



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO
PROGRAMA DE MAESTRÍA Y DOCTORADO EN INGENIERÍA
INGENIERÍA ELÉCTRICA – PROCESAMIENTO DIGITAL DE SEÑALES

SEGMENTACIÓN DE CAVIDADES CARDÍACAS EN IMÁGENES DE
ULTRASONIDO.

TESIS
QUE PARA OPTAR POR EL GRADO DE:
MAESTRO EN INGENIERÍA

PRESENTA:
ING. ALAN DE JESÚS CERVANTES GUZMÁN

TUTORA PRINCIPAL
DRA. JIMENA OLVERES MONTIEL
FACULTAD DE INGENIERÍA

CIUDAD UNIVERSITARIA, CD. MX., ENERO, 2022.



Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

JURADO ASIGNADO:

Presidente: Dr. Boris Escalante Ramírez

Secretario: Dr. Edgar Garduño Ángeles

1^{ra}. Vocal: Dra. Jimena Olveres Montiel

2^{do}. Vocal: Dr. Gibrán Fuentes Pineda

3^{er}. Vocal: Médico C. E. Octavio Barragán García

Lugar o lugares donde se realizó la tesis: Posgrado de la Facultad de Ingeniería

TUTORA DE TESIS:

Dra. Jimena Olveres Montiel

FIRMA

Agradecimientos

A la Dra. Jimena Olveres Montiel y al Dr. Boris Escalante Ramírez por brindarme la oportunidad de pertenecer al LAPI, por su guía y su dedicación que me brindaron a lo largo de mi estancia en la maestría.

A mis seres queridos, sobre todo a mi madre, quienes me brindaron su apoyo incondicional para seguir adelante en los tiempos de pandemia global en el que se realizó el presente trabajo.

Al Consejo Nacional de Ciencia y Tecnología (CONACYT) por el apoyo económico brindado.

A los proyectos PAPIIT: TA101121 y IV100420. Al proyecto SECTEI/202/2019.

Resumen

Las imágenes médicas juegan un papel fundamental dentro de la medicina porque permiten obtener información anatómica y ayudan a realizar diagnósticos más precisos de distintas enfermedades que afectan estructuras internas del cuerpo humano. El ultrasonido 2D es una modalidad de imagen médica que se caracteriza por tener alta accesibilidad para la población, y por ser una técnica no invasiva y libre de radiación. Sin embargo, es también una modalidad de imagen con importantes limitaciones como son el bajo contraste y el ruido que suelen complicar la tarea de segmentación dentro de los procedimientos de análisis de imágenes.

Gracias a los esfuerzos realizados en el área de las ciencias computacionales en los últimos años, ha sido posible desarrollar algoritmos que facilitan la tarea de segmentación, obteniendo así resultados más precisos en mucho menos tiempo a comparación de las técnicas manuales que aún siguen utilizándose en el mundo.

La presente tesis muestra la utilización de un algoritmo de Aprendizaje Profundo llamado U-Net para realizar la tarea de segmentación de las cuatro cavidades cardíacas (ventrículo izquierdo, ventrículo derecho, aurícula izquierda y aurícula derecha) en sus dos fases: sístole y diástole, en imágenes de ecocardiografía 2D apical de 4 cámaras. Esta técnica arrojó buenos resultados: 91.96% en coeficiente de Soresen-Dice, 85.17% en *Intersection over Union* y 93.52% en *Mean Pixel Accuracy*. Además de los resultados cuantitativos, se analizaron los resultados cualitativos obtenidos y se realizaron pruebas con imágenes de ecocardiografía obtenidas de una base de datos clínicos ajena a la base de datos utilizada para entrenar y validar el algoritmo.

Este trabajo forma parte del proyecto SECTEI 202/2019: Sistema Inteligente para el diagnóstico médico asistido para enfermedades cardiovasculares (SYSCARDIO).

Palabras claves: Segmentación, aprendizaje profundo, red neuronal artificial, ecocardiografía, cavidades cardíacas, imágenes médicas.

Abstract

Medical images play an important role in the medical area, they are usually used to acquire anatomic information and make diagnosis of different illnesses. 2D Ultrasound is a medical image modality that is characterized by high accessibility to most people and it is based on non invasive and radiation free techniques. Nevertheless, these images are often impaired by noise and low contrast, two features that complicate segmentation tasks in the image analysis procedure.

Thanks to the recent progress of computational sciences, it has been possible to develop novel algorithms that improve the segmentation task, thereby getting more accurately results in less time in comparison with the current manual techniques still worldwide used.

This thesis presents a convolutional neural network called U-Net used to segment of the cardiac cavities (left ventricle, left atrium, right ventricle and right atrium) in apical four chamber view echocardiography images. This technique shows a good performance with 91.16% Dice Coefficient, 85.17% intersection over union and 93.52% mean pixel accuracy. Moreover, additional results with an independent data base of clinical images are provided

This work takes part of the project SECTEI 202/2019: Intelligent system for assisted medical diagnosis (SYSCARDIO).

Keywords: Segmentation, Deep Learning, Artificial Neural Network, echocardiography, cardiac cavities, medical images.

Índice general

1. Introducción.....	1
1.1 Motivación.....	2
1.2 Justificación.....	3
1.3 Objetivos.....	6
1.4 Organización de la tesis.....	7
2. Imágenes médicas.....	8
2.1 Modalidades más comunes de la imagenología médica.....	9
2.1.1 Ultrasonido.....	9
2.1.2 Tomografía por emisión de fotones únicos (SPECT).....	10
2.1.3 Resonancia magnética.....	11
2.1.4 Tomografía computarizada multicapa (MSCT).....	13
2.2 Contexto actual de la imagenología médica.....	14
3. Ecocardiografía.....	16
3.1 Principio de funcionamiento.....	16
3.1.1 Sonido y espectro sonoro.....	16
3.1.2 El ultrasonido en la imagenología médica.....	17
3.2 Métodos de aproximación más utilizados.....	20
3.2.1 Vía transtorácica.....	20
3.2.2 Vía transesofágica.....	21

3.3	Modalidades de la ecocardiografía.....	21
3.3.1	Modo M.....	21
3.3.2	Doppler.....	22
3.3.3	Ecocardiografía bidimensional.....	25
3.3.4	Ecocardiografía 3D.....	30
4.	Redes Neuronales Artificiales.....	32
4.1	Perceptrón multicapa.....	34
4.2	Funciones de pérdida, activación y optimización más utilizadas.....	37
4.3	Redes neuronales artificiales y la tarea de clasificación.....	38
4.3.1	Redes neuronales convolucionales y la segmentación semántica.....	40
4.4	U - Net.....	42
4.4.1	Arquitectura.....	43
4.4.2	Convolución.....	44
4.4.3	Submuestreo.....	46
4.4.4	Convolución transpuesta.....	48
5.	Desarrollo y metodología.....	50
5.1	Conjunto de datos.....	50
5.2	Herramientas de desarrollo.....	51
5.3	Arquitectura propuesta.....	52
5.4	Entrenamiento.....	59
5.4.1	Datos.....	59
5.4.2	Función de pérdida.....	61
5.4.3	Optimizador.....	62
5.4.4	Hiperparámetros.....	62
6.	Resultados.....	64
6.1	Métricas utilizadas.....	64
6.2	Gráficas de pérdida y precisión.....	66
6.3	Matriz de confusión.....	68

6.4	Resultados visuales.....	70
6.5	Validación con datos clínicos.....	73
7.	Discusión y conclusiones.....	77
7.1	Conclusiones.....	77
7.2	Trabajo a futuro.....	78
	Bibliografía.....	79

Índice de figuras

Fig. 1.1: Contexto de los trabajos científicos dirigidos a imágenes médicas.....	4
Fig. 2.1.: Algunas de las más importantes modalidades que existen actualmente.....	9
Fig. 2.2: Imagen de ultrasonido fetal.....	10
Fig. 2.3: Comparación de imágenes SPECT de cerebro.....	11
Fig. 2.4: Imagen de resonancia magnética de cerebro.....	13
Fig. 2.5: Representación tridimensional de la pelvis a partir de imágenes de tomografía computarizada multicorte.....	14
Fig. 2.6: Población mundial mayor a 60 años y su proyección hacia 2050	15
Fig. 3.1: Caracterización de una onda de sonido.....	17
Fig. 3.2: Espectro sonoro.....	17
Fig. 3.3: Ecógrafo y transductor.....	18
Fig. 3.4: Principio de funcionamiento del transductor de un ecógrafo.....	19
Fig. 3.5: Imagen de ecocardiografía.....	20
Fig. 3.6: Ventanas ecocardiográficas.....	20
Fig. 3.7: Imagen de ecocardiografía en Modo M.....	22
Fig. 3.8: Efecto Doppler.....	23
Fig. 3.9: Parámetros de la ecuación de Doppler.....	24
Fig. 4.1: Analogía entre neuronas biológicas y las neuronas artificiales.....	32
Fig. 4.2: Problemas de compuertas <i>AND</i> y <i>OR</i>	34
Fig. 4.3: Problema de compuerta <i>XOR</i>	35
Fig. 4.4: Un ejemplo sencillo de un perceptrón multicapa.....	36
Fig. 4.5 Representación gráfica de aprendizaje supervisado y aprendizaje no supervisado.....	39

Fig. 4.6: Ejemplos de clasificación binaria, multiclase y multi etiqueta.....	40
Fig. 4.7: Ejemplo de segmentación semántica multiclase.....	41
Fig. 4.8 Arquitectura de la red <i>U-Net</i>	43
Fig. 4.9: Ejemplos de <i>zero padding</i> y <i>mirror padding</i>	45
Fig. 4.10 Kernel con diferentes stride.....	46
Fig. 4.11 Filtros con diferentes tamaños.....	46
Fig. 4.12: Ejemplo de operación <i>average pooling</i>	47
Fig. 4.13: Ejemplo de operación max pooling.....	48
Fig. 4.14: Ejemplo de operación de convolución transpuesta.....	49
Fig. 5.1: Algunos ejemplos del conjunto de datos y sus respectivas máscaras.....	51
Fig. 5.2: Estructura del bloque de contracción, <i>bottleneck</i> y expansión.....	54
Fig. 5.3: Bloque final de U-Net.....	56
Fig. 5.4: Esquema de arquitectura U-Net utilizada.....	56
Fig. 5.5 Casos de <i>underfitting</i> , un <i>fit</i> apropiado y <i>overfitting</i>	60
Fig. 6.1: Histórico de pérdidas.....	67
Fig. 6.2: Histórico de métricas a lo largo del entrenamiento.....	68
Fig. 6.3: Matriz de confusión del modelo de aprendizaje desarrollado en este trabajo.....	70
Fig. 6.4: Instancias detectadas como un primer caso de formas regulares.....	71
Fig. 6.5: Instancias detectadas como un segundo caso. Formas ligeramente irregulares en la aurícula izquierda.....	72
Fig. 6.6: Instancias detectadas como un tercer caso.....	72
Fig. 6.7: Instancias detectadas como un cuarto caso.....	73
Fig. 6.8: Una de las 204 nuevas imágenes de ecocardiografía apical de cuatro cámaras.....	74
Fig. 6.9: Imagen después del proceso del módulo extra de centrado y <i>cropping</i>	74
Fig. 6.10: La red neuronal busca mantener la forma de cada cavidad además de buscar ajustarse a los bordes, siempre y cuando estos estén presentes dentro de la imagen.....	75

Fig. 6.11: En estos casos de mayor complicación, la red neuronal artificial aún es capaz de buscar mantener las formas de estas cavidades y aprovechar al máximo los bordes que puedan existir para ajustarse a la hora de segmentar.....76

Índice de tablas

Tabla 3.1: Planos ecocardiográficos: estructuras cardíacas observables y su utilidad clínica.....	28
Tabla 3.2: Planos ecocardiográficos: su representación gráfica e imagen de ultrasonido.....	30
Tabla 4.1: Funciones de activación más utilizadas.....	37
Tabla 4.2: Funciones de pérdida más utilizadas.....	38
Tabla 5.1: Estructura de la arquitectura U-Net y el total de parámetros.....	58
Tabla 6.1: Métricas obtenidas en la etapa de validación.....	66

Capítulo 1

Introducción

En mayo de 2017, la OMS (Organización Mundial de la Salud) documentó que las ECV (Enfermedades Cardiovasculares) fueron la primera causa de muerte en el mundo en 2016, donde se estimó que alrededor de 17.9 millones de personas murieron en ese año, representando el 31% del total de las muertes globales. Además, de estas 17.9 millones de muertes por ECV, tan solo los ataques al corazón y derrames cerebrales representan el 85%. Otro dato importante que nos proporciona la OMS es que el 75% de las muertes por ECV tomaron lugar en países de ingresos económicos bajo y medio [1].

