



**Escuela Superior
de Ingeniería y Tecnología**
Universidad de La Laguna

Trabajo de Fin de Grado

Grado en Ingeniería Informática

Aplicación de Inteligencia Artificial para la segmentación de imágenes médicas de la columna vertebral

*Artificial Intelligence Application for medical image segmentation of
the spine*

Tania Évora Vargas Martínez

La Laguna, 14 de junio de 2025

D. **Rafael Arnay del Arco**, profesor Titular de Universidad adscrito al Departamento de Ingeniería Informática y Sistemas de la Universidad de La Laguna, como tutor

C E R T I F I C A

Que la presente memoria titulada:

“Aplicación de Inteligencia Artificial para la segmentación de imágenes médicas de la columna vertebral”

ha sido realizada bajo su dirección por Dña **Tania Évora Vargas Martínez**.

Y para que así conste, en cumplimiento de la legislación vigente y a los efectos oportunos, firman la presente en La Laguna a 14 de junio de 2025

Agradecimientos

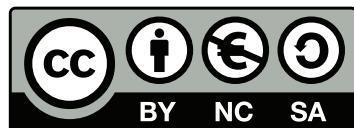
Quiero agradecer a mi tutor, Rafael Arnay del Arco, por su guía y paciencia durante el desarrollo del proyecto. Su experiencia y consejos han sido de gran ayuda para avanzar en cada fase del trabajo. Además, le agradezco especialmente por haber propuesto este tema que me ha permitido explorar una línea que me emociona mucho, como es la aplicación de la inteligencia artificial en el ámbito médico.

También quiero agradecer al Dr. Julio Manuel Plata Bello, por su colaboración y tiempo. Sus aportaciones clínicas y las explicaciones sobre las métricas utilizadas han sido clave para definir los objetivos del trabajo, además de aportar una visión profesional desde la práctica médica. Agradezco también el material proporcionado para la validación externa, que me permitió realizar más pruebas y mejorar aspectos del proyecto. Aunque nuestra comunicación fue limitada debido a la carga de trabajo, valoro profundamente su implicación en el proyecto.

A mis compañeros y amigos, gracias por el apoyo, las discusiones constructivas y la motivación, que hicieron todo el proceso más llevaderos.

Finalmente, a mi familia, por su apoyo incondicional y su comprensión durante toda mi etapa universitaria. Su confianza y ánimo han sido esenciales para llegar hasta aquí.

Licencia



© Esta obra está bajo una licencia de Creative Commons
Reconocimiento-NoComercial-Compartirlgual 4.0
Internacional.

Resumen

Este trabajo aborda la segmentación de la columna vertebral en radiografías laterales, con el objetivo de facilitar la extracción de métricas clínicas relevantes, como medidas angulares y translacionales. Para ello, se desarrolla un sistema basado en inteligencia artificial y aprendizaje profundo, y se evalúan tres enfoques: un modelo de segmentación general y dos modelos especializados por región anatómica.

El entrenamiento se realizó con el conjunto de datos NHANES II, complementando las segmentaciones obtenidas con técnicas de posprocesado para mejorar la precisión. Los resultados muestran que, si bien el modelo general ofrece un rendimiento aceptable en la mayoría de los casos, los modelos especializados ofrecen mejor rendimiento en áreas morfológicamente más complejas, como la región cervical, y que el posprocesado es clave para obtener métricas más precisas.

Palabras clave: Segmentación de imagen médica, columna vertebral, aprendizaje profundo, radiografía, métricas clínicas.

Abstract

This project approaches the segmentation of the spine in lateral radiographs, with the aim of facilitating the extraction of relevant clinical metrics, such as angular and translational measurements. To this end, an artificial intelligence and deep learning-based system is developed, and three approaches are evaluated: a general segmentation model and two region-specific models based on anatomical location.

The training was conducted using the NHANES II dataset, and the obtained segmentations were refined through post-processing techniques to improve accuracy. The results show that, while the general model performs reasonably well in most cases, the specialized models achieve better performance in anatomically complex regions, such as the cervical spine. Moreover, post-processing plays a key role in obtaining more precise clinical measurements.

Keywords: Medical image segmentation, spine, deep learning, radiography, clinical metrics.

Índice general

Capítulo 1 Introducción	1
1.1 Explicación del problema médico	1
1.2 Objetivos	1
1.3 Antecedentes y trabajos relacionados	3
1.4 Estructura de la memoria	3
Capítulo 2 Materiales y métodos	4
2.1 Materiales	4
2.1.1 Conjunto de datos NHANES II	4
2.1.2 Conjunto de pruebas en condiciones reales	5
2.2 Metodología	6
2.2.1 Fase 1: Entrenamiento de los modelos de segmentación	6
2.2.2 Fase 2: Posprocesado de las segmentaciones	9
2.2.3 Fase 3: Cálculo de métricas morfológicas clínicas	12
Capítulo 3 Resultados	16
3.1 Rendimiento de la segmentación	16
3.1.1 Resultados cuantitativos	16
3.1.2 Resultados cualitativos	17
3.2 Calidad de las métricas clínicas	21
3.2.1 Resultados cuantitativos	21
3.2.2 Resultados cualitativos	26
3.3 Análisis general	34
3.3.1 Limitaciones y casos problemáticos	35
3.3.2 Potencial aplicación médica	35
Capítulo 4 Conclusiones y líneas futuras	37
4.1 Limitaciones del estudio	37
4.2 Líneas futuras de desarrollo	38
Capítulo 5 Summary and Conclusions	39
5.1 Study Limitations	39
5.2 Future Work	40
Capítulo 6 Presupuesto	41

Índice de figuras

1.1	Esquemas anatómico de la columna vertebral en vistas anterior, posterior y lateral, destacando las regiones y elementos clave como las etiquetas vertebrales [1].	2
1.2	Esquemas ilustrativos de las medidas morfológicas clínicas: (Izquierda) Medida angular. (Derecha) Medida traslacional.	2
2.1	Imágenes del conjunto NHANES II. (Izquierda) Región cervical. (Derecha) Región lumbar.	5
2.2	Arquitectura original de U-Net propuesta por Ronneberger et al., utilizada ampliamente en tareas de segmentación biomédica.	7
2.3	Flujo visual del posprocesado para una imagen de la región cervical.	11
3.1	Segmentación con el modelo general en imagen lumbar del conjunto NHANES II.	18
3.2	Segmentación con el modelo general en imagen cervical del conjunto NHANES II.	18
3.3	Segmentación con el modelo general en imagen lumbar del conjunto clínico.	19
3.4	Segmentación con el modelo general en imagen cervical del conjunto clínico.	20
3.5	Comparación de segmentación en una imagen cervical del conjunto NHANES II	20
3.6	Comparación de segmentación en una imagen lumbar del conjunto NHANES II	21
3.7	Histograma de errores angulares - modelo general por región.	23
3.8	Histograma de errores angulares - modelo especializado en cervicales	23
3.9	Histograma de errores angulares - modelo especializado en lumbares	24
3.10	Histograma de errores traslacionales - modelo general por región.	25
3.11	Histograma de errores traslacionales - modelo cervical por región.	26
3.12	Histograma de errores traslacionales - modelo lumbar por región.	27
3.13	Medidas angulares en región cervical. (Izquierda) Original. (Derecha) Modelo.	28
3.14	Medidas angulares en región lumbar. (Izquierda) Original. (Derecha) Modelo.	28
3.15	Medidas traslacionales en región cervical. (Izquierda) Original. (Derecha) Modelo.	29
3.16	Medidas traslacionales en región lumbar. (Izquierda) Original. (Derecha) Modelo.	30
3.17	Medidas angulares en región cervical.	31
3.18	Medidas angulares en región lumbar.	31
3.19	Medidas traslacionales en región cervical.	31
3.20	Medidas traslacionales en región lumbar.	31
3.21	Comparación de modelos sobre cervicales - (Izquierda) Modelo general. (Derecha) Modelo especializado.	32
3.22	Comparación de modelos sobre lumbares - (Izquierda) Modelo general. (Derecha) Modelo especializado.	33
3.23	Comparación de modelos sobre cervicales - (Izquierda) Modelo general. (Derecha) Modelo especializado.	33
3.24	Comparación de modelos sobre lumbares - (Izquierda) Modelo general. (Derecha) Modelo especializado.	34

3.25 Ejemplo de selección errónea del grupo candidato a columna debido a una alineación geométrica más favorable	36
3.26 Ejemplo en el que el posprocesado no identifica ningún grupo válido por no cumplirse los criterios mínimos de selección.	36

Índice de tablas

3.1	Rendimiento sobre el conjunto de test (10 % del conjunto NHANES II), antes y después del posprocesado.	16
3.2	Comparación de medidas angulares en la región cervical (conjunto NHANES II) . . .	22
3.3	Comparación de medidas angulares en la región lumbar (conjunto NHANES II) . . .	22
3.4	Comparación de medidas traslacionales en la región cervical (conjunto NHANES II) .	24
3.5	Comparación de medidas traslacionales en la región lumbar (conjunto NHANES II) .	25
6.1	Presupuesto de Equipos y Licencias	41
6.2	Coste de Mano de Obra	41
6.3	Coste Total del Proyecto	41

Capítulo 1

Introducción

Las afecciones de la columna vertebral representan un problema de salud importante a nivel mundial, afectando la calidad de vida de millones de personas. Estas patologías pueden derivar en dolor crónico, limitaciones de movilidad e incluso discapacidades graves si no se detectan y tratan a tiempo.

En el contexto del análisis de imágenes de la columna vertebral, las técnicas de inteligencia artificial, especialmente aquellas basadas en aprendizaje profundo, han automatizado el proceso, abriendo nuevas posibilidades para mejorar la precisión y la eficiencia en la segmentación de estructuras anatómicas.

1.1. Explicación del problema médico

La identificación precisa de vértebras individuales se aplica en el diagnóstico de diversas patologías vertebrales, como escoliosis, fracturas o desplazamientos vertebrales. Para el diagnóstico y seguimiento de estas afecciones, se emplean comúnmente las radiografías de columna; sin embargo, la segmentación manual y el cálculo de métricas morfológicas son procesos lentos, complejos y propensos a errores.

La implementación de métodos automáticos de segmentación no solo mejora la consistencia y reduce el tiempo de análisis, sino que también proporciona información objetiva que facilita el diagnóstico y seguimiento clínico. Para una mejor comprensión de las regiones anatómicas y los elementos de la columna a los que se hace referencia en este trabajo, la Figura 1.1 presenta un esquema detallado de sus diferentes secciones y estructuras.

1.2. Objetivos

El objetivo principal de este Trabajo de Fin de Grado fue el desarrollo de un sistema capaz de extraer métricas clínicas morfológicas a partir de imágenes radiográficas de la columna vertebral. Estas métricas permiten cuantificar la alineación y el desplazamiento entre vértebras, y son de utilidad en contextos de diagnóstico médico. Se consideraron dos tipos de métricas principales, ilustradas en la Figura 1.2

- **Métricas angulares:** estiman la orientación relativa entre vértebras consecutivas, calculando el ángulo entre sus ejes principales.

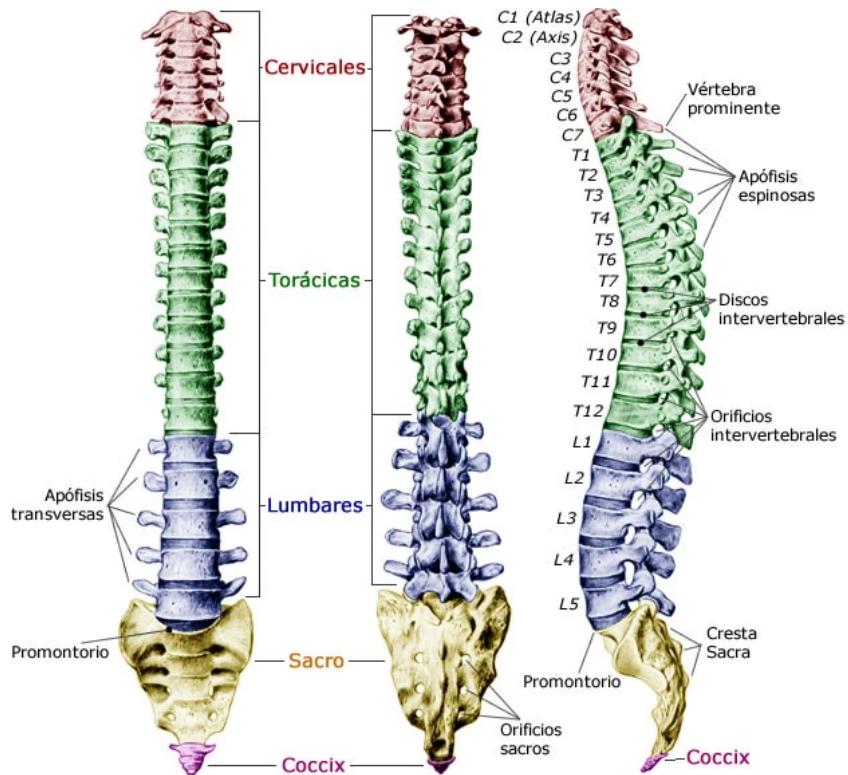


Figura 1.1: Esquemas anatómico de la columna vertebral en vistas anterior, posterior y lateral, destacando las regiones y elementos clave como las etiquetas vertebrales [1].

- **Métricas translacionales:** cuantifican el desplazamiento lateral entre vértebras consecutivas, proyectando cada vértebra sobre un eje intermedio que representa la dirección conjunta entre ambas.

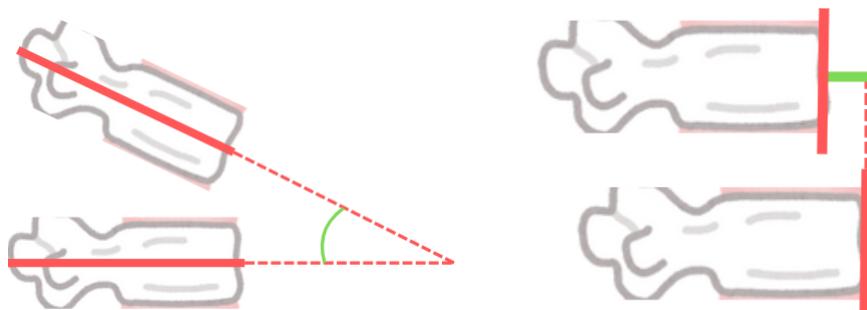


Figura 1.2: Esquemas ilustrativos de las medidas morfológicas clínicas: (Izquierda) Medida angular. (Derecha) Medida translacional.

Para alcanzar este objetivo general, se establecieron dos objetivos específicos:

- **Segmentación vertebral:** diseñar y entrenar modelos de redes neuronales convolucionales capaces de segmentar con precisión las vértebras en radiografías.
- **Posprocesado de segmentaciones:** aplicar técnicas de posprocesado con el fin de mejorar la calidad de las segmentaciones y facilitar el cálculo de las métricas clínicas.

1.3. Antecedentes y trabajos relacionados

Diversos enfoques basados en el aprendizaje profundo han abordado la segmentación de imágenes médicas. Yao Chen y colaboradores plantean un modelo llamado VertXNet [2], que combina las arquitecturas U-Net y Mask R-CNN, y mejora así la precisión en la segmentación de cuerpos vertebrales en imágenes de rayos X. Otro enfoque es el propuesto por Pengfei Cheng y colaboradores [3], quienes desarrollaron un sistema de localización y segmentación automática de vértebras en tomografías computarizadas (TC). Su metodología utiliza un modelo Dense-U-Net en dos etapas: la primera localiza las vértebras detectando los centroides, y la segunda segmenta dentro de una región definida por el centroide. Por último, Jing Liu y su equipo [4] proponen un método basado en WGAN para la segmentación de imágenes de resonancia magnética (RM). Su enfoque combina una red ResU-Net mejorada con transformadores basados en clústeres.

Aunque estos trabajos han logrado avances significativos, la mayoría se enfoca en herramientas como TC o RM, donde las condiciones de imagen son más controladas. En contraste, las radiografías de columna presentan desafíos inherentes como la superposición de tejidos, el bajo contraste y la variabilidad anatómica entre pacientes, que complican significativamente la segmentación automática precisa. Estas dificultades subrayan la necesidad de complementar la segmentación con técnicas de posprocesado que mejoren la calidad de las máscaras antes de cualquier análisis posterior, como el cálculo de métricas clínicas.

En esta línea, varios trabajos han abordado no solo la segmentación, sino la posterior obtención de métricas clínicas. Por ejemplo, el trabajo de Liu et al. [5] destaca por el uso de redes de segmentación para el análisis de escoliosis, demostrando cómo el refinamiento posterior de las segmentaciones permite obtener mediciones más fiables para esta patología. Otro estudio relevante es el de Renato R. Maaliw [6], que propone el sistema ScolioNet para la estimación automática del ángulo de Cobb a partir de imágenes segmentadas. No obstante, al igual que en los casos de segmentación, muchos de estos estudios emplean conjuntos de datos privados, lo que limita la reproducibilidad y dificulta el avance en la investigación aplicada a radiografías de columna.

1.4. Estructura de la memoria

Esta memoria se organiza en varios capítulos. En primer lugar, se introduce el problema médico, seguido de una revisión de los trabajos relacionados y la definición del marco del proyecto. Posteriormente, se describe la metodología seguida, incluyendo los materiales empleados y las estrategias de análisis aplicadas. A continuación, se presentan y discuten los resultados obtenidos. Finalmente, se exponen las conclusiones, las limitaciones y posibles líneas de investigación futuras.

Capítulo 2

Materiales y métodos

En este capítulo se describen los materiales utilizados y la metodología seguida para el desarrollo del sistema propuesto. Se comienza detallando los conjuntos de datos empleados en el entrenamiento, validación y evaluación del modelo. A continuación, se describen las distintas fases del proceso metodológico, desde el desarrollo del modelo de segmentación hasta la validación y análisis de métricas clínicas.

2.1. Materiales

2.1.1. Conjunto de datos NHANES II

Para el entrenamiento, validación y evaluación del modelo de segmentación y el desarrollo de las fases posteriores, se utilizó el conjunto de datos NHANES II Spinal X-ray [7], un recurso público que forma parte del estudio National Health and Nutrition Examination Survey (NHANES), disponible para su descarga y uso con fines de investigación. Este conjunto contiene imágenes de radiografías laterales de columna vertebral anotadas por especialistas, y ha sido empleado en trabajos previos de segmentación vertebral.

El conjunto incluye un total de **543 radiografías** correspondientes a las regiones **cervical** y **lumbar**. Cada muestra se encuentra almacenada en un archivo en formato **.pkl**, que contiene los siguientes elementos:

- Un identificador único de la imagen.
- La imagen radiográfica en escala de grises, representada como un array numérico.
- Un diccionario de máscaras de segmentación, donde cada clave corresponde a una vértebra específica.

Las claves del diccionario varían según la región anatómica:

- Región lumbar: *s1, t1, l1, l2, l3, l4, l5* y *vertebrae*
- Región cervical: *c2, c2_lower, c3, c4, c5, c6, c7* y *vertebrae*.

