10 épocas. Presenta un coeficiente Dice de 0.8107, lo que indica una superposición del 0.8107 entre la predicción y la máscara real, con un IoU de 0.6863 y una exactitud global a nivel de píxel (pixel accuracy) de 0.9908. Aunque este valor es elevado, debe interpretarse con cautela debido al desbalance entre clases y al predominio de píxeles correspondientes al fondo frente a los pertenecientes a la lesión. Por esta razón, el Coeficiente Dice y el IoU se consideran métricas más representativas del desempeño de segmentación.

La precisión obtenida fue de 0.8126, con una sensibilidad (Recall) de 0.8213 y una tasa de falsos negativos (FNE) de 0.1787. En un contexto clínico, este resultado implica que aproximadamente una fracción de 0.1787 de los píxeles correspondientes a tejido lesionado no fueron identificados, lo cual representa una limitación para el cálculo automático exacto del área de la herida, especialmente en lesiones pequeñas o con bordes difusos.

5.- DISCUSIÓN

La presente investigación se planteó con el objetivo fundamental de superar las limitaciones de granularidad observadas en los sistemas de detección de objetos convencionales aplicados a las UPD. Trabajos previos basados en YOLOv8l o clasificadores como ViT, si bien lograban identificar la presencia de la lesión, presentan la limitación de que las cajas delimitadoras pueden incluir tejido sano y fondo irrelevante alrededor de la lesión. Este enfoque de detección, a pesar de ofrecer tiempos de inferencia rápidos, impedía el cálculo automatizado del área de la herida, un biomarcador crítico para la evolución clínica. En contraste, los resultados obtenidos mediante la arquitectura U-Net con backbone ResNet34 muestran que la segmentación semántica permite una delimitación de la región patológica a nivel de píxel, proporcionando información morfológica más detallada que la obtenida mediante cajas delimitadoras. La propuesta logra discriminar píxel a píxel la región patológica, alcanzando un Coeficiente Dice de 0.8107 y una exactitud global a nivel de píxel de 0.9908, con un tiempo de entrenamiento de 342.69 segundos para 10 épocas.

Para contextualizar los resultados obtenidos, se tomaron como referencia los valores publicados para la competencia DFUC2022. Es importante señalar que esta comparación tiene un carácter exclusivamente referencial, ya que los resultados del presente estudio fueron obtenidos sobre una partición propia del conjunto de datos, utilizando el 90% de las imágenes para entrenamiento y el 10% restante para validación, mientras que los resultados reportados para el desafío DFUC2022 corresponden a su protocolo oficial de evaluación. Por tanto, las diferencias entre las métricas no deben interpretarse como una comparación experimental directa entre ambos modelos. El modelo ganador del desafío, basado en una arquitectura HarDNet-MSEG modificada, reportó un Coeficiente Dice de 0.7287, un índice IoU aproximado de 0.6252 y una FNE cercana a 0.234. El modelo propuesto obtuvo un Dice de 0.8107 y un IoU de 0.6863; estos valores son superiores numéricamente a los reportados para HarDNet-MSEG, pero, debido a las diferencias en los conjuntos y protocolos de evaluación, deben interpretarse únicamente como una referencia respecto a los resultados publicados.

Desde una perspectiva metodológica, los resultados obtenidos muestran que una arquitectura U-Net con Transfer Learning mediante ResNet34 puede alcanzar un desempeño competitivo en las métricas de segmentación evaluadas sin recurrir a arquitecturas de mayor complejidad. En las configuraciones experimentales analizadas, la combinación de AdamW con una tasa de aprendizaje de 0.0001 produjo los mejores valores de Dice e IoU. Estos resultados resaltan la influencia que ejercen tanto la selección del optimizador como sus hiperparámetros sobre el proceso de entrenamiento y el desempeño final del modelo.

