

EchoNet Análisis de Ecocardiograma del Ventrículo Izquierdo

Fausto Alejandro Palma Cervantes^[A01639224], Alejandro Pizarro Chavez^[A01633784], Diego Rosas^[A01634154], Alan Ricardo Vilchis Arceo^[A01640260],
and Héctor Manuel Cárdenas Yáñez^[A01634154]

Instituto Tecnológico y de Estudios Superiores de Monterrey,
Campus Guadalajara, Zapopan, Jalisco, MX

Abstract. Este artículo explora el conjunto de datos de ecocardiogramas EchoNetc, que contiene múltiples ecocardiogramas que muestran el ventrículo izquierdo del corazón junto con dos archivos separados que registran los nombres y las coordenadas del ventrículo izquierdo de cada video dividido por fotogramas. El objetivo principal de este trabajo es evaluar modelos de máscara y landmarks (puntos de referencia) para poder detectar el ventrículo izquierdo del corazón de un dataset de ecocardiogramas. Este análisis incluye el modelo CNN, la arquitectura U-Net y los puntos de referencia.

Keywords: ecocardiograma · aprendizaje automático · máscaras · landmarks (puntos de referencia) · modelo CNN · arquitectura U-net

1 Introducción

Exo es una empresa de dispositivos e imágenes médicas que desarrolla dispositivos de ultrasonido portátiles e inteligencia artificial para aplicaciones terapéuticas y de imágenes [1]. Trabajando junto a Exo en uno de sus muchos proyectos, este trabajo se centra en analizar ecocardiogramas a través de diferentes métodos. Un ecocardiograma es una prueba de ultrasonido que verifica la estructura y función de el corazón. Es un esquema gráfico del movimiento del corazón mediante el uso de ultrasonido (ondas sonoras de alta frecuencia) de una varilla manual colocada en el pecho para tomar fotografías de las válvulas y cámaras de su corazón, que ayudan a evaluar la acción de bombeo de su corazón. [2]

Para comprender la importancia de estos análisis, primero debemos comprender la importancia del ventrículo izquierdo del corazón en el cuerpo. El ventrículo izquierdo es la cámara más grande y fuerte del corazón, bombea sangre oxigenada a través de la válvula aórtica (AoV) hacia la aorta (Ao), la arteria principal que transporta sangre oxigenada al resto del cuerpo. El análisis del ventrículo izquierdo ayuda a estudiar el flujo sanguíneo, la presión arterial y la oxigenación de la sangre, que, en conjunto, nos ayudan a detectar enfermedades como la hipertrofia del ventrículo izquierdo. Al analizar y entrenar un modelo de inteligencia artificial, podemos crear un modelo capaz de analizar un conjunto

de datos de ecocardiogramas y aprender a detectar síntomas físicos de enfermedades del ventrículo izquierdo. En general, este modelo puede ayudar a los médicos a prevenir enfermedades relacionadas con el corazón durante las etapas primarias y a diagnosticar enfermedades o anomalías. [3]

El objetivo principal de este trabajo es evaluar modelos de máscara y de landmarks (puntos de referencia) para analizar ecocardiogramas del ventrículo izquierdo del corazón y determinar cuál es más eficiente. El conjunto de datos que se nos proporcionó incluye múltiples videos de ecocardiogramas que se muestran desde el corazón. El programa debe poder analizar el vídeo e identificar correctamente el ventrículo izquierdo del corazón como se muestra en la imagen a continuación. Exo proporcionó estas imágenes de la base de datos de ecocardiogramas.

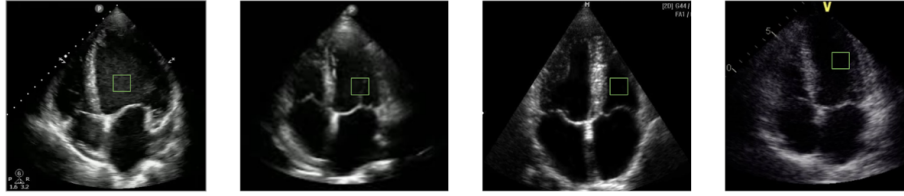


Fig. 1. Conjunto de ecocardiogramas que identifican el ventrículo izquierdo del corazón.

