



Pontificia Universidad
JAVERIANA
Cali

**SEGMENTACIÓN AUTOMÁTICA, CON TÉCNICAS DE APRENDIZAJE
PROFUNDO, DE IMÁGENES DE ULTRASONIDO PARA DETECTAR ESTRUCTURAS
NERVIOSAS**

Programa de Maestría en Ingeniería

Presentado por:

LAURA ALEJANDRA BEDOYA MERA
RUBEN DARIO ESCANDON GUZMAN

Dirigido por:

JULIAN GIL GONZALES (CON TÍTULO, Ph.D. EN INGENIERÍA)

Pontificia Universidad Javeriana Cali
Facultad de Ingeniería y Ciencias
Diciembre de 2025

Segmentación automática, con técnicas de aprendizaje profundo, de imágenes de ultrasonido para detectar estructuras nerviosas.

Laura Alejandra Bedoya Mera, Ruben Dario Escandon Guzman

laubedo25@javerianacali.edu.co, ruben.escandon@javerianacali.edu.co

Pontificia Universidad Javeriana, Facultad de Ingeniería, Cali, Colombia

Resumen

La ecografía es una herramienta esencial en la anestesia regional para identificar estructuras como el nervio del plexo braquial; sin embargo, el bajo contraste, el ruido y la variabilidad dificulta su localización precisa. Este estudio propone un método de segmentación automática basado en aprendizaje profundo, utilizando un conjunto de datos y aplicando el algoritmo CLAHE para mejorar el contraste de las imágenes, además de depurar aquellas sin región de interés. Se entrenaron las arquitecturas U-Net, U-Net++ y FPN con y sin preprocesamiento, evaluando su desempeño mediante Dice e IoU. Los resultados evidencian que U-Net obtuvo el mejor rendimiento en ambos escenarios, alcanzando un Dice de 0.6704 e IoU de 0.5042 con preprocesamiento, y manteniendo desempeño superior sin él. La comparación visual confirma que U-Net produce segmentaciones más definidas y anatómicamente coherentes. Estos hallazgos destacan la importancia del preprocesamiento y consolidan a U-Net como una arquitectura robusta para la segmentación automática del plexo braquial en imágenes de ultrasonido, con potencial para apoyar procedimientos de anestesia regional.

Palabras claves: Segmentación automática, imágenes de ultrasonido, plexo braquial, aprendizaje profundo, U-Net, U-Net++, FPN, CLAHE, ecografía en anestesia regional, redes convolucionales.

Abstract

Ultrasound imaging is an essential tool in regional anesthesia for identifying anatomical structures such as the brachial plexus; however, low contrast, noise, and operator variability hinder its precise localization. This study proposes an automatic segmentation method based on deep learning, using a public dataset and applying the CLAHE algorithm to enhance image contrast, along with filtering out masks without a region of interest. The architectures U-Net, U-Net++, and FPN were trained with and without preprocessing, and their performance was evaluated using Dice and IoU metrics. The results show that U-Net achieved the best performance in both scenarios, reaching a Dice of 0.6704 and an IoU of 0.5042 with preprocessing, while maintaining superior performance without it. The visual comparison confirms that U-Net produces more defined and anatomically coherent segmentations.

These findings highlight the importance of preprocessing and position U-Net as a robust architecture for the automatic segmentation of the brachial plexus in ultrasound images, with potential to support regional anesthesia procedures.

Keywords: Automatic segmentation, ultrasound imaging, brachial plexus, deep learning, U-Net, U-Net++, FPN, CLAHE, regional anesthesia ultrasound, convolutional neural networks.

Introducción

La administración de anestesia ya sea regional (RA) o general, es un factor crítico que depende de múltiples elementos para reducir las complicaciones y mejorar los resultados postoperatorios del paciente [1]. A diferencia de la anestesia general, que se realiza para inducir la pérdida de la conciencia del paciente, la RA tiene uno de los mayores beneficios el cual es que puede proveer control del dolor en una parte específica del cuerpo sin depender fuertemente de la medicación con opioides y sin perder la conciencia [2]. Las ventajas de la RA son notables dado que se asocia con la disminución de efectos adversos tales como náuseas, vómitos. [1-3], constipación, sedación, movilización retrasada, y depresión respiratoria [3]. Además, los pacientes sometidos a RA experimentan un menor tiempo de estancia en hospitalización y los riesgos de infección se reducen [4]. Estas características hacen que

sea uno de los métodos preferidos en comparación con la anestesia general.

Uno de los aspectos críticos para el éxito de la RA es la ubicación precisa de la estructura nerviosa objetivo (procedimiento conocido como bloqueo de nervios periféricos PNB). Diferentes estrategias se han propuesto en la literatura y en la práctica clínica para dicha localización, incluyendo: elicitación de parestesia, neuroestimulación y recientemente la utilización de imágenes de ultrasonido (US) [5]. Las dos primeras técnicas ofrecen indicios de la ubicación del nervio objetivo, sin embargo, se consideran como técnicas ciegas dado que no es posible observar todas las estructuras circundantes (vasos, arterias, aguja, etc.), lo cual puede ser perjudicial dado que al no observar la trayectoria de la aguja se puede incurrir en inyección intraneural, un evento que puede llevar a un riesgo de lesión nerviosa [6-7]. De hecho, estudios muestran cómo las técnicas tradicionales (elicitación de parestesia y neuroestimulación) ciegas exhiben una tasa de fallo que van desde el 10% a 25% para punzadas y una tasa de fallo en el bloqueo de 40% [8]. Además, las técnicas tradicionales están asociadas con lesión en nervios y envenenamiento por introducción de líquido anestésico en el torrente sanguíneo [9].

Por otro lado, en los últimos años, las técnicas de bloqueo de nervio por imágenes de ultrasonido (USG) han emergido como el método preferido para localización del nervio [8], dado a que es una técnica poco

invasiva, accesible, segura para el paciente y que permite tener una visualización en tiempo real de todas las estructuras anatómicas que intervienen en el PNB [10-11]. Por tanto, esto conlleva a reducir el uso de docenas de anestésicos locales [12]. Estudios indican una tasa de éxito excediendo el 95% para anestesia en cirugía inguinal pediátrica [8]. Otro estudio ha mostrado una tasa de éxito de 72% en bloqueo del nervio tibial con ultrasonido y 22% con tradicional aproximación [13].

No obstante, este tipo de técnica de bloqueo de nervio por imágenes US trae desafíos que deben ser abordados. Primero, las imágenes de US tienen muchos artefactos, ruido, bajo contraste, y regiones difusas. Segundo, las técnicas de localización de nervio por imágenes de US requieren anestesiólogos que no solo tengan un profundo entendimiento de la estructura anatómica sino también una amplia experiencia en el uso de las imágenes por ultrasonido. Tercero, ciertos pacientes con obesidad, edemas, y atrofia muscular, la calidad de las US son pobres, y el área del nervio es difícil de identificar [14]. Los anteriores desafíos hacen que la identificación de estructuras nerviosas en imágenes de ultrasonido sea una tarea desafiante.