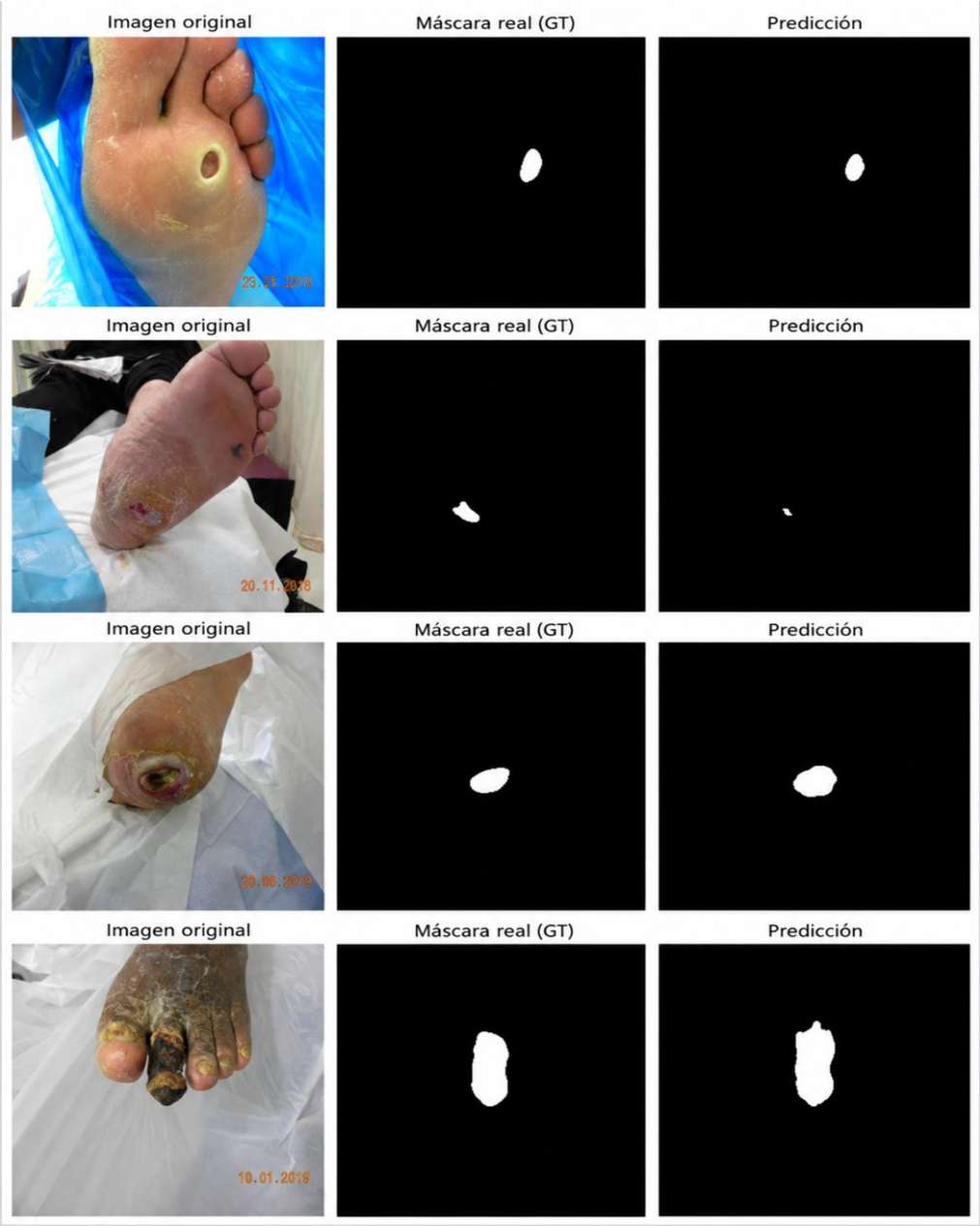


Figura 4

Comparación cualitativa de la segmentación. De izquierda a derecha: imagen original, máscara real (Ground Truth) y máscara predicha por el modelo U-Net-ResNet34 con la configuración óptima (AdamW, LR=0.0001, Batch=16, 10 épocas).

No obstante, un análisis detallado de las métricas de error revela áreas de mejora. El modelo propuesto presenta una tasa de falsos negativos (FNE) de 0.1787, lo que indica que aproximadamente el 17.87% de los píxeles correspondientes a tejido lesionado no fueron identificados. Como referencia, en los resultados publicados del DFUC2022 se reportan valores de FNE cercanos a 0.22-0.24; sin embargo, esta comparación debe interpretarse con cautela debido a las diferencias en los protocolos de evaluación. Desde una perspectiva clínica, la omisión de píxeles pertenecientes a la lesión puede provocar una subestimación del área real de la úlcera, especialmente en lesiones pequeñas o con bordes difusos, lo que podría afectar la cuantificación objetiva de su evolución. Este comportamiento sugiere que una de las principales áreas de mejora se encuentra en la delimitación precisa de los bordes de la lesión. Como trabajo futuro, sería conveniente evaluar estrategias orientadas a

reducir estos errores, como funciones de pérdida sensibles a los bordes, mecanismos de atención o técnicas de postprocesamiento morfológico.