La clave *vertebrae* corresponde a un array numérico que contiene la máscara que combina todas las vértebras segmentadas en la imagen. Esta fue la única utilizada en este trabajo como objetivo de segmentación, permitiendo simplificar el problema a una clasificación binaria de los píxeles como fondo o vértebra. En la figura 2.1 se muestra un ejemplo de las radiografías de este conjunto con su correspondiente máscara.

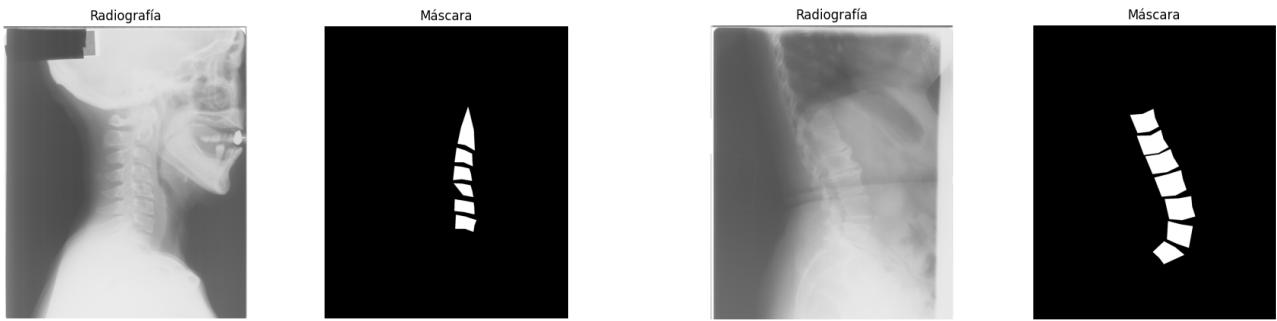


Figura 2.1: Imágenes del conjunto NHANES II. (Izquierda) Región cervical. (Derecha) Región lumbar.

Antes de ser utilizadas por el modelo, las imágenes y sus máscaras asociadas fueron procesadas mediante los siguientes pasos:

- Extracción desde los archivos .pkl.
- Conversión a formato .png para facilitar la manipulación.
- Binarización de la máscara combinada (*vertebrae*).
- Redimensionamiento de las imágenes y máscaras a una resolución estándar de 512x512 píxeles.

2.1.2. Conjunto de pruebas en condiciones reales

Como parte del proceso de pruebas, se utilizó un conjunto adicional de imágenes radiográficas proporcionadas por el médico colaborador, el Dr. Julio Manuel Plata Bello. Este conjunto consta de un total de **90 imágenes radiográficas laterales** de las regiones **cervical** y **lumbar**. Inicialmente se recibieron 103 imágenes, pero algunas fueron descartadas durante el preprocesado debido a problemas de formato o errores al convertirlas.

A diferencia del conjunto NHANES II, estas imágenes **no incluyen anotaciones** de referencia, por lo que su uso estuvo limitado a una validación cualitativa y visual del sistema en funcionamiento. Las imágenes se encontraban en formato .jpg, y presentaban diferencias notables en comparación con el conjunto de entrenamiento, especialmente en cuanto a calidad, orientación anatómica y condiciones de captura. Desde el punto de vista técnico, estas radiografías presentaban:

- Variaciones de brillo, contraste y resolución.
- Presencia de inclinaciones y rotaciones anatómicas (posturas de extensión y flexión).
- Capturas con la columna no alineada al eje principal de la imagen, dificultando una interpretación directa.

Estas diferencias permiten analizar la capacidad del modelo para generalizar sobre datos reales en condiciones distintas a las del conjunto de entrenamiento. El análisis se realizó mediante inspección visual del resultado de la segmentación y su posterior procesamiento, con especial atención al impacto que estas variaciones tienen sobre la limpieza y el cálculo de métricas clínicas.

Antes de su uso, las imágenes fueron convertidas de .jpg a .png, aplicando una transformación a escala de grises para asegurar la compatibilidad con el modelo entrenado. Posteriormente, se les

aplicó un redimensionamiento a una resolución estándar de 512 x 512 píxeles. Este proceso permitió homogeneizar las características de entrada respecto a las del conjunto NHANES II.

2.2. Metodología

La metodología se divide en tres fases principales: desarrollo del modelo de segmentación, posprocesado de las segmentaciones obtenidas y cálculo de métricas morfológicas clínicas. Cada una de estas fases se describe a continuación. Todo el código fuente que implementa el flujo descrito en esta sección está disponible públicamente [8].

2.2.1. Fase 1: Entrenamiento de los modelos de segmentación

Esta primera fase se centró en el desarrollo del modelo de segmentación, cuyo objetivo es recibir como entrada una imagen radiográfica y devolver una máscara binaria en la que cada píxel se clasifica como fondo o vértebra. Para ello, se empleó un conjunto de datos previamente anotado, se diseñó y entrenó un modelo basado en redes convolucionales, y se evaluaron las métricas estándar en segmentación de imágenes para medir su rendimiento.

Desarrollo y flujo de trabajo

El desarrollo de esta fase se implementó basándose en una arquitectura U-Net, ampliamente utilizada en el ámbito de la imagen médica por su capacidad para preservar detalles espaciales y segmentar estructuras anatómicas complejas.

Para abordar el problema específico de la segmentación vertebral y evaluar diferentes aproximaciones, se diseñaron y entrenaron tres modelos de segmentación distintos:

- **Modelo general:** Un único modelo entrenado para segmentar vértebras de cualquier región de la columna.
- **Modelos especializados por región:** Dos modelos específicos, uno entrenado exclusivamente para la segmentación de vértebras cervicales y otro para vértebras lumbares.

El proceso de desarrollo para la implementación de estos modelos fue el siguiente:

1. **Definición de la arquitectura del modelo:** Se seleccionó la red neuronal convolucional U-Net por su robustez y eficiencia en tareas de segmentación biomédica, tal como se implementa en la librería Segmentation Models PyTorch (SMP) [9]. Esta arquitectura se caracteriza por una estructura de encoder-decoder simétrica: la parte de codificación reduce progresivamente la resolución espacial, capturando características contextuales, mientras que la de decodificación recupera la resolución original mediante capas de transposición y concatenación, permitiendo así la segmentación precisa de regiones pequeñas como las vértebras. La arquitectura empleada se basó en el diseño original de Ronneberger et al. [10], y se muestra en la Figura 2.2.

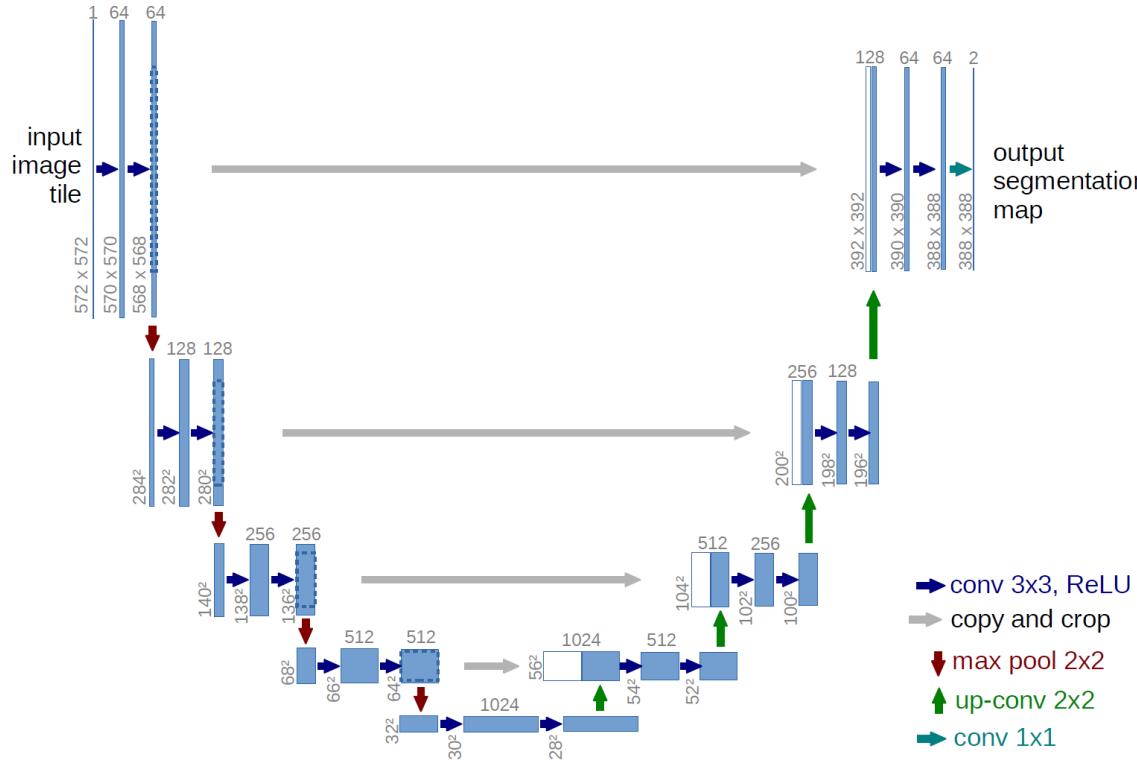


Figura 2.2: Arquitectura original de U-Net propuesta por Ronneberger et al., utilizada ampliamente en tareas de segmentación biomédica.

2. **Configuración del encoder:** Se utilizó una arquitectura ResNet34 [11] preentrenada en ImageNet [12] como encoder para los modelos U-Net. Si bien se exploraron otras combinaciones de encoders inicialmente, esta configuración particular proporcionó resultados de segmentación adecuados para llevar a cabo la extracción de métricas posteriores, que es el objetivo principal de este trabajo. Por esta razón, no se realizó una comparación exhaustiva de las posibles combinaciones de arquitecturas y encoders.
3. **Preparación y preprocessado de datos:** Se adaptó el conjunto de datos de radiografías con anotaciones vertebrales a la entrada de los modelos. Para ello, se extrajeron imágenes y máscaras desde archivos .pk1, convirtiéndolos a formato .png, y se redimensionaron a 512x512 píxeles. Además, se aplicó una normalización de tipo estándar con media 0 y desviación típica 1, y posteriormente se convirtieron las imágenes a tensores. Para el modelo general, se utilizaron 543 imágenes completas del conjunto NHANES II (2.1.1), abarcando tanto vértebras cervicales como lumbares. Estas se dividieron siguiendo una proporción 70-20-10 para entrenamiento, validación y test, respectivamente, resultando en 380 imágenes para entrenamiento, 108 para validación y 55 para test. En caso de los modelos especializados:
 - Para el modelo de la región cervical, se emplearon 275 imágenes exclusivas de esta región. La distribución fue de 192 imágenes para entrenamiento, 55 para validación y 28 para test.
 - Para el modelo de la región lumbar, se utilizaron 268 imágenes exclusivas de esta región, distribuidas en 187 imágenes para entrenamiento, 53 para validación y 28 para test.
4. **Aplicación de Data Augmentation:** Para aumentar la diversidad del conjunto de datos y

mejorar la capacidad de generalización del modelo, reduciendo el riesgo de sobreajuste debido al tamaño limitado del conjunto NHANES II y la variabilidad anatómica entre pacientes, se aplicaron transformaciones aleatorias durante el entrenamiento. Estas incluyeron rotaciones de 90°, reflejos horizontales y verticales, además de transformaciones suaves como desplazamientos, escalados, rotaciones, y ajustes de brillo y contraste. Aunque en el conjunto NHANES II las imágenes presentan una orientación consistente (columna en posición vertical), se optó por introducir variabilidad artificial para favorecer la robustez del modelo frente a imágenes con diferente orientación, como las del conjunto clínico externo, donde la columna puede aparecer diagonal o en posición horizontal.

5. **Configuración y Entrenamiento del Modelo:** El modelo fue entrenado para segmentación binaria (fondo o vértebra), utilizando la función de pérdida Dice Loss, ya que es adecuada en contextos con clases desbalanceadas, como es el caso del problema que nos ocupa, con regiones correspondientes a las vértebras relativamente pequeñas si las comparamos con el fondo. Se empleó un tamaño de lote de 16, el optimizado Adam y una tasa de aprendizaje de 0.0001. El modelo general fue entrenado durante 70 épocas, alcanzando resultados adecuados para la extracción de características clínicas. En cuanto los modelos especializados, el modelo lumbar alcanzó su mejor rendimiento tras 60 épocas, mientras que el modelo cervical requirió 65 épocas para lograr resultados comparables. No se realizó una búsqueda exhaustiva de hiperparámetros, ya que esta configuración permitió obtener resultados adecuados para la segmentación, que permitieron una posterior extracción de características.

Herramientas

El desarrollo de la fase de segmentación se llevó a cabo empleando las siguientes herramientas y librerías:

- **PyTorch:** Framework de aprendizaje profundo para la construcción y entrenamiento de la red neuronal.
- **Segmentation Models PyTorch (SMP):** Librería de alto nivel utilizada para la implementación eficiente de la arquitectura U-Net y la integración del encoder seleccionado.
- **Jupyter Notebook:** Entorno interactivo utilizado para el desarrollo, experimentación, prototipado y visualización de resultados del código.
- **Google Colab:** Plataforma basada en la nube que proporcionó el punto de partida inicial y el entorno de trabajo para las primeras exploraciones.

Métricas de evaluación

Para evaluar el rendimiento del modelo, se han empleado dos métricas:

- **IoU por imagen:** se calcula la intersección sobre la unión de forma individual para cada imagen y se obtiene el promedio de todos los valores.
- **IoU por dataset:** se calcula una única IoU acumulada. Refleja el rendimiento global del modelo sobre el conjunto completo.

La intersección sobre la unión (IoU) se define como el cociente entre la intersección y la unión de los píxeles predichos y los reales. Se expresa mediante la fórmula:

$$\text{IoU} = \frac{|A \cap B|}{|A \cup B|}$$

donde A representa el conjunto de píxeles predichos como vértebra y B el conjunto de píxeles realmente etiquetados como vértebras.

Estas métricas permiten analizar el rendimiento del modelo tanto de manera global como en casos individuales, proporcionando una visión detallada de su eficiencia en la segmentación.

2.2.2. Fase 2: Posprocesado de las segmentaciones

Tras la primera fase de segmentación, las máscaras obtenidas a menudo presentan imperfecciones como píxeles ruidosos aislados, bordes irregulares o vértebras unidas debido a su proximidad o a la segmentación del modelo. En particular, esto se observó en la región cervical con mayor frecuencia, lo que dificultaba la extracción precisa y comprometía las medidas clínicas. Por ello, el posprocesado es una etapa importante para refinar estas máscaras, logrando separar individualmente cada vértebra, corregir estas imperfecciones y mejorar su representación para el análisis morfológico clínico.

Desarrollo y flujo de trabajo

El desarrollo de esta fase se realizó de forma progresiva y exploratoria. Se comenzó con estrategias simples, y se fueron incorporando técnicas más robustas basadas en geometría, clustering y análisis de dirección. A continuación, se detalla el flujo de trabajo final adoptado, ilustrado con un ejemplo visual sobre la región cervical (Figura 2.3).

La imagen de la izquierda, en la fila superior de la Figura 2.3 muestra la imagen radiográfica de partida, mientras que la imagen central de la misma fila presenta la segmentación generada por el modelo, que incluye algunas de las imperfecciones mencionadas.

1. **Refinado morfológico:** Consiste en la aplicación de operaciones basadas en la forma de las imágenes. Se aplicaron operaciones morfológicas clásicas de procesamiento de imágenes:

- Erosión: Reduce el tamaño de los objetos y elimina pequeños detalles o ruido.
- Dilatación: Expande el tamaño de los objetos y rellena pequeños agujeros.
- Apertura: Es una erosión seguida de una dilatación, útil para eliminar ruido y suavizar bordes sin alterar el tamaño del objeto significativamente.
- Cierre: Es una dilatación seguida de una erosión, útil para llenar pequeños agujeros dentro de los objetos.

El objetivo fue suavizar bordes y eliminar pequeñas manchas ruidosas. Esta etapa, aunque útil, no fue suficiente por sí sola para eliminar todas las imperfecciones, especialmente en la zona cervical. Como se observa en la imagen de la derecha de la fila superior (Figura 2.3), tras este paso desaparecen ciertos puntos y se aprecia una mayor limpieza en los bordes de algunos elementos.

2. **Filtrado por área:** Se descartaron regiones cuya superficie estuviera fuera de los rangos típicos observados en vértebras reales, especialmente las regiones pequeñas [13]. Estos rangos fueron calculados a partir de un análisis estadístico de las máscaras reales del conjunto NHANES II, diferenciando entre vértebras cervicales y lumbares. En la imagen de la izquierda de la fila inferior (Figura 2.3) puede verse cómo un elemento situado en el centro de la imagen desaparece por no alcanzar el área mínima establecida.
3. **Evaluación geométrica inicial:** En una primera aproximación, se calcularon los centroides de cada región segmentada y se descartaron aquellos demasiado alejados de las otras regiones. Esta técnica se descartó al observar que no era robusta frente a columnas inclinadas o grupos con ruido cercano.
4. **Separación de regiones grandes:** Para dividir vértebras que aparecían unidas, se utilizó una combinación de:
 - Transformada de distancia euclídea: que asigna a cada píxel de una región su distancia al borde más cercano.
 - peak_local_max: que localiza los máximos locales de la transformada de distancia, normalmente ubicados en el centro de las vértebras.
 - Algoritmo Watershed: que simula una inundación controlada desde esos centros, generando líneas divisorias que separan vértebras fusionadas.
- En la Figura 2.3 puede observarse, en la imagen de la izquierda en la fila inferior, por ejemplo, cómo el elemento grande en la parte superior central de la imagen ha sido dividido en dos regiones distintas.
5. **Clustering con DBSCAN:** Se aplicó el algoritmo DBSCAN sobre los centroides de las vértebras ya separadas, para identificar agrupaciones densas. DBSCAN no requiere conocer de antemano el número de vértebras, y es tolerante a regiones ruidosas aisladas. Como muestra la imagen central de la fila inferior (Figura 2.3), se han identificado tres grupos distintos (representados en rojo, verde y azul), cada uno de los cuales podría representar la columna vertebral.
6. **Estimación de dirección con PCA:** Para cada grupo de vértebras potenciales, se aplicó análisis de componentes principales (PCA) para estimar la dirección predominante de la columna. Los centroides de proyectaron y ordenaron según esta dirección para evaluar la coherencia estructural del grupo. Cabe destacar que esta técnica solo se aplicó a la región cervical. La morfología de la región cervical es más lineal y el PCA puede estimar su alineación de forma adecuada. En contraste, la región lumbar presenta una morfología una curvatura natural que hace que la suposición de alineación lineal del PCA no sea apropiada para estimar su dirección principal.
7. **Filtrado de centroides desalineados:** A partir de la vértebra más baja, se recorre la secuencia ordenada de centroides descartando aquellos que no siguen la dirección principal estimada. Esta validación se basa en el cálculo de ángulos entre vectores consecutivos, usando el coseno como medida de alineación. Este paso se aplica únicamente a la región cervical debido a su morfología.
8. **Selección del grupo candidato a columna:** Todos los grupos detectados se evaluaron mediante una puntuación que combinaba:

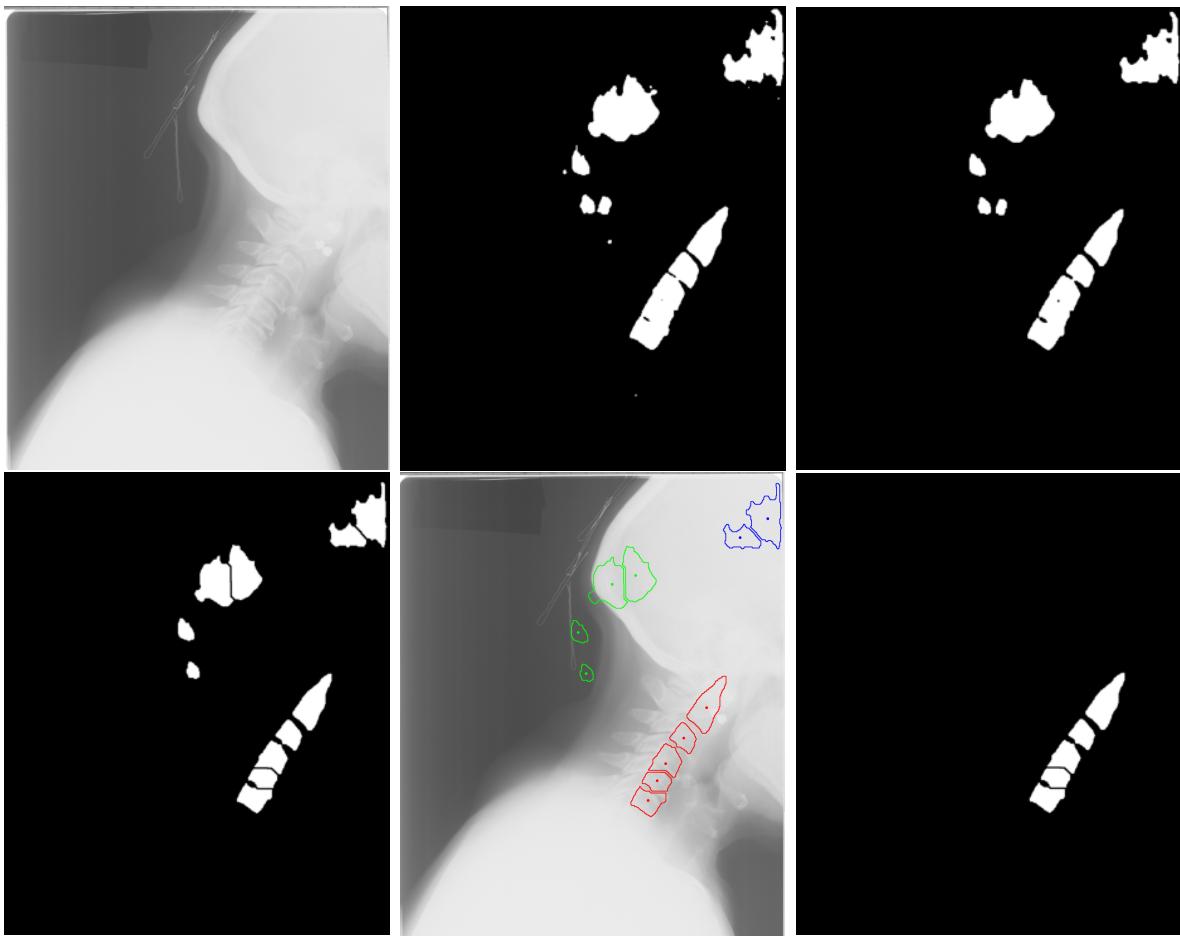


Figura 2.3: Flujo visual del posprocesado para una imagen de la región cervical.

- Número de vértebras del grupo.
- Desviación estándar de las distancias entre centroides.
- Promedio del coseno de ángulo entre vértebras consecutivas.

mediante una función definida como:

$$\text{Puntuación}(G) = n^2 - \sigma_d$$

donde n es el número de vértebras del grupo y σ_d la desviación estándar de las distancias entre centroides consecutivos. Además, se exigió que:

$$n \geq n_{\min} \quad y \quad \sigma_d < \sigma_{\max}$$

con $n_{\min} = 3$ y $\sigma_{\max} = 70$. Si ningún grupo superaba los umbrales, se seleccionaba aquel con mayor puntuación. En la imagen de la derecha en la fila inferior (Figura 2.3), se muestra el grupo rojo como el grupo final seleccionado tras la evaluación de todas las condiciones anteriores.

Este flujo de posprocesado, en su diseño general y específicamente la idea de la separación de regiones grandes, se inspiró parcialmente en el enfoque propuesto por Liu et al. [5], tras detectar que las estrategias morfológicas y geométricas iniciales resultaban poco eficientes.

Herramientas

El desarrollo se realizó en Python, empleando:

- **OpenCV**: para operaciones morfológicas y contornos.
- **Scikit-image**: para transformada de distancia, detección de picos locales y Watershed.
- **Scikit-learn**: para DBSCAN y PCA.
- **Matplotlib**: para inspección visual y validación del proceso.

Métricas de evaluación

El posprocesado se validó principalmente mediante inspección visual, ya que la interpretación estructural de la columna requiere evaluar la alineación y el contexto morfológico. No obstante, se realizó:

- Verificación del número de vértebras detectadas tras el posprocesado.
- Comprobación de la consistencia del área segmentada respecto a los valores estadísticos de las máscaras reales.
- Evaluación de la regularidad de las distancias entre vértebras y alineación entre centroides.

2.2.3. Fase 3: Cálculo de métricas morfológicas clínicas

Una vez obtenidas y procesadas las máscaras de segmentación vertebral, el objetivo principal de esta fase es calcular dos tipos de medidas clínicas que resultan de interés para el diagnóstico médico: medidas angulares y medidas traslacionales. Estas métricas permiten cuantificar la alineación y la posición relativa de las vértebras consecutivas, facilitando la identificación de posibles patologías. Como salida, se genera una imagen visual con las anotaciones superpuestas sobre la radiografía original y un vector con los valores numéricos de las medidas obtenidas.

Desarrollo y flujo de trabajo

Esta fase se centra en la extracción y análisis de métricas morfológicas clínicas a partir de las máscaras de segmentación vertebral procesadas, para su posterior comparación con las métricas obtenidas a partir de las máscaras de referencia del conjunto NHANES II. Se calculan dos tipos de medidas por separado, tal como se mencionó en la introducción:

- Medidas angulares: asociadas a la orientación relativa entre vértebras.
- Medidas traslacionales: asociadas al desplazamiento lateral entre vértebras consecutivas.

Cálculo de medidas angulares. El objetivo es calcular el ángulo relativo de cada par consecutivo de vértebras. Para ello, se siguen los siguientes pasos:

1. **Determinación de la orientación de cada vértebra:** Para cada vértebra, se utiliza la función `cv2.minAreaRect` de la librería OpenCV para obtener el rectángulo de área mínima que la englobe completamente. Este rectángulo proporciona dos ejes que representan las orientaciones principales de la vértebra. La selección del eje que representa la orientación anatómica de la vértebra se realiza de forma iterativa, comenzando desde la vértebra más inferior y ascendiendo. Para cada par de vértebras consecutivas (V_i, V_{i+1}) , donde V_i es la vértebra inferior:

- Se calcula un vector director de la columna entre los centroides de V_i y V_{i+1} . Este vector $(C_i C_{i+1})$ indica la dirección local de la columna.
- Para V_i , de los dos ejes proporcionados por `cv2.minAreaRect`, se selecciona aquel que sea más perpendicular al vector director $(C_i C_{i+1})$. Este eje se considera la orientación principal de la vértebra V_i (θ_{i+1}), ya que las vértebras deben estar transversalmente alineadas respecto al eje longitudinal de la columna. Este procedimiento es relevante en imágenes con inclinaciones distintas, ya que permite adaptar dinámicamente la orientación vertebral a la dirección de la columna.
- Este mismo criterio se aplica a V_{i+1} utilizando el mismo vector director.

De esta manera, se obtiene un ángulo θ_{i+1} que representa la orientación de cada vértebra respecto al eje horizontal.

2. **Cálculo del ángulo entre vértebras:** Dados los vectores de orientación de dos vértebras consecutivas \vec{u}_i y \vec{u}_{i+1} , se calcula la diferencia angular utilizando el producto escalar: $\alpha_i = \text{arc cos}(|\vec{u}_i \cdot \vec{u}_{i+1}|)$. Esto garantiza un valor absoluto en el rango $[0^\circ, 90^\circ]$.

Cálculo de medidas translacionales: Estas miden el desplazamiento lateral entre vértebras consecutivas, según su orientación local.

1. **Obtención del eje intermedio:** El eje intermedio entre dos vértebras consecutivas $(V_i$ y $V_{i+1})$ se define como la recta que pasa por un punto central P_m y tiene como dirección un ángulo intermedio α , calculado a partir de los ángulos individuales θ_i y θ_{i+1} de las vértebras mediante un promedio vectorial:

$$\vec{v}_m = \frac{\vec{u}_i + \vec{u}_{i+1}}{\|\vec{u}_i + \vec{u}_{i+1}\|} \quad \text{donde} \quad \vec{u}_k = \begin{bmatrix} \cos(\theta_k) \\ \sin(\theta_k) \end{bmatrix}$$

Este enfoque permite obtener una dirección intermedia coherente incluso cuando los ángulos presentan signos opuestos, evitando así errores derivados de discontinuidades angulares en el rango $[-90^\circ, 90^\circ]$.

El punto P_m por el que pasa este eje intermedio se calcula como el punto medio entre el punto más alto del contorno de la vértebra inferior (V_i) y el punto más bajo del contorno de la vértebra superior (V_{i+1}). Esta selección garantiza que el eje se sitúe entre ambas vértebras, proporcionando una base coherente para proyectar las distancias laterales.

2. **Selección del punto lateral:** Para cada vértebra del par analizado (V_i y V_{i+1}), se toma el punto más a la derecha de su contorno, que se considera representativo del borde lateral externo de la vértebra. Este punto será el proyectado sobre el eje intermedio calculado anteriormente, con el fin de estimar la distancia lateral entre ambas vértebras.
3. **Proyección del punto sobre el eje intermedio:** Para calcular la distancia de desplazamiento, se proyecta cada punto lateral (obtenido en el paso anterior) sobre una recta perpendicular al eje intermedio. Sea P el punto lateral a proyectar y d_m el vector director del eje intermedio (paso 1). La recta de proyección tiene un vector director p que es perpendicular a d_m . El punto proyectado P' sobre el eje intermedio se obtiene resolviendo el sistema:

$$P + t\vec{p} = Q + s\vec{d}_m$$

donde Q es un punto conocido que pertenece al eje intermedio (por ejemplo, P_m). Este punto P' es el que marca la posición lateral de la vértebra sobre el eje de referencia.

4. **Medida translacional:** Finalmente, la medida translacional se calcula como la distancia entre las proyecciones de los puntos laterales de las dos vértebras consecutivas sobre el eje intermedio:

$$\text{Distancia} = \|P'_{i+1} - P'_i\|$$

Esta distancia esta expresada en píxeles, debido a la ausencia de metadatos de la resolución física de las imágenes.

Herramientas

Al igual que en fases anteriores, el desarrollo se realizó en Python, empleando las siguientes librerías:

- **OpenCV:** Para operaciones de procesamiento de contornos, dibujo y geometría computacional.
- **NumPy:** Para operaciones de álgebra lineal y manejo eficiente de vectores y matrices.
- **Matplotlib:** Para visualización de resultados gráficos, incluyendo las anotaciones sobre las radiografías.
- **OS/CSV:** Para la gestión de archivos y almacenamiento de resultados en formato CSV.

Métricas de evaluación

En esta fase, la validación se realizó en dos niveles. Primero, se llevó a cabo una validación cualitativa visual, comparando las imágenes anotadas con medidas obtenidas de las máscaras predichas con aquellas derivadas de las máscaras reales. Adicionalmente, se realizó una comparación cuantitativa directa entre las medidas calculadas a partir de las máscaras predichas y las máscaras reales, evaluando cada tipo de métrica (angular y translacional) de forma individual. Para cada imagen, se comparó el vector de medidas obtenido de la segmentación con el vector de medidas de referencia (elemento a elemento).

Las principales métricas utilizadas en este análisis cuantitativo fueron:

- **Error absoluto medio (MAE)**: Promedio de los valores absolutos de las diferencias entre las medidas obtenidas y las reales. Se utilizó como métrica principal de referencia por su fácil interpretación.
- **Error cuadrático medio (MSE)**: Promedio de los cuadrados de las diferencias entre las medidas obtenidas y las reales. Penaliza con mayor intensidad los errores grandes, ayudando a detectar casos con alta desviación.
- **Raíz del error cuadrático medio(RMSE)**: Raíz cuadrada del MSE, con las mismas propiedades, pero expresa en las mismas unidades que las medidas originales. Ofrece una interpretación más directa en término clínicos.
- **Coeficiente de determinación (R^2)**: Indica la proporción de la varianza total de las medidas reales que es explicada por el modelo. Un valor cercano a 1 indica una fuerte correlación entre las predicciones y los valores reales.

Además, se generaron representaciones gráficas como histogramas del MAE y diagramas de caja (*boxplots*) para cada tipo de medida y región (cervical y lumbar). Estas visualizaciones sirvieron para detectar outliers y errores sistemáticos que no eran fácilmente perceptibles en la inspección visual de las imágenes.

En algunos casos, el número de vértebras detectadas por el modelo no coincide con el número presente en la máscara real, dificultando una comparación directa de las métricas. Para resolverlo, se implementó una comparación iterativa. Este proceso comienza desde la vértebra inferior y compara las medidas consecutivas de ambas secuencias (real y predicha) hasta que una de ellas se agota. Las métricas de error se calcularon únicamente sobre estos pares coincidentes.

Capítulo 3

Resultados

En este capítulo se presentan los resultados obtenidos durante el desarrollo del sistema. Los resultados se organizan en dos bloques principales: el primero se centra en el rendimiento del modelo de segmentación, y el segundo en la calidad de las métricas clínicas obtenidas a partir de las segmentaciones procesadas. En ambos casos se incluyen análisis cuantitativos y cualitativos, así como una comparación entre el modelo general y los modelos específicos entrenados para las regiones cervical y lumbar. Por último, se discuten los casos problemáticos y las posibles implicaciones clínicas del sistema.

3.1. Rendimiento de la segmentación

3.1.1. Resultados cuantitativos

Para evaluar el rendimiento de los modelos de segmentación, se utilizaron las métricas estándar: la **IoU por imagen** y la **IoU global**. Los resultados que se presentan corresponden exclusivamente al subconjunto de test, definido como el 10 % del conjunto NHANES II, y utilizado para evaluar el rendimiento final de cada modelo.

La Tabla 3.1 muestra los valores de IoU obtenidos antes y después del posprocesado, incluyendo también la desviación típica del IoU por imagen para reflejar la dispersión entre casos.

Tabla 3.1: Rendimiento sobre el conjunto de test (10 % del conjunto NHANES II), antes y después del posprocesado.

Modelo	Sin posprocesado			Con posprocesado		
	IoU imagen (%)	σ	IoU global (%)	IoU imagen (%)	σ	IoU global (%)
General	54,70	24,49	63	76,40	13,51	76,79
Cervical	40,30	13,12	38,31	79,31	7,29	79,19
Lumbar	86,40	6,28	86,18	84,86	9	84,54

Los resultados muestran una mejora significativa tras aplicar el posprocesado en los modelos general y cervical. En el caso del modelo general, la IoU por imagen aumenta más de 21 puntos porcentuales, acompañada de una reducción notable en la desviación típica, lo que indica una mayor coincidencia en las predicciones.

El modelo cervical experimentó una mejora aún más marcada, con un incremento de casi 39 puntos porcentuales en la IoU por imagen. Esto demuestra la eficiencia del posprocesado para limpiar segmentaciones que inicialmente podrían ser muy imprecisas.

Por otro lado, el modelo lumbar presentó un rendimiento alto incluso sin posprocesado, superando el 86 % en IoU por imagen. El posprocesado introdujo una ligera reducción, probablemente asociada a la eliminación de pequeñas regiones válidas o pequeñas distorsiones que afectaron la métrica, aunque los valores se mantuvieron estables y elevados.

Cabe destacar que no se observaron diferencias relevantes entre la métrica de IoU por imagen y la IoU global en ninguno de los modelos. Esto indica una detección estable y homogénea a lo largo de las distintas imágenes, sin casos extremos con IoU cercanos a cero. En contextos donde existen errores de segmentación graves en imágenes con regiones de segmentación pequeñas, la IoU por imagen tiende a disminuir drásticamente, mientras que la global apenas se ve afectada. La ausencia de este tipo de casos sugiere una segmentación robusta y consistente en todo el conjunto evaluado.

3.1.2. Resultados cualitativos

La evaluación cualitativa se realizó mediante la inspección visual de los resultados de segmentación sobre dos conjuntos de imágenes: el conjunto NHANES II y el proporcionado por el médico colaborador. En cada caso, se analizó el rendimiento de los tres modelos, mostrando tanto resultados satisfactorios como problemáticos.

Conjunto NHANES II

En las Figuras 3.1 y 3.2 se muestran dos ejemplos de segmentación con el modelo general sobre imágenes del conjunto NHANES II. En ambos casos, se observa la radiografía original (izquierda), la máscara esperada (centro) y la segmentación obtenida por el modelo (derecha).

La Figura 3.1 corresponde a una imagen de la región lumbar. Se observa una buena segmentación, con vértebras bien definidas, separadas y sin presencia de ruido. En general, las imágenes lumbares de este conjunto presentan menor ruido y permiten una segmentación más precisa, aunque en algunos casos puntuales pueden aparecer vértebras unidas o ausentes.

En cambio, la Figura 3.2 muestra un caso más complejo de región cervical. Aquí, la segmentación presenta ruido considerable, posiblemente debido a la presencia de otras estructuras óseas como el cráneo o la mandíbula. Además, algunas vértebras aparecen fusionadas o deformadas, y su morfología se aleja de la esperada en la máscara real. Aunque también se han observado buenos resultados en esta región, es más frecuente encontrar problemas en este tipo de imágenes cervicales.

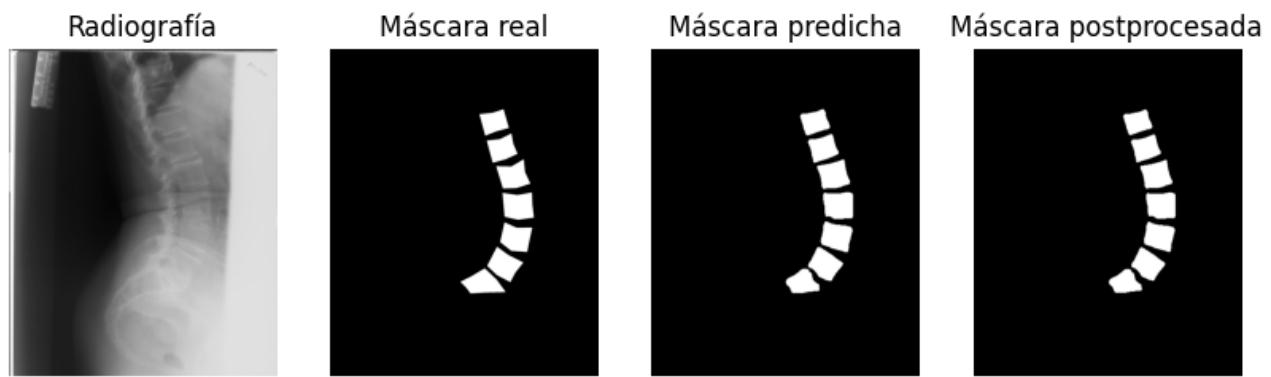


Figura 3.1: Segmentación con el modelo general en imagen lumbar del conjunto NHANES II.