Una aproximación para resolver los desafíos mencionados previamente son los modelos de segmentación automática basadas en técnicas de aprendizaje automático, específicamente aquellos basados en Deep

Learning DL, los cuales son fundamentales en el procesamiento de imágenes de US en procesos de RA [15-16]. El objetivo primordial de los modelos que utilizan DL es la identificación automática del nervio objetivo directamente en las imágenes de ultrasonido con el fin de guiar a los especialistas en procedimiento de PNB [17].

La aplicación de los modelos de DL tiene la capacidad de optimizar la interpretación de las imágenes [18]. Además, del reconocimiento exitoso de estructuras anatómicas y realizar la detección y el seguimiento automático de la estructura nerviosa [19]. Esto ayuda a reducir o incluso evitar posibles complicaciones en el procedimiento clínico.

Desde el punto de vista técnico, la identificación de nervios puede tratarse como un problema de segmentación semántica, el cual hace referencia a un problema de clasificación que se hace a nivel de píxeles [20]. Dentro de los estudios alrededor de la detección de nervios periféricos se destaca que gran parte de estos se basan en la arquitectura U-Net. Dingcheng logró segmentar el nervio del plexo braquial y hacer comparativa con varias arquitecturas de CNN, U-Net sin pre-procesamiento en las imágenes obtuvo una Intersection over Union (IoU) 64% y 68.50% con preprocesamiento. U-Net++ obtuvo un IoU de 60.9% sin pre-procesamiento y 64.3% con preprocesamiento [20]. Johan Berggreen entrenó un modelo U-Net usando 1410

imágenes de 48 participantes para lograr una segmentación del nervio femoral con una media de Intersection over Union (IoU) de 74% [13]. Kengo demostró una alta precisión en la clasificación del nervio tibial para cada posición del tobillo más del 77% y una IoU en promedio de 81% por un extracción automática con U-Net [14]. Ando demostró un IoU de 71% para U-Net y una IoU de 62% para Seg-Net con un Dice de 83% y 76% para U-Net y Seg-Net correspondientemente para la segmentación del nervio medio para diagnóstico de túnel carpiano [21]. Daoud realizó una comparativa de varios modelos para segmentar un riñón en imágenes de ultrasonido y comunicó una IoU de 76% para U-Net, una IoU 73% para SegNet y a IoU de 89% para SegFormer B5 [22].

De acuerdo con lo anterior, en este estudio se abordará la segmentación del plexo braquial, el cual es una complicada estructura anatómica [10]. Aquí se propone el desarrollo de un sistema de segmentación automática de imágenes US, basado en arquitecturas de aprendizaje profundo como U-Net y SegNet, complementadas con técnicas de explicabilidad en inteligencia artificial (XAI), específicamente Grad-CAM++ con el propósito de mejorar la interpretación y visualización de las regiones de interés identificadas por los modelos. La metodología contempla el uso de un dataset público del nervio de plexo braquial por USG, acompañado de procesos sistemático de preprocesamiento, selección

y diseño de aprendizaje profundo, así como la evaluación del desempeño mediante métricas estándares con IoU, Dice que permitan cuantificar con precisión y robustez la capacidad de los modelos de DL para segmentar y delimitar las estructuras nerviosas de interés, con el fin de contribuir con el mejoramiento y desarrollo de soluciones confiables y transparentes que fortalezcan la práctica de la RA y el manejo del dolor en el contexto de la salud moderna.

Materiales y Métodos

Aquí describen las arquitecturas de aprendizaje profundo empleadas para el entrenamiento y evaluación del modelo, así como los métodos de validación utilizados, con el fin de analizar en detalle las imágenes de ultrasonido.

1. Base de datos

En este estudio se utilizó la base de datos que contenía imágenes de US del nervio de plexo braquial, el cual fue obtenido de la plataforma Kaggle [23], empleada en investigaciones de segmentaciones de nervios periféricos [24].

El conjunto de datos está conformado por 5.635 imágenes de US en formato TIFF Figura 1 y sus correspondientes 5.635 máscaras de segmentación en formato TIFF, las cuales delimitan la región de interés (ROI) donde está el nervio objetivo Figura 2.

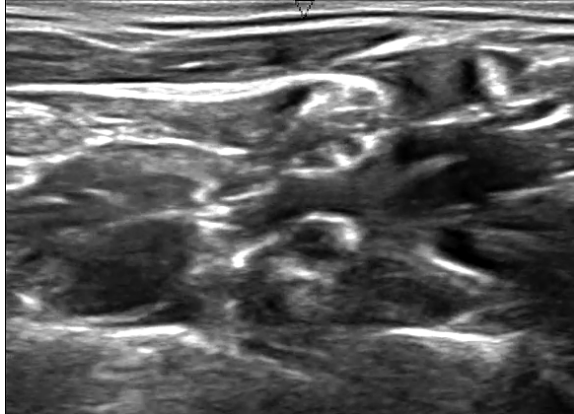


Figura 1. Imagen ecográfica que muestra la visualización del nervio del plexo braquial en un corte transversal.

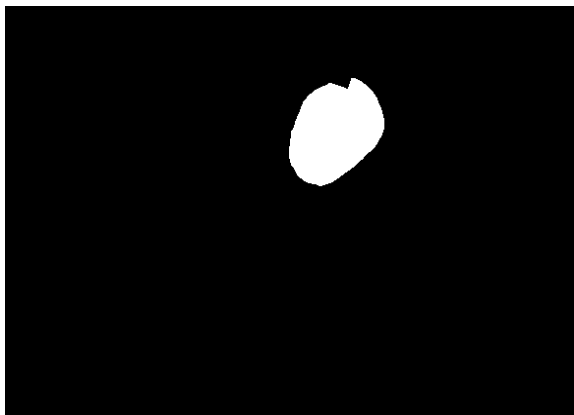


Figura 2. Máscara de segmentación que delimita la región de interés (ROI) correspondiente a la ubicación del nervio del plexo braquial en la imagen de ultrasonido.

Ambos conjuntos de imágenes poseen una resolución 420×580 píxeles (alto \times ancho) y provienen de 47 sujetos, con un rango aproximado de 115 a 580 imágenes por individuo.

Esta base de datos ha sido diseñada para evaluar y entrenar modelos de aprendizaje profundo aplicados a la segmentación automática de nervios periféricos, permitiendo validar el desempeño de diferentes arquitecturas bajo condiciones reales de viabilidad anatómica.

2. Arquitecturas para segmentación semántica

2.1 Redes convolucionales

Las redes neuronales convolucionales (CNN, por sus siglas en inglés) son un tipo especializado de red neuronal diseñado principalmente para el procesamiento y la clasificación de imágenes [25]. Estas arquitecturas destacan por su capacidad para identificar patrones espaciales —como bordes, texturas y formas— y por superar las limitaciones de las redes neuronales artificiales tradicionales (ANN), las cuales presentan dificultades al trabajar con imágenes de gran tamaño y complejidad.

Las CNN están construidas mediante bloques que integran capas de neuronas organizadas en tres dimensiones: la dimensionalidad espacial de la imagen de entrada alto y ancho y la profundidad, es decir, al volumen de características [25]. Cada bloque aplica la operación matemática denominada *convolución*, la cual permite extraer características relevantes de la imagen de manera progresiva y jerárquica.