En México, el panorama no está muy alejado de las estadísticas globales, ya que las ECV se han presentado como primera causa de muerte en el periodo de 2009 a 2019, con un crecimiento del 48% en el 2019 con respecto al 2009, incluso por encima de otras enfermedades graves comunes en nuestro país como la diabetes, enfermedades crónicas de riñones, cirrosis, entre otras [2]. En 2018, la Secretaría de Salud de México comunicó que un alto porcentaje de muertes por ECV se atribuyen al IAM (Infarto Agudo al Miocardio) y sólo en 2017 hubieron más de 92 mil muertes por IAM. También se mencionó que nuestro país tiene una tasa de mortalidad por IAM tres veces más alta (28.1 vs 7.5 muertes por cada 100 personas) que el promedio de los países de la OCDE (Organización para la Cooperación y el Desarrollo Económico) en pacientes de 45 años de edad y más. De igual manera, se comunicó que uno de cada cuatro pacientes fallece por IAM, por lo que la misma secretaría gubernamental comenzó a afrontar de manera directa al problema diseñando un programa nacional para reducir la tasa de mortalidad [3].

Todas las estadísticas anteriores reflejan el gran problema mundial de salud pública que son las ECV, por lo que es necesario realizar detecciones tempranas de estas enfermedades, así como tratamientos oportunos para así reducir la tasa de mortalidad causada por estas, se busca esto sobre todo en personas con alto riesgo de padecerlas y que generalmente son personas con abuso al tabaco, dietas poco sanas, obesidad, inactividad física, abuso de alcohol, hipertensión, diabetes y, sobretodo, personas que pertenecen al sector geriátrico de la población.

Por otro lado, de manera general, en el campo de la medicina, al momento de buscar diagnósticos sobre alguna posible enfermedad o su examinación, es común y bastante usual que el médico tenga que solicitar la adquisición de imágenes médicas (en alguna de sus modalidades) con el fin de realizar dicho diagnóstico o examinación de forma precisa. Existen varios tipos de estas imágenes, tales como: ultrasonidos, resonancia magnética, tomografía computarizada, rayos X, entre otros más. Dentro de las tareas más comunes en el análisis de imágenes médicas es la de detección de estructuras anatómicas y estimación de parámetros numéricos medibles que brinden información importante o relevante para la toma de una decisión respecto al diagnóstico. Es por ello que es tan importante que esta tarea se deba realizar con una precisión y fiabilidad suficientemente buena para tomar una decisión clínica acertada a pesar de la incertidumbre que pueda introducir la variabilidad poblacional.

1.1 Motivación

El problema radica, una vez que se conoce la importancia de realizar buenos diagnósticos (una tarea que puede llegar a tomar su tiempo en realizar adecuadamente), en que se está presentando un aumento de pacientes por enfermedades graves, ya que la cantidad de imágenes analizadas ha ido aumentando cada año de una manera muy apresurada por el crecimiento en la cantidad de pacientes, lo que se traduce en menos tiempo de análisis de las imágenes médicas, disminuyendo la precisión del diagnóstico al mínimo clínicamente requerido, así como también la esperanza de vida de los pacientes.

Un estudio realizado *The Mayo Clinic* en Rochester, Minnesota, Estados Unidos y publicado por la *Academic Radiology*, menciona que en 1999 se analizaron poco más de 9 millones de imágenes médicas en sólo esa clínica, mientras que, 11 años después, en 2010 se analizaron más de 94 millones de imágenes. El estudio muestra que en promedio se analizaban 2.9 imágenes por minuto en 1999, mientras que en 2010 esta cifra aumentó hasta a 16.1 imágenes por minuto, lo que significa que un especialista radiólogo necesitaría analizar en promedio una imagen cada 3 o 4 segundos, lo cual es una tarea prácticamente imposible para las prácticas actuales [4].

1.2 Justificación

En la actualidad, a pesar del gran auge tecnológico que ha surgido con las herramientas computacionales, aún existe una gran cantidad de especialistas que realizan todo el proceso de examinación de imágenes médicas de forma manual alcanzando el requerimiento clínico en un gran periodo de tiempo y con mayor variabilidad. Por otro lado, un especialista que utiliza una herramienta computacional para realizar el proceso le toma mucho menos tiempo para alcanzar el requerimiento clínico, con menos variabilidad, e incluso logran alcanzar una mayor precisión, es decir, realizan mucho mejor su trabajo [44]. Dicho lo anterior, aunado a la problemática del aumento de imágenes a las que está sometida un radiólogo a examinar, es que surge la necesidad de crear herramientas computacionales tanto semiautomáticas como automáticas para ayudar al radiólogo a conseguir su objetivo de manera más sencilla, rápida y segura.

Con una población geriátrica en crecimiento, y también pacientes con algún tipo de cáncer, la adquisición de imágenes para este sector de personas se vuelve más restringida, ya que la exposición a radiación puede representar un riesgo para ellas, por lo que surgen opciones más viables y libres de radiación como el ultrasonido y la resonancia magnética [5]. El ultrasonido tiene una serie de ventajas sobre la resonancia magnética entre las cuales se encuentran principalmente su bajo costo y su gran accesibilidad, por lo que lo hace el principal contendiente como prueba radiológica para este tipo de situaciones [5]. Sin olvidar que la prueba de ultrasonido especializado en corazón también lleva como nombre Ecocardiografía.

Gracias a que cada vez más se han estado realizando trabajos sobre imágenes médicas del corazón en los últimos años, la cantidad de datos públicos de imágenes cardíacas ha ido aumentando, en especial, el ultrasonido de corazón, del cual, a partir de 2018, de solo estar contabilizados mucho menos de 250 casos pasó a contabilizarse más de 500 casos (Figura 1.1) [6]. Lo anterior demuestra el creciente interés por el cual se busca trabajar con estas imágenes para diversas tareas, entre las cuales se encuentra el desarrollo de herramientas computacionales. Sin embargo, a pesar de este incremento, aún se suelen presentar limitaciones al acceso de bases de datos de imágenes médicas anotadas que puedan servir para la realización de trabajos de investigación.

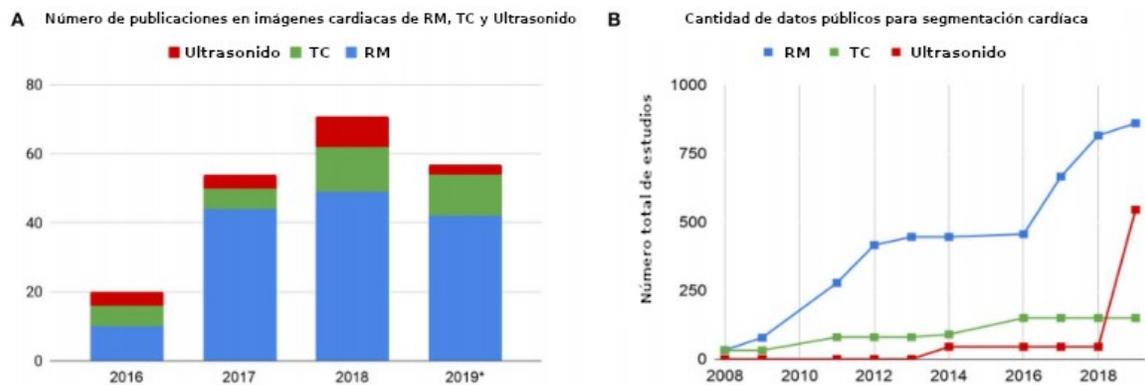


Fig. 1.1: Contexto de los trabajos científicos dirigidos a imágenes médicas durante los años 2016 a mediados de 2019 (A). Cantidad de datos públicos disponibles desde 2008 hasta 2018. Se puede observar un aumento en las tres modalidades, sobre todo en Ultrasonido. Figura tomada de: [6], bajo licencia CC BY.

Dentro del desarrollo de herramientas computacionales se encuentra el uso de modelos computacionales de aprendizaje profundo, los cuales (en el área médica) ha ido aumentando en el paso de los últimos años. Surgieron diversos trabajos, pero es importante detallar el trabajo de Ronneberger et. al. quienes en 2015 proponen una arquitectura totalmente innovadora dentro del campo de las redes neuronales convolucionales. Su trabajo *U-Net: Convolutional Networks for Biomedical Image Segmentation* [7], propone una arquitectura convolucional de aprendizaje profundo que nace a raíz de trabajos previos como el de Ciresan et. Al., *Deep Neural Networks Segment Neuronal Membranes in Electron Microscopy Images* de 2012 y el trabajo de Long et. Al., *Fully Convolutional Networks for Semantic Segmentation* de 2014, en el que de estos últimos tomaron varios elementos importantes y buscaron atacar las debilidades que presentaban a la hora de realizar

su respectiva tarea de segmentación deseada. Con la *U-Net*, Ronneberger et. al. probaron la efectividad de su arquitectura propuesta a la hora de realizar segmentación en tres tareas completamente dirigidas hacia imágenes biomédicas como imágenes de microscopio electrónico de membranas neuronales e imágenes de células. Los resultados que obtuvieron fueron totalmente satisfactorios logrando así ganar el *ISBI cell tracking challenge* en 2015. Los buenos resultados y la gran efectividad que esta arquitectura logra, es gracias a que atiende uno de los principales problemas dentro del aprendizaje profundo: la dificultad de encontrar conjuntos de datos extensos que cuentan con miles de ejemplos (cantidad que en imágenes médicas aún es complicado de conseguir), el cual ataca buscando aprovechar un conjunto de datos reducido (con cientos de ejemplos) realizando aumento de datos a través de una serie de transformaciones para tratar ese problema, brindando a la vez propiedades de invarianza ante desplazamientos y rotaciones [7].

La *U-Net* ha probado en diversas publicaciones científicas ser uno de los modelos computacionales de aprendizaje profundo más efectivos en trabajos donde se utilicen imágenes médicas, tal el caso de Zyuzin *et al.* que en 2018, con su trabajo *Identification of the left ventricle endocardial border on two-dimensional ultrasound images using the convolutional neural network U-Net* [8], presentaron un trabajo de segmentación del ventrículo izquierdo utilizando *U-Net* como modelo de aprendizaje profundo en imágenes de ecocardiografía 2D y muestran la importancia de la segmentación de las cavidades cardíacas para el cálculo de métricas importantes para los radiólogos en la obtención de diagnósticos. En dicho trabajo, presentan su propia arquitectura *U-Net*, así como toda la metodología utilizada que implica el desarrollo de un modelo computacional de aprendizaje profundo. Los resultados que obtuvieron fueron bastante cercanos a las anotaciones realizadas por el especialista en su respectivo conjunto de datos, sin embargo, su trabajo se enfoca solamente en la segmentación y estudio del ventrículo izquierdo. Dentro de la perspectiva de la recopilación realizada por Chen *et al.* en 2020 [6], se muestra una predominancia por parte de la *U-Net* en los trabajos de segmentación cardíaca en imágenes de resonancia magnética, tomografía computarizada y ecocardiografía, sin embargo, ninguno de los trabajos ahí recopilados aborda la segmentación de las cuatro cavidades cardíacas en imágenes de ecocardiografías de plano apical 4 cámaras.

Como justificación final, es de destacar que esta tesis parte de los resultados del proyecto SECTEI 202/2019: Sistema inteligente para el diagnóstico médico asistido para enfermedades cardiovasculares (SYSCARDIO), un sistema computacional que integra métodos de segmentación y estimación de movimiento de cavidades cardíacas que permite calcular parámetros clínicos relacionados a la evaluación del desempeño de la función cardíaca y que consta de una interfaz gráfica y un visualizador. De la misma manera, este proyecto se realizó en coordinación con cardiólogos del Centro Médico Nacional Siglo XXI y del Centro Médico ABC Santa Fe.

1.3 Objetivos

Es por lo anteriormente mencionado en el apartado 1.2 que se ha decidido establecer los siguientes objetivos:

- Aportar un método computacional eficiente y robusto de segmentación de cuatro cavidades cardíacas en imágenes de ecocardiografía apical en cuatro cámaras como parte de una herramienta integral de análisis de este tipo de imágenes.
- Implementar la red neuronal convolucional *U-Net* en lenguaje de programación *Python* con la ayuda de bibliotecas especializadas en Aprendizaje Profundo para realizar la segmentación de las cuatro cavidades cardíacas.
- Obtener resultados que sean tanto numéricos como visuales que permitan el adecuado análisis y evaluación del desempeño del modelo en la tarea asignada de segmentación.
- Someter al modelo a una validación con datos clínicos.