Figura 3.2: Segmentación con el modelo general en imagen cervical del conjunto NHANES II.

Conjunto clínico (conjunto de pruebas en condiciones reales)

En este caso, las imágenes solo cuentan con la radiografía original y la segmentación obtenida por el modelo, ya que no disponen de máscaras de referencia. En las Figuras 3.3 y 3.4 se muestran dos ejemplos correspondientes al modelo general aplicado sobre este conjunto.

La Figura 3.3 corresponde a una imagen de la región lumbar. La segmentación es precisa, con vértebras bien definidas y poco ruido. Las imágenes lumbares proporcionadas por el médico presentan, en general, características morfológicas similares a las del conjunto de NHANES II, lo que ha permitido al modelo generalizar de forma razonable. No obstante, debido a las variaciones en la calidad y contraste de las imágenes, en algunos casos se observa un aumento del ruido o imperfecciones en las segmentaciones.

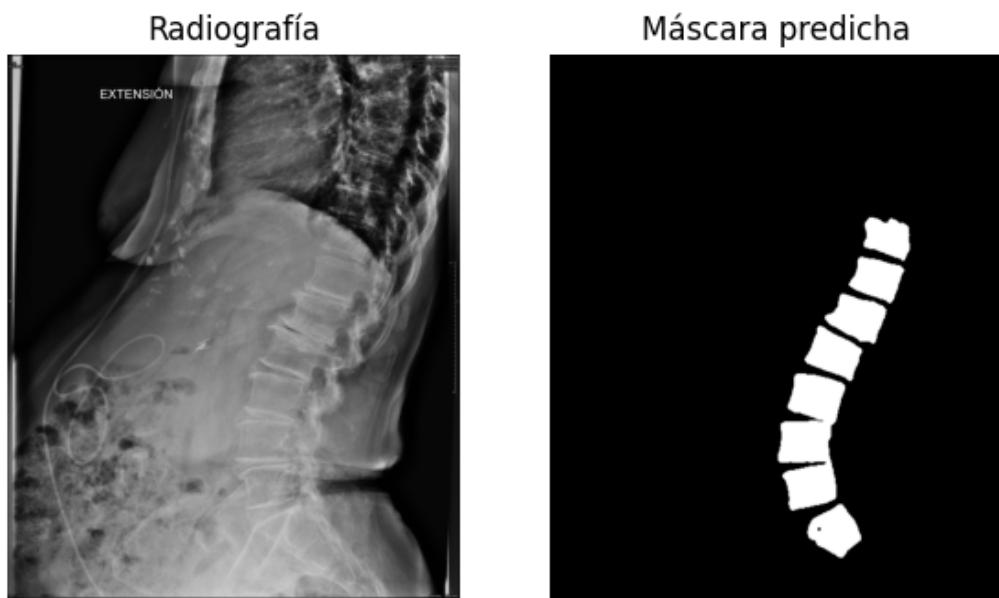


Figura 3.3: Segmentación con el modelo general en imagen lumbar del conjunto clínico.

En contraste, la Figura 3.4 muestra una imagen de la región cervical, donde el modelo falla claramente. En este conjunto, todas las imágenes correspondientes a la región cervical presentaron resultados deficientes, con ruido significativo, vértebras ausentes o mal definidas, y estructuras que no corresponden con la anatomía esperada. Estas imágenes muestran una morfología diferente respecto al conjunto NHANES II, con vértebras más pequeñas y orientaciones más variables, lo que parece haber afectado negativamente al rendimiento del modelo. Este resultado resalta la importancia de considerar la variabilidad entre conjuntos y refuerza la necesidad de validar los modelos en contextos clínicos diversos.

Comparación entre modelos: general vs especializado

Con el objetivo de evaluar el impacto de la especialización regional, se presenta una comparación visual entre las segmentaciones obtenidas con el modelo general y con los modelos específicos entrenados exclusivamente con imágenes cervicales y lumbares.

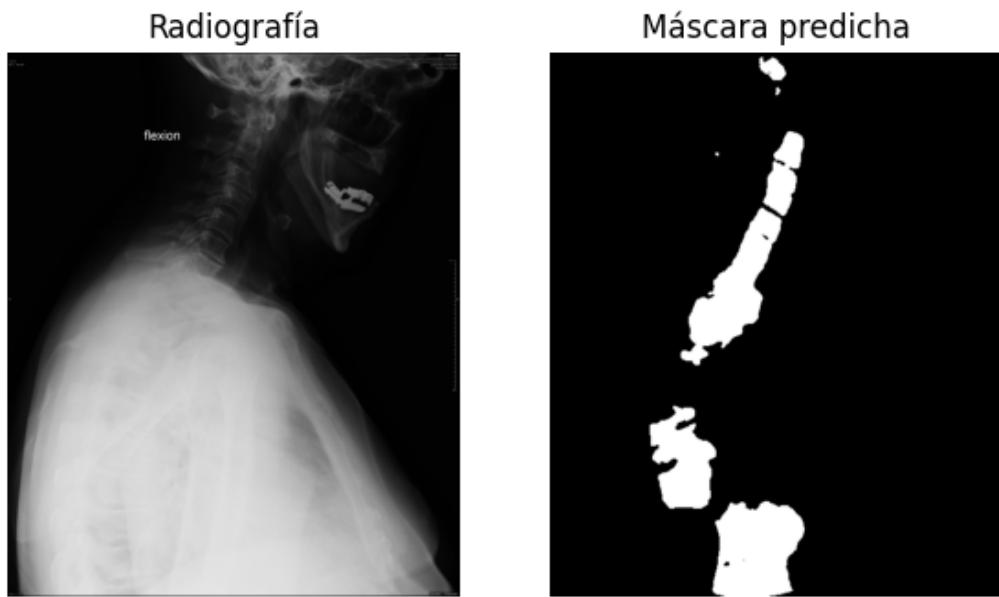


Figura 3.4: Segmentación con el modelo general en imagen cervical del conjunto clínico.

Región cervical: En la Figura 3.5, se muestra una imagen del conjunto NHANES II correspondiente a la región cervical. A la izquierda se encuentra la radiografía original, al centro la segmentación obtenida con el modelo general y a la derecha la segmentación generada con el modelo especializado en cervicales.

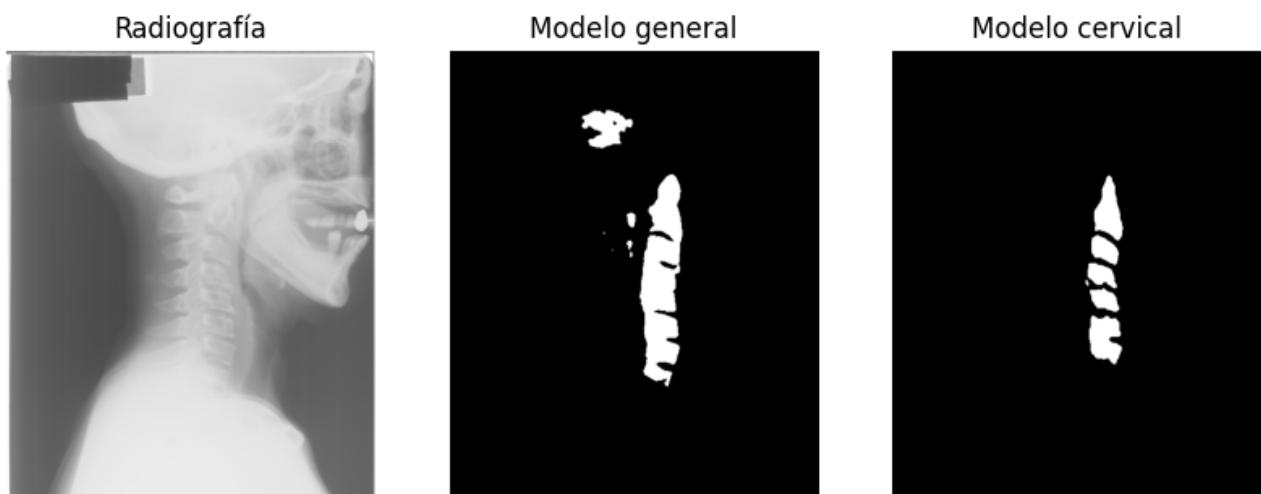


Figura 3.5: Comparación de segmentación en una imagen cervical del conjunto NHANES II

Se observa que, si bien el modelo especializado no obtuvo los mejores resultados cuantitativos, logra una separación entre vértebras y una reducción del ruido en comparación con el modelo general. Este comportamiento sugiere que la especialización con el modelo específico por región puede favorecer la precisión cuando existe una mayor variabilidad morfológica, como ocurre en la región cervical. En particular, el modelo especializado identifica con mayor claridad estructuras complejas como la vértebra C1, que el modelo general tiende a fusionar con otras o a segmentar de forma poco coherente, dificultando su correcta identificación en etapas posteriores.

Región lumbar: En la Figura 3.6 se muestra una imagen del conjunto NHANES II correspondiente a la región lumbar. A la izquierda se presenta la radiografía original; en el centro, la segmentación obtenida con el modelo general; y a la derecha, la segmentación generada con el modelo especializado en lumbares.

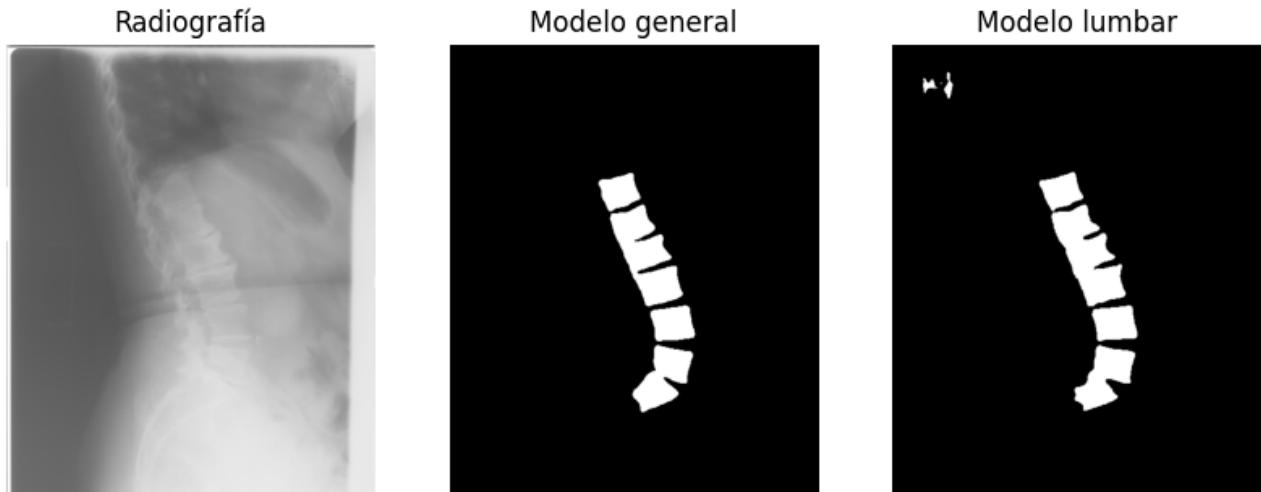


Figura 3.6: Comparación de segmentación en una imagen lumbar del conjunto NHANES II

En este caso, el modelo general ofrece una segmentación más precisa, detectando correctamente todas las vértebras con menor grado de unión entre regiones y sin generar mayor ruido. Por el contrario, el modelo especializado tiende a fusionar vértebras y producir ruido adicional en algunas imágenes. Este comportamiento se ha repetido en otros casos del conjunto, lo que sugiere que el modelo general, al haber sido entrenado con una mayor diversidad de datos, presenta una mayor capacidad de generalización en esta región.

3.2. Calidad de las métricas clínicas

Una vez aplicadas las técnicas de segmentación y posprocesado descritas previamente, se analizó la calidad de las métricas clínicas derivadas, comparando los resultados obtenidos con las máscaras predichas frente a aquellas extraídas de las máscaras reales (cuando estaban disponibles). A continuación, se presentan los resultados cuantitativos de dicho análisis, seguido de una evaluación cualitativa visual que permite observar el impacto real de estas métricas en imágenes representativas.

3.2.1. Resultados cuantitativos

Para evaluar la precisión de las medidas extraídas a partir de las segmentaciones, se calcularon diversas métricas de error tomando como referencia las máscaras reales del conjunto NHANES II. En cada apartado, se compararon las métricas de error obtenidas para los distintos modelos.

Medidas angulares

La Tabla 3.2 muestra los resultados obtenidos para la región cervical. Se observa que el modelo especializado mejora sustancialmente las métricas respecto al modelo general, destacando una reducción del MAE de más de un grado y una mejora notable del R^2 (de 0,74 a 0,85), lo cual indica una mejor capacidad explicativa. Esto sugiere que el modelo cervical es más preciso para capturar la orientación relativa entre vértebras en esta región. Es importante aclarar que las métricas de error fueron calculadas a nivel de imagen. Es decir, para cada par de máscaras (real y predicha), se compararon y se obtuvo un valor medio de error por imagen, y finalmente se promedió este valor a lo largo del conjunto.

Tabla 3.2: Comparación de medidas angulares en la región cervical (conjunto NHANES II)

Modelo	MAE	MSE	RMSE	R^2
General	5,51	40,72	6,38	0,74
Especializado	4,22	24,55	4,95	0,85

Por el contrario, en la región lumbar (Tabla 3.3) se observa el comportamiento inverso al de las cervicales: el modelo general presenta mejores métricas de error que el especializado. Este resultado coincide con las observaciones previas realizadas en la comparación de modelos (3.1.2). El MAE es inferior en aproximadamente 0,7 grados, y el R^2 alcanza un valor de 0,95 frente al 0,90 del modelo lumbar.

Tabla 3.3: Comparación de medidas angulares en la región lumbar (conjunto NHANES II)

Modelo	MAE	MSE	RMSE	R^2
General	3,81	22,69	4,76	0,95
Especializado	4,54	35,26	5,94	0,90

Los histogramas de errores angulares (MSA) para los tres modelos, se presentan en las siguientes figuras para analizar la distribución del MAE. En la Figura 3.7, correspondiente al modelo general, se observa que los errores en la región lumbar (color naranja) tienden a concentrarse entre 1 y 6 grados, mientras que los errores en la región cervical (color azul) se distribuyen más ampliamente, con mayor frecuencia superando los 10 grados. Esta mayor dispersión en cervicales está alineada con los resultados métricos obtenidos.

La Figura 3.8 muestra el histograma del modelo especializado en cervicales. En comparación con el modelo general, se observa una reducción considerable de los errores más altos, con más de 40 casos concentrados por debajo de 5 grados, y una frecuencia mucho menor en errores superiores a 10 grados. Esto confirma la mejora cuantitativa observada previamente.

Por el contrario, el histograma del modelo especializado en lumbares (Figura 3.9) muestra un pico de la distribución de los errores ligeramente mayor, concentrando la mayoría de los errores por debajo de 7 grados, mostrando una distribución más compacta y centrada en torno a valores más bajos. Frente al modelo general que presenta una distribución más dispersa.

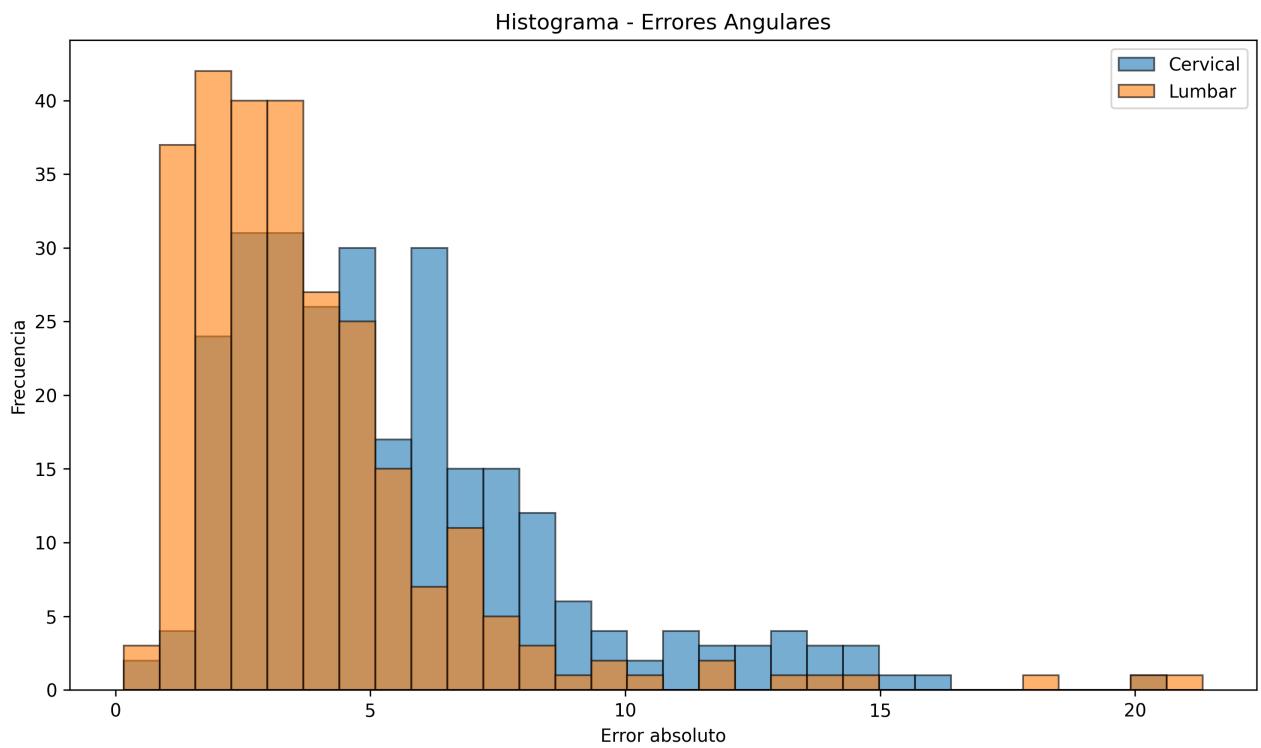


Figura 3.7: Histograma de errores angulares - modelo general por región.