En términos generales, una CNN está compuesta por tres tipos principales de capas: *capas convolucionales*, *capas de pooling* y *capas completamente conectadas* Figura 3 [25].

2.1.1 Capas convolucionales

Las capas convolucionales determinan la activación de cada neurona a partir de pequeñas regiones de la imagen de entrada. En ellas se aplican filtros especializados, conocidos como *kernels*, los cuales son pequeñas matrices usualmente 3x3, que mediante la operación de convolución realizan el cálculo del producto punto entre sus pesos de los kernels y la porción de la imagen y permiten aprender y detectar patrones visuales cada vez más complejos tales como bordes, texturas, y formas [25]. Adicionalmente, las capas convolucionales tienen una propiedad conocida como profundidad que se refiere al número de kernels que contienen y es una medida que determina la cantidad de mapas de características se desean aprender [25].

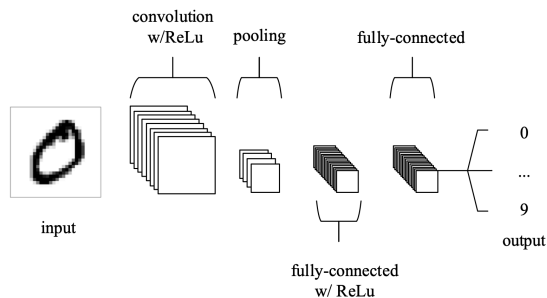


Figura 3. Una simple red convolucional de cinco capas. 1 capa de convolución, 1 capa de pooling, 2 capas completamente conectadas y 1 de salida.

2.1.2 Capa pooling

La capa *pooling* realiza un submuestreo sobre la dimensión espacial de la imagen de entrada, reduciendo su tamaño y, con ello, la cantidad de parámetros y el costo computacional. Este proceso conserva la información más relevante de cada región, contribuyendo a mejorar la eficiencia y la capacidad de generalización del modelo [25].

2.1.3 Capas completamente conectadas

Las capas completamente conectadas funcionan de manera similar a las de una red neuronal artificial tradicional. Su objetivo es integrar las características extraídas por las capas anteriores y generar los puntajes o probabilidades finales asociados a cada clase, permitiendo realizar la clasificación [25].

2.2 Pooling

El principal objetivo de las capas pooling es reducir la dimensionalidad de cada mapa de característica extraída de la imagen de entrada y que viene de las capas convolucionales, a través de ventanas conocidas como *max-pooling* y *average-pooling* [25]. En un *max-pooling* la ventana o kernel es de 2x2 y realiza un paso de (2,2) tomando el más alto dígito desde el área o ventana cubierta por el kernel, mientras el *average* toma el dígito promedio contenido en esa ventana [25].

2.3 Arquitecturas de CNNs

Las redes convolucionales han ganado amplia popularidad debido al surgimiento de numerosas arquitecturas, cada una diseñada para superar limitaciones previas y aportar mejoras significativas en el procesamiento de imágenes. En esta sección se describen algunas de las arquitecturas más representativas y relevantes para el presente estudio.

2.3.1 Arquitecturas VGG

Clasificar imágenes usando CNN puede ser hecho con diferentes arquitecturas de red, algunas de las más representativas son las Grupo de Geometría Visual (VGG) [26]. VGG fue estudiada en 2014 y demostró que el desempeño de la CNN puede ser mejorado aumentando la profundidad de la topología de la red, la arquitectura VGG usa pequeños filtros de convolución de (3x3) en vez de grandes kernels como de (5x5) [27]. VGG16 y su variante VGG19 Figura 4 son arquitecturas ampliamente usadas, ambas son similares pero se diferencian en el número de capas, VGG16 cuenta con 16 y VGG19 cuenta con 19 [26]. VGG16 demuestra ser más preciso que VGG19, por ejemplo en un estudio comparativo para clasificar múltiples enfermedades de la piel VGG16 tuvo una precisión de 89% versus 87% de VGG19 [26].

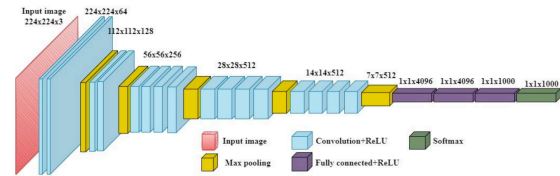


Figura 4. Representación de una arquitectura VGG-19 y que pertenece a la familia de arquitecturas de tipo VGG. Las cajas azules son las capas convolucionales, las cajas amarillas son max-pooling, las cajas púrpuras son capas completamente conectadas y la última capa es una capa soft-max [28].

2.3.2. Arquitecturas U-shaped

Este tipo de arquitectura surgió para recuperar la pérdida de información espacial causada por las operaciones realizadas por las capas de pooling [29]. Si bien la U-Net es la primera en su tipo, numerosas variantes se han desarrollado, empujando más allá de los límites de desempeño [29]. Aquí abordaremos algunas de ellas.

2.3.2.1. U-Net

La arquitectura U-Net fue desarrollada en 2015 y ha sido ampliamente usada en muchas aplicaciones de visión por computador [30-31]. La U-Net es comúnmente usada para segmentar imágenes médicas y biomédicas debido a las capacidades de entrenamiento del modelo bajo datasets con limitados tamaños [22]. La idea es suplementar la usual contracción de la red con capas

sucesivas donde las capas de pooling son reemplazadas por capas que expanden la resolución espacial de las salidas de las capas convolucionales [32].

Esta arquitectura está compuesta principalmente por dos fases, la fase de contracción y la fase de expansión. Primero, la fase de contracción logra reducir las dimensiones espaciales de la imagen de entrada mientras aprende características. Luego, la fase de expansión, expande el mapa de característica y produce un significativo mapa de segmentación usando los patrones aprendidos de la etapa previa [31]. Gracias a su característica de re-ensamble encoder-decoder, la cual es simétrica [22], termina teniendo una forma de U Figura 5 y le da honor a su nombre a este tipo de arquitectura [32].

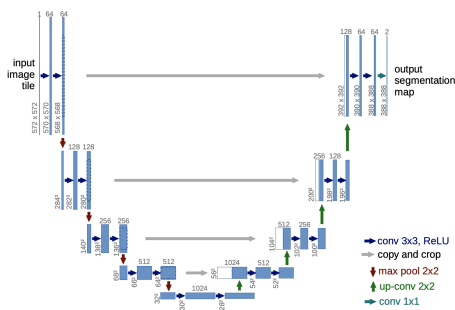


Figura 5. Arquitectura U-Net. Cada caja azul corresponde a un mapa característico multi-canal. La profundidad está descrita en la parte superior de cada caja. Las flechas describen diferentes operaciones. Las cajas blancas representan la copia de mapas de características. Los tamaños espaciales están en la esquina inferior izquierda de cada caja [32].