1.4 Organización de la tesis

A lo largo de este trabajo se estarán presentando diversos conceptos básicos relacionados a la imagenología médica, ecocardiografía y redes neuronales artificiales.

Dentro del capítulo 2 se presentarán las diferentes modalidades de la imagenología médica, así como el contexto actual por el que pasa esta muy importante área de la medicina. Por otro lado, el capítulo 3 se enfocará en la ecocardiografía y sus más importantes modalidades: Modo M, Doppler, ecocardiografía 2D y 3D.

A partir del capítulo 4 el enfoque será completamente hacia las redes neuronales artificiales y eventualmente se presentará la red neuronal *U-Net*, describir su origen, mencionar por quiénes fue desarrollado y en qué consiste la arquitectura de esta red neuronal artificial.

Por último, se mostrará todo el proceso de desarrollo y la metodología que se siguió para segmentar las cavidades cardíacas en imágenes de ecocardiografía 2D en su modalidad apical 4 cámaras.

Capítulo 2

Imágenes médicas

Las imágenes médicas son el resultado de un conjunto de métodos o pruebas de adquisición de imágenes tanto bidimensionales como tridimensionales de algún órgano o parte del cuerpo del que exista cierto interés de grado médico, con el propósito de obtener información cualitativa y cuantitativa suficiente sobre el objeto de estudio para posteriormente llevar a cabo un adecuado análisis de este y que permita tomar una decisión clínica sumamente importante. Generalmente, la adquisición de estas imágenes se lleva a cabo cuando existen ciertas sospechas de un funcionamiento anormal de algún órgano o parte del cuerpo, he ahí su importancia del que tanto se insiste.

Existen distintas modalidades dentro del área de imágenes médicas, entre las cuales se encuentran: ultrasonido, resonancia magnética, tomografía por emisión de positrones, tomografía por emisión de fotones únicos y la tomografía computarizada multicorte, cada una con su particularidad que las caracterizan una de otra, con sus respectivas ventajas y desventajas que son necesarias conocer a fondo.

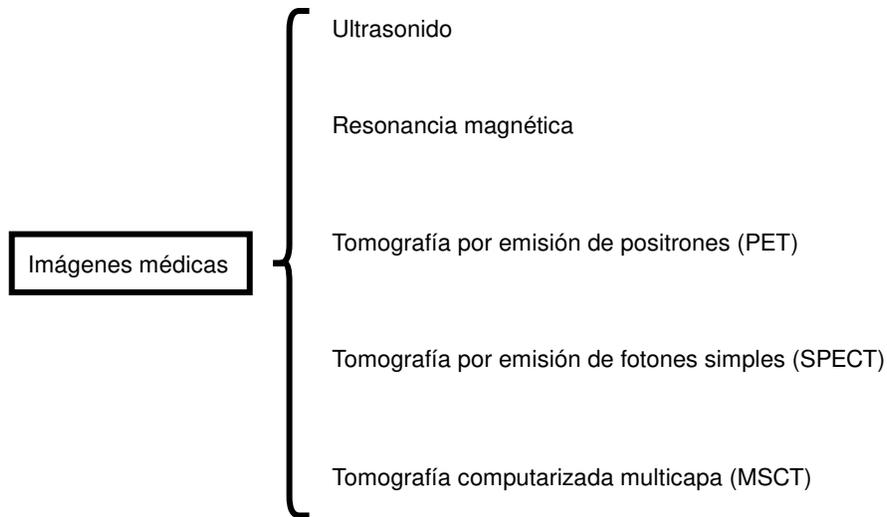


Fig. 2.1.: Algunas de las más importantes modalidades o tipos de imágenes médicas que existen actualmente.

2.1 Modalidades más comunes de la imagenología médica.

2.1.1 Ultrasonido

Las imágenes de ultrasonido son, básicamente, obtenidas mediante una respuesta a las ondas acústicas emitidas por un transductor ultrasónico que se posiciona encima de la piel con la ayuda de un líquido acoplador. Dicha respuesta acústica se manifiesta en el dispositivo receptor después de que esta haya pasado por el receptor (usualmente ubicado también dentro de la sonda o transductor que emite la onda acústica inicial) seguido de una conversión de energía mecánica a energía eléctrica generando así pulsos eléctricos que son interpretados por la máquina de manera digital para producir las imágenes. Aunque las imágenes que se crean casi siempre son para fines estructurales o anatómicos, también suelen existir modalidades para estudiar la fisiología de alguna parte anatómica de interés. Por ejemplo, el estudio de Doppler color permite visualizar la dirección y la velocidad del flujo sanguíneo en estructuras como arterias o el corazón.

La prueba de ultrasonido es una de las más utilizadas en el campo de imagenología [55], es también (junto a rayos X) una de las más económicas y, por si fuera poco, no necesita de algún tipo de radiación o agente radiactivo para la obtención de imágenes. Sin embargo, la baja relación señal-ruido (SNR), obstáculos

anatómicos y la dependencia de la experiencia del operador son los problemas más comunes con los que cuenta las imágenes de esta modalidad.



Fig. 2.2: Imagen de ultrasonido fetal, uno de los más conocidos. Figura tomada de: <https://www.ginecologaengualajara.com.mx>

2.1.2 Tomografía por emisión de fotones únicos (SPECT)

En ambas pruebas se utiliza una sustancia compuesta de moléculas biológicas que transportan isótopos que emiten fotones (para la prueba SPECT), a esta sustancia se le suele referir casualmente como marcador y se caracteriza por ser una sustancia radiactiva.

Es aplicada una pequeña dosis de esta sustancia en el paciente de forma intravenosa y tarda en ser absorbida entre 20 o 30 minutos y una hora, por otro lado, cuando se trata de alguna prueba para pulmones se aplica un marcador en aerosol [47]. La función del marcador es la de emitir cierta cantidad de radiación que ayuda al escáner a obtener una visualización digital del área de interés y así mismo obtener su correspondiente fisiología, esto debido a que los tejidos que se encuentren más activos absorberán mayor cantidad del marcador poniendo en evidencia las posibles anomalías que pudieran estar presentándose, tal como el caso de tumores.

Por otro lado, una de las razones por la que el tiempo de adquisición varía, es por la zona del cuerpo que se va a examinar. También cabe mencionar que habitualmente es necesario hacer uso de otra modalidad que sea más dedicada a la morfología, esto para facilitar a una localización anatómica exacta de la estructura de interés, por ejemplo, la tomografía computarizada de rayos x, en caso de no contar con la posibilidad de generar imágenes 3D con esta modalidad [9].

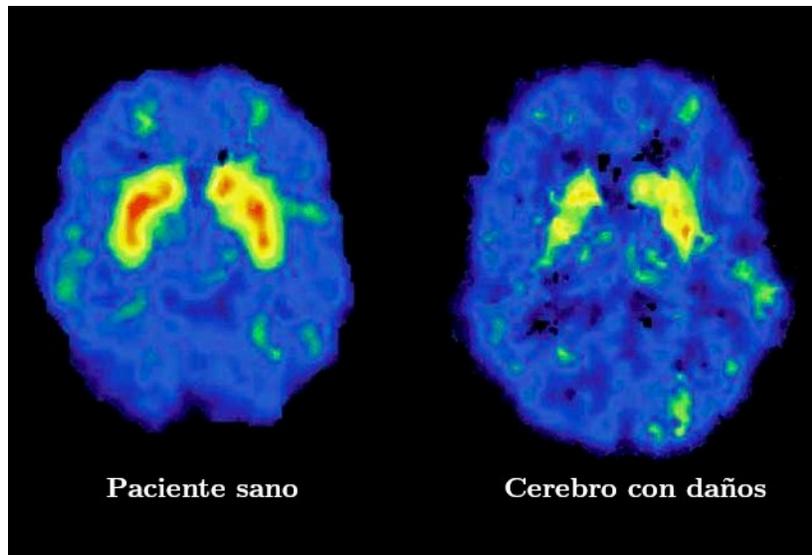


Fig. 2.3: Comparación de imágenes SPECT de cerebro.

2.1.3 Resonancia magnética

La resonancia magnética es una prueba que no utiliza radiación sino más bien aprovecha las propiedades magnéticas naturales del cuerpo [48] para hacer uso de fuertes campos magnéticos y ondas de radio para producir las imágenes de la zona de interés. La adquisición resulta ser tridimensional y estar compuesta de una extensa serie de cortes por lo que la resolución espacial transversal logra ser buena y permite analizar partes del cuerpo que son difíciles de observar con otras modalidades, aunque desgraciadamente para lograr esto se necesita de la inversión de un gran lapso de tiempo durante la prueba.

Para propósitos de obtención de imágenes, el hidrógeno es mayormente utilizado debido a que existe una abundancia de este en el cuerpo. El hidrógeno es

un elemento que cuenta con un solo protón en su núcleo que se comporta como una pequeña barra magnética y que bajo circunstancias normales giran con sus ejes aleatoriamente alineados [48]. Una vez que el cuerpo es sometido a un gran campo magnético, los ejes de los protones comienzan a tener un movimiento de precesión sobre la línea del campo magnético generando la frecuencia de Larmor, sin embargo, el movimiento de precesión de cada protón está desfasado con respecto a los movimientos del resto y una vez se aplican ondas de radios, muchos de estos se sincronizan provocando una resonancia que continúa manifestándose aun después de que las ondas de radios ya no estén presentes. Gracias a esto, se crean vectores magnéticos en los protones con un componente perpendicular al campo magnético que cuenta con un movimiento rotatorio y que generan señales electromagnéticas que son captadas a través de los sensores ubicados en la máquina envolvente. Finalmente, se llega a un estado de relajación donde la resonancia se disipa y el movimiento de precesión de cada protón vuelven a desfasarse.

Es destacable que, cuando más rápido se logre este efecto durante la adquisición, mejor calidad de imagen se obtiene y para lograr esto, es necesario hacer uso de una sustancia llamada contrastante, la cual sirve para mejorar la claridad de la imagen aumentando su contraste y así poder observar con mayor facilidad algunas zonas de interés [10].

La importancia de esta prueba recae en el gran contraste que se logra en las imágenes sin la necesidad de algún agente radioactivo. Sin embargo, como contrapunto está el costo elevado de la prueba sobre otras como la tomografía computarizada y el ultrasonido, por las que suele estar 1.5 y 10 veces más costosa, respectivamente.



Fig. 2.4: Imagen de resonancia magnética de cerebro. Es considerablemente apreciable el buen contraste con el que cuenta este tipo de imágenes.

2.1.4 Tomografía computarizada multicorte (MSCT)

La tomografía computarizada multicorte, además de aprovechar el mismo principio de funcionamiento con rayos X de su versión unicorte, utiliza arreglos de detectores formados en más de 64 filas que van registrando la cantidad de radiación que penetra al paciente [45]. Su invención permitió reducir de manera significativa el lapso de tiempo para realizar la prueba: desde la base del cuello hasta el periné en máximo 20 segundos [11], sin embargo, la dosis de radiación suministrada al paciente es alta a comparación de la tomografía computarizada unicorte y el precio de costo y mantenimiento es alto.

Aunque de ambas pruebas se suelen obtener imágenes con muy buen nivel de detalle, otras de las mayores ventajas de las pruebas de MSCT sobre la tomografía computarizada unicorte son la obtención de imágenes volumétricas en un periodo menor de tiempo y una mayor resolución axial [45].



Fig. 2.5: Representación tridimensional de la pelvis a partir de imágenes de tomografía computarizada multicorte. Figura tomada de: [11].

2.2 Contexto actual de la imagenología médica

La Organización de Naciones Unidas (ONU) reportó que la población mayor a 60 años creció de 382 millones en 1980 a 962 millones de personas en 2017 y, a su vez, advierte que la misma crecerá aún más, estimando 2,100 millones de personas de la tercera edad en 2050 alrededor del mundo (Figura 2.6), representando un porcentaje de 9%, 12% y 20% de la población total, respectivamente [12]. Este fenómeno ha llevado a que diferentes organizaciones se preocupen y realicen sus propios estudios, la mayoría de estas organizaciones se encuentran dedicadas a la salud pública y se comenta que mientras que la población geriátrica vaya en aumento es muy posible que el número de pacientes también lo haga. Para entrar en contexto (y a manera de ejemplo del impacto del crecimiento de población anciana) en el portal *The Cancer Atlas*, una organización que se encarga de generar información, datos y estadísticas mundiales sobre esta enfermedad, menciona que se estima que pasarán de 18 millones de casos, documentados en 2018, a tener 29 millones de casos para 2040 debido al fenómeno de envejecimiento de la población [13]. Cabe señalar, que es muy evidente que el cáncer no sea la única enfermedad en reflejar un aumento de casos en los próximos años.