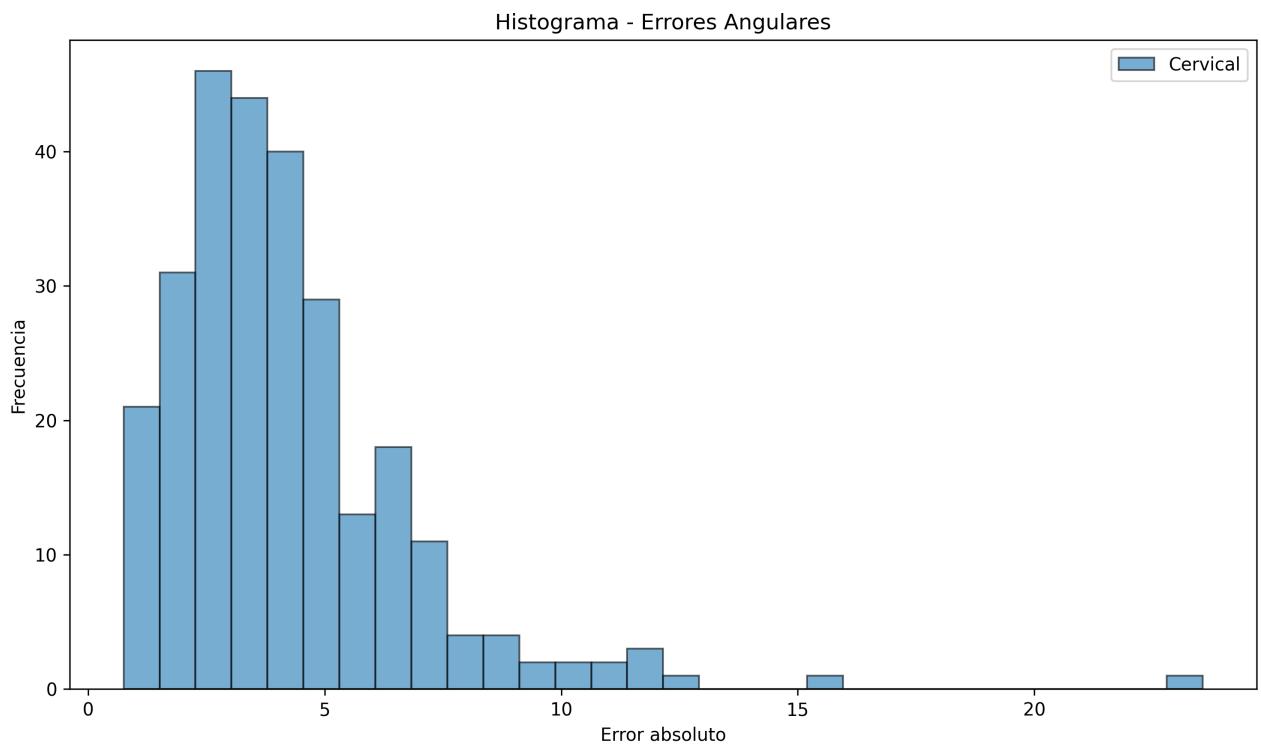


Figura 3.8: Histograma de errores angulares - modelo especializado en cervicales

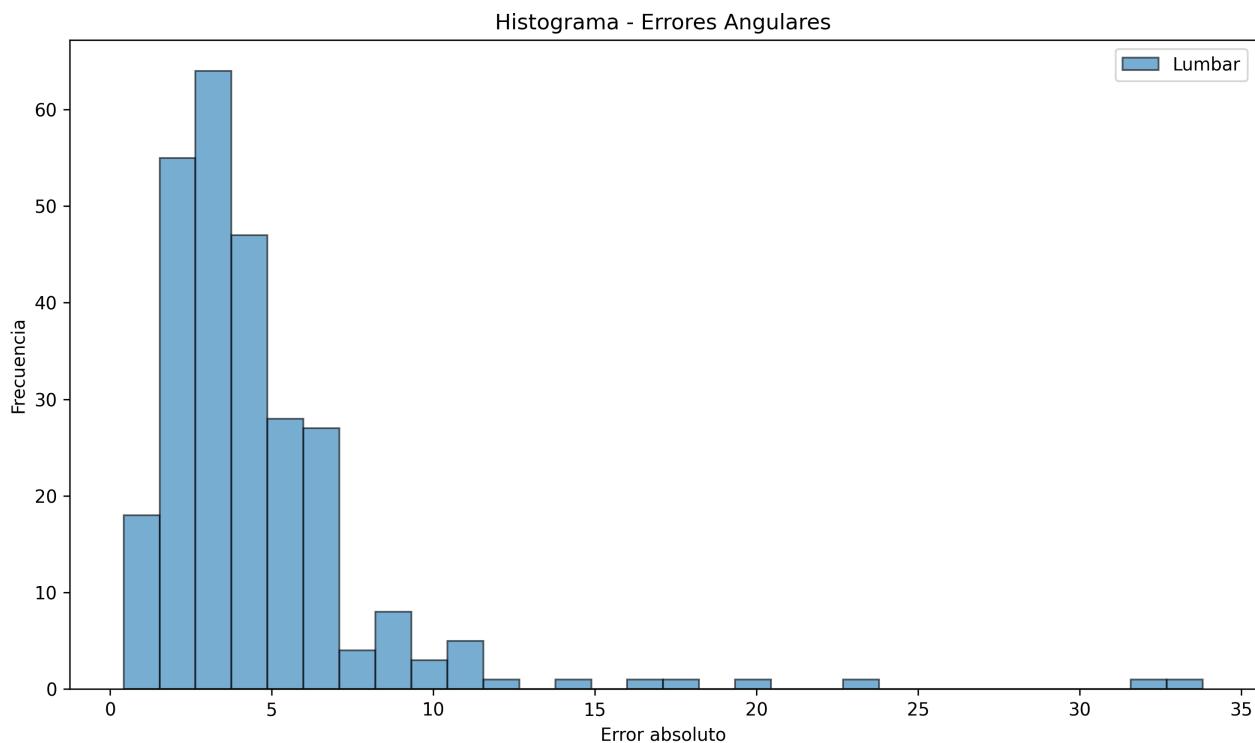


Figura 3.9: Histograma de errores angulares - modelo especializado en lumbares

Medidas traslacionales

De forma análoga al análisis anterior, la Tabla 3.4 muestra los resultados obtenidos para la región cervical al comparar las medidas traslacionales obtenidas a partir de las segmentaciones frente a las máscaras reales. En este caso, el modelo especializado vuelve a superar al modelo general en todas las métricas de error, con una reducción destacada del MSE de 71,56 frente a los 47,79 del modelo general; y un incremento en el coeficiente de determinación R^2 , que pasa de 0,69 a 0,80. Este resultado sugiere una mayor capacidad del modelo especializado para estimar correctamente los desplazamientos entre vértebras en esta región.

Tabla 3.4: Comparación de medidas traslacionales en la región cervical (conjunto NHANES II)

Modelo	MAE	MSE	RMSE	R^2
General	7,45	71,56	8,46	0,69
Especializado	6,17	47,79	6,91	0,80

En el caso de la lumbar (Tabla 3.5), se repite el patrón ya observado en las métricas angulares: el modelo general muestra un comportamiento más estable y preciso que el especializado. Aunque las diferencias no son tan marcadas en métricas como el MAE con menos de un píxel de diferencia; sí resulta que se perciben en el MSE, que penaliza con mayor intensidad los errores grandes, mostrando una diferencia de 17,81 píxeles. El R^2 alcanza un valor superior de 0,85 en el modelo general frente al 0,73 del especializado, lo que sugiere que para regiones con menor complejidad morfológica, un modelo general puede ser suficiente.

Tabla 3.5: Comparación de medidas traslacionales en la región lumbar (conjunto NHANES II)

Modelo	MAE	MSE	RMSE	R^2
General	9,1512	127,6659	11,2989	0,8468
Especializado	9,7657	145,4849	12,0617	0,7278

Se presentan a continuación los histogramas de errores traslacionales (MSA) para ambas regiones anatómicas, calculados tanto para el modelo general como para los especializados. Estas visualizaciones permiten analizar la distribución de errores y la presencia de valores atípicos, proporcionando un complemento visual a las métricas analizadas.

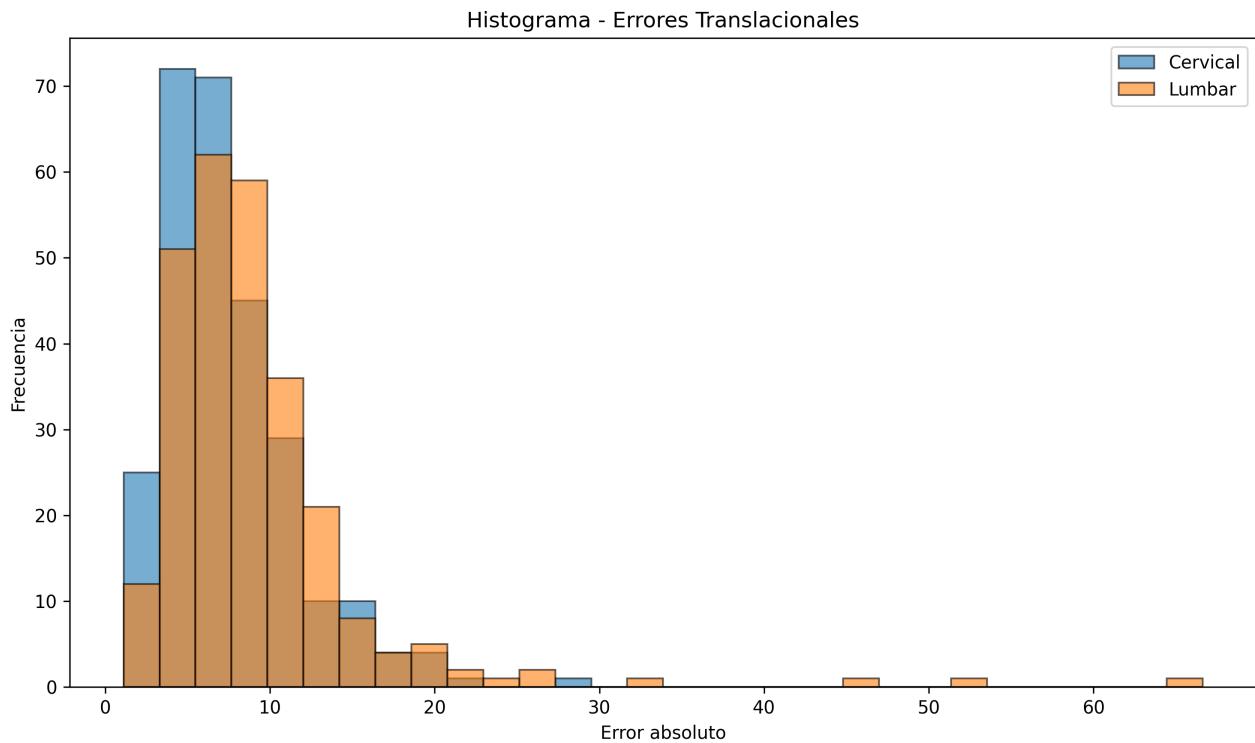


Figura 3.10: Histograma de errores translacionales - modelo general por región.

En la Figura 3.10, correspondiente al modelo general, se observa que la mayor parte de los errores de la región lumbar (color naranja) se concentran entre los 4 y 10 píxeles, con una barra larga que se extiende hasta más de 60 casos para algunos errores. En contraste, los errores en la región cervical (color azul) se distribuyen de forma más uniforme entre los 2 y 12 píxeles, aunque con menos frecuencia en los extremos altos, que como mucho alcanzan los 40 casos. Esta mayor dispersión y presencia de outliers en la región lumbar indica una mayor inestabilidad en la predicción de desplazamientos en esta zona por parte del modelo general.

Al observar la Figura 3.11, correspondiente al modelo especializado en cervicales, se nota una mejora clara respecto al modelo general. La mayoría de los errores se agrupan entre 4 y 8 píxeles, con un número mucho menor de errores por encima de 10 píxeles. Esta distribución más compacta refleja un comportamiento más estable y preciso, lo que refuerza los resultados cuantitativos obtenidos

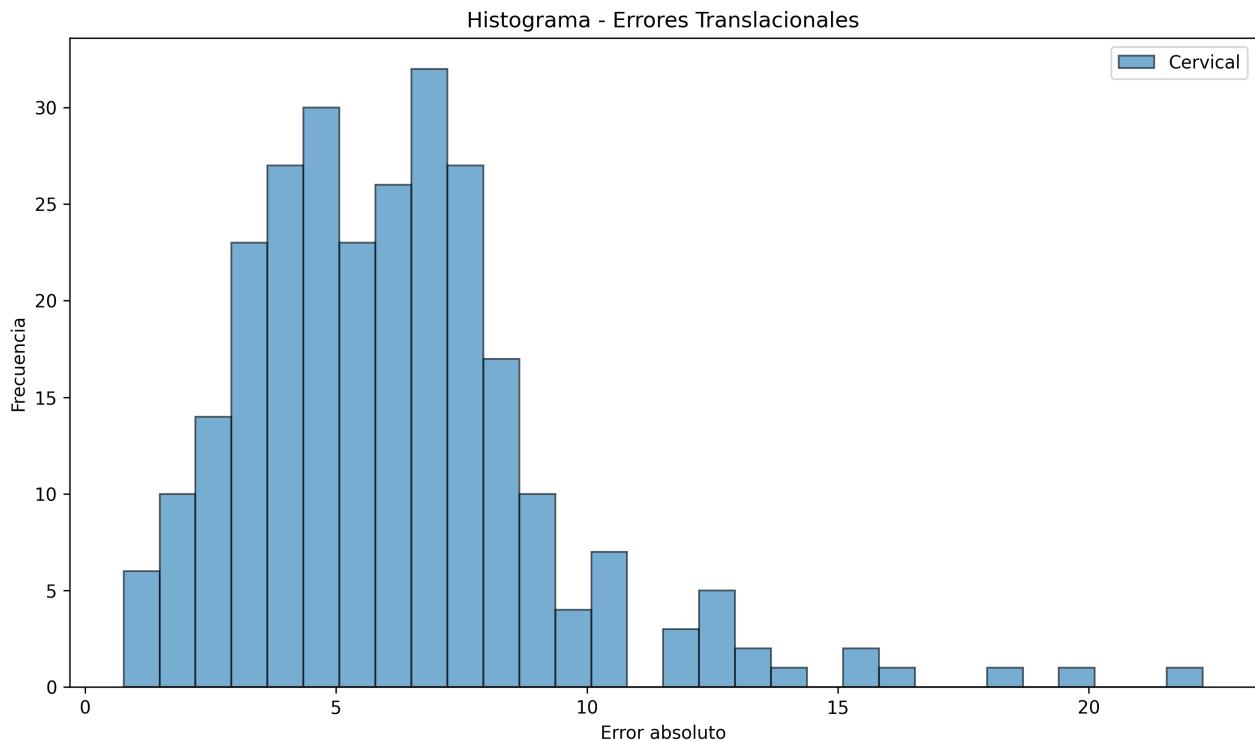


Figura 3.11: Histograma de errores translacionales - modelo cervical por región.

previamente en esta región.

Por último, la Figura 3.12 muestra los errores translacionales para el modelo especializado en la región lumbar. Los errores se concentran entre 6 y 12 píxeles, siguen siendo frecuentes; se observa una reducción de la dispersión, pero con una mayor concentración de errores alrededor de 9 píxeles (alcanzando los 70 casos).

3.2.2. Resultados cualitativos

En esta sección se muestran visualmente las medidas clínicas (angulares y translacionales) superpuestas sobre las imágenes radiográficas, permitiendo evaluar visualmente el sistema. Se analizan ejemplos de los conjuntos NHANES II y clínico, así como comparaciones entre los modelos entrenados.

Conjunto NHANES II

En este caso se analizan los resultados visuales de las medidas clínicas extraídas a partir de las segmentaciones generadas por el modelo general, comparadas con las medidas obtenidas desde las máscaras reales del conjunto NHANES II. En todos los casos, se representa a la izquierda la máscara real y a la derecha la segmentación automática.

- **Medidas angulares:**

- Contornos de las vértebras en amarillo.
- Eje principal en rojo.

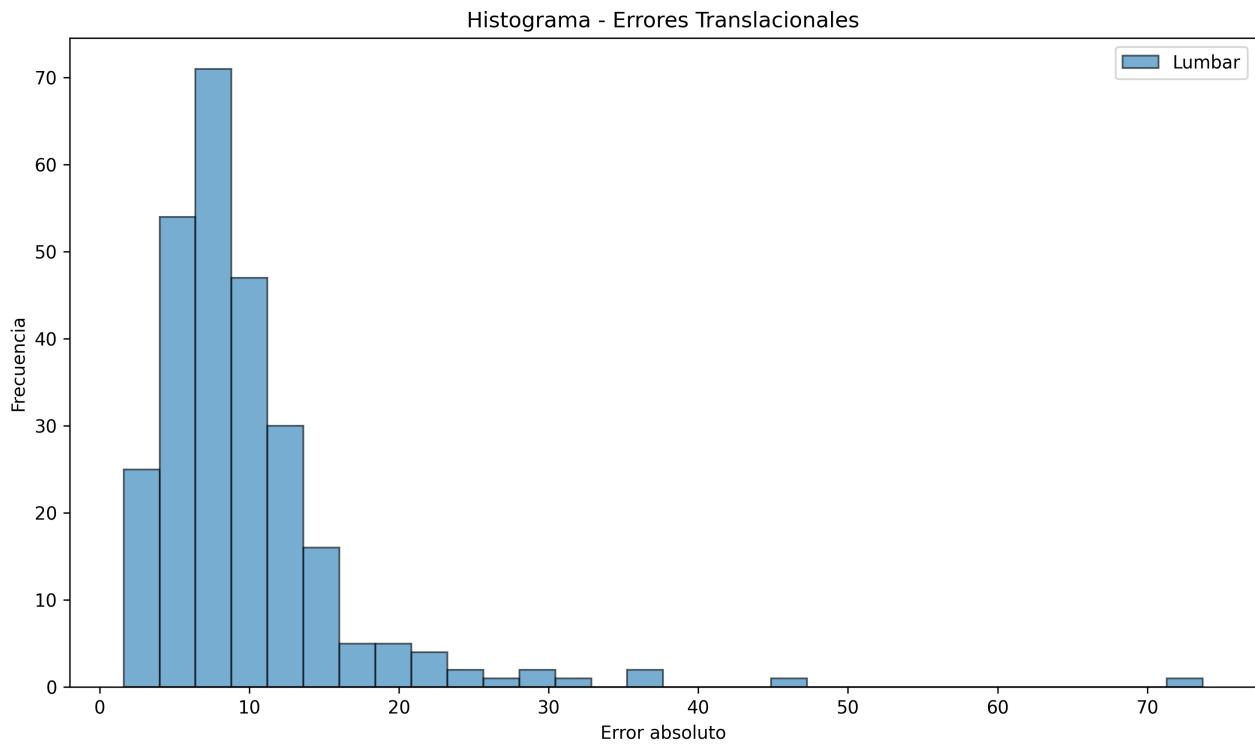


Figura 3.12: Histograma de errores traslacionales - modelo lumbar por región.

- Ángulo intervertebral en azul, representado entre $[0^\circ, 90^\circ]$ para preservar la morfología natural de la columna.

- **Medidas translacionales:**

- Contornos de las vértebras en amarillo.
- Eje intermedio entre vértebras en naranja.
- Proyección de cada vértebra en rojo, perpendicular desde la vértebra hacia el eje.
- Distancia en píxeles como una línea azul entre las proyecciones y valores numéricos.