2 Marco Teórico

La aplicación de la tecnología de inteligencia artificial (IA) en imágenes cardiovasculares se ha convertido en un punto de interés en los últimos años, dado que puede reducir los costos y el tiempo del tratamiento y ayudar a evitar pruebas innecesarias. En el diagnóstico y tratamiento de enfermedades cardíacas, las técnicas de IA se han aplicado a la vectocardiografía, electrocardiografía y ecocardiografía, entre otras. Cabe resaltar que hablando específicamente de la implementación de IA con ecocardiografía, varias empresas han logrado avances significativos.

2.1 Red Neuronal Convolutacional (CNN)

Una red neuronal convolutacional (CNN o ConvNet) es una arquitectura de red para aprendizaje profundo que aprende directamente de los datos. Esta red puede resultar beneficiosa para identificar patrones en imágenes para reconocer objetos, clases y categorías. Una CNN consta de una capa de entrada, una capa de salida y varias capas ocultas intermedias. Estas capas realizan operaciones que

modifican los datos, para comprender sus características particulares. Las 3 capas más comunes son convolución, activación (ReLU) y agrupación. [4]

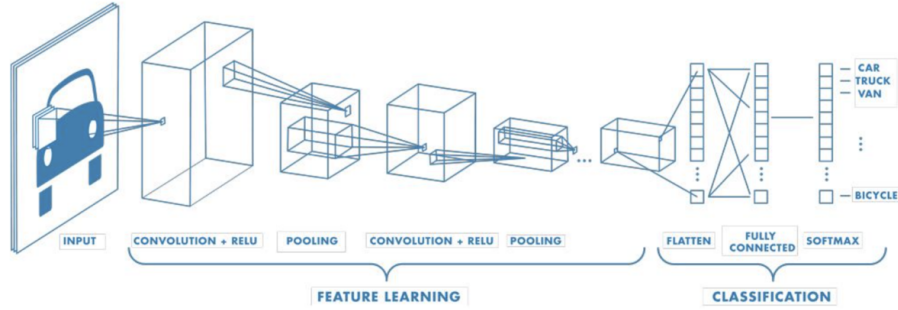


Fig. 2. Una red con múltiples capas convolucionales. Se aplican filtros a las imágenes de entrenamiento con diferentes resoluciones y el resultado resultante de la convolución de cada imagen se utiliza como entrada para la siguiente capa [5].

Sin embargo, a diferencia de una red neuronal tradicional, una CNN tiene pesos y valores de sesgo compartidos, que son los mismos para todas las neuronas ocultas en una capa determinada. Esto significa que todas las neuronas ocultas detectan las mismas características, como bordes o formas, en diferentes regiones de la imagen. Esto hace que la red sea tolerante a la traducción de objetos en una imagen. [4]

2.2 Architecture U-Net

La segmentación de imágenes mediante U-Net permite asignar una etiqueta a cada píxel de la imagen, recreando este mismo proceso a múltiples áreas hasta cubrir toda la imagen. Esto nos proporciona más información sobre una imagen que la detección de objetos, la cual dibuja un cuadro delimitador alrededor del objeto detectado. Resulta inclusive más eficiente que la clasificación de imágenes donde se asigna una etiqueta al objeto. U-Net es una técnica de segmentación semántica, esto consiste en dividir una imagen digital en varias regiones denominadas segmentos para su análisis, para la segmentación de imágenes médicas. La arquitectura del modelo es bastante simple: un codificador (para reducir la resolución) y un decodificador (para aumentar la resolución) con conexiones de salto. [6]