En conjunto, las métricas obtenidas muestran que el modelo propuesto alcanza un desempeño favorable en la segmentación de las úlceras de pie diabético dentro de la partición de validación utilizada. Los valores de Coeficiente Dice e IoU obtenidos son numéricamente superiores a algunos de los resultados publicados para el DFUC2022; no obstante, las diferencias en los protocolos de evaluación impiden establecer una superioridad experimental directa. La utilización de una arquitectura U-Net con Transfer Learning mediante ResNet34 permitió obtener resultados competitivos manteniendo una arquitectura relativamente simple, lo que evidencia el potencial de este enfoque para continuar investigando sistemas automatizados de segmentación y monitorización de las úlceras de pie diabético.

6.- CONCLUSIONES

Se desarrolló un modelo de segmentación semántica automatizada para úlceras de pie diabético orientado a obtener una delimitación de la lesión a nivel de píxel, evitando las limitaciones inherentes al uso de cajas delimitadoras en los métodos de detección. Con un Coeficiente Dice de 0.8107, el sistema propuesto permite aislar la morfología de la herida excluyendo el ruido de fondo y el tejido adiposo que generaba falsos positivos en modelos de clasificación anteriores.

La evaluación sistemática de hiperparámetros mostró que, entre las configuraciones evaluadas, el optimizador AdamW combinado con una tasa de aprendizaje de 0.0001 alcanzó el mejor rendimiento para la arquitectura U-Net-ResNet34, con los mayores valores de Dice e IoU obtenidos en los experimentos realizados.

La eficiencia computacional observada durante los experimentos, junto con el uso de ResNet34 como backbone, muestra el potencial de esta arquitectura para el desarrollo de herramientas automatizadas de apoyo al análisis de úlceras de pie diabético. No obstante, su implementación en entornos clínicos reales requerirá estudios adicionales que evalúen aspectos como el tiempo de inferencia, el rendimiento sobre datos externos y su integración en los flujos de trabajo clínicos.

Aunque el modelo alcanzó una exactitud global a nivel de píxel del 0.9908, este valor debe interpretarse con cautela debido al desbalance entre las clases y al predominio de píxeles correspondientes al fondo. Por esta razón, el Coeficiente Dice y el IoU constituyen métricas más representativas del desempeño del modelo en la segmentación de la lesión. Asimismo, la tasa de falsos negativos obtenida evidencia que la delimitación de los bordes continúa siendo un desafío técnico, por lo que como trabajo futuro se propone evaluar funciones de pérdida sensibles a los bordes, mecanismos de atención y técnicas de postprocesamiento orientadas a reducir estos errores.

AGRADECIMIENTOS

Este trabajo ha sido parcialmente financiado por los Fondos de Proyectos con código PN223LH004-039 y PT241VC001-009. También agradecemos el apoyo de los organizadores del desafío DFU 2022 que brindaron acceso a la base de datos.

REFERENCIAS

1. Arias-Rodríguez FD, Jiménez-Valdiviezo MA, del Cisne-Ríos-Criollo K, Murillo-Araujo GP, Toapanta-Allauca DS, Rubio-Laverde KA, et al. Pie diabético. Actualización en diagnóstico y tratamiento. Revisión bibliográfica. Angiología. 2023;75(4):242-58.

2. Krichen M. Convolutional Neural Networks: A Survey. Computers. 2023;12(8):151.

3. Palanisamy B, Hassija V, Chatterjee A, Mandal A, Chakraborty D, Pandey A, et al. Transformers for Vision: A Survey on Innovative Methods for Computer Vision. IEEE Access. 2025;13:95496-523.

4. Ruiz-González Y, Cáceres EM, García Lorenzo MM, Andreu AM, Benavidez CG. Detection of Diabetic Foot Ulcers Using Deep Learning Models. En: Figueroa-García JC, López-Sotelo JA, Moreno-Trujillo JF, Gaona-García EE, editores. Applied Computer Sciences in Engineering. Cham: Springer Nature Switzerland; 2025. p. 308-19.

5. Rizk M, Bayad I. Bringing Intelligence to SAR Missions: A Comprehensive Dataset and Evaluation of YOLO for Human Detection in TIR Images. IEEE Access. 2025;13:17208-35.

6. Ren W, Qian T, Li G, Li X. RT-DETR-MC: A Novel Fusion of Multi-Scale Features and Gating Mechanism for Lettuce Disease Detection. En: Proceedings of the 2025 6th International Conference on Computer Information and Big Data Applications. New York, NY, USA: Association for Computing Machinery; 2025. p. 196-202. (CIBDA '25).

7. Samad SA, Gitanjali J. Augmenting DenseNet: Leveraging Multi-Scale Skip Connections for Effective Early-Layer Information Incorporation. IEEE Access. 2024;12:141344-60.