Medidas angulares. En la Figura 3.13, se observa cómo la segmentación logra capturar de forma general la morfología y posición de las vértebras cervicales. Sin embargo, se aprecian pequeñas variaciones en el contorno y orientación de algunas vértebras que repercuten directamente en la estimación de los ángulos. Por ejemplo, el ángulo entre las dos vértebras superiores disminuye considerablemente en la segmentación, posiblemente debido a una rotación leve en el eje principal de la vértebra superior (C1). Estas diferencias, aunque visualmente sutiles, explican la presencia de errores angulares observados en los histogramas y métricas de error.

En el caso de la región lumbar (Figura 3.14), la segmentación automática reproduce con bastante precisión la orientación y forma de las vértebras. Las diferencias angulares entre las medidas extraídas de la máscara real y de la segmentación automática son mínimas. Este buen alineamiento refuerza los resultados cuantitativos, donde el modelo general obtenía mejores métricas en esta región.

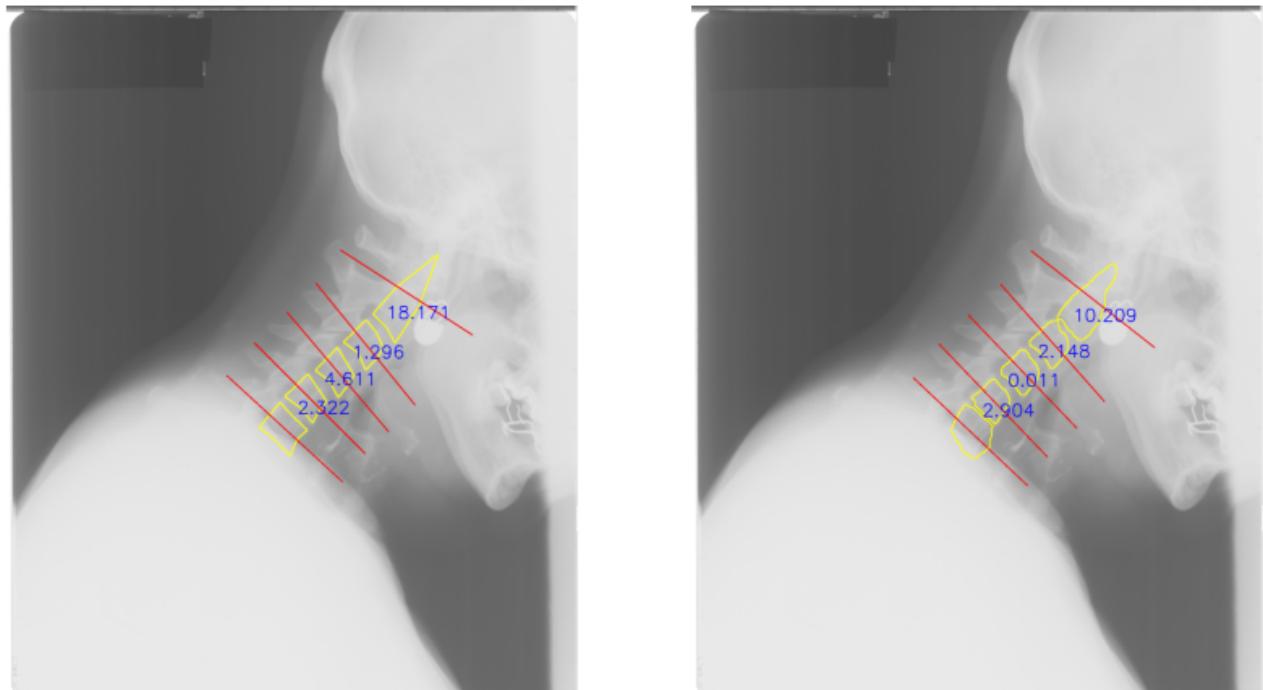


Figura 3.13: Medidas angulares en región cervical. (Izquierda) Original. (Derecha) Modelo.

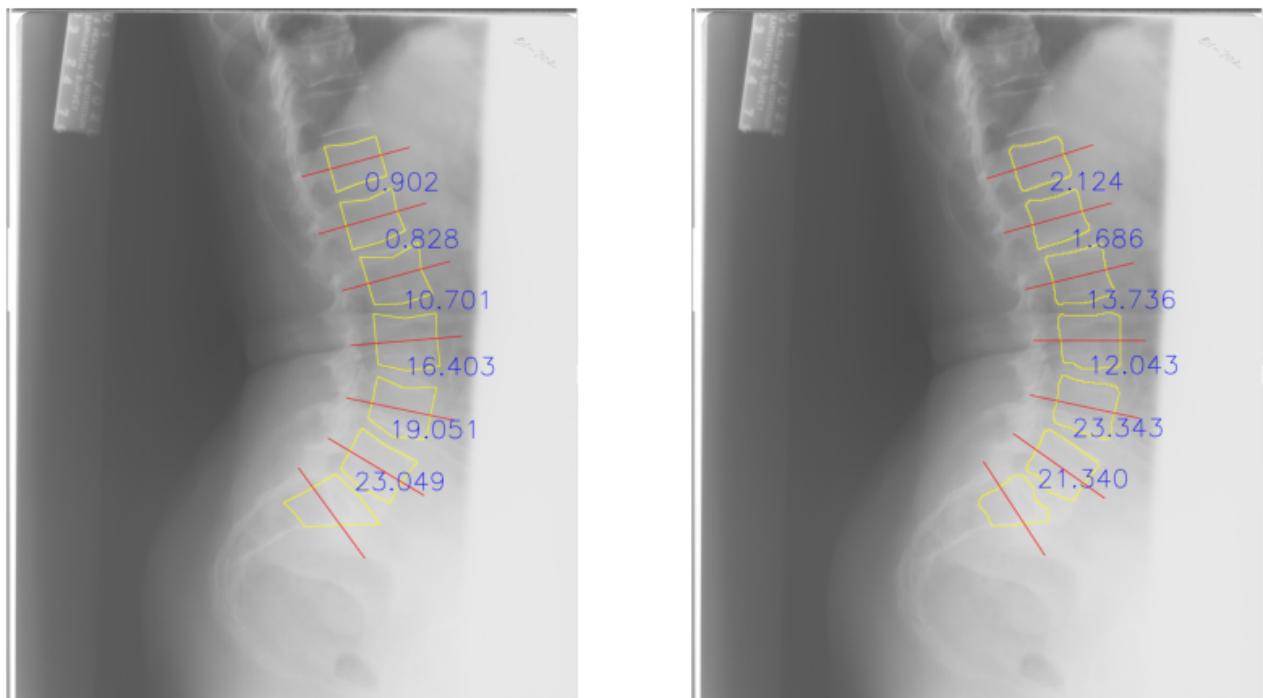


Figura 3.14: Medidas angulares en región lumbar. (Izquierda) Original. (Derecha) Modelo.

Medidas translacionales. En la Figura 3.15, se observan varias diferencias en cuanto al desplazamiento, encontrando entre 1 hasta 16 píxeles de diferencia en algunos casos. Este tipo de errores explica la mayor dispersión de errores en los histogramas y el valor alto del MAE en esta región para el modelo general.

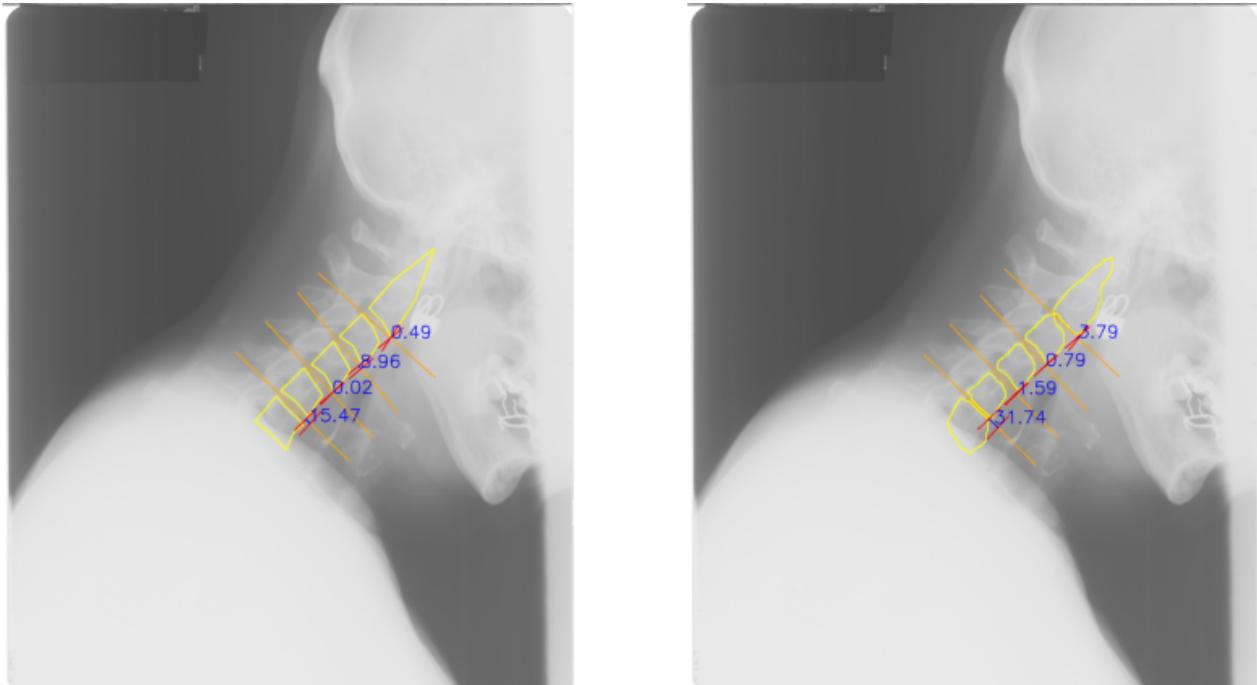


Figura 3.15: Medidas translacionales en región cervical. (Izquierda) Original. (Derecha) Modelo.

En la región lumbar (Figura 3.16), las proyecciones se alinean con mayor precisión. El eje intermedio está bien estimado, y los desplazamientos laterales entre vértebras son coherentes. Salvo pequeñas diferencias en la orientación de la vértebra más baja, la segmentación permite replicar las medidas originales con alta fiabilidad. Esta estabilidad concuerda con los resultados numéricos, donde el modelo general alcanza una mayor precisión.

Conjunto de pruebas en condiciones reales

En este apartado se muestran los resultados cualitativos de las métricas clínicas calculadas a partir de las segmentaciones del modelo general, aplicadas a radiografías proporcionadas por el médico colaborador. Como estas imágenes no cuentan con máscaras de referencia, el análisis se centra exclusivamente en las interpretaciones de los resultados generados.

Se han seleccionado ejemplos de ambas regiones anatómicas (cervical y lumbar) para las dos medidas principales: angulares y translacionales. El esquema visual mantiene la misma convención de colores descrita previamente, con la única diferencia de que las letras se pintan en verde para facilitar la visualización, ya que el color azul resulta poco legible en estas radiografías, que son más oscuras.

Medidas angulares. En la Figura 3.17, correspondiente a la región cervical, se observa que las vértebras están bien localizadas, pero el contorno y orientación de algunas de ellas presentan irregularidades que afectan al cálculo del eje principal. Esto repercute en los ángulos entre vértebras, con

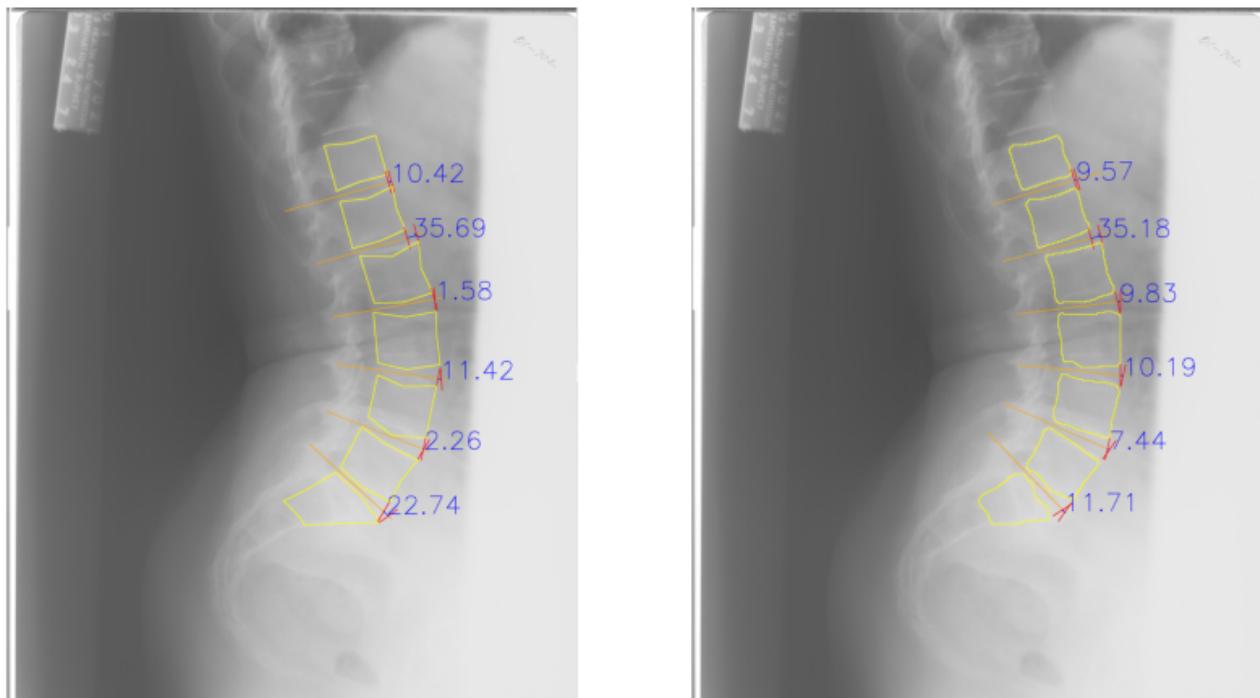


Figura 3.16: Medidas translacionales en región lumbar. (Izquierda) Original. (Derecha) Modelo.

valores algo dispersos y desviaciones morfológicas. A pesar de ello, la representación general de la curvatura cervical se mantiene coherente.

En contraste, la Figura 3.18 muestra un caso lumbar en el que la segmentación resulta más limpia y definida. Las vértebras están bien separadas, y los ángulos obtenidos presentan una progresión regular, coherente con las curvaturas esperadas. Esta diferencia cualitativa refuerza la idea de que la región lumbar tiende a presentar menos complejidad y mayor estabilidad en los resultados, incluso con conjuntos externos.

Medidas translacionales. La Figura 3.19 muestra una segmentación cervical donde ocurre lo mismo que en el caso anterior, donde el contorno y orientación de algunas de las vértebras presentan irregularidades. Esto introduce errores en las proyecciones laterales y, por tanto, en las distancias medidas. Aun así, las diferencias no parecen excesivas, y el sistema conserva la dirección general de la columna.

Por su parte, en la Figura 3.20, los resultados son significativamente más estables. El eje intermedio entre vértebras está bien definido y las líneas de proyección permiten una estimación coherente de las distancias laterales. La distribución de las medidas es más homogénea, lo que indica que el sistema generaliza mejor en esta región anatómica, incluso fuera del conjunto de entrenamiento.

Comparación de modelos: general vs especializados

Para evaluar de forma cualitativa el impacto del entrenamiento especializado en la calidad de las métricas clínicas frente al entrenamiento general, se presentan a continuación comparaciones visuales entre los resultados obtenidos sobre las imágenes del conjunto clínico.

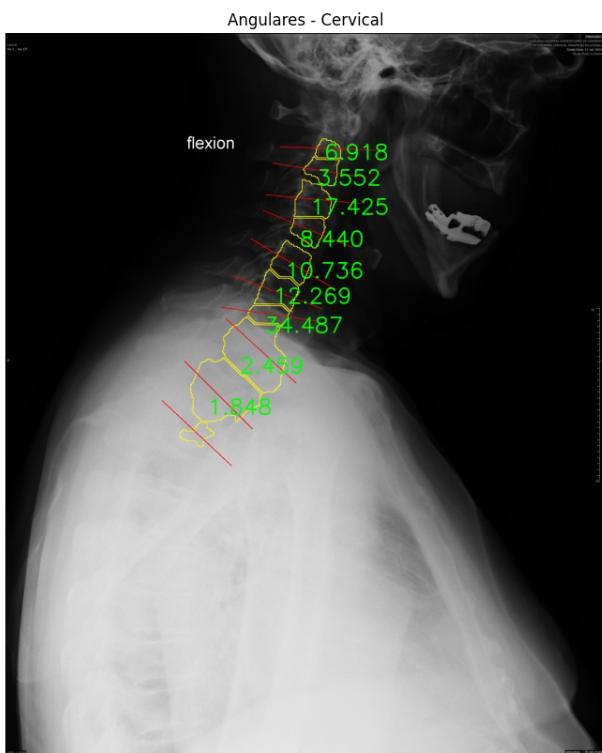


Figura 3.17: Medidas angulares en región cervical.

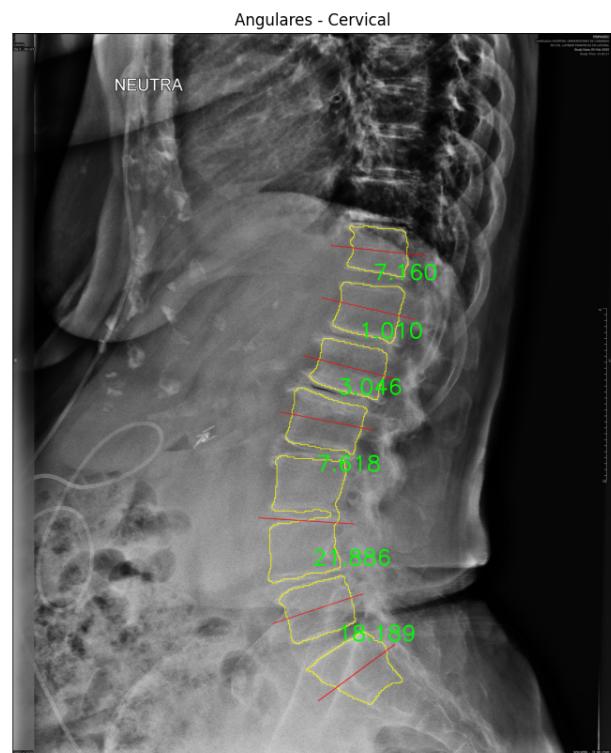


Figura 3.18: Medidas angulares en región lumbar.