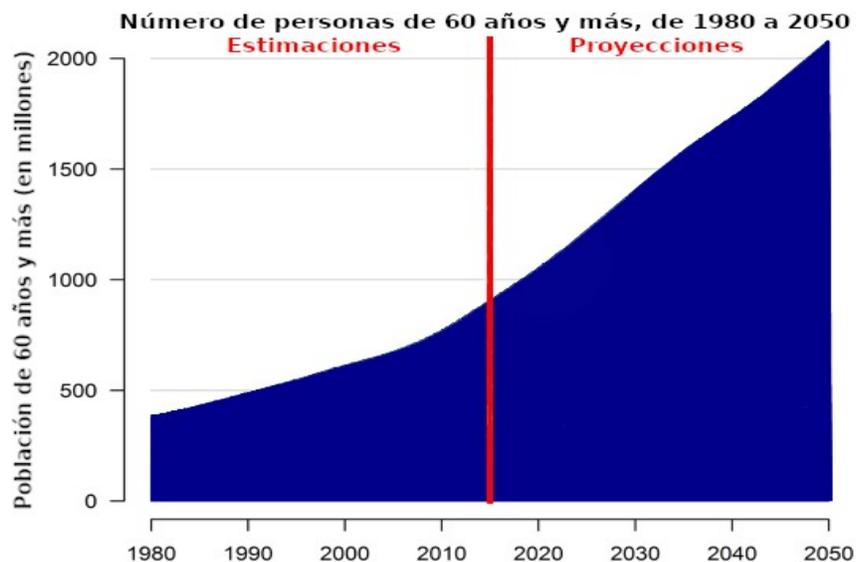


Fig. 2.6: Población mundial mayor a 60 años y su proyección hacia 2050.

Por otro lado, en los últimos años, gracias al desarrollo tecnológico que ha logrado el campo de la imagenología médica, tanto los diagnósticos como tratamientos médicos han sido más eficientes y efectivos ante enfermedades como el cáncer para evitar que puedan llegar a ser terminales.

Dicho lo anterior, ante el inminente crecimiento tanto de población geriátrica, así como de enfermedades, la importancia y la demanda del campo de imagenología médica aumentará todavía más. Así como lo demuestra un estudio realizado en *The Mayo Clinic* en Rochester, Minnesota, Estados Unidos y publicado por la revista *Academic Radiology*, el cual respalda la importancia y la gran demanda que ha tenido la imagenología médica en los últimos años. El estudio menciona que en 1999 se analizaron poco más de 9 millones de imágenes médicas en sólo esa clínica, mientras que, 11 años después, en 2010 se analizaron más de 94 millones de imágenes [4]. A pesar de que el estudio se realizó tan sólo en una clínica, no hay duda de que es algo que se replica en cualquier otra ya sea en menor o mayor grado, pero con una tendencia creciente.

Capítulo 3

Ecocardiografía

La ecocardiografía es una prueba médica de adquisición de imágenes del corazón de un paciente a través de la tecnología del ultrasonido, es por esto que a la ecocardiografía también se le suele conocer, aunque de manera menos formal, como ultrasonido de corazón. Al ser una prueba médica, la ecocardiografía se utiliza tanto para obtener parámetros cualitativos como cuantitativos del corazón y así conocer el estado de su funcionamiento y estructura para ayudar al especialista a obtener un diagnóstico para el paciente. Comparte ventajas que el ultrasonido tiene como modalidad en la imagenología médica como la no invasividad, el bajo costo y un corto lapso de tiempo para la realización de la prueba. Además, la ecocardiografía también tiene la gran ventaja de ser una prueba con gran disponibilidad y portabilidad.

3.1 Principio de funcionamiento

3.1.1 Sonido y espectro sonoro

El fenómeno físico del sonido se da cuando una serie de movimientos vibratorios creados de alguna manera por una fuente, se logra transmitir a través de un medio elástico (generalmente el aire) transmitiendo también así dicha energía generada inicialmente por la fuente [14]. En la Figura 3.1 se puede observar la caracterización gráfica de una onda longitudinal de cierta frecuencia f :

$$f = \frac{1}{T} \tag{3.1}$$

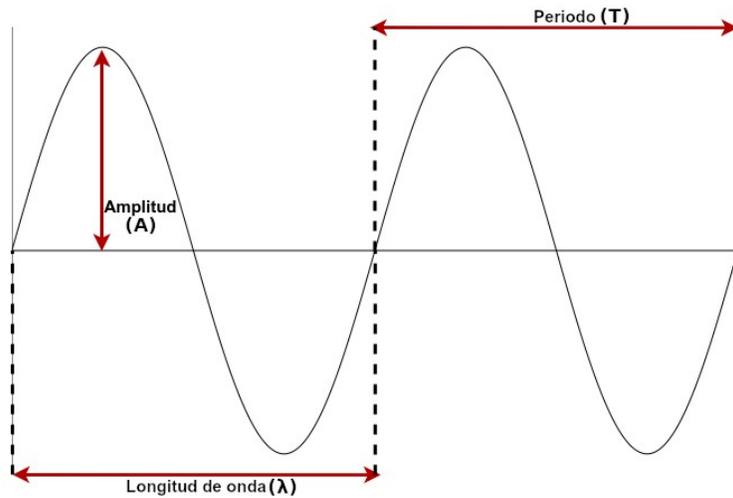


Fig. 3.1: Caracterización de una onda de sonido.

Por otro lado, el espectro sonoro (Figura 3.2) se suele dividir en tres partes: las frecuencias del infrasonido que pertenecen a todas las frecuencias menores a 20 Hz, las frecuencias audibles (20 Hz a 20 kHz) y las frecuencias de ultrasonido que están por encima de los 20 kHz.



Fig. 3.2: Espectro sonoro.

3.1.2 El ultrasonido en la imagenología médica

Las ondas de sonido pertenecientes a las de una frecuencia mayor de los 20 kHz son denominadas ondas de ultrasonido y están fuera del rango de las ondas de sonido perceptibles para el oído humano. Este tipo de ondas sónicas tienen diversas aplicaciones, una de las más conocidas es su aplicación médica de adquisición de imágenes. En la ecocardiografía, las ondas de ultrasonido generadas están en un rango de 1 a 10 MHz [15].

La historia del ultrasonido comienza cuando Lazzaro Spallanzani demostró que los murciélagos utilizaban sonido no audible para navegar [16]. Los primeros trabajos con el ultrasonido comienzan a partir de 1880 con el descubrimiento de la piezoelectricidad y fueron posteriormente seguidos de una serie de invenciones tecnológicas para fines militares e industriales hasta que, en 1946, el francés fisioterapeuta André Denier propuso que el ultrasonido podría ser usado para visualizar órganos internos del cuerpo humano [17]. El comienzo de su uso médico data desde la década de 1950 y se le acredita a Inge Edler y Carl Helmuth Hertz como los primeros en lograr obtener una imagen de ecocardiografía [18].



Fig. 3.3: A la izquierda, Ecógrafo. En la derecha, cable transductor. Figura tomada de <https://www.digimed.com.ar>

Para obtener imágenes de ultrasonido, el ecógrafo (Figura 3.3) genera una serie de pulsos electromagnéticos que viajan a través de un cable hacia un periférico llamado transductor (Figura 3.3). Dentro del transductor se encuentran cristales piezoeléctricos, estas piezas son sometidas a un campo eléctrico que les permite crear haces de ondas ultrasónicas y que son conducidas al cuerpo con la ayuda de un gel de acoplamiento. Conforme el ultrasonido atraviesa los tejidos del cuerpo, se puede crear cierta impedancia acústica que, dependiendo de la densidad del tejido que atravesase, puede crear alguno de los siguientes tres efectos ecoicos:

- **Anecoicidad:** Presente en líquidos y, por lo tanto, el sonido se refleja débilmente y la mayor parte de la energía se pierde. Visualmente se observa en negro o en tonos bajos de gris.

- Hipoecoicidad: Presente en tejidos blandos, se refleja parte del sonido. Visualmente son tonos medios de gris.
- Hiperecoicidad: Presente en tejidos con mayor densidad donde el sonido se refleja en su mayoría o casi su totalidad. Se identifican con los tonos más altos de gris.

Una vez que sucede la reflexión de las ondas acústicas de ultrasonido, muchas de estas regresan hacia donde se encuentra el transductor para seguir con una conversión de energía mecánica a energía eléctrica. Los cristales dentro del transductor se ven sometidos a cambios de presión generados por las ondas de ultrasonido (energía mecánica) creando un campo eléctrico (energía eléctrica) que regresará a través por el cable hacia el ecógrafo.

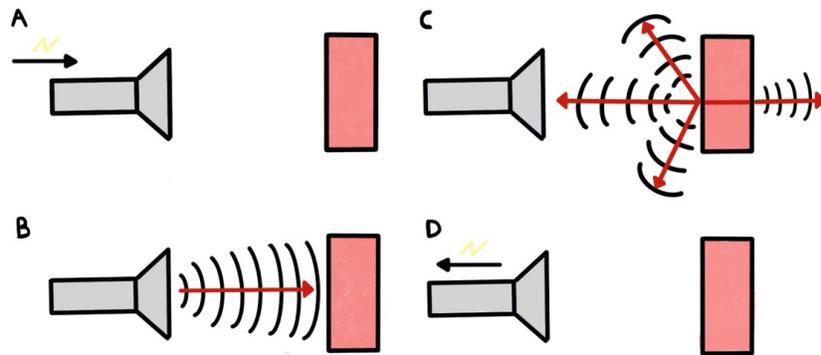


Fig. 3.4: Principio de funcionamiento del transductor de un ecógrafo. A) Una señal eléctrica viaja por el cable. B) El transductor genera ondas de ultrasonido. C) El tejido absorbe una parte de los ultrasonidos, el resto rebota y el transductor los recibe. D) Una señal eléctrica viaja de regreso por el cable hacia el ecógrafo para posteriormente ser interpretada.

Finalmente, el ecógrafo interpreta, a una velocidad promedio de entre 3,000 y 5,000 pulsos de repetición por segundo, los pulsos eléctricos recibidos como puntos que se van desplegando en la pantalla hasta obtener algo como lo que se observa en la Figura 3.5 [19].



Fig. 3.5: Imagen de ecocardiografía. Son señalados los tejidos hiperecoicos, anecoicos e hipoecoicos.

3.2 Métodos de aproximación más utilizados

3.2.1 Vía transtorácica

Este método de aproximación es el más común y también bastante simple ya que sólo es necesario posicionar el transductor sobre la piel de alguna de las *ventanas* (Figura 3.6). Existe la desventaja de que las estructuras óseas del tórax puedan intervenir en la obtención de una buena imagen que sea de utilidad para el especialista.

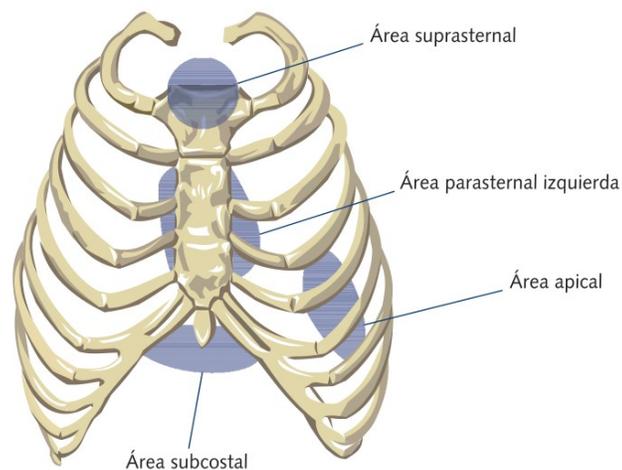


Fig. 3.6: Ventanas ecocardiográficas. Figura tomada de: [15].

3.2.2 Vía transesofágica

A diferencia de la aproximación transtorácica, la aproximación por vía transesofágica utiliza un tipo de transductor adecuado para su posterior introducción hacia el esófago por medio de la boca del paciente, que debe de estar preparado y sedado para dicho procedimiento. Se le considera semi-invasivo pero, a pesar de esto, realmente no presenta ningún riesgo para el paciente además de que sólo se accede a esta aproximación cuando se necesita examinar de manera más detallada una zona del corazón, o bien, porque la imagen generada por vía transtorácica no brinda la información suficiente para el especialista ya que una imagen adquirida por vía transesofágica es de mayor calidad que el de por vía transtorácica, dado que se evitan estructuras óseas del tórax.