En detalle la U-Net en su fase de contracción sigue la típica arquitectura de una red convolucional, consiste en la repetida aplicación de bloques de 2 kernels de (3x3), cada una seguida por una función de activación ReLU y una operación de max-pooling de (2x2) con un paso de 2 para submuestrear [32].

Por otra parte, la fase de expansión consiste en bloques donde sobre-muestran la mapa característica con una convolución transpuesta de (2x2) que reduce a la mitad el número de canales, una concatenación con la correspondiente copia recortada del mapa de característica desde la fase de contracción, y dos convoluciones de (3x3). Finalmente, en el último bloque de la fase de expansión, tiene una capa convolucional (1x1) con capa de activación para reducir el número de canales al número de clases deseadas para un total de 23 capas en la red [32].

2.3.2.2 U-Net++

La arquitectura U-Net++ se construye sobre el estándar encoder-decoder diseño de la U-Net. Aquí se sustituyen las simples conexiones directas que se tienen en una U-Net con densas conexiones de bloques de capas convolucionales [22]. Además, U-Net++ integra U-Nets en diferentes profundidades dentro de la estructura, donde todas las U-Nets parcialmente comparten el mismo encoder y sus decoders están interconectados Figura 6.

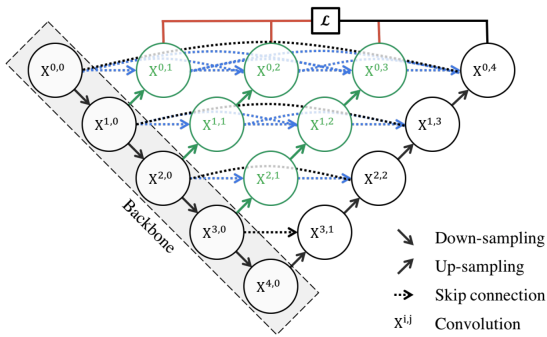


Figura 6. Arquitectura U-Net++. En color negro se indica la U-Net arquitectura original, en verde y azul se muestran los bloques densos de convolucion que salta el camino original. Lo rojo indica la supervisión profunda [33].

Por lo tanto, esto mejora la agregación de características y la representación. También, a través de una profunda supervisión, todas las U-Nets dentro de la U-Net++ son entrenadas concurrentemente, lo cual, mejora la eficiencia y el proceso de entrenamiento [22].

2.3.2.3. SegNet

Por su parte, SegNet se presenta como una arquitectura encoder-decoder para la segmentación semántica en la que el decoder utiliza los índices de pooling del encoder para reconstruir mapas de alta resolución, esto significa usar el mapa de posiciones donde el kernel de max-pooling o average-pooling obtuvo su valor [34]. Por lo tanto, esto reduce su número de parámetros requeridos para el entrenamiento y la memoria [21]. El encoder está basado en los pesos

pre-entrenados de las arquitecturas VGG16 y VGG19 con el dataset ImageNet el cual cuenta con millones de imágenes y miles de clases para clasificar. En cuanto a su arquitectura Figura 7, cuenta con un encoder de 13 capas convolucionales los cuales corresponden a las primeras 13 capas convolucionales de una red VGG [35]. Luego el decoder tiene también 13 capas. Al final del decoder tiene un clasificador soft-max que produce las clases de cada píxel independientemente [35].

2.3.2.4. Feature pyramid network FPN

FPN emplea una arquitectura asimétrica con dos flujos o fases para la detección de objetos y tareas de segmentación llamados abajo-arriba (bottom-up) y arriba-abajo (top-bottom) [36]. Este esquema de segmentación representa una forma de pirámide Figura 8.

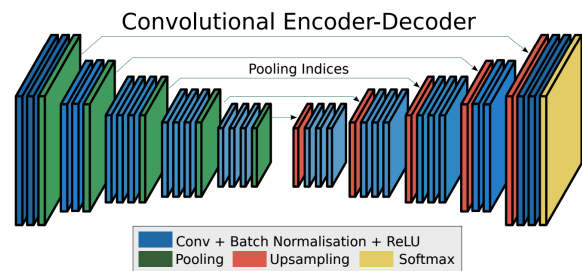


Figura 7. Una ilustración de una arquitectura Seg-Net. No hay capas completamente conectadas. Cada bloque del decoder sobre-muestra su entrada usando los índices de pooling que vienen de su respectivo bloque del encoder [35].

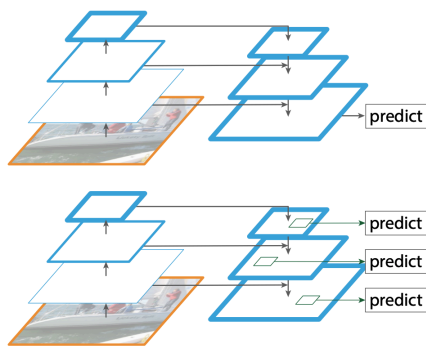


Figura 8. Ilustración de las fases “abajo-arriba” y “arriba-abajo” en una arquitectura FPN. La fase abajo-arriba es la etapa donde las características son extraídas y la resolución disminuye, mientras la fase arriba-abajo intenta recuperar la resolución espacial mientras se enriquece la información semántica a través de conexiones laterales [37].

Donde la captura semántica de una simple imagen se realiza desde todos los niveles, sin comprometer la fuerza representacional, desempeño y memoria. La fase inicial de abajo-arriba involucra procesar la imagen a través de múltiples capas convolucionales. Esto extrae características de bajo nivel, esquinas, bordes. La salida de cada capa es un mapa de características y mientras esto va subiendo su resolución espacial va disminuyendo [36].

Luego, lateralmente cada salida es subsecuentemente procesada a través de una red conexiones laterales por medio de capas convolucionales 1x1 y que combina estos mapas de características a diferentes escalas. La fase arriba-abajo juega un rol

importante en la arquitectura, esto ayuda a mantener la consistencia semántica de los mapas de características generadas por la fase abajo-arriba que son típicamente de baja resolución espacial, por lo cual, hace más difícil la localización precisa de objetos. La fase de arriba-abajo gracias a estas conexiones laterales puede enriquecer cada mapa de característica con una mayor resolución espacial, mantener enriquecido la semántica y mejorando la precisión de la detección del objeto gracias a esta retroalimentación [36].

3. Diseño de experimentos

3.1. Estudio de la base de datos

Durante esta etapa se analizaron las máscaras correspondientes del plexo braquial con el fin de verificar la presencia de ROI. En promedio las máscaras presentaron un tamaño de 116 x 91 píxeles (altura x ancho). Luego, mediante un algoritmo de detección de ROI se buscó la presencia de píxeles encendidos para terminar detectando 3.312 máscaras que no contenían dicha región segmentada. Por lo que fueron consideradas sin información anatómica relevante.

Estas máscaras junto con sus respectivas imágenes de US fueron excluidas del conjunto de datos, conservando únicamente aquellas que contenían una ROI válida. Como resultado se obtuvo un conjunto de 2.323 imágenes de US con su correspondiente máscara, el cual fue

empleado para el entrenamiento del modelo de segmentación.