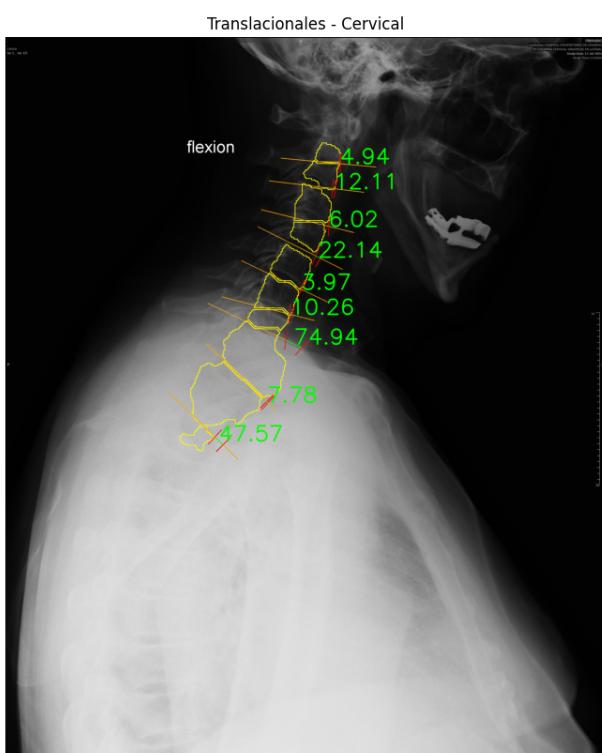


Figura 3.19: Medidas translacionales en región cervical.

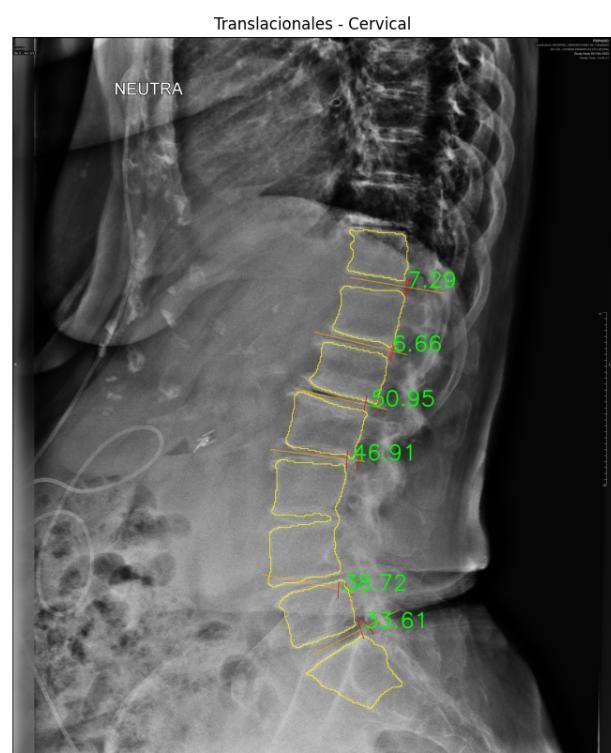


Figura 3.20: Medidas translacionales en región lumbar.

A diferencia del conjunto NHANES II, este conjunto presenta mayor variabilidad morfológica, lo que permite observar más claramente las diferencias en la estimación de medidas clínicas.

Medidas angulares. En la Figura 3.21 se presenta la comparación entre el modelo general (izquierda) y el modelo especializado en cervicales (derecha). Aunque ambos modelos logran obtener medidas razonables, se aprecia una diferencia clara en la calidad de la segmentación y, por consiguiente, en la precisión de las medidas. Esta diferencia es especialmente visible en la vértebra C1, que aparece claramente delimitada en la segmentación del modelo especializado. Este comportamiento refuerza las observaciones cuantitativas previas, donde el modelo especializado en cervicales demostró un mejor rendimiento tanto en segmentación como en las métricas clínicas derivadas.

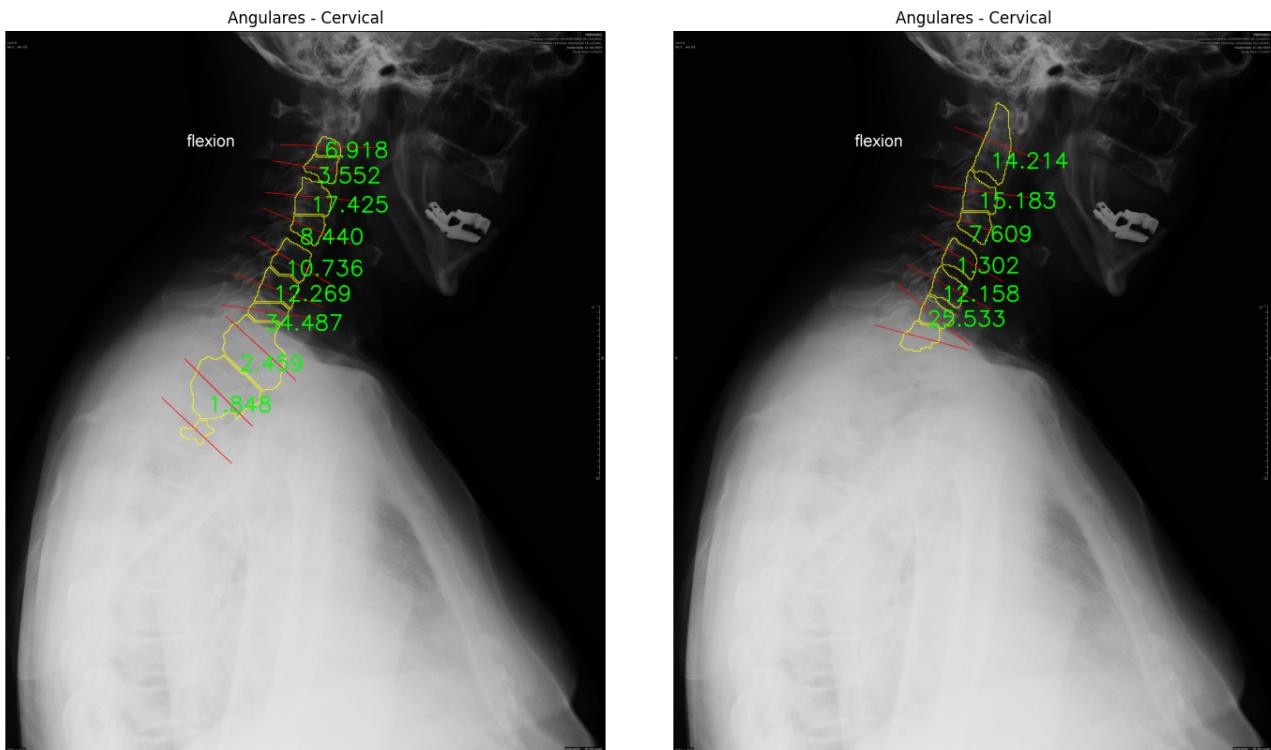


Figura 3.21: Comparación de modelos sobre cervicales - (Izquierda) Modelo general. (Derecha) Modelo especializado.

En la figura 3.22, correspondiente a la región lumbar, ambos modelos obtienen resultados satisfactorios. En ambos casos, las vértebras se encuentran correctamente separadas, salvo una o dos fusiones puntuales, que difieren entre modelos. Estas pequeñas diferencias afectan a los valores individuales de las medidas, pero no se traducen en una ventaja clara para ninguno de los modelos.

Medidas traslacionales. En la Figura 3.23 se compara el resultado del modelo general (izquierda) con el del modelo especializado en cervicales (derecha). Al igual que en las medidas angulares, ambos modelos logran calcular las métricas sin errores evidentes de trazado. Sin embargo, se aprecia una mayor coherencia en la alineación vertebral y ejes intermedios en el modelo especializado. Esto se traduce en medidas más consistentes, evitando valores atípicos como los observados con el modelo general, donde algunos desplazamientos son exageradamente altos.

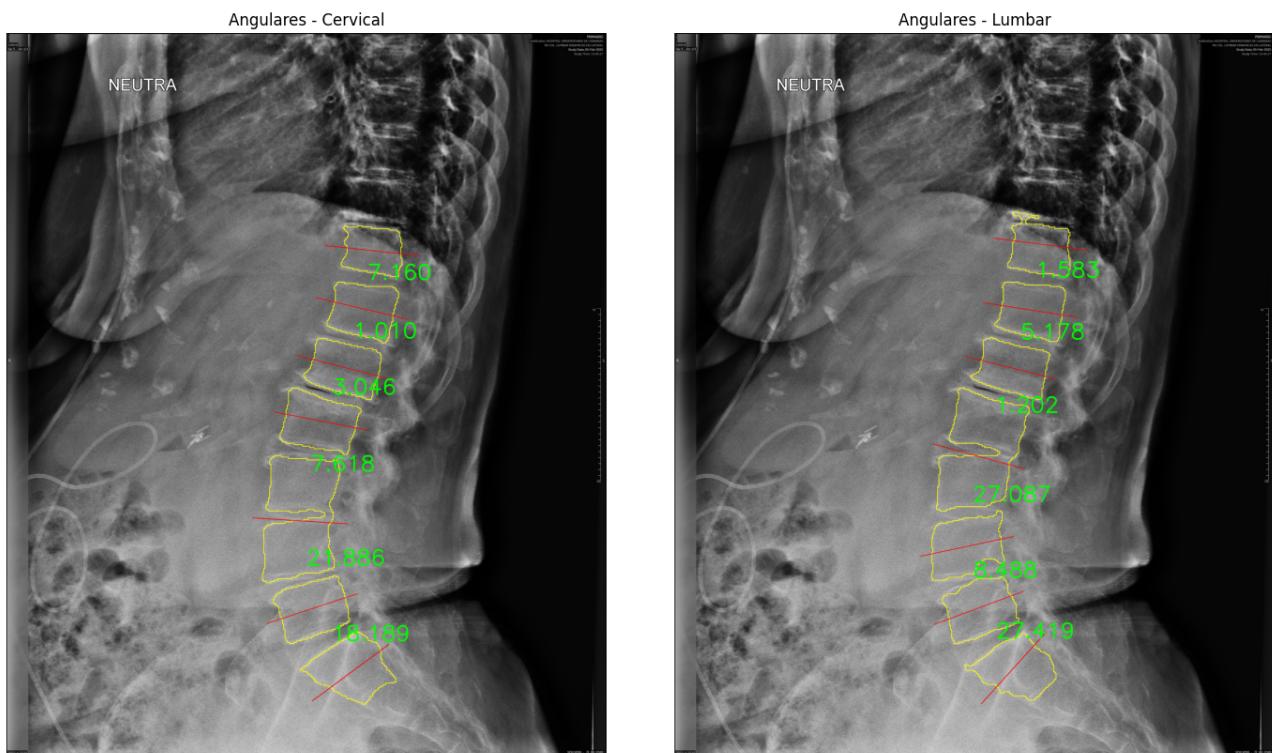


Figura 3.22: Comparación de modelos sobre lumbares - (Izquierda) Modelo general. (Derecha) Modelo especializado.

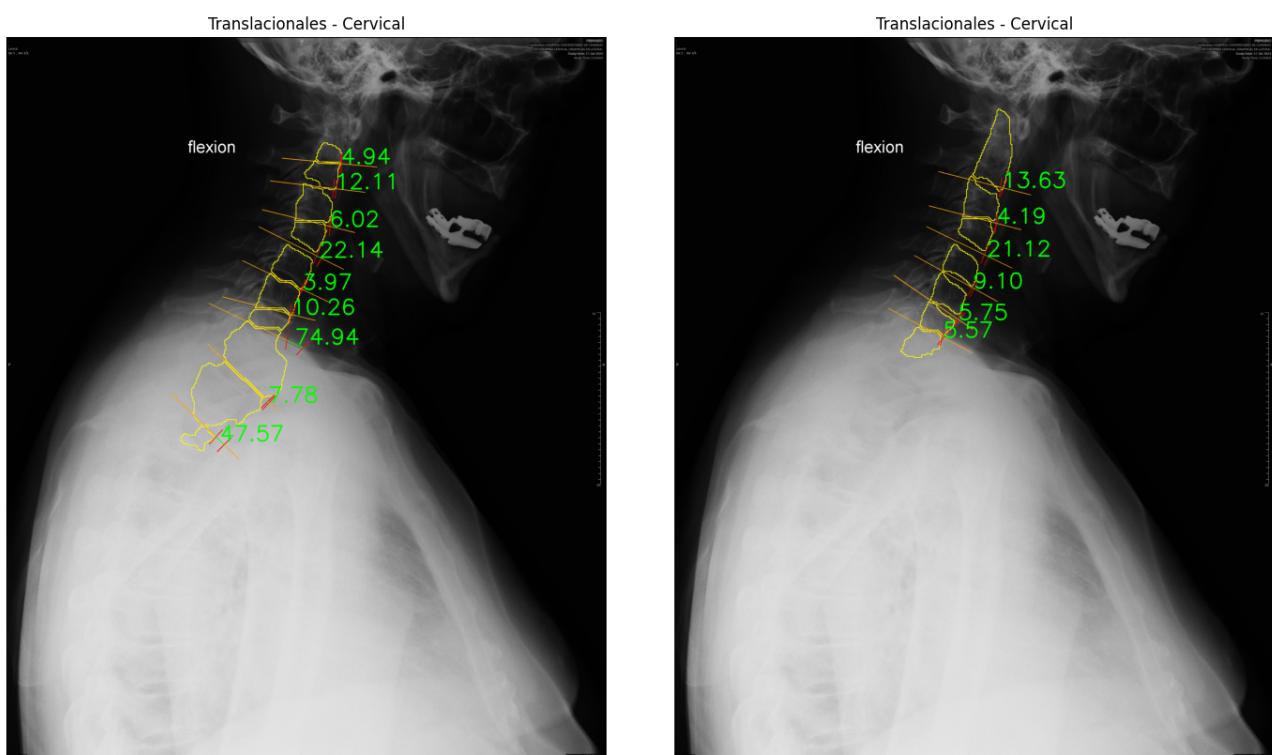


Figura 3.23: Comparación de modelos sobre cervicales - (Izquierda) Modelo general. (Derecha) Modelo especializado.

En la Figura 3.24, se presentan los resultados para la región lumbar. En este caso, ambas segmentaciones producen medidas razonables y con ligeras diferencias. A pesar de que en algunos puntos el modelo especializado introduce mayores desplazamientos, estos no parecen deberse a fallos del trazado, sino más bien a pequeñas variaciones en la segmentación, como ya se había observado.

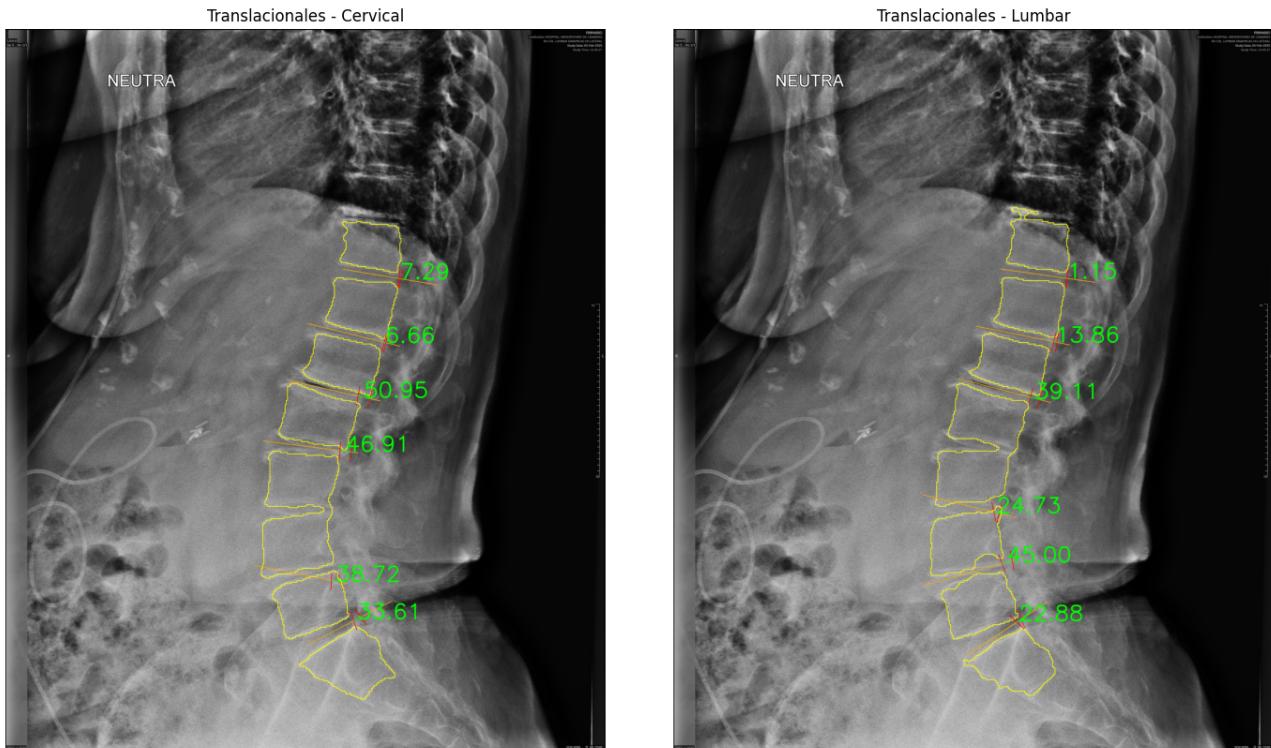


Figura 3.24: Comparación de modelos sobre lumbares - (Izquierda) Modelo general. (Derecha) Modelo especializado.

3.3. Análisis general

Los resultados obtenidos a lo largo del estudio muestran que la calidad de las métricas clínicas depende en gran medida de la precisión en la segmentación inicial. Si bien las métricas agregadas, como el MAE o el RMSE, permiten cuantificar el rendimiento medio del sistema, el análisis visual ha permitido identificar casos de error y evaluar con mayor claridad los resultados finales.

En la región cervical, el modelo especializado logró mejoras claras respecto al modelo general, tanto en términos de segmentación como en la precisión de las medidas clínicas. Esto se evidenció no solo en los valores medios de error, sino también en la reducción de la dispersión observada en los histogramas y en las imágenes con medidas clínicas superpuestas. En cambio, en la región lumbar, el modelo general se comportó de forma más estable, mostrando menor dispersión en los errores y menor sensibilidad a casos problemáticos.

Este patrón sugiere que la complejidad morfológica y la variabilidad de cada región influyen directamente en la necesidad o no de especialización del modelo. Mientras que en regiones como la lumbar un modelo general es suficiente, en zonas como la cervical, donde hay vértebras pequeñas y estructuras óseas adyacentes, la especialización es beneficiosa.

Cabe destacar que en varios casos, las diferencias en las métricas no reflejan necesariamente errores en el cálculo de las medidas clínicas, sino imperfecciones en la segmentación previa. En estos casos, incluso pequeñas uniones o deformaciones afectan a la extracción de ángulos o desplazamientos, lo que refuerza la importancia de mantener una segmentación morfológicamente coherente.