8. Yap MH, Cassidy B, Pappachan JM, O'Shea C, Gillespie D, Reeves N. Analysis Towards Classification of Infection and Ischaemia of Diabetic Foot Ulcers. En: 2021 IEEE EMBS International Conference on Biomedical and Health Informatics (BHI). 2021. p. 1-4.

9. Wu X, Liu R, Wen Q, Ao B, Li K. DFUC2021 Dataset Classification based on Deep Semi-supervised Learning Methods. En: 2022 2nd International Conference on Consumer Electronics and Computer Engineering (ICCECE). Guangzhou, China: IEEE; 2022. p. 499-502.

10. Yang Z, Yu H, Zhang J, Tang Q, Mian A. Deep learning based infrared small object segmentation: Challenges and future directions. Information Fusion. 2025; 118:103007.

11. Koonce B. ResNet 34. En: Convolutional Neural Networks with Swift for Tensorflow: Image Recognition and Dataset Categorization. Berkeley, CA: Apress; 2021. p. 51-61.

12. Suri JS, Bhagawati M, Agarwal S, Paul S, Pandey A, Gupta SK, et al. UNet Deep Learning Architecture for Segmentation of Vascular and Non-Vascular Images: A Microscopic Look at UNet Components Buffered With Pruning, Explainable Artificial Intelligence, and Bias. IEEE Access. 2023;11:595-645.

13. Zhuang F, Qi Z, Duan K, Xi D, Zhu Y, Zhu H, et al. A Comprehensive Survey on Transfer Learning. Proceedings of the IEEE. 2021, vol. 109, no. 1, pp. 43-76 , doi: 10.1109/JPROC.2020.3004555

14. Gómez GG, Santiago DV, Díaz-Romañach MLB, Gutiérrez WOS, Figueroa AA, Pérez CV, et al. Mejoramiento de contraste y segmentación en imágenes de úlceras del pie diabético. Revista Cubana de Angiología y Cirugía Vascular. 2019;20(3).

15. Shenoy K. P P. Image Segmentation using Chan-Vese Active Contours. 2025. 10.48550/arXiv.2506.19344.

16. Abd al-munaim HA, Araheemah WA. Comparative Evaluation of FCM, K-Means, and GMM for Brain MRI Segmentation. En: Proceedings of the 2026 2nd International Conference on Computing and Emerging Sciences. New York, NY, USA: Association for Computing Machinery; 2026. p. 114-20. (ICCES '26).

17. Rother C, Kolmogorov V, Blake A. «GrabCut»: interactive foreground extraction using iterated graph cuts. ACM Trans Graph. 2004;23(3):309-14.

18. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. En: Navab N, Hornegger J, Wells WM, Frangi AF, editores. Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. Cham: Springer International Publishing; 2015. p. 234-41.

19. Goyal M, Reeves ND, Davison AK, Rajbhandari S, Spragg J, Yap MH. DFUNet: Convolutional Neural Networks for Diabetic Foot Ulcer Classification. IEEE Trans Emerg Top Comput Intell. 2020;4(5):728-39.

20. Kendrick C, Cassidy B, Reeves ND, Pappachan JM, O'Shea C, Chandrabalan V, et al. Diabetic Foot Ulcer Grand Challenge 2022 Summary. En: Yap MH, Kendrick C, Cassidy B, editores. Diabetic Foot Ulcers Grand Challenge. Cham: Springer International Publishing; 2023. p. 115-23.

21. Kamsari M, Sadeghi S, de Oliveira GG, Alves AM, Sarshar NT, Anari S, et al. The role of data augmentation and attention mechanisms in UNet and ConvNeXt architectures for optimizing breast tumor segmentation. Sci Rep. 2025;15(1):45268.

22. Suárez Sáez N del C. HPPrintingSegmentation: extrae e imprime desde tu imaginación [Tesis de Grado]. Universidad de Oviedo; 2025. (Grado en Ingeniería Informática en Tecnologías de la Información).

23. U-Net. En: Wikipedia, la enciclopedia libre [Internet]. 2024 [citado 17 de diciembre de 2025]. Disponible en: https://es.wikipedia.org/w/index.php?title=U-Net&oldid=158704581.

24. Yojana K, Thillai Rani L. OCT layer segmentation using U-NET semantic segmentation and RESNET34 encoder-decoder. Measurement: Sensors. 1 de octubre de 2023;29:100817. doi:10.1016/j.measen.2023.100817

25. Lou Y, Shen Y, Ahmed Khan M, Zhao M, Dai F, Shan X, et al. Bidirectional accelerated adaptive moment estimation for deep neural networks. Knowledge-Based Systems. 2026; 341:115784.