3.3 Modalidades de la ecocardiografía

3.3.1 Modo M

El modo M nació como la primera modalidad de ecocardiografía gracias al trabajo *The Use of Ultrasonic Reflectoscope for Continuous Movements of the Heart Wall* de Inge Edler y Carl Helmuth Hertz en 1954 [20]. A partir de ese entonces y a lo largo de la década de 1960, se empezó a trabajar en el desarrollo del modo M para diversas aplicaciones. Una de estas aplicaciones fue la propuesta por Feigenbaum y Dodge en 1968 para la medición de dimensiones del ventrículo izquierdo [21]. En la década de los 70's, el modo M comenzó a tener más reconocimiento en la medicina y fue considerado como una técnica de diagnóstico muy buena [22].

En esta modalidad, el transductor se sitúa sobre el tercer, cuarto o quinto espacio intercostal para enviar el ultrasonido en forma de haz único. Este haz realmente no abarca un espacio considerable del corazón, por lo que, la medición que realiza es unidimensional, es decir, captura los movimientos en el espacio de las estructuras cardíacas sobre las que esté situada el transductor durante el ciclo cardíaco (tiempo) como si de un oscilógrafo se tratase (Figura 3.7). Dentro de las imágenes obtenidas se puede medir el tamaño de las cavidades, el grosor de las paredes de los ventrículos, etc [23].

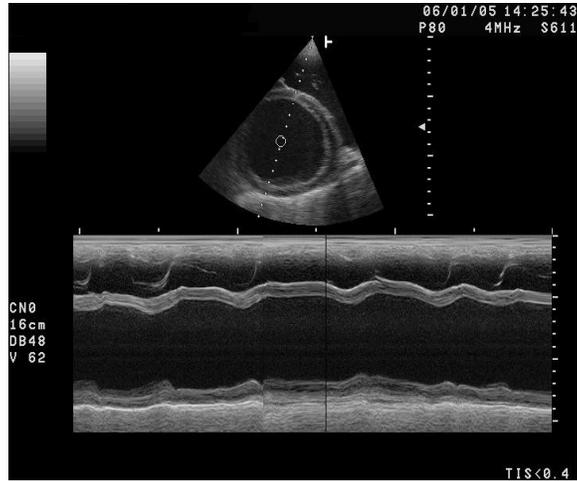


Fig. 3.7: Imagen de ecocardiografía en Modo M.

A pesar de que el modo M ha sido actualmente superado por la ecocardiografía bidimensional y la ecocardiografía Doppler, aún sigue teniendo relevancia médica por su capacidad de proporcionar información adicional o incluso única que las otras modalidades no puedan proporcionar.

3.3.2 Doppler

En física, existe un fenómeno acústico llamado efecto Doppler, que fue llamado así gracias al físico austriaco Christian Andreas Doppler quien propuso la existencia de este efecto en 1842 [24]. Este fenómeno es el resultado de una interacción de una fuente emisora de ondas acústicas y un objeto observador. Cuando, por ejemplo, la fuente emisora de ondas acústicas a una frecuencia f_e se aleja a una velocidad v del observador, la frecuencia que este último percibe f_r es menor a la frecuencia f_e , mientras que, en caso contrario, cuando la fuente emisora se acerca al observador, este percibe una frecuencia f_r mayor a la frecuencia emitida f_e (Figura 3.8).

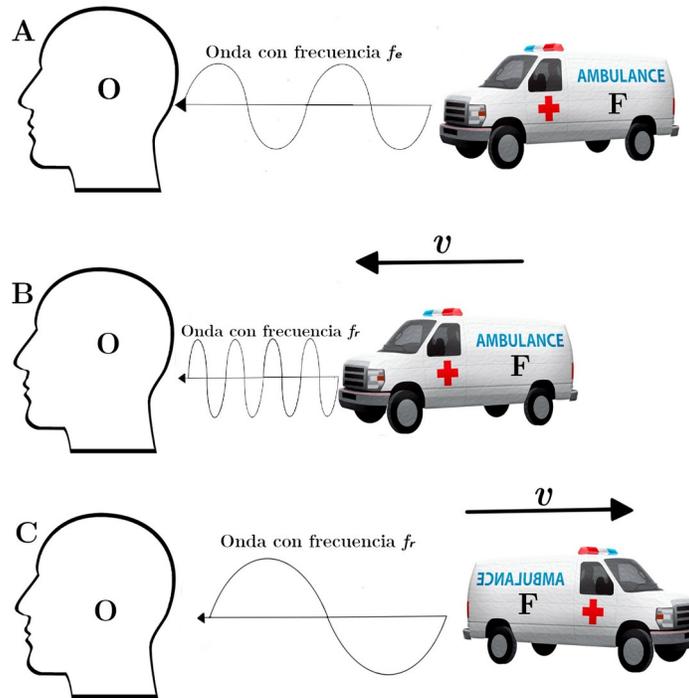


Fig. 3.8: Efecto Doppler. A) La fuente F se encuentra en reposo emitiendo un sonido a una frecuencia f_e . B) Cuando la fuente emisora F se acerca al observador O con una velocidad v , la frecuencia f_r del sonido con la que es recibida es mayor que f_e , en cambio, C) cuando la fuente emisora F se aleja a una velocidad v del observador O, la frecuencia f_r del sonido con la que es recibida es menor que f_e .

La aplicación de este fenómeno se desarrolló como modalidad dentro de la ecocardiografía y su uso ha sido imprescindible para obtener la dirección y velocidad del flujo sanguíneo dentro de las cámaras cardíacas y así poder evaluar mediante la cuantificación, el funcionamiento de las partes cardíacas involucradas en el proceso.

Cuando el transductor del ecógrafo emite las ondas de ultrasonido hacia algún torrente sanguíneo, los glóbulos rojos o eritrocitos actúan como elementos reflectores del ultrasonido por lo que actuarían como el elemento emisor de las ondas acústicas creadas por el transductor mientras que, el mismo transductor, actuaría como el elemento observador [40]. La ecuación que describe la velocidad del flujo sanguíneo está dada por la ecuación 3.1.

$$v = \frac{(f_e - f_r)k}{2f_e \cos\theta} \quad (3.2)$$

Donde v es la velocidad del flujo sanguíneo dado en cm/seg en un instante de tiempo; f_e es la frecuencia emitida en Hertz por el transductor; f_r es la frecuencia de recepción en Hertz; k es la constante de velocidad de propagación del ultrasonido en la sangre ($1.54 \times 10^5 \text{ cm/seg}$); y $\cos\theta$ es el coseno del ángulo formado por el haz de ultrasonido y la dirección del flujo. En la Figura 3.9 se pueden observar gráficamente los parámetros que intervienen en la ecuación de Doppler.

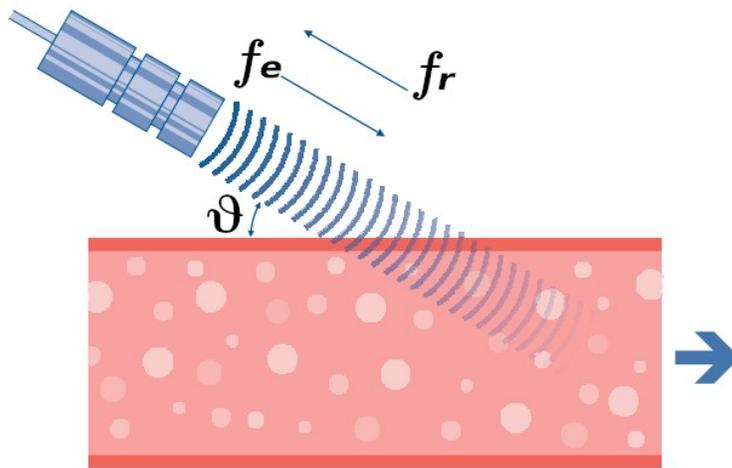


Fig. 3.9: Parámetros de la ecuación 3.1. Figura tomada de: [40].

Cabe señalar que se recomienda que el ángulo de incidencia entre el haz de ultrasonido y el torrente sanguíneo no rebase los 20° , debido que, superando este ángulo, la velocidad del flujo se subestima de manera importante [25].

Existen variantes del modo Doppler, las cuales son:

- Doppler pulsado: Permite estudiar el flujo sanguíneo en un punto determinado de las cámaras. Cabe destacar que esta variante no puede registrar altas velocidades del flujo [23].

- Doppler continuo o tisular: Permite analizar el registro de los flujos a velocidades altas que en el Doppler pulsado no es posible realizar. Esto gracias a que recoge la suma de las velocidades que registre [23].
- Doppler color: Es la variante que trabaja en conjunto con la ecocardiografía bidimensional donde se puede visualizar, con la ayuda de un mapa de calor, la velocidad y la dirección del flujo sanguíneo (rojo se acerca, azul se aleja). Lo anterior, con la finalidad de encontrar movimientos anormales dentro del corazón.

3.3.3 Ecocardiografía bidimensional

La ecocardiografía bidimensional nace como un gran avance tecnológico en la ecocardiografía, ya que provee una imagen bidimensional que permite observar las estructuras del corazón en tiempo real. La ecocardiografía 2D (como también se suele llamar) a diferencia del modo M, utiliza un transductor con un juego de cristales piezoeléctrico que permiten obtener sectores mucho más amplios del corazón. Se mide la densidad ecoica de las estructuras del corazón para obtener su visualización.

La exploración se puede realizar tanto en el área paraesternal como en la apical, subcostal o supraesternal, áreas también llamadas ventanas básicas de estudio en la ecocardiografía (Figura 3.6). Estas ventanas están establecidas para evitar estructuras óseas y los pulmones, con la finalidad de evitar que el ultrasonido que se emita del transductor se refleje o atenúe en exceso, respectivamente.

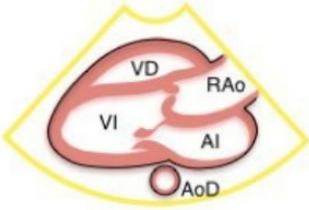
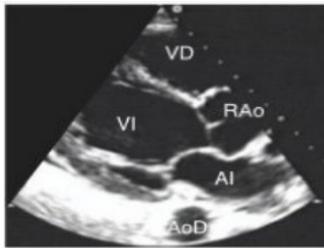
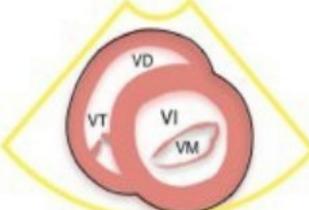
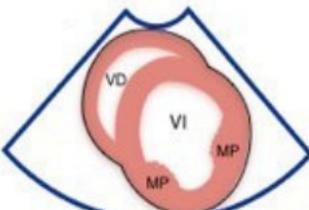
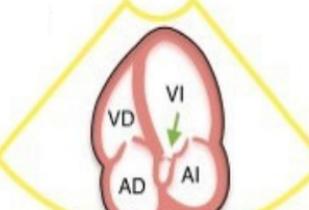
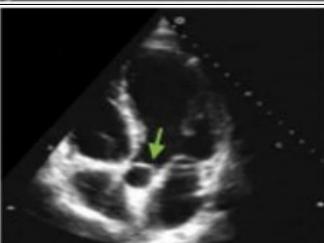
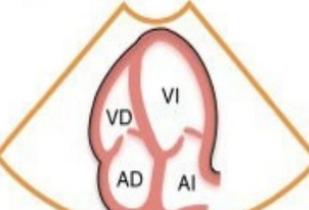
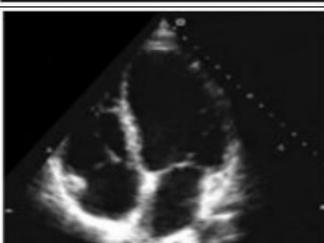
Con la ayuda de planos ortogonales dentro de un espacio tridimensional, se pueden obtener distintos planos ecocardiográficos para cada ventana de análisis. Con la información que brinda cada corte se puede realizar un estudio completo de la anatomía y fisiología del corazón y así poder encontrar posibles anomalías. Dichos planos ecocardiográficos, así como su utilidad clínica se presentan en las siguientes dos tablas (3.1 y 3.2) [35]