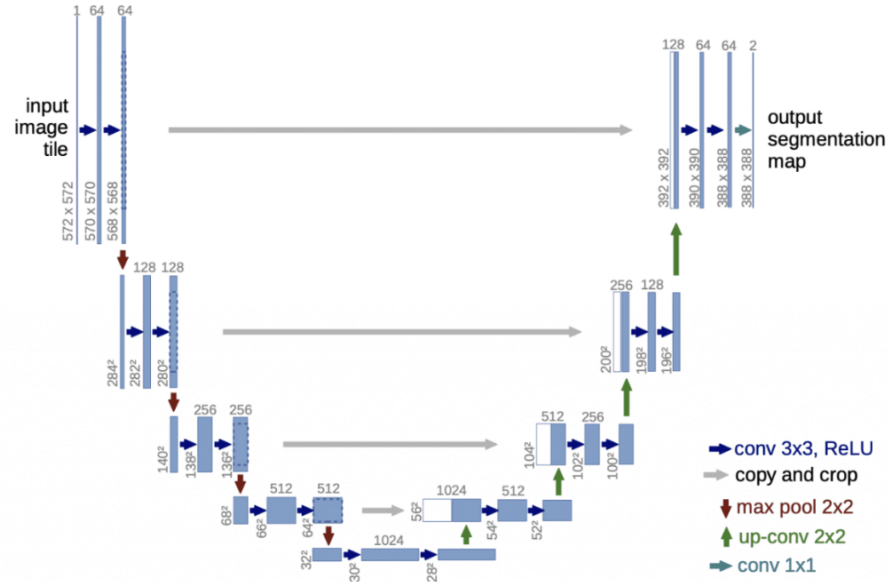


Fig. 3. Arquitectura U-Net [7]

2.3 Puntos de Referencia (Landmarks Detection)

En informática, la detección de landmarks es el proceso de localizar coordenadas importantes para poder identificar secciones relevantes de una imagen. Este modelo se utiliza principalmente con fines de navegación, por ejemplo, en visión robótica para crear mapas para imágenes satelitales. [6] Sin embargo, investigaciones recientes han extendido estos métodos a otros campos como el reconocimiento facial y las aplicaciones médicas anatómicas. En medicina, se han utilizado puntos de referencia para la identificación de tumores u órganos en imágenes médicas. Estas partes del cuerpo pueden considerarse puntos de referencia, y la implementación de un modelo que las detecte puede ayudar en el diagnóstico y la planificación del tratamiento. [8]

3 Trabajos Similares

En el artículo [9], los autores crean una investigación sobre la tecnología de inteligencia artificial en conjuntos de datos de CXR de pulmones. En su enfoque, decidieron implementar máscaras a partir de los datos de CXR del pulmón. Estas máscaras se utilizan para alimentar su modelo, que funciona de la siguiente manera. Se entrenó un codificador automático de eliminación de ruido convolucional en el conjunto de datos RSNA CXR. El conjunto de datos se dividió a nivel de paciente en 75, 13 y 12 por ciento para entrenamiento, validación y

prueba, respectivamente. El modelo UMS-RepCXR-Denoising tiene cuatro capas convolucionales (3×3) en el codificador para comprimir la entrada a su representación espacial latente. Al implementar otros métodos para disminuir el error cuadrático medio, acelerar el entrenamiento del modelo, lo que resulta en una convergencia más rápida y, en general, mejorar la precisión general del modelo, llegaron a un modelo funcional. [9]

Otro ejemplo es el artículo de Circulation [10]. Este artículo tiene como objetivo explicar el proceso y resultado del desarrollo de un modelo de clasificación de enfermedades utilizando redes neuronales convolucionales. Se entrenaron más de 20 redes neuronales convolucionales alimentándose con máscaras capaces de analizar los ventrículos izquierdo y derecho del corazón. Los resultados fueron increíblemente prometedores con una precisión del 90 por ciento para analizar y predecir enfermedades o anomalías cardíacas. [10]

A diferencia de las investigaciones previas, el objetivo principal de este artículo es crear un modelo de inteligencia artificial capaz de detectar el ventrículo inferior izquierdo del corazón en tiempo real. Implementando un modelo U-Net de cinco capas con múltiples neuronas con la finalidad de entrenar el modelo con un dataset de ecocardiogramas. Los resultados se analizarán a comparación de los videos para poder determinar si el acuracy del modelo es correcto.