En este trabajo, se aplicó un método de normalización mediante escalado lineal [30] a las imágenes de las máscaras. De esta manera, los valores de la matriz, originalmente en el rango de 0 a 255, fueron transformados al intervalo [0, 1], donde 0 representa el valor mínimo y 255 el valor máximo.

3.2. Pre-procesamiento

3.2.1. CLAHE implementación

En el siguiente proceso, se utilizó el algoritmo de ecualización de histograma adaptativo con limitación de contraste (CLAHE), una técnica que mejora la visibilidad de imágenes digitales con bajo contraste o iluminación deficiente [20–31]. Este método previene una mejora excesiva del contraste y reduce la amplificación del ruido [20]. Para su implementación, se empleó el toolkit incorporado de procesamiento de imágenes de OpenCV, configurando un “clip limit” de 2.0 y un tamaño de grilla de 8×8.

3.3. Entrenamiento del modelo

Los modelos fueron entrenados utilizando el entorno de ejecución Python 3.12.12, PyTorch, `segmentation_models.pytorch` 0.5.0. Se aplicaron algunas transformaciones de `data-augmentation` sobre el dataset por medio de

`Albumentations` 2.0.8 para incrementar el tamaño y la diversidad del conjunto de entrenamiento a través de pequeños cambios [22], tales como rotaciones ligeras y aleatorias sobre la horizontal, cambios aleatorios sobre el brillo y contraste, y normalizaciones. Esto alivia el fenómeno del sobre-entrenamiento por pequeños volúmenes de datos [20-22]. Un tamaño de lote (“batch size”) de 16. En este estudio los modelos se construyeron usando el encoder de la de una arquitectura “ResNet-34”, luego para inicializar los pesos del modelo se utilizaron los obtenidos por el entrenamiento con el dataset ImageNet. También, se configuró la entrada del modelo para recibir imágenes en escala de grises y las clases que el modelo debe segmentar a la salida son 2 los cuales son el background y la ROI (donde está ubicado el nervio periférico).

La base de datos se dividió en dos subconjuntos: 70% de las imágenes (1.626 muestras) se destinaron al entrenamiento del modelo, mientras que el 30% restante (697 muestras) se empleó para la fase de prueba. Por otro lado, el modelo fue entrenado durante 25 epochs dado que se emplearon pesos pre-entrenados, se utilizó la función de pérdida de Dice como métrica de costo, el optimizador Adam para el ajuste de los parámetros del modelo [20] y una tasa de aprendizaje (“learning rate”) de 0.001.

3.4. GradCAM implementación

En la última parte del entrenamiento del modelo se realizó el algoritmo el mapeo de activación de clase ponderado por gradiente (GradCAM) el cual ayuda a identificar tendencias o errores en el entrenamiento del modelo, tales sobre-entrenamiento sobre características irrelevantes o una incorrecta clasificación de los objetos [38]. Para su implementación se usó la librería `pytorch_grad_cam`, de cada modelo se toma la capa que capture la información espacial de las características de alto nivel, en la U-Net fue el ultimo bloque del decoder, para U-Net++, FPN, y Seg-Net fue el penúltimo bloque del decoder.

3.5. Evaluación del modelo

La evaluación del modelo se realizó utilizando Coeficiente de Similitud Dice (DSC) [39]. Es una métrica altamente empleada en el ámbito de la segmentación de imágenes médicas, por su capacidad para cuantificar el grado de similitud entre las regiones segmentadas por el modelo y las máscaras de referencia. El DSC se define matemáticamente como:

$$Dice = DSC = \frac{2|X \cap Y|}{|X| + |Y|} \quad (1)$$

donde X representa el conjunto de píxeles correspondientes a la segmentación predicha por el modelo y Y los píxeles de la segmentación real. Su valor oscila entre 0 y 1, siendo 1 el resultado ideal que indica una

superposición perfecta entre ambas máscaras.

También, en este estudio se empleó la métrica de evaluación intersección sobre unión (IoU) para evaluar el desempeño del modelo. Esto mide la similitud entre los dos conjuntos. Es una de las métricas más comúnmente usadas para medir la medición de segmentación de imágenes [20]. Definida

$$IoU = \frac{Intersection}{Union} \quad (2)$$

Donde la intersección es el solapamiento de la parte de la imagen predicha y la imagen de referencia, y la unión es la combinación de la imagen predicha y la imagen de referencia [20]. Adicionalmente, se empleó la métrica de precisión para acompañar las demás métricas y ayudar a explicar los resultados

$$Precision = \frac{TP}{TP + FP} \quad (3)$$

Donde TP se refiere a los verdaderos positivos, y FP son los falsos positivos.

Resultados

El preprocesamiento de las imágenes se efectuó mediante el algoritmo CLAHE, con el fin de optimizar el contraste local y resaltar detalles anatómicos de interés. La Figura 9 muestra un ejemplo representativo del resultado obtenido tras aplicar esta técnica.

En la Tabla 1 se presentan los valores de desempeño obtenidos al entrenar las arquitecturas U-Net, U-Net++, FPN y Seg-Net utilizando imágenes que fueron sometidas a técnicas de preprocesamiento. Bajo estas condiciones, el modelo U-Net alcanzó el mejor rendimiento, obteniendo un Dice de 0.6704 y un IoU de 0.5042 y una precisión de 79.54%, superando a U-Net++ y FPN. Los modelos U-Net++ y FPN mostraron valores inferiores, con un Dice de 0.6019 y 0.6188, respectivamente, y un IoU de 0.4305 y 0.4480, manteniendo un comportamiento similar entre ellos pero por debajo de U-Net. Finalmente, en el último lugar Seg-Net obtuvo un Dice de 0.5125 y un IoU de 0.3442 muy por debajo de otros modelos.

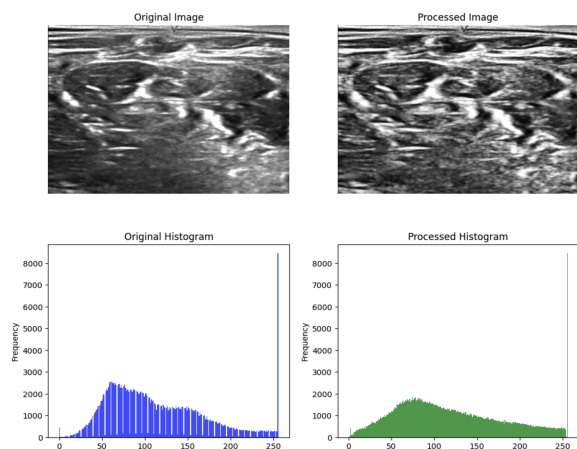


Figura 9. Aplicación del algoritmo CLAHE: a la izquierda se muestra la imagen de ultrasonido junto con su histograma original; a la derecha se presenta la imagen resultante tras el preprocesamiento y el histograma corregido con aumento de contraste local.