26. Zhou P, Xie X, Lin Z, Yan S. Towards Understanding Convergence and Generalization of AdamW. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2024;46(9):6486-93.

27. Swathika R, Kumar SMD, Srinidhi NN, Harshitha BR. Harnessing Learn Rate Schedule for Adaptive Deep Learning in LoRaWAN-IoT Localization. IEEE Access. 2024;12:72034-50.

28. Baek TW, Ahn YM, Ye CS. Comparative Analysis of Image Segmentation Performance Based on Batch Size Variations in UNet Based Deep Learning Models. Korean Journal of Remote Sensing. 2024;40(6_1):895-906.

29. Kurniawan A, Erlangga E, Tanjung T, Ariani F, Aprilinda Y, Endra RY. Review of Deep Learning Using Convolutional Neural Network Model. 2024;49-55.

30. Vlăsceanu GV, Tarbă N, Voncilă ML, Boiangiu CA. Selecting the Right Metric: A Detailed Study on Image Segmentation Evaluation. BRAIN Broad Research in Artificial Intelligence and Neuroscience. 2024;15(4):295-318.

31. Rezatofighi SH, Tsoi N, Gwak J, Sadeghian A, Reid ID, Savarese S. Generalized Intersection over Union: A Metric and A Loss for Bounding Box Regression. CoRR [Internet]. 2019;abs/1902.09630. Disponible en: http://arxiv.org/abs/1902.09630

32. Fränti P, Mariescu-Istodor R. Soft precision and recall. Pattern Recognition Letters. 2023;167:115-21.

33. Apte AP, Iyer A, Thor M, Pandya R, Haq R, Jiang J, et al. Library of deep-learning image segmentation and outcomes model-implementations. Physica Medica. 2020; 73:190-6.

34. Barillaro L. Deep Learning Platforms: PyTorch. En: Ranganathan S, Cannataro M, Khan AM, editores. Encyclopedia of Bioinformatics and Computational Biology (Second Edition). Oxford: Elsevier; 2025. p. 167-70.

35. Bojer CS, Meldgaard JP. Kaggle forecasting competitions: An overlooked learning opportunity. International Journal of Forecasting. 2021, 37(2):587-603.

CONFLICTO DE INTERESES

No existe conflicto de intereses entre los autores, ni con ninguna institución a la que cada uno está afiliado, ni con ninguna otra institución.
Las opiniones expresadas aquí son únicamente responsabilidad de los autores y no representan la posición de la Institución o las instituciones a las que están afiliados.

CONTRIBUCIONES DE LOS AUTORES

Dailys Pilar Oliver Borrero: Investigación, Metodología, Software, Visualización, Redacción- borrador original, Análisis formal.

María Matilde García Lorenzo: Conceptualización, Redacción – revisión y edición, Supervisión.

Yusely Ruiz-Gonzalez: Adquisición de fondo, Recursos, Administración de proyecto, Supervisión, Curación de datos, Redacción – revisión y edición, Validación -Verificación.

Rosalís Amador García: Supervisión, Redacción – revisión y edición.

AUTORES

Dailys Pilar Oliver Borrero, Lic. en Ciencias de la Computación, Universidad Central "Marta Abreu" de Las Villas, Villa Clara, Cuba, dpob0305@gmail.com, No. ORCID: 0009-0009-3418-6628, Aprendizaje automatizado, XAI, Evaluación de XAI, Segmentación de imágenes y exploración de diferentes arquitecturas CNN y Transformers.

María Matilde García Lorenzo, Lic. Cibernética Matemática, Doctora en Ciencias Técnicas, Universidad Central "Marta Abreu" de Las Villas, Villa Clara, Cuba, mmgarcia@uclv.edu.cu; mmgarcialorenzo@gmail.com, No. ORCID: 0000-0002-1663-5794, Aprendizaje automatizado, XAI, Evaluación de XAI

Yusely Ruiz González, Ing. en Telecomunicaciones y Electrónica, Doctora en Ciencias Técnicas, Universidad Central "Marta Abreu" de Las Villas, Villa Clara, Cuba, yuselyr@uclv.edu.cu, No. ORCID: 0000-0002-6011-1947, Procesamiento digital de señales e imágenes, visión computacional e inteligencia artificial.

Rosalís Amador García, Lic. en Ciencias de la Computación, Master en Ciencias, Universidad Central "Marta Abreu" de Las Villas, Villa Clara, Cuba, rosalisamadorgarcia@gmail.com , No. ORCID: 0009-0003-9334-3079, Aprendizaje automatizado, XAI, Evaluación de XAI

Esta revista se publica bajo una Licencia Creative Commons Atribución-No Comercial-Sin Derivar 4.0 Internacional