4 Metodología

4.1 Dataset

Los dos archivos proporcionados contienen información esencial para comprender mejor los ecocardiogramas y su análisis. Muestran los nombres de todos los archivos de video y la cantidad de cuadros dentro de cada video. Por último, hay un archivo que contiene un conjunto de coordenadas utilizadas para crear un contorno del ventrículo izquierdo del corazón para cada video y cuadro específico, como se muestra en la Imagen 4.

Esto es fundamental, dado que jugará un papel vital a la hora de diseñar las mascarillas. Se espera utilizar toda esta información para crear un modelo de predicción de IA utilizando máscaras y puntos de referencia. Al analizar los resultados y las predicciones, debemos determinar qué modelo es más eficiente para este escenario.

5 Equipo de Trabajo

En el transcurso del desarrollo de los modelos, se implementaron dos configuraciones de hardware con el propósito específico de llevar a cabo las tareas computacionales y el procesamiento de los datos.

FileName	X1	Y1	X2	Y2	Frame
0X100009310A3BD7FC.avi	51.26	15.35	64.93	69.13	46
0X100009310A3BD7FC.avi	50.04	17.17	53.37	16.32	46
0X100009310A3BD7FC.avi	49.16	20.41	57.09	18.39	46
0X100009310A3BD7FC.avi	48.54	23.58	60	20.67	46
0X100009310A3BD7FC.avi	47.92	26.75	62.9	22.94	46
0X100009310A3BD7FC.avi	47.96	29.76	65.81	25.22	46
0X100009310A3BD7FC.avi	48.17	32.72	68.25	27.62	46
0X100009310A3BD7FC.avi	48.37	35.69	70.39	30.09	46
0X100009310A3BD7FC.avi	48.58	38.65	72.52	32.56	46
0X100009310A3BD7FC.avi	49.01	41.56	74.15	35.17	46
0X100009310A3BD7FC.avi	49.83	44.36	75.71	37.78	46
0X100009310A3BD7FC.avi	50.65	47.17	77.27	40.4	46
0X100009310A3BD7FC.avi	51.46	49.98	78.64	43.07	46
0X100009310A3BD7FC.avi	52.55	52.72	79.49	45.87	46
0X100009310A3BD7FC.avi	53.7	55.44	80.33	48.68	46
0X100009310A3BD7FC.avi	54.37	58.29	80.61	51.62	46
0X100009310A3BD7FC.avi	51.15	62.12	80.61	54.64	46
0X100009310A3BD7FC.avi	49.34	65.6	80.43	57.7	46
0X100009310A3BD7FC.avi	50.35	68.36	80.07	60.81	46
0X100009310A3BD7FC.avi	57.51	69.56	79.7	63.91	46
0X100009310A3BD7FC.avi	71.94	68.9	79.33	67.02	46

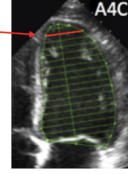


Fig. 4. Imagen de los puntos de coordenadas que delinean el ventrículo izquierdo del corazón para cada archivo y cuadro de video.