Por su parte, la Tabla 2 presenta los resultados del entrenamiento realizado sin aplicar técnicas de preprocesamiento a las imágenes. Aunque la U-Net continúa mostrando el mejor desempeño general, sus métricas (Dice = 0.5333; IoU = 0.3636) evidencian disminución respecto al escenario con preprocesamiento. En el caso de U-Net++, el rendimiento también se ve afectado, alcanzando un Dice de 0.4746 y un IoU de 0.3403. Por los resultados obtenidos en la Tabla 1 sugiere que la etapa de preprocesamiento puede resaltar información relevante para la identificación de estructuras nerviosas y ayudar a los modelos en su segmentación.

Finalmente, FPN y Seg-Net son (Dice = 0.4977; IoU = 0.3310) y (Dice = 0.4933 y IoU = 0.3366) respectivamente, lo que indica una menor variabilidad para la Seg-Net ante cambios en la calidad de las imágenes de entrada. En conjunto, los resultados muestran que el preprocesamiento contribuye a mejorar el desempeño de las arquitecturas, especialmente en el caso de U-Net, que presenta el comportamiento más estable y superior en ambas métricas.

Tabla 1. Desempeño de los modelos entrenados utilizando imágenes con técnicas de preprocesamiento aplicadas.

Con pre-procesamiento			
Modelo	Dice	IoU	Precisión
U-Net	0.6704	0.5042	0.7954
U-Net++	0.6019	0.4305	0.7410
FPN	0.6188	0.4480	0.7165
Seg-Net	0.5125	0.3442	0.8152

Tabla 2. Desempeño de los modelos entrenados utilizando imágenes sin aplicación de técnicas de preprocesamiento.

Sin pre-procesamiento			
Modelo	Dice	IoU	Precisión
U-Net	0.5333	0.3636	0.7556
U-Net++	0.4746	0.3403	0.7400
FPN	0.4977	0.3310	0.7429
Seg-Net	0.4933	0.3366	0.7312

Discusión

En la Figura 10 se observa la comparación visual entre la imagen original, la máscara real y la predicción generada por cada modelo. El modelo U-Net logró una segmentación más ajustada a la forma y localización del nervio, mostrando bordes bien definidos y una relativa coincidencia con la máscara real. En contraste, U-Net++ presentó una ligera desviación en tamaño de la región segmentada, lo que sugiere una menor precisión espacial. Por su parte, el modelo FPN evidenció un desempeño

intermedio, y la Seg-Net se comportó aceptable con leves diferencias respecto a la máscara de referencia.

Finalmente, los resultados predichos por los modelos se confirman con herramientas de interpretabilidad como GradCAM Figura 11, U-Net ofrece una mejor capacidad de generalización y una delineación más precisa del nervio en las imágenes US, en comparación a U-Net++, FPN, y Seg-Net, son menos delineadas o precisas, lo que coincide con las métricas cuantitativas presentadas previamente.

En cuanto a los valores de Dice e IoU, estos difieren un poco comparado en estudios previos [20]. En particular, la métrica IoU se encuentra aproximadamente 18 puntos por debajo del promedio alcanzado por otros autores. Esta diferencia pudo deberse a dos relevantes configuraciones comparadas con estudios previos. Primero, la configuración del encoder utilizado en este estudio fue “ResNet-34” mientras que en estudio previo emplearon VGG como arquitectura para el encoder [20]. Segundo, la ausencia de un segundo dataset para hacer transferencia de aprendizaje. Comparado con estudio previo en el cual emplearon el dataset ImageNet + un dataset de nervio femoral [20], esto pudo haber contribuido en este resultado.

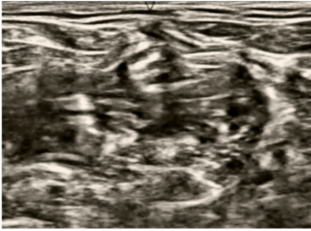


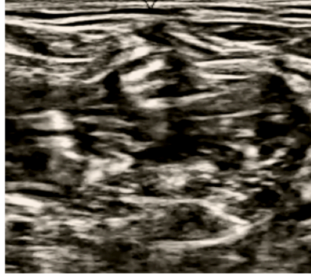
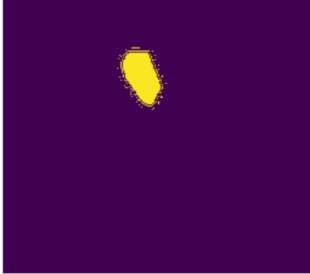

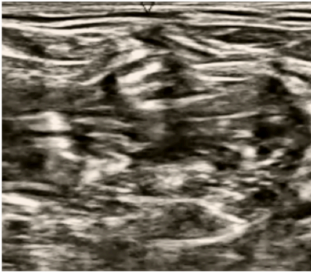


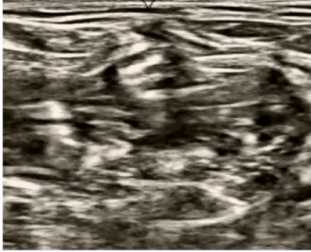
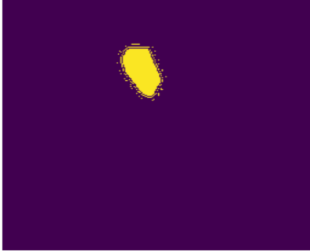

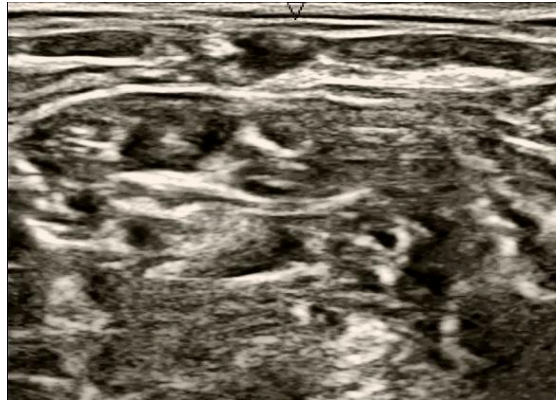
Modelo	Imagen original	Máscara Real	Predicción del Modelo
U-Net			
U-Net++			
FPN			
SegNet			

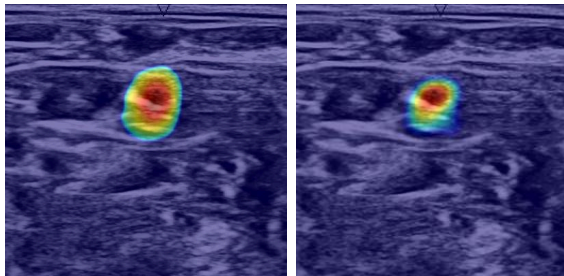
Figura 10. Comparación visual entre las segmentaciones generadas por los modelos U-Net, U-Net++, FPN y Seg-Net frente a las máscaras de referencia en imágenes de ultrasonido.

El uso de la métrica Dice permitió evaluar de manera precisa la capacidad del modelo para delinear las estructuras anatómicas del nervio en las imágenes ecográficas. Un valor elevado de esta métrica refleja una segmentación más coherente con los contornos reales, aspecto crucial para aplicaciones clínicas como la anestesia regional y el diagnóstico por imágenes.

Finalmente, los resultados evidencian que el modelo alcanza una alta concordancia espacial con las regiones de interés (ROI), demostrando la efectividad del entrenamiento y la capacidad de generalización del modelo frente a variaciones de textura, brillo y forma presentes en las imágenes ecográficas.