Ventana de análisis	Plano ecocardiográfico	Se visualiza	Utilidad clínica (Para valorar)
Paraesternal	Longitudinal	<ul style="list-style-type: none"> • Válvula aórtica. • Aurícula izquierda. • Válvula mitral. • Ventrículo izquierdo. • Ventrículo derecho. 	<ul style="list-style-type: none"> • Contractilidad global y segmentaria del ventrículo izquierdo. • Movilidad y morfología de los velos aórticos y mitrales. • Diámetro de la aorta descendente. • Información estructural de las paredes y cavidades cardíacas. • Sospecha de presencia de derrame pericárdicos.
	Transversal	<ul style="list-style-type: none"> • Válvula aórtica con sus tres velos. • Arteria pulmonar. • Aurícula derecha. • Ventrículo derecho. • Válvula tricúspide. • Aurícula izquierda. 	<p>A nivel de grandes vasos:</p> <ul style="list-style-type: none"> • Área y morfología de la válvula aórtica. • Sospecha de una comunicación interauricular. • Presencia de masas, trombos o dispositivos intracavitarios. • Diámetro de la arteria pulmonar y su relación con la raíz aórtica. • Presencia de insuficiencia tricúspidea. • Presencia de derrame pericárdico. <p>A nivel de la válvula mitral:</p> <ul style="list-style-type: none"> • Relación de tamaño de ambos ventrículos. • Sobrecarga del septo interventricular. • Contractilidad regional. • Valoración global subjetiva de la función sistólica del ventrículo izquierdo. • Morfología de la válvula mitral. • Presencia de derrame pericárdico. <p>A nivel de los músculos papilares:</p> <ul style="list-style-type: none"> • Se valora lo mismo del

			plano a nivel de la válvula mitral, pero desde otra perspectiva.
Apical	5 cámaras	<ul style="list-style-type: none"> • Ventrículo izquierdo. • Ventrículo derecho. • Aurícula izquierda. • Aurícula derecha. • Válvula aórtica. 	<ul style="list-style-type: none"> • Información estructural del TSVI, de la válvula aórtica y del aparato subvalvular. • Estimación del volumen sistólico y del gasto cardíaco. • Estenosis e insuficiencia aórtica.
	4 cámaras	<ul style="list-style-type: none"> • Ventrículo izquierdo. • Ventrículo derecho. • Aurícula izquierda. • Aurícula derecha. • Tabique Interventricular. • Tabique interauricular. • Válvula mitral. • Válvula tricúspide. 	<ul style="list-style-type: none"> • Medida de los ejes menores de los ventrículos derecho e izquierdo y su relación. • Tamaño de paredes y cavidades cardíacas. • Estructura de aparatos valvulares y subvalvulares. • Función diastólica. • Estimación de las presiones de llenado del ventrículo izquierdo. • Contractilidad global y segmentaria. • Presencia de estenosis e insuficiencia mitral. • Presencia de insuficiencia tricúspidea. • Estimación de la presión arterial pulmonar sistólica. • Presencia de masas, trombos y dispositivos intracavitarios. • Patología del pericardio. • Guía de punción de derrames pericárdico. • Aorta descendente. • Presencia de derrames pleurales y colapsos pulmonares de la llingula.
	2 cámaras	Desde otra perspectiva: <ul style="list-style-type: none"> • Aurícula izquierda. • Ventrículo izquierdo. 	<ul style="list-style-type: none"> • Contractilidad global y segmentaria. • Tamaño de las paredes y cavidades cardíacas. • Estructura de la válvula mitral. • Estimación del volumen sistólico y del gasto

			cardíaco.
	3 cámaras	<ul style="list-style-type: none"> • Ventrículo izquierdo. • Aurícula izquierda. • Válvula aórtica. 	<ul style="list-style-type: none"> • Información sobre la estructura del TSVI, la válvula aórtica y el aparato subvalvular. • Estimación del volumen sistólico y del gasto cardíaco. • Información sobre la valoración de estenosis e insuficiencia aórtica.
Subcostal	4 cámaras	<ul style="list-style-type: none"> • Ventrículo derecho. • Aurícula derecha. • Ventrículo izquierdo. • Aurícula izquierda. 	<ul style="list-style-type: none"> • Presencia y cantidad de líquido pericárdico. • Relación entre las cavidades cardíacas. • Medida de la pared libre del ventrículo derecho. • Función sistólica global del ventrículo izquierdo. • Contractilidad del ventrículo derecho. • Presencia de regurgitación tricúspidea. • Realización de una estimación semicuantitativa de la presión venosa central.
Supraesternal	Longitudinal	<ul style="list-style-type: none"> • Cayado aórtico. • Arteria carótida izquierda. • Arteria subclavia izquierdas. • Arteria pulmonar 	<ul style="list-style-type: none"> • Patología de la aorta. • Flujo aórtico.

Tabla 3.1: Planos ecocardiográficos: estructuras cardíacas observables y su utilidad clínica.

Plano ecocardiográfico	Imagen gráfica	Imagen ecocardiografía
Paraesternal Longitudinal		
Paraesternal transversal 90°		
Paraesternal transversal 45° - 60°		
Paraesternal transversal <45°		
Apical 5 cámaras		
Apical 4 cámaras		

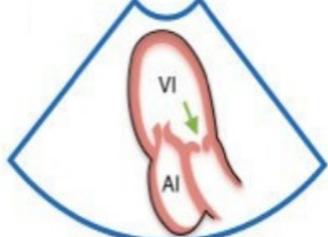
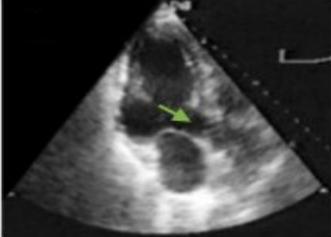
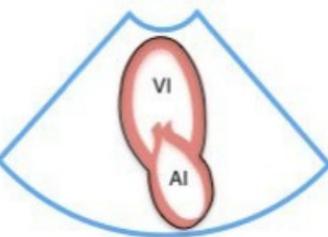
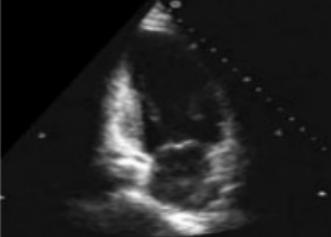
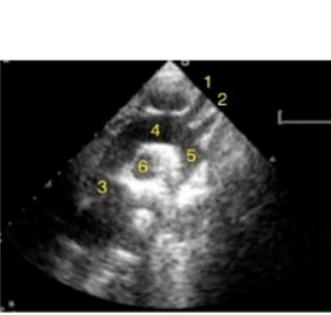
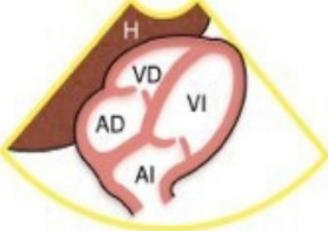
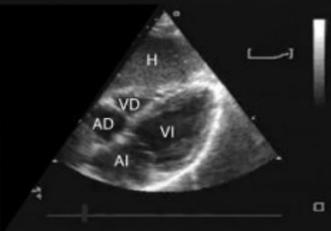
Apical 3 cámaras		
Apical 2 cámaras		
Supraesternal longitudinal		
Subcostal 4 cámaras		

Tabla 3.2: Planos ecocardiográficos: su representación gráfica e imagen de ultrasonido.

3.3.4 Ecocardiografía 3D

La ecocardiografía tridimensional nace a partir de la necesidad de poder estudiar y analizar la compleja anatomía del corazón de una manera más intuitiva, algo que en las técnicas de modo M y ecocardiografía bidimensional es complicado.

El análisis de imágenes de ecocardiografía tridimensional proporciona información completa sobre las estructuras cardíacas desde cualquier punto de vista

espacial. Por otra parte, dentro de las limitaciones con las que cuenta esta técnica es la necesidad de personal médico experto en la adquisición de estas imágenes [26].

Entre las aplicaciones que tiene esta técnica se encuentran las siguientes:

- Obtención del volumen del ventrículo izquierdo, así como su fracción de eyección.
- Se puede usar junto a la ecocardiografía de estrés para una detección más sencilla y directa de enfermedades como la enfermedad de la arteria coronaria.
- Evaluación de la masa del ventrículo izquierdo.
- Obtención del volumen del ventrículo derecho, así como su fisiología.
- Obtención del volumen y fisiología de la aurícula izquierda.
- Visualización de otras estructuras cardíacas.

Gracias a que esta técnica es una de visualización tridimensional, permite acceso en, al menos, la gran mayoría de las estructuras cardíacas del corazón y datos cuantitativos precisos [27]. Incluso hay publicaciones como la realizada en 2007 por Jenkins *et al.* [42], donde comparan la ecocardiografía tridimensional con la resonancia magnética y se logra conseguir grandes cifras de correlación, lo que significa que se pueden obtener resultados cuantificables casi igual de buenos en esta técnica que los de la resonancia magnética y por un costo mucho menor [43].

Capítulo 4

Redes Neuronales Artificiales

Aportaciones científicas del campo de la neurociencia y neurobiología sobre la inteligencia humana, el comportamiento de las neuronas y la estructura del cerebro, sirvieron de inspiración para que las redes neuronales artificiales nacieran y que abrieron la posibilidad de mejorar la disciplina de la inteligencia artificial [53]. La sola estructura que tienen las redes neuronales artificiales llegan a asemejarse, en cierta medida, a la estructura de una red neuronal biológica, esto debido a la inspiración biológica que toman (Figura 4.1), y es por ello que suelen estar constituidas de unidades lógicas llamadas neuronas. Sin embargo, en cuanto al principio de funcionamiento, las redes neuronales artificiales llegan a tomar solo algunas características del funcionamiento del sistema nervioso, tales como: la interconexión de unidades neuronales, la creación de sinápsis y la operación en paralelo [46].

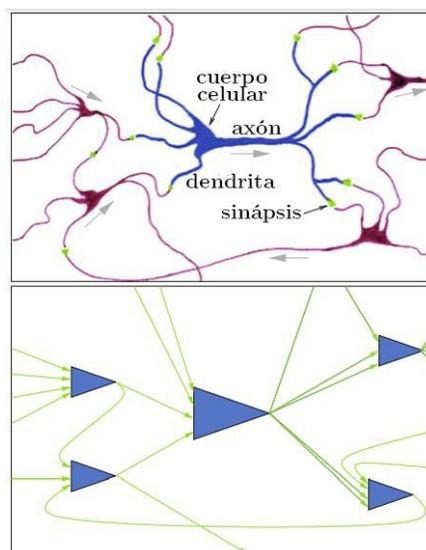


Fig. 4.1: Analogía entre neuronas biológicas (parte superior) y las neuronas artificiales (parte inferior).

Las redes neuronales artificiales lograron establecer una nueva tendencia en las ciencias de la computación y gracias a su no linealidad, y a que realizan operaciones en paralelo de manera distribuida, son altamente capaces de buscar modelos que describan un fenómeno que ayude a realizar tareas específicas y resolver problemas de gran complejidad, que serían complicados de replicar en el cómputo secuencial [54].

Entre las tareas o aplicaciones más comunes de las redes neuronales artificiales se encuentran [49], [50], [51]:

- Modelado y control de procesos.
- Traducción automática de textos.
- Reconocimiento facial.
- Verificación de firmas y análisis de manuscritos.
- Diagnósticos médicos.
- Reconocimiento de voz.
- Detección de fraudes.
- Calificación crediticia.
- Proyección financiera.

- Pronóstico del tiempo.

4.1 Perceptrón multicapa

En 1943 McCulloch y Pitts modelaron la unidad básica de una red neuronal artificial con la llamada *Neurona de McCulloch y Pitts* [28], tiempo después, basado en el trabajo de McCulloch y Pitts, Rosenblatt presentó el perceptrón en 1958, dicho perceptrón demostró ser capaz de resolver problemas linealmente separables como la caracterización de las compuertas *AND* y *OR* (Figura 4.2) [29].

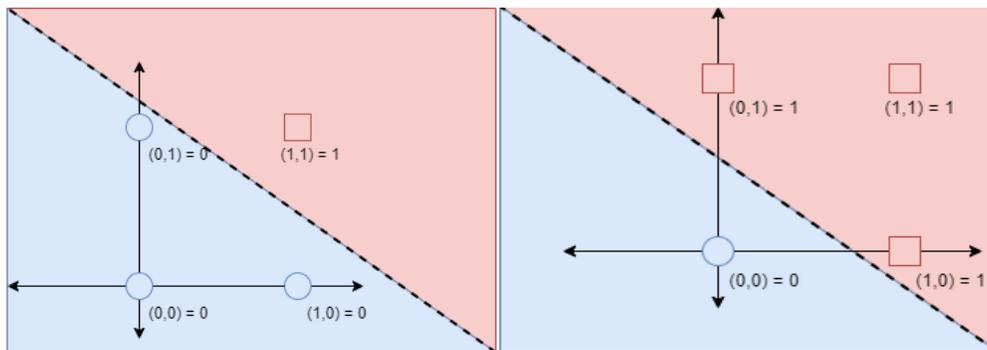


Fig. 4.2: Problemas de compuertas *AND* (izquierda) y *OR* (derecha). Se puede observar que son linealmente separables.

Un perceptrón simple está compuesto por:

- Una capa de entrada.
- Una capa de salida.
- Una función de activación.
- Pesos y un *bias*.

- Una función de error.