5.1 Equipo 1: Equipo Personal

Memoria RAM: 32 GB

GPU: NVIDIA GeForce RTX 3060

- Reloj de aceleración: 1882 MHz
- Reloj de memoria: 15000 MHz efectivo
- Núcleos CUDA: 3584
- Tipo de bus: PCIe 4.0
- Detalle de la memoria: 12288 MB GDDR6
- Ancho de bits de memoria: 192 bits
- Ancho de banda de memoria: 360 GB/s

CPU: AMD Ryzen 7 3700X 8-Core Processor

5.2 Equipo 2: Google Colab

Memoria RAM: 12.7 GB

GPU: NVIDIA T4

- Reloj de aceleración: 1300 MHz
- Reloj de memoria: 12000 MHz efectivo
- Núcleos CUDA: 2560
- Tipo de bus: PCIe 3.0
- Detalle de la memoria: 16384 MB GDDR6
- Ancho de bits de memoria: 256 bits
- Ancho de banda de memoria: 320 GB/s

CPU: vCPU (núcleo virtual). Google Colab no proporciona información detallada sobre el modelo o la marca de la CPU específica que se asigna a sus instancias virtuales.

6 Generador de Máscaras

Para la generación de la máscara, el código realiza operaciones de procesamiento de imágenes en los frames seleccionados de una serie de videos de ecocardiogramas. Primero, se crea un directorio para almacenar las imágenes de salida, eliminando y recreando el directorio si ya existe. Luego, utiliza un bucle para iterar sobre cada par de frames y listas de vértices correspondientes, definidos previamente. En cada iteración, se crea una región de interés (ROI) basada en los vértices especificados y se genera una máscara que resalta esta región en el frame original. Se aplican operaciones morfológicas, como dilatación y cierre, para mejorar la máscara. Finalmente, tanto la máscara como la imagen enmascarada resultante se guardan en un directorio de salida, con nombres de archivo únicos basados en el índice del bucle. Este código es crucial para el análisis de datos médicos, especialmente en el contexto de EchoNet-Dynamic, con el objetivo de extraer y procesar información relevante de imágenes ecocardiográficas dinámicas.

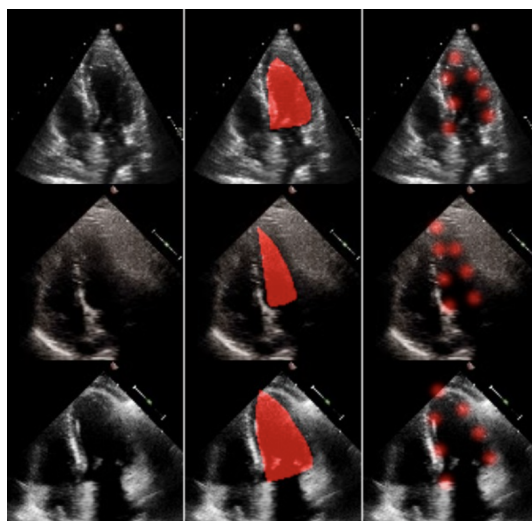


Fig. 5. Imagen de la generación de máscaras y puntos de importancia para entrenamiento. De izquierda a derecha, se muestra el frame original, la máscara y los puntos de importancia.

Posterior a la creación del dataset de entrenamiento, se generó el uso de checkpoints para guardar el progreso de entrenamiento en cada uno de los dos modelos, esto con la finalidad de tener un respaldo en caso de cualquier falla / desconexión. Adicionalmente, los checkpoints son de utilidad para poder identificar en qué punto los modelos tuvieron el mejor desempeño.

7 Puntos de Importancia

La generación de puntos de importancia va de la mano con el archivo de “Volume-Traicings”. En una primera instancia, el modelo toma las líneas perpendiculares que proporciona las coordenadas del archivo. La línea perpendicular con mayor diferencia marca el punto más alto. Con las líneas perpendiculares restantes se calcula la mayor área posible y se consiguen los dos puntos inferiores, creando un triángulo. Los puntos restantes se obtienen midiendo la distancia de los puntos inferiores con el superior y dividiendo entre tres la distancia para crear dos puntos por lado. De esta manera, se crean la máscara del ventrículo izquierdo y finalmente se guarda esta información en un arreglo. Cabe destacar que cada landmark es calculado en un canal destino, por lo que al momento de entrenar la red debemos especificar las dimensiones de la imagen así como los canales.