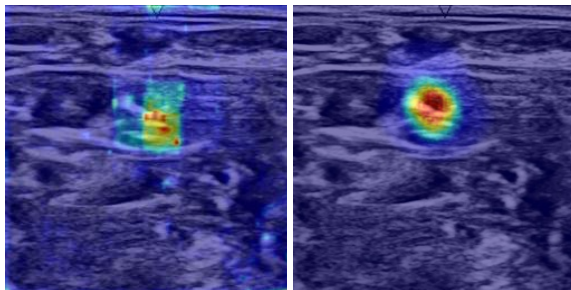


(a)



(b)

(c)



(d)

(e)

Figura 11. Visualización Grad-CAM++ generada a partir de la predicción de los modelos entrenados en este estudio, (a) imagen original, (b) U-Net, (c) U-Net++, (d) FCN. y (e) Seg-Net Donde se destacan las regiones de mayor activación utilizadas por la red para identificar la ubicación del nervio en la imagen de ultrasonido.

Conclusión

En este trabajo se desarrolló e implementó un sistema de segmentación automática del nervio del plexo braquial en imágenes de ultrasonido, empleando arquitecturas de aprendizaje profundo como U-Net, U-Net++, FPN, y Seg-Net. Los resultados experimentales demostraron que el modelo U-Net obtuvo el mejor desempeño tanto en las métricas cuantitativas (Dice = 0.6704, IoU = 0.5042, precisión de 79.54%) como en la evaluación visual, mostrando una delimitación más precisa y coherente con las estructuras anatómicas reales.

La aplicación de técnicas de preprocesamiento, especialmente la normalización lineal y la ecualización adaptativa de histograma (CLAHE), contribuyó a mejorar la calidad de las imágenes y optimizar la capacidad de detección de las regiones de interés. Sin embargo, los valores obtenidos de Dice e IoU se ubicaron por debajo de los reportados en estudios previos, posiblemente debido a la elección del encoder "ResNet-34" y la ausencia de un dataset extra para ser ajustado específicamente para este tipo de imágenes médicas, y el transfer learning no logró una completa adaptación al dominio ecográfico.

A pesar de estas limitaciones, los resultados obtenidos evidencian la viabilidad del uso de modelos de segmentación profunda en la detección y análisis del nervio del plexo braquial, constituyendo una base sólida

para el desarrollo de herramientas de apoyo en anestesia regional y diagnóstico clínico asistido por inteligencia artificial.

Como trabajo futuro, se propone incrementar el tamaño y la diversidad del conjunto de datos, examinar diferentes profundidades en la red, así como explorar otros encoders especializados en imágenes médicas como VGG, DenseNet, MobileNet MixVision Transformer y sus variantes, aplicar transferencia de aprendizaje para mejorar el desempeño y técnicas avanzadas de ajuste fino (fine-tuning) que permitan mejorar la precisión y robustez del modelo ante variaciones en textura, ruido y contraste en las imágenes de ultrasonido.

Disponibilidad de código

Los algoritmos y código usado para este estudio están abiertamente disponibles y pueden ser encontrados aquí <https://drive.google.com/drive/folders/14sYWigg0wL8kiwCxjZABiEUETMRI-r4J?usp=sharing>

Referencias

- [1] L. Mason, T. Jackman. Comparative Analysis of Regional versus General Anesthesia. *Journal of Student Research*. 9. 2. 2020. 6 páginas.
- [2] Y.-Y. K. Chen, K. A. Boden, K. L. Schreiber. The role of regional anaesthesia and multimodal analgesia in the prevention of chronic postoperative pain: a narrative review. *Anaesthesia*. 76. S1. 2021. 11 páginas.
- [3] D. Marianello, F. Ginetti, F. Sanfilippo, C. Biuzzi, C. Catelli, E. Modica, F. Silva, A. Cartocci, L. Luzzi, R. Corzani, P. Paladini, S. Scolletta, F. Franchi. *Clinical Medicine*. 14. 16. 2025. 12 páginas.
- [4] J. L. Levene, E. J. Weinstein, M. S. Cohen, D. A. Andrae, J. Y. Chao, M. Johnson, C. B. Hall, M. H. Andrae. Local Anesthetics and Regional Anesthesia versus Conventional Analgesia for Preventing Persistent Postoperative Pain in Adults and Children: A Cochrane Systematic Review and Meta-analysis Update. *J Clin Anesth*. 55. 2019. 28 páginas.
- [5] W. Zehao, Y. Wang. Development of Guidance Techniques for Regional Anesthesia: Past, Present and Future. *Journal of Pain Research* 2021. 14. 1631 - 1641.
- [6] J. Berggreen, A. Johansson, J. Jahr, S. Möller and T. Jansson. Deep Learning on Ultrasound Images Visualizes the Femoral Nerve with Good Precision. *Healthcare* 2023, 11, 184. 9 páginas.
- [7] L. Mirea, A.-M. Dumitriu, C. Cobilinschi, R. Ene, R. Ungureanu. Integrating Peripheral Nerve Blocks in Multiple Trauma Care: Current Evidence and Clinical Challenges. *Clinical Medicine*. 14. 15. 2025. 13 páginas.
- [8] P. Li, W. Tang. Efficacy of Ultrasound-Guided Iliohypogastric and Ilioinguinal Nerve Block for Anesthesia in Pediatric Inguinal Surgery. *Journal of Pain Research*. 18. 2025. 1891-1899.
- [9] K. R. Gujarati, L. Bathala, V. Venkatesh, R. S. Mathew, P. K. Yalavarthy.

Transformer-Based Automated Segmentation of the Median Nerve in Ultrasound Videos of Wrist-to-Elbow Region. *IEEE Transactions On Ultrasonics, Ferroelectrics, And Frequency Control*. 71. 1. 2023. 56-69.

[10] G. Pacella, R. Natella, F. Bruno, M. Bruno, D. Franco, D. Giuseppe Romano, M. Zappia. Brachial Plexopathies: A Comprehensive Radiologic Method Integrating Ultrasound and MRI. *Clinical Medicine*. 14. 17. 2025. 17 páginas.

[11] L. C. Thomas, S.K. Graham, Kristie D. Osteen, H. Scuderi Porter, B. D. Nossaman. Comparison of Ultrasound and Nerve Stimulation Techniques for Interscalene Brachial Plexus Block for Shoulder Surgery in a Residency Training Environment: A Randomized, Controlled, Observer-Blinded Trial. *The Ochsner Journal* 11:246–252, 2011. 7 páginas.

[12] H. Akelma, E. Çelik, Y. Ipek, M. Ali Turgut, M. Rasit Tanırcan, A. Aktan, M. Z. Karahan. Ultrasound-Guided Regional Anesthesia in Permanent Pacemaker Implantation: An Observational Study. *Medicina*. 61. 6. 2025. 13 páginas.

[13] J. Berggreen, A. Johansson, J. Jahr, S. Möller, T. Jansson. Deep Learning on Ultrasound Images Visualizes the Femoral Nerve with Good Precision. *Healthcare*. 11. 2. 2023. 9 páginas.