Sin embargo, en 1969 Minsky y Papert demostraron las limitaciones del perceptrón de Rosenblatt con el problema linealmente no separable de la compuerta *XOR* (Figura 4.3) [30]. Dicho problema se resolvería varios años después cuando se creó lo que se convertiría en el sucesor del perceptrón, el cual permitió resolver problemas de ese tipo.

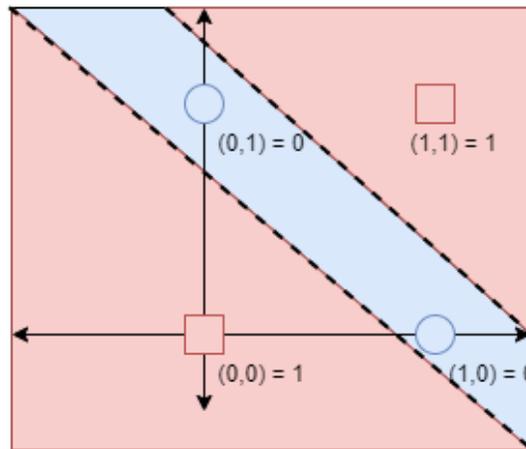


Fig. 4.3: Problema de compuerta *XOR*. A diferencia de las compuertas *AND* y *OR* (Fig. 4.2), este problema no es linealmente separable.

El perceptrón multicapa partió con las bases del perceptrón simple y, por tanto, está compuesto en parte por todo de lo que el perceptrón simple lo está: función de activación, función de error, capa de entrada, capa de salida, pesos y *biases*. Sin embargo, un par de elementos le otorgaron al perceptrón multicapa tener la relevancia que aún tiene en la actualidad: La adición de capas ocultas y el algoritmo de retropropagación.

La capa oculta se encarga de extraer características útiles a un nivel superior que sirven para resolver la tarea en cuestión, a través de un procesamiento donde se crean relaciones no lineales con la ayuda de la función de activación, por lo que esta capa es la clave para aumentar la capacidad y flexibilidad de la red neuronal. Cabe destacar que la red neuronal artificial puede tener más de una capa oculta y

esto permite que sea capaz de aprender de los datos de entrada en varios niveles de abstracción.

Por otra parte, la importancia del algoritmo de retropropagación radica en el hecho de que se apoya en la función de pérdida para emplear el error entre el resultado producido al final de la red con el resultado esperado, creando así una señal que sirve para poder actualizar los pesos de todas las neuronas en dirección contraria (hacia atrás) dentro de la red y así buscar reducir el error en la siguiente iteración de entrenamiento. Este algoritmo ayudó bastante a resolver problemas linealmente no separables como el problema de la compuerta *XOR* [31].

El perceptrón multicapa también es conocido como red totalmente conectada, debido a que todas sus capas están conectadas entre sí, es decir, la salida de una neurona sirve como entrada para todas las neuronas de la siguiente capa (Figura 4.4).

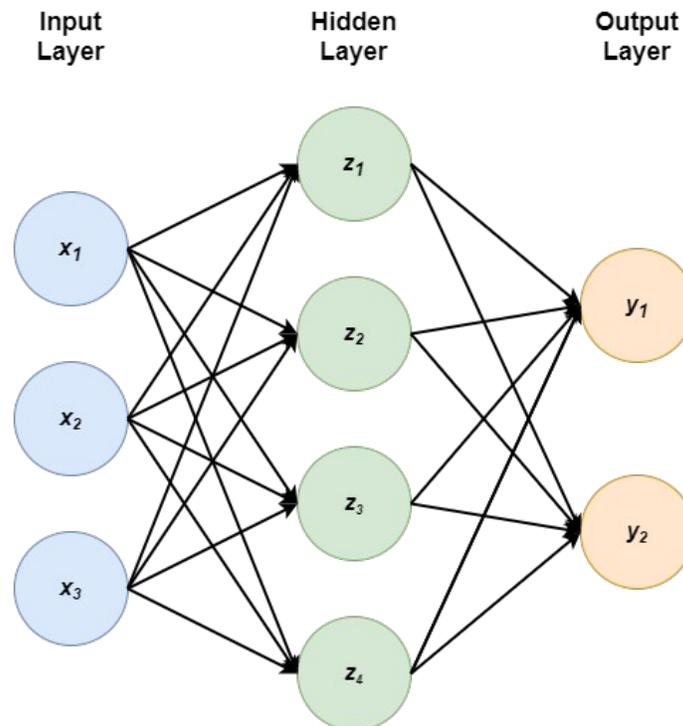


Fig. 4.4: Un ejemplo sencillo de un perceptrón multicapa. Además de una capa de entrada y una de salida, cuenta con una capa intermedia denominada como capa oculta.

4.2 Funciones de pérdida, activación y optimización más utilizadas

Cada una de estas son importantes para la red neuronal artificial debido a que brindarán cierta ayuda al entrenamiento. Por una parte, la función de pérdida calcula el error que existe entre la salida esperada y la salida que arroja la red neuronal y que posteriormente será utilizado para calcular los gradientes que servirán para actualizar los pesos de la red neuronal artificial. Mientras que la función de activación procesa la información recibida que será enviada a la siguiente capa de neuronas y ayudará a ajustar los pesos. Por último, la función de optimización ayuda a minimizar la función de pérdida.

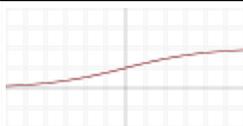
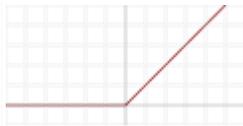
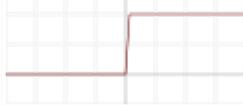
Función	Gráfica	Ecuación
Sigmoide		$\sigma(x) = \frac{1}{1+e^{-x}}$
Tangente hiperbólica		$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$
ReLU		$ReLU(x) = \begin{cases} x & \text{si } x \geq 0 \\ 0 & \text{si } x < 0 \end{cases}$
Step		$Step(x) = \begin{cases} 1 & \text{si } x \geq 0 \\ 0 & \text{si } x < 0 \end{cases}$

Tabla 4.1: Funciones de activación más utilizadas.

Función	Ecuación	Aplicación principal
Error Medio Cuadrático	$ECM(y, \hat{y}) = \frac{1}{N} \sum_{i=1}^N (\hat{y}_i - y_i)^2$	Regresión logística
Entropía Cruzada Binaria	$ECB(y, \hat{y}) = - \sum_{i=1}^N [y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)]$	Clasificación binaria
Entropía Cruzada Categórica	$ECC(x, clase) = - \log \left(\frac{\exp(x[clase])}{\sum_j \exp(x[j])} \right)$	Clasificación multiclase

Tabla 4.2: Funciones de pérdida más utilizadas.

4.3 Redes neuronales artificiales y la tarea de clasificación

Las redes neuronales artificiales son algoritmos de aprendizaje que suelen estar dentro de alguna de las siguientes categorías: aprendizaje supervisado, aprendizaje no supervisado, aprendizaje semi supervisado o aprendizaje por refuerzo. Cabe destacar que ninguna categoría es mejor que otra, sino que cada categoría tiene sus respectivas ventajas y diferentes aplicaciones, también tareas o áreas en los que cada uno se pueda desenvolver de mejor manera. Una de estas tareas es la clasificación, presente en mayor medida en algoritmos de aprendizaje supervisado y no supervisado.

El aprendizaje supervisado se trata de aquel algoritmo en el que los datos con los que lo alimentas, en este caso a la red neuronal artificial, incluyen una etiqueta de pertenencia (Figura 4.5) con lo que se crea un modelo predictivo capaz de realizar clasificación o regresión de acuerdo a las características que representan dichas etiquetas.

Por otra parte, el algoritmo de aprendizaje no supervisado se diferencia del supervisado por contar con un conjunto de datos sin etiquetar, por lo que el modelo que se crea clasifica de acuerdo a las características que el algoritmo sea capaz de agrupar y que puedan compartir los ejemplos (Figura 4.5).

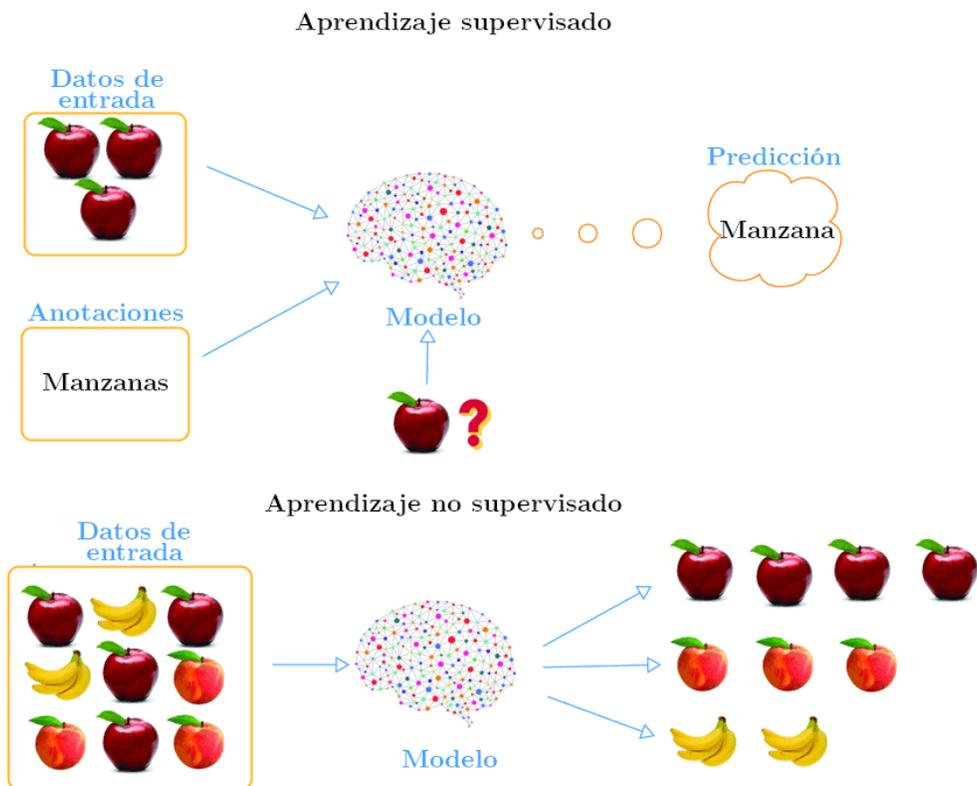


Fig. 4.5 Representación gráfica de aprendizaje supervisado (parte superior) y aprendizaje no supervisado (parte inferior). Cabe destacar que la principal diferencia es que el aprendizaje supervisado cuenta con etiquetas que establecen la pertenencia de los datos a una o varias clases, mientras que en el aprendizaje no supervisado esta característica está ausente.

Una vez que ya se comentó las particularidades del aprendizaje supervisado y no supervisado, es turno de revisar en qué consiste la tarea de clasificación.

La clasificación, como ya se había comentado, es una de las tareas más comunes de los algoritmos de aprendizaje supervisado y no supervisado y consiste básicamente en asignar al resultado una clase (de entre un número limitado de clases). Existen diferentes tipos de clasificación: clasificación binaria, clasificación multiclase y clasificación multi etiqueta [32].

- Clasificación binaria: El modelo determina si una instancia pertenece o no a la clase en cuestión.
- Clasificación multiclase: El modelo determina la pertenencia de una instancia dentro de una clase. Existen más de dos clases.
- Clasificación multi etiqueta: El modelo establece que una instancia pertenece simultáneamente a varias clases.



Fig. 4.6: Ejemplos de clasificación binaria (fondo color azul), multiclase (fondo color morado) y multi etiqueta (fondo color rojo).

4.3.1 Redes neuronales convolucionales y la segmentación semántica

La tarea de segmentación consiste en dividir una imagen en grupos de píxeles (segmentos) representativos, lo cual se logra asignando cada píxel a una clase con la que el algoritmo considere que se comparta mayores características (Figura 4.7), por lo que esta tarea se convierte en una de clasificación píxel por píxel. También

cabe destacar que la segmentación usualmente ayuda a localizar objetos dentro de una imagen.

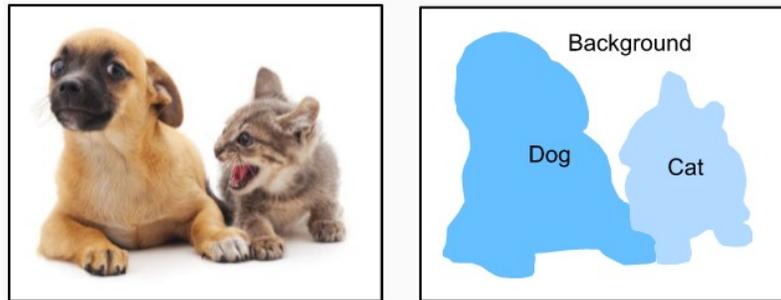


Fig. 4.7: Ejemplo de segmentación semántica multiclase. Figura tomada de: [38].