8 Arquitectura U-Net

En la preparación para entrenar el modelo U-Net primera se crea el Dataset. Debido a que los datos están divididos en tres categorías (TRAIN, VAL, TEST), se guardaron los datos en tres diferentes carpetas, esto resulta en que tenemos tres diferentes carpetas cada una con máscaras e imágenes de ecocardiogramas. Después, se guarda esta información en un arreglo y el programa lo transforma en un archivo. Esto ayudará a disminuir el tiempo para los siguientes entrenamientos y análisis. Finalmente, se hace el entrenamiento de la red U-Net con cinco capas y diferente cantidad de neuronas en cada capa. El entrenamiento de la red de landmarks resultó ser más tardado que la de máscaras, lógicamente ya que contienen más canales. La arquitectura de la capa contiene cinco métricas para evaluar la precisión del algoritmo: accuracy, precision, recall y dice score distribution.

Por fines de investigaciones, se decidió entrenar el modelo en dos escenarios distintos. Para los modelos de máscaras y puntos de importancia, el entrenamiento de la arquitectura U-Net se llevara acabo en un canal y en tres canales para ambos modelos. Con esto se busca encontrar los beneficios y áreas de oportunidad para determinar cuál es más eficiente.

9 Predicciones Modelos Máscaras y Landmarks

9.1 Predicciones en 1 Canal

A continuación se observan las predicciones de los modelos en 1 canal. Las predicciones de “Testing Data” se basan a partir del dataset inicial proporcionado por Exo Net. Por otro lado las predicciones de “Apical Dataset” provienen de otro dataset con mayor ruido proporcionado por el socio fromador pero externas al dataset inicial. A simple vista, se nota que las predicciones de máscaras y landmarks van conforme al ecocardiograma del ventrículo izquierdo. Los resultados están especificados y analizados en el punto 9.

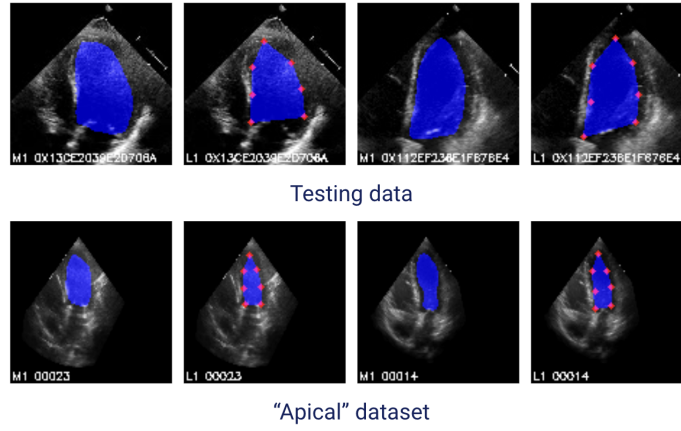


Fig. 6. Resultados de predicción de los modelos de máscaras y landmarks con un canal para testing dataset y apical dataset.

9.2 Predicciones en 3 Canales

A continuación se observan las predicciones de los modelos en 1 canal. Las predicciones de "Testing Data" se basan a partir del dataset inicial proporcionado por Exo Net. Por otro lado las predicciones de "Apical Dataset" provienen de otro dataset con mayor ruido proporcionado por el socio fromador pero externas al dataset inicial. A simple vista, se nota que las predicciones de máscaras y landmarks van conforme al ecocardiograma del ventrículo izquierdo. Sin embargo, cabe mencionar que los resultados de accuracy, precision y dice score mejoraron significativamente.