[14] K. Kawanishi, A. Kakimoto, K. Anegawa, M. Tsutsumi, I. Yamaguchi, S. Kudo. Automatic Identification of Ultrasound Images of the Tibial Nerve in Different Ankle Positions Using Deep Learning. *Sensors*. 23. 10. 2023. 9 páginas.

[15] V. Bernard, O Maupain, T. Giral, J. Bowness. L. Mercadal. Evaluation of AI-based nerve segmentation on ultrasound: relevance of standard metrics in the clinical setting. *British Journal of Anaesthesia*. 134. 5. 2025. 1497 - 1502.

[16] D. Viderman, M. Dossov, S. Seitenov. M. Lee. Artificial intelligence in ultrasound-guided regional anesthesia: A scoping review. *Front. Med*. 9:994805. 11 páginas

[17] E. D. Pelletier, S. D. Jeffrie, N. Suissa, I. Sarty, N. Malka, K. Song, A. Sinha, T. M. Hemmerling. From Consensus to Standardization: Evaluating Deep Learning for Nerve Block Segmentation in Ultrasound Imaging. *A & A Practice* 19(8):p e02040, August 2025. 8 páginas.

[18] N. Ca, G. Wang, L. Xu, Y. Zhou, H. Chong, Y. Zhao, J. Wang, W. Yan, B. Zhang. N.Liu. Examining the impact perceptual learning artificial-intelligence-based on the incidence of paresthesia when performing the ultrasound-guided popliteal sciatic block: simulation-based randomized study. *BMC Anesthesiol* 22, 392 (2022). 9 páginas.

[19] J. S. Bowness, R. Morse, O. Lewis, J. Lloyd, D. Burckett-St Laurent, B. Bellew, A. J.R. Macfarlane, A. Pawa, A. Taylor, J. A. Noble, H. Higham. Variability between human experts and artificial intelligence in identification of anatomical structures by ultrasound in regional anaesthesia: a framework for evaluation of assistive artificial intelligence. *British Journal of Anaesthesia*, 132 (5): 1063e1072 (2024). 10 páginas.

- [20] D. Tian, B. Zhu, J. Wang, L. Kong, B. Gao, Y. Wang, D. Xu, R. Zhang, Y. Yao. Brachial Plexus Nerve Trunk Recognition From Ultrasound Images: A Comparative Study of Deep Learning Models. IEEE Engineering In Medicine And Biology Society Section. 10. 2022. 82003 - 82014.
- [21] S. Ando, P. Y. Loh. Convolutional Neural Network Approaches in Median Nerve Morphological Assessment from Ultrasound Images. Journal of Imaging. 10. 13. 2024. 16 páginas.
- [22] M. I. Daoud, F. Abunameh, K. Shweikeh, S. K. Alzamer, M. Z. Ali, R. Alazrai. A Comparative Study of Deep Learning Semantic Segmentation Models for Kidney Segmentation in Ultrasound Images Using the Open Kidney Ultrasound Dataset. IEEE Access. 13. 2025. 144417-144433
- [23] Kaggle. Ultrasound Nerve Segmentation. 2016. Available online: <https://www.kaggle.com/c/ultrasound-nerve-segmentation/data> (acesado 21 October 2025).
- [24] C. A. Jimenez-Castaño, A. M. Álvarez-Meza, O. D. Aguirre-Ospina, D. A. Cárdenas-Peña, Á. A. Orozco-Gutiérrez. Random Fourier Features-Based Deep Learning Improvement with Class Activation Interpretability for Nerve Structure Segmentation. Sensors. 21. 22. 2021. 19 páginas.
- [25] K. O'Shea. R. Nash. An Introduction to Convolutional Neural Networks. Neural and Evolutionary Computing. 2015. 10 paginas.
- [26] R. Puspitaa, A. Elok Amaliaa, C. Rahayu. A comparative evaluation of CNN architectures for skin disease classification: VGG16 and VGG19. Procedia Computer Science. 269. 2025. 331-339.
- [27] X. Zhao, L. Wang, Y. Zhang, X. Han, M. Deveci, M. Parmar. A review of convolutional neural networks in computer vision. Artificial Intelligence Review. 57. 2024. 47 páginas.
- [28] T.-Hai Nguyen, T.-Nghia Nguyen, B.-Viet Ngo. A VGG-19 Model with Transfer Learning and Image Segmentation for Classification of Tomato Leaf Disease. AgriEngineering. 4. 4. 2022. 871-887.
- [29] B. Wang, H. Shi, Z. Zhao, S. Zhang. PSCT-Net: A parallel symmetric CNN-transformer hybrid network for medical image segmentation. Medical Engineering and Physics. 146. 2025. 13 páginas.
- [30] R. Zannah, M. Bashar, R. B. Mushfiq, A. Chakrabarty, S. Hossain, Y. J. Jung. Semantic Segmentation on Panoramic Dental X-Ray Images Using U-Net Architectures. IEEE Access. 12. 2024. 44598-44612.
- [31] Y. Shen, J. Li, W. Zhu, K. Yu, M. Wang, Y. Peng, Y. Zhou, L. Guan, X. Chen. Graph Attention U-Net for Retinal Layer Surface Detection and Choroid Neovascularization Segmentation in OCT Images. IEEE Transactions On Medical Imaging. 42. 11. 2023. 3140-3154.
- [32] O. Ronneberger, P. Fischer, T. Brox. U-Net: Convolutional Networks for Biomedical Image Segmentation. Computer Vision and Pattern Recognition. 2015. 8 páginas.
- [33] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, J. Liang. UNet++: A Nested U-Net Architecture for Medical Image

Segmentation. Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support. DLMIA ML-CDS 2018 2018. Lecture Notes in Computer Science. 11045. 2018. 3-11.

[34] V. Badrinarayanan, A. Kendall, R. Cipolla. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. Senior Member, IEEE 2016. 14 páginas.

[35] V. Badrinarayanan, A. Kendall, R. Cipolla. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. Computer Vision and Pattern Recognition. 2022. 14 páginas.

[36] H. Sharen, M. Jawahar, L. Jani Anbarasi, V. Ravi, N. Saleh Alghamdi, W. Suliman. FDUM-Net: An enhanced FPN and U-Net architecture for skin lesion segmentation. Biomedical Signal Processing and Control. 91. 2024. 15 páginas.

[37] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, S. Belongie. Feature Pyramid Networks for Object Detection. Computer Vision and Pattern Recognition. 2016. 10 páginas.

[38] S. N. Almuayqil, S. A. El-Ghany, A. A. Abd El-Aziz, Mohammed Elmogy. KidneyNet: A Novel CNN-Based Technique for the Automated Diagnosis of Chronic Kidney Diseases (CKD) from CT Scans. Electronics. 13. 24. 2024. 37 páginas.

[39] S. Davar, T. Fevens. Enhanced U-Net Architecture for Brain Tumor Localization and Segmentation in T1-Weighted M. IEEE Transactions On Circuits And Systems. 72. 8. 2025. 993-997.