Existen diversos algoritmos para realizar segmentación, pero sin duda, los algoritmos que mejor desempeño han mostrado para este tipo de tarea son las redes neuronales artificiales, más específicamente las llamadas redes neuronales convolucionales.

Las redes neuronales convolucionales marcaron un antes y un después en el campo de clasificación de imágenes puesto que en 2012 *AlexNet* (una red neuronal convolucional) batió récord de clasificación de *ILSVRC* quedando demostrado que este tipo de redes neuronales pueden lograr muy buenos resultados en tareas de clasificación de imágenes. En 2014, es publicado el trabajo que popularizó el uso las redes neuronales convolucionales para tareas de segmentación semántica: *Fully Convolutional Networks for Semantic Segmentation* por Long et. al. [33].

En dicho trabajo se presenta una arquitectura convolucional que cuenta de manera resumida con capas convolucionales para la extracción de características de las imágenes, capas de *subsampling* que reducen la resolución espacial y capas de convolución transpuesta como operación de *upsampling* y recuperar información espacial (el “dónde”) que se había perdido anteriormente.

Esta arquitectura destaca por reemplazar la última capa totalmente conectada de las redes neuronales convolucionales comunes con una red de capas de

convolución transpuesta y también por hacer uso del llamado *skip connections* para realizar la inferencia final.

4.4 U - Net

Gracias al éxito que las redes neuronales convolucionales comenzaron a tener en tareas de visión computacional y la influencia que tuvo la red completamente convolucional de Long et. Al., en 2015 Ronneberger, Fischer y Brox publicaron su trabajo *U-Net: Convolutional Networks for Biomedical Image Segmentation* [7] en el cual presentan una nueva y particular arquitectura en forma de U (Figura 4.8), diseñado especialmente para realizar tareas de segmentación en imágenes biomédicas.

En dicho trabajo se hace énfasis en un par de detalles como la importancia de la localización del objeto segmentado dentro de la imagen, es decir, se especifica que la tarea que realiza la red *U-Net* se trata de una tarea de clasificación píxel a píxel. Por otra parte, y aún más importante, la red *U-Net* aprovecha la técnica de aumento de datos para afrontar los reducidos conjuntos de imágenes biomédicas con los que frecuentemente se cuentan en la práctica y donde dicha limitante es realmente insuficiente en otras redes neuronales convolucionales en los que se necesitan una gran cantidad de imágenes de entrenamiento, afectando sus desempeños.

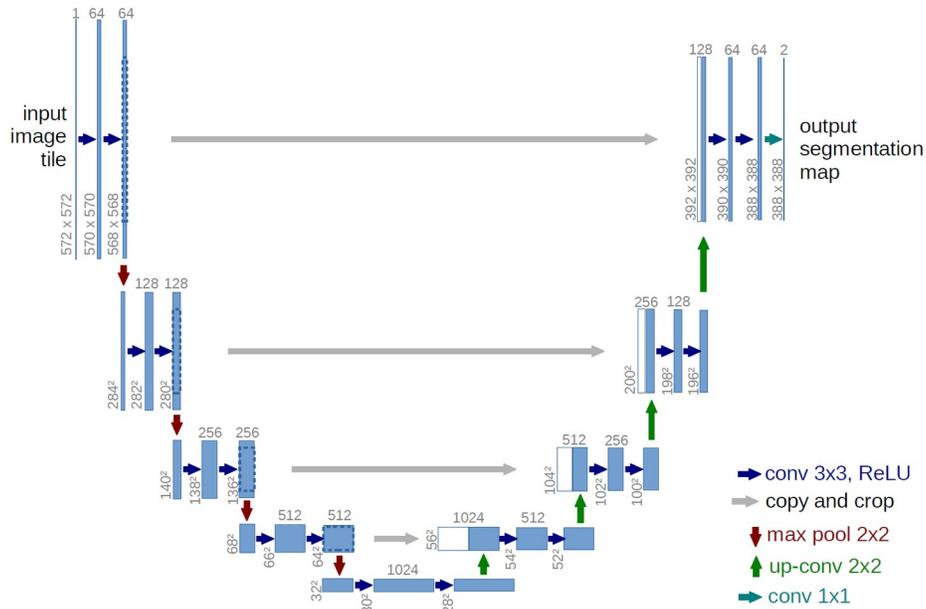


Fig. 4.8 Arquitectura de la red *U-Net*. Se puede observar de manera notable su forma en U. Figura tomada de: [7].

4.4.1 Arquitectura

La arquitectura original de la *U-Net* consiste principalmente en tres secciones, de las que destacan dos, una de ellas es la que definen como la parte de contracción [7] que se sitúa del lado izquierdo. Esta parte, básicamente sigue la misma arquitectura de una típica red neuronal convolucional y consiste en aplicar una serie de capas de convolución con *kernel* de tamaño 3×3 (y la función de activación *ReLU*) seguido de una operación de submuestreo por cada bloque de contracción. Al final se logra extraer características, así como reducir la resolución espacial y, a su vez, aumentar el número de canales de características (profundidad).

La sección derecha, que es definida por los autores como la parte de expansión, es muy similar a la parte de contracción, en cada bloque de expansión también existen capas de convolución. Sin embargo, en esta sección la resolución espacial original de la imagen de entrada se va recuperando gradualmente con capas de *upsampling* al final de cada capa de expansión y, por otro lado, la profundidad se va reduciendo.

En esta sección también existe una particularidad muy importante dentro de la arquitectura, la cual es el *long skip connection* que realiza y que consiste en la concatenación entre mapas de características de la sección de contracción con la de expansión de manera simétrica, esto permite que se aprenda de mejor manera la información contextual de la relación que existe entre un píxel con sus vecinos, algo que funciona a manera de complemento con las capas de convolución transpuesta. El *long skip connection* también permite que se pueda recuperar detalles más finos en resoluciones mayores, como por ejemplo los bordes.

Por último, al final de la red, se aplica una operación de convolución sencilla de 1×1 para que el número de mapas de características sea el mismo que el número de clases que se desean segmentar.

4.4.2 Convolución

La convolución es una operación matemática que toma dos funciones para crear una tercera como resultado. Tiene diferentes utilidades, una de ellas es para conocer la respuesta al impulso de un sistema lineal. En el tiempo continuo la expresión de la convolución es la siguiente (Ecuación 4.1):

$$y(t) = (f(\tau) * g(\tau)) = \int_{-\infty}^{\infty} f(\tau)g(t-\tau)d\tau \quad (4.1)$$

Donde $f(\tau)$ es una función en tiempo continuo y $g(\tau)$ es una segunda función en tiempo continuo que convolucionará a la primera.

Mientras que, por otro lado, su expresión en tiempo discreto (Ecuación 4.2) es:

$$y[n] = (f[m] * g[m]) = \sum_{-\infty}^{\infty} f[m]g[n-m] \quad (4.2)$$

$$y[i, j] = (f[m, n] * g[m, n]) = \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} f[m, n]g[i-m, j-n] \quad (4.3)$$

Sin embargo, la convolución no es exclusiva para señales unidimensionales (Ecuación 4.3). En el área de procesamiento digital de imágenes, esta operación es comúnmente utilizada para aplicar un filtro espacial sobre la imagen en cuestión, estos filtros generalmente son cuadrados y de menor cantidad de píxeles que la imagen y son de dimensiones impares (aunque no existe ninguna limitante para utilizar filtros con dimensiones pares). Para realizar la convolución se necesitan definir ciertos parámetros. Dichos parámetros son:

- *Padding*: Agrega valores por encima de la orilla de la imagen con el fin de operar los valores que se encuentran en estas regiones. Los tipos de *padding* más comunes son el *zero-padding* (el cuál agrega ceros) y *mirror padding*.

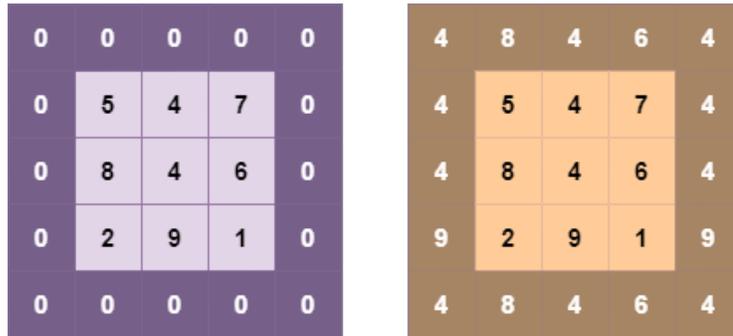


Fig. 4.9: Ejemplos de *zero padding* (izquierda) y *mirror padding* (derecha).

- *Stride*: Es la manera en la que el filtro convolucionará la imagen, es decir, se crea un determinado paso o salto que el filtro tomará para desplazarse sobre la imagen. Para un *stride* $s > 1$, el filtro se desplazará s píxeles hacia la derecha a través del eje horizontal y, posteriormente, este mismo paso será utilizado para desplazarse hacia abajo a través del eje vertical (hacia abajo).

8	5	4	7	6	5	9
3	1	4	1	2	2	8
3	6	1	7	4	9	5
9	2	4	1	5	3	8
7	6	1	2	8	3	6
7	9	5	4	9	1	3
4	6	5	7	9	2	1

Fig. 4.10 *Kernel* en su posición origen (línea punteada en negro). Cuando el caso es de *stride* de 1, el *kernel* original se desplaza hacia el *kernel* con línea punteada en rojo (dirección x). Por otro lado, cuando el caso es *stride* 2, el *kernel* original se desplaza hacia el *kernel* con línea punteada en verde (dirección y).

- Tamaño del filtro: Como ya se había mencionado, los filtros son usualmente cuadrados y de tamaño impar.

2	8	3
7	4	1
5	6	9

3	6	9	8	1
2	5	4	7	1
7	8	4	6	3
5	2	9	1	1
8	2	6	7	9

3	5	4	7	6	5	9
3	8	4	1	2	2	8
3	6	1	7	4	9	5
9	2	4	1	5	3	8
7	6	1	2	8	3	6
7	9	5	4	9	1	3
4	6	5	7	9	2	1

Fig. 4.11 Ejemplos de filtros con diferentes tamaños. De izquierda a derecha: filtro de tamaño 3×3 , 5×5 y 7×7 .

- Número de filtros: Afectan directamente a la profundidad de la salida y sirven para extraer características de diferentes formas.

4.4.3 Submuestreo

En las redes neuronales convolucionales, el submuestreo es una operación que reduce la resolución espacial de una entrada (su ancho y su largo) sin afectar la

profundidad de esta. También tiene como propósito reducir el costo computacional, así como controlar el sobreajuste en el modelo y es sistemáticamente aplicado posterior a una función *ReLU* que se aplique a la salida de una capa de convolución. La operación se realiza desplazando un *kernel* (comúnmente) de tamaño 2×2 sobre cada canal o mapa de características con un stride de 2 píxeles.

Existen diferentes maneras de realizar un submuestreo o *pooling*. Las más utilizadas son:

- *Average Pooling*: Es aquel que calcula el promedio de los valores numéricos dentro de una ventana.

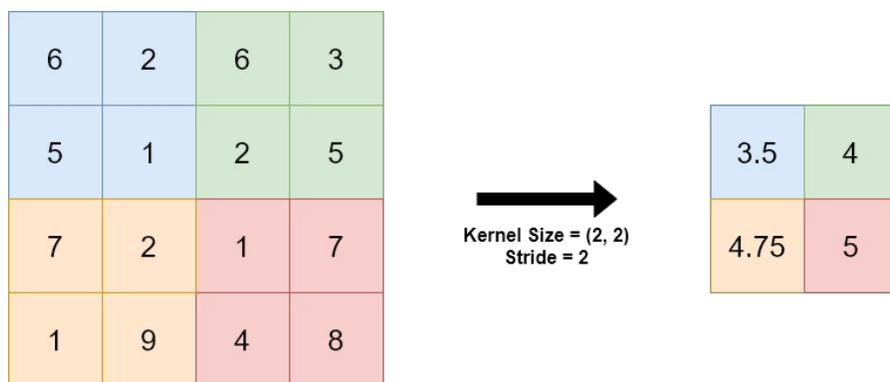


Fig. 4.12: Ejemplo de operación *average pooling*.

- *Max Pooling*: Selecciona el valor máximo entre los valores numéricos que se encuentren dentro de la ventana.