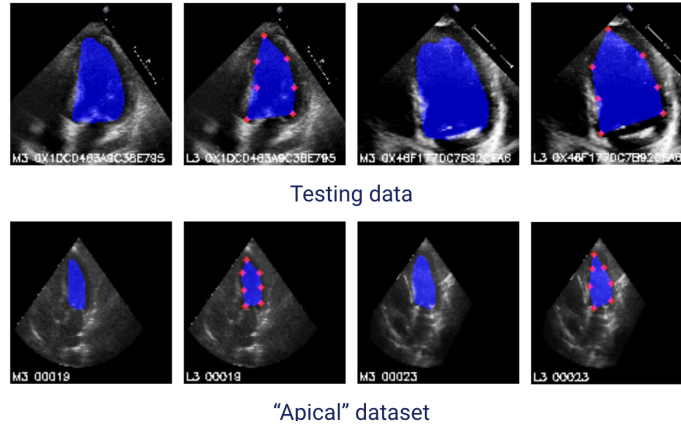


Fig. 7. Resultados de predicción de los modelos de máscaras y landmarks con tres canal para testing dataset y apical dataset.

10 Resultados y Trabajo Futuro

Después de la creación y entrenamiento de los cuatro modelos, empleando la métrica de puntaje Dice, se identificó cuál de ellos exhibió la mayor precisión en sus predicciones. Conforme se detalla en la imagen 8, el modelo que registró la menor exactitud fue el de landmarks con tres canales, seguido por el modelo de máscaras con tres canales. En contraste, el segundo mejor rendimiento provino del modelo de landmarks con un solo canal, situando al modelo de máscaras con un canal como el más preciso entre los desarrollados. Al profundizar en estos resultados, se deduce que la utilización de tres canales impacta significativamente en el desempeño del modelo. Por ende, esta opción sería descartada para futuras mejoras del modelo.

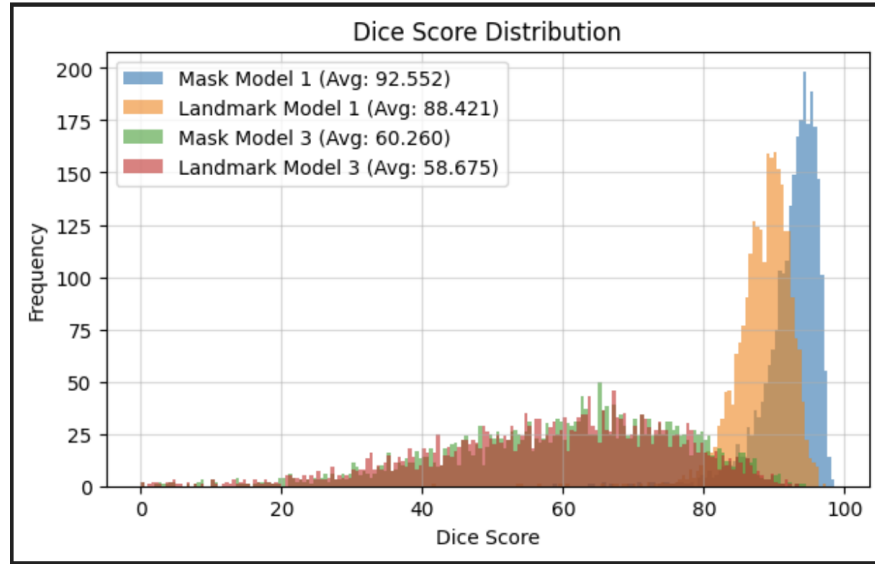


Fig. 8. Resultados de Dice score de los cuatro modelos desarrollados

En el contexto de una posible mejora para el modelo, se puede explorar una variedad de enfoques que podrían contribuir a lograr una mayor precisión en las predicciones. Uno de estos aspectos considerados es la incorporación de conjuntos de datos externos al conjunto original, lo cual permitiría un entrenamiento más completo del modelo. Esto resulta especialmente beneficioso para mejorar la capacidad predictiva del modelo frente a datos externos, los cuales son frecuentes en el ámbito médico.