3.3.1. Limitaciones y casos problemáticos

Durante el desarrollo del sistema se identificaron varias limitaciones que afectaron al rendimiento final, especialmente en escenarios clínicos reales o en casos con segmentaciones complejas. A continuación se destacan las más relevantes:

- **Validación externa con imágenes clínicas reales.** Uno de los principales desafíos fue la aplicación de los modelos entrenados con el conjunto NHANES II sobre el conjunto clínico proporcionado por el médico colaborador. Aunque en la región lumbar el modelo general mostró una generalización aceptable, en la región cervical el rendimiento fue considerablemente inferior. Las diferencias en orientación anatómica, calidad de imagen, contraste y morfología entre ambos conjuntos afectaron notablemente la segmentación automática. En particular, el modelo general tendió a fusionar o deformar vértebras, lo que generó errores elevados en las medidas clínicas derivadas. Esto pone de manifiesto la necesidad de entrenar o ajustar los modelos sobre datos representativos del entorno clínico objetivo.
- **Dependencia del posprocesado.** Precisamente por las dificultades en la segmentación, es necesario un posprocesado previo al cálculo de medidas clínicas. Aunque en términos generales los criterios aplicados permitieron mejorar la coherencia estructural de las máscaras, también se identificaron casos donde dichos criterios fallaron o resultaron demasiado restrictivos. En la Figura 3.25, se observa un ejemplo donde el sistema seleccionó como columna vertebral un grupo no ideal debido a su mejor alineación aparente, descartando la secuencia anatómicamente correcta. Por otro lado, en la Figura 3.26, el posprocesado no logra identificar ningún grupo válido, eliminando completamente la segmentación por no cumplirse los umbrales mínimos exigidos. Este tipo de casos refleja la necesidad de contar con mecanismos adaptativos o tolerantes a variaciones extremas para evitar pérdidas de información clínica relevante.

3.3.2. Potencial aplicación médica

A pesar de las limitaciones señaladas, el sistema desarrollado presenta un buen potencial para su aplicación como herramienta de apoyo clínico. La automatización de medidas angulares y traslacionales a partir de imágenes radiográficas podría suponer una reducción importante del tiempo necesario para evaluar alineaciones vertebrales.

En su estado actual, el sistema ha demostrado que es capaz de replicar con buena aproximación medidas clínicas a partir de segmentaciones automáticas, al menos en imágenes similares a las de entrenamiento. En contextos controlados, podría ser útil como preanálisis automatizado, que posteriormente sería validado o corregido por un especialista.

Para su futura adopción en entornos clínicos, sería necesario mejorar la robustez frente a variabilidad interpaciente y trabajar con imágenes anotadas clínicas reales para ajustar el sistema a condiciones de adquisición más diversas. Además, sería relevante integrar un módulo de confianza o validación automática que advierta al usuario en caso de segmentaciones dudosas.

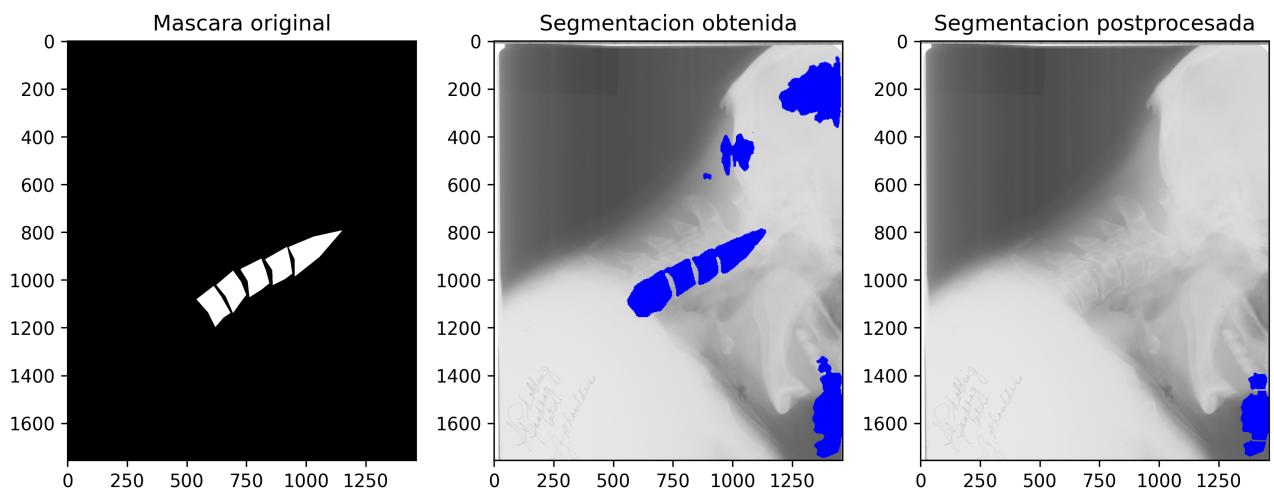


Figura 3.25: Ejemplo de selección errónea del grupo candidato a columna debido a una alineación geométrica más favorable

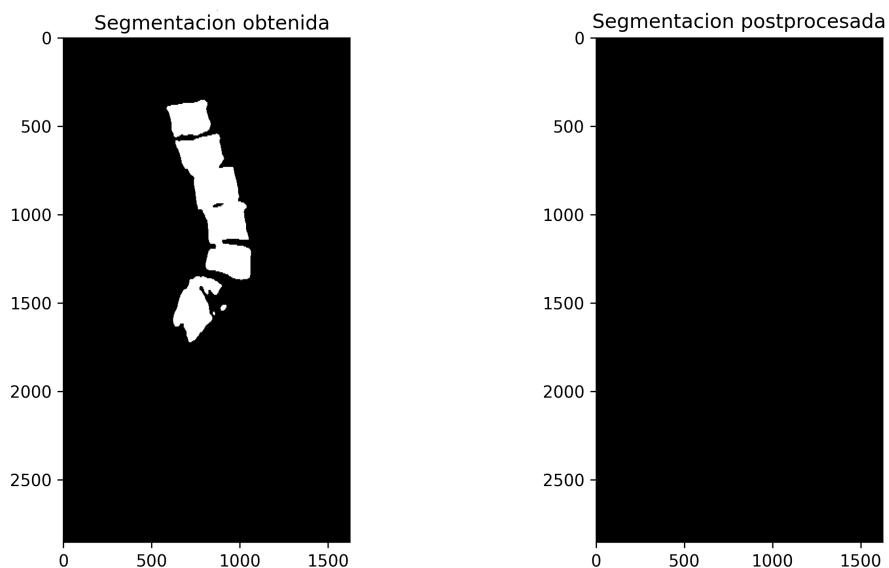


Figura 3.26: Ejemplo en el que el posprocesado no identifica ningún grupo válido por no cumplirse los criterios mínimos de selección.

Capítulo 4

Conclusiones y líneas futuras

A lo largo de este trabajo se ha desarrollado un sistema para la segmentación vertebral en radiografías laterales y la extracción de métricas clínicas morfológicas. El flujo de trabajo abarcó la preparación y limpieza de los datos, el entrenamiento de los modelos con arquitectura U-Net, el posprocesado de las segmentaciones para mejorar la calidad estructural y el cálculo de medidas clínicas angulares y traslacionales.

Los resultados obtenidos demuestran que el sistema es capaz de identificar la columna vertebral y extraer medidas con una precisión razonable. En regiones con mayor variabilidad morfológica, como la cervical, el uso de un modelo especializado aportó claras ventajas: el error medio absoluto (MAE) en medidas angulares se redujo de $5,51^\circ$ a $4,22^\circ$ y el coeficiente de determinación R^2 aumentó de 0,74 a 0,85. En medidas traslacionales, el MAE disminuyó de 7,45 a 6,17 píxeles, lo que refuerza su mayor precisión en esta zona.

Por el contrario, en la región lumbar, donde la morfología es más homogénea y presenta menor ruido, el modelo general ofreció mejores resultados que el especializado. En este caso, el MAE angular fue de $3,81^\circ$ frente a $4,54^\circ$, y el R^2 alcanzó un valor de 0,95, superando al modelo especializado (0,90). Este patrón sugiere que, para regiones anatómicas con menor complejidad, un modelo general puede ser suficiente.

Respecto a la capacidad de generalización del sistema sobre imágenes clínicas externas, se observó un comportamiento satisfactorio en la mayoría de los casos, especialmente en la región lumbar, donde el modelo general logró obtener segmentaciones coherentes. En el caso de la región cervical, el modelo general presentó segmentaciones deficientes; sin embargo, el modelo especializado logró mejorar significativamente los resultados, con mejor definición de las vértebras y una menor presencia de ruido. Esta diferencia refuerza la necesidad de emplear modelos especializados en regiones anatómicas con mayor variabilidad morfológica.

Finalmente, la evaluación cualitativa, tanto sobre el conjunto NHANES II como sobre imágenes clínicas reales, permitió verificar visualmente el impacto que tienen las segmentaciones en la estimación de medidas clínicas. Estas visualizaciones confirmaron que incluso pequeñas imperfecciones en los contornos o uniones entre vértebras pueden repercutir en la calidad de las métricas angulares y traslacionales, lo que refuerza la importancia del posprocesado y la coherencia morfológica.

4.1. Limitaciones del estudio

A pesar de los avances alcanzados, el trabajo presenta algunas limitaciones relevantes:

- **Limitada variabilidad en los conjuntos de entrenamiento.** El modelo se entrenó exclusivamente con el conjunto NHANES II, lo cual limita su capacidad de generalización ante variaciones clínicas como la calidad de imagen, el ángulo de adquisición, o la morfología específica de ciertas poblaciones. La falta de disponibilidad pública de otros conjuntos etiquetados ha limitado la capacidad de generalización del modelo.
- **Tiempo y recursos acotados.** Dado el carácter académico del proyecto, algunas fases, como la mejora del posprocesado o la exploración de arquitecturas alternativas y su parametrización, no pudieron desarrollarse en profundidad.
- **Evaluación clínica limitada.** La validación externa se realizó de forma preliminar, sin una retroalimentación directa del entorno clínico. La colaboración con el personal médico se vio restringida por la limitada disponibilidad de este personal y los plazos de entrega.

4.2. Líneas futuras de desarrollo

Existen múltiples direcciones en las que este trabajo podría ser extendido y mejorado:

- **Ampliación del conjunto de datos.** Colaborar con equipos médicos o instituciones hospitalarias es importante para obtener nuevos conjuntos de imágenes etiquetadas que permitan entrenar modelos con mayor diversidad anatómica y calidad.
- **Exploración de nuevas arquitecturas.** Además de U-Net, podrían evaluarse arquitecturas modernas como U-Net++, DeepLabv3+ o Swin UNet, que podrían mejorar la precisión en segmentaciones complejas. Aunque en este trabajo se optó por U-Net con ResNet34 preentrenado en ImageNet por su buen rendimiento inicial, no se realizó una comparativa exhaustiva con otras configuraciones.
- **Ajuste y exploración de hiperparámetros.** Aunque se seleccionaron hiperparámetros razonables durante el entrenamiento, no se realizó una búsqueda exhaustiva. Podría aplicarse una optimización de estos, para mejorar el rendimiento de los modelos sin necesidad de modificar la arquitectura.
- **Mejora del posprocesado.** El flujo actual podría optimizarse con estrategias más robustas para la validación de alineación, segmentación de vértebras fusionadas y selección automática del mejor grupo candidato.
- **Generalización a otras regiones y vistas.** Ampliar el sistema para incluir otras regiones de la columna, como la torácica, o permitir su funcionamiento con radiografías anteroposteriores o con distintas orientaciones.
- **Desarrollo de una interfaz médica.** Si el sistema fuera a aplicarse en un entorno clínico real, sería conveniente implementar una interfaz interactiva que permita a los profesionales cargar una radiografía y visualizar los resultados del análisis de forma clara y útil.
- **Colaboración activa con profesionales sanitarios.** Para abordar las limitaciones actuales y convertir esta herramienta en un asistente real de diagnóstico, sería imprescindible involucrar directamente a especialistas durante las fases de desarrollo, evaluación y validación clínica.

Capítulo 5

Summary and Conclusions

Throughout this project, a system has been developed for vertebral segmentation in lateral radiographs and the extraction of morphological clinical metrics. The workflow covered data preparation and cleaning, training of models with U-Net architecture, post-processing of the segmentation to improve structural quality, and the computation of angular and translational clinical measurements.

The results demonstrate that the system is capable of identifying the spinal column and extracting measurements with reasonable accuracy. In regions with greater morphological variability, such as the cervical spine, the use of a specialized model provided clear advantages: the mean absolute error (MAE) in angular measurements was reduced from 5.51° to 4.22° , and the coefficient of determination R^2 increased from 0.74 to 0.85. In translational measurements, the MAE decreased from 7.45 to 6.17 pixels, reinforcing the improved precision in this area.

In contrast, in the lumbar region, where morphology is more homogeneous and less noisy, the general model performed better than the specialized one. In this case, the angular MAE was 3.81° versus 4.54° , and the R^2 reached 0.95, outperforming the specialized model (0.90). This pattern suggests that, for anatomically simpler regions, a general model may be sufficient.

Regarding the system's generalization capacity on external clinical images, satisfactory performance was observed in most cases, especially in the lumbar region, where the general model was able to produce coherent segmentations. In the cervical region, however, the general model yielded deficient segmentations; nonetheless, the specialized model significantly improved the results, offering better vertebra delineation and reduced noise. This difference highlights the importance of using specialized models in anatomical regions with greater morphological variability.

Finally, the qualitative evaluation, conducted on both the NHANES II dataset and the real clinical images, allowed for a visual verification of how segmentation impacts the estimation of clinical measurements. These visualizations confirmed that even small imperfections in vertebral contours of connections can affect the quality of angular and translational metrics, reinforcing the importance of post-processing and morphological coherence.

5.1. Study Limitations

Despite the progress made, the project presents several relevant limitations:

- **Limited variability in training datasets.** The model was trained exclusively on the NHANES II dataset, which limits its ability to generalize to clinical variations such as image quality, acquisition angle, or the specific morphology of certain populations. The lack of publicly available labeled datasets has constrained the model's generalization capacity.

- **Limited time and resources.** Due to the academic nature of the project, some phases, such as improving the post-processing stage or the exploration and tuning of alternative architectures, could not be developed in depth.
- **Limited clinical evaluation.** External validation was conducted in a preliminary manner, without direct feedback from a clinical setting. Collaboration with medical staff was constrained by their limited availability and project deadlines.

5.2. Future Work

There are several directions in which this work could be extended and improved:

- **Expansion of the dataset.** Collaborating with medical teams or hospital institutions is important to obtain new labeled image datasets that allow training of models with greater anatomical diversity and image quality.
- **Exploration of new architectures.** Beyond U-Net, modern architectures such as U-Net++, DeepLabv3+, or Swin UNet could be evaluated to improve segmentation accuracy in complex cases.
- **Hyperparameter Tuning and Exploration.** Although reasonable hyperparameters were selected during training, an exhaustive search was not performed. Optimization of these could be applied to improve model performance without requiring architectural modifications.
- **Improvement of post-processing.** The current workflow could be optimized with more robust strategies for alignment validation, segmentation of fused vertebrae, and automatic selection of the best candidate group.
- **Generalization to other regions and views.** Expanding the system to include other regions of the spine, such as the thoracic area, or enabling its operation with anteroposterior radiographs or different orientations would enhance its applicability.
- **Development of a medical interface.** If the system were to be used in a real clinical environment, it would be advisable to implement an interactive interface that allows professionals to upload a radiograph and clearly visualize the analysis results.
- **Active collaboration with healthcare professionals.** To address current limitations and turn this tool into a real diagnostic assistant, it would be essential to involve specialists directly during development, evaluation, and clinical validation phases.

Capítulo 6

Presupuesto

Tabla 6.1: Presupuesto de Equipos y Licencias

Descripción	Cantidad	Coste (€)
Portátil	1	62,41
Servicios en la Nube	5 meses	14,95
Subtotal de Equipos y Licencias		77,36

Tabla 6.2: Coste de Mano de Obra

Descripción	Horas	Coste (€)
Precio por Hora		20,00
Total de Horas de Trabajo	729	
Costo Total del Trabajo Humano		14.580,00

Tabla 6.3: Coste Total del Proyecto

Descripción	Coste Total (€)
Subtotal de Equipos y Licencias	77,36
Costo Total del Trabajo	14.580,00
Coste Total del Proyecto	14.657,36

Bibliografía

- [1] Wiki Integración IV. *Anatomía columna vertebral*. URL: https://integracion-iv.fandom.com/es/wiki/Anatom%C3%ADA_columna_vertebral.
- [2] Yao Chen et al. «VertXNet: an ensemble method for vertebral body segmentation and identification from cervical and lumbar spinal X-rays». En: *Scientific Reports* 14 (2024). DOI: [10.1038/s41598-023-49923-3](https://doi.org/10.1038/s41598-023-49923-3).
- [3] Pengfei Cheng et al. «Automatic vertebrae localization and segmentation in CT with a two-stage Dense-U-Net». En: *Scientific Reports* 11 (2021). DOI: [10.1038/s41598-021-01296-1](https://doi.org/10.1038/s41598-021-01296-1).
- [4] Jing Liu et al. «WGAN-based multi-structure segmentation of vertebral cross-section MRI using ResU-Net and clustered transformer». En: *Scientific Reports* 14 (2024). DOI: [10.1038/s41598-024-79244-y](https://doi.org/10.1038/s41598-024-79244-y).
- [5] Dong Liu et al. «Lenke Classification of Scoliosis Based on Segmentation Network and Adaptive Shape Descriptor». En: *Applied Sciences* 13 (2023). DOI: [10.3390/app13063905](https://doi.org/10.3390/app13063905).
- [6] Renato R. Maaliw III. «SCOLIONET: An Automated Scoliosis Cobb Angle Quantification Using Enhanced X-ray Images and Deep Learning Models». En: *Journal of Imaging* 9 (2023). DOI: [10.3390/jimaging9120265](https://doi.org/10.3390/jimaging9120265).
- [7] Yuanhan Mo et al. *NHANES II Spinal X-ray with Extra Labels For Reference Vertebral Bodies (C2, S1)*. Repositorio de datos en Zenodo. 2023. DOI: [10.5281/zenodo.10223910](https://doi.org/10.5281/zenodo.10223910). URL: <https://doi.org/10.5281/zenodo.10223910>.
- [8] Tania Évora Vargas Martínez. *Aplicación de Inteligencia Artificial para la segmentación de imágenes médicas de la columna vertebral*. GitHub. 2025. URL: https://github.com/Tania-Vargas/TFG-segmentacion_columna.git.
- [9] Pavel Yakubovskiy. *Segmentation Models PyTorch*. GitHub repository. 2019. URL: https://github.com/qubvel-org/segmentation_models.pytorch.
- [10] Olaf Ronneberger, Philipp Fischer y Thomas Brox. «U-Net: Convolutional Networks for Biomedical Image Segmentation». En: *Medical Image Computing and Computer-Assisted Intervention (MICCAI)* (2015). DOI: [10.48550/arXiv.1505.04597](https://doi.org/10.48550/arXiv.1505.04597).
- [11] Kaiming He et al. «Deep Residual Learning for Image Recognition». En: *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*. 2016, págs. 770-778. DOI: [10.48550/arXiv.1512.03385](https://doi.org/10.48550/arXiv.1512.03385).
- [12] Li Fei-Fei et al. *ImageNet*. 2021. URL: <https://www.image-net.org/index.php>.
- [13] Tongxue Zhou, Su Ruan y Stéphane Canu. «A review: Deep learning for medical image segmentation using multi-modality fusion». En: *Array* 3-4 (2019), pág. 100004. ISSN: 2590-0056. DOI: <https://doi.org/10.1016/j.array.2019.100004>. URL: <https://www.sciencedirect.com/science/article/pii/S2590005619300049>.