Además, se podrían realizar pruebas utilizando arquitecturas de redes neuronales convolucionales (CNN) completamente distintas y experimentar con la

implementación de una función de pérdida personalizada. Estas estrategias, combinadas con otras que puedan surgir durante el proceso de mejora, tienen el potencial de generar un modelo aún más preciso. La clave reside en la experimentación cuidadosa, conservando los elementos útiles y descartando aquellos que no contribuyan significativamente a la mejora de las predicciones. Este enfoque iterativo y reflexivo puede resultar fundamental para alcanzar un rendimiento óptimo en la tarea específica del modelo.

11 Conclusión

Este proyecto representó una valiosa inmersión en el ámbito de la Inteligencia Artificial y la Ciencia de Datos. Desde la preparación de los conjuntos de datos hasta el entrenamiento de los modelos, experimentamos y ajustamos nuestras estrategias a lo largo de todo el proceso. Aunque algunos modelos no alcanzaron la precisión deseada y enfrentamos desafíos, las diversas pruebas y modificaciones que implementamos nos proporcionaron insights valiosos sobre los factores que impactan positiva o negativamente en nuestro modelo.

Considerando el tiempo limitado disponible para el desarrollo del proyecto y el nivel inicial de conocimientos en IA que todos teníamos, estamos satisfechos con los resultados obtenidos. En un futuro, con un mayor tiempo de dedicación y la implementación de optimizaciones y experimentos adicionales, creemos que el modelo podría experimentar mejoras sustanciales. Este proyecto ha sentado las bases para un continuo aprendizaje y refinamiento en el emocionante campo de la Inteligencia Artificial.

Por último, la distribución del equipo se ve representada en la siguiente tabla. Es importante mencionar que todos los integrantes estuvieron activamente involucrados en todas las áreas para la solución de este proyecto. Sin embargo, los encargados de cada área están designados a continuación.

Activity	Team Members
Dataset Creation	Alan Vilchis & Fausto Palma
Model Creation	Alan Vilchis, Fausto Palma & Alejandro Pizarro
Model Comparison	Diego Rosas & Héctor Cárdenas
Research Paper	Diego Rosas & Héctor Cárdenas
Presentation	Alejandro Pizarro

Fig. 9. Tabla de responsabilidades de cada integrante en el proyecto

References

1. Exo. (2022). Crunchbase. <https://www.crunchbase.com/organization/exo-systems>
2. Echocardiogram. (s.f). Celeveland Clinic. <https://my.clevelandclinic.org/health/diagnostics/16947-echocardiogram>
3. Centro para el Control y la Prevención de Enfermedades. (2020). Cómo funciona el corazón. <https://www.cdc.gov/ncbddd/spanish/heartdefects/howtheheartworks.html#:text=El>
4. MathWorks. (2023). ¿Qué son las redes convolicionales? <https://es.mathworks.com/discovery/convolutional-neural-network-matlab.html#:text=Una>
5. MathWorks. (2023). [Image]. ¿Qué son las redes convolicionales? <https://es.mathworks.com/discovery/convolutional-neural-network-matlab.html#:text=Una>
6. Maynard-Reid, M. (2022). U-Net Image Segmentation in Keras. PyImagesSearch. <https://pyimagesearch.com/2022/02/21/u-net-image-segmentation-in-keras/>
7. Maynard-Reid, M. (2022). [Image] U-Net Image Segmentation in Keras. PyImagesSearch. <https://pyimagesearch.com/2022/02/21/u-net-image-segmentation-in-keras/>
8. Antoniadis, P. (2023). Introduction to Landmark Detection. Baeldung. <https://www.baeldung.com/cs/landmark-detection#:text=Landmark>
9. Zamzmi, G. (2021). UMS-Rep: Unified modality-specific representation for efficient medical image analysis. Science Direct. <https://www.sciencedirect.com/science/article/pii/S2352914821000617>
10. Zhang, J. (2018). Fully Automated Echocardiogram Interpretation in Clinical Practice. Circulation. <https://www.ahajournals.org/doi/10.1161/CIRCULATIONAHA.118.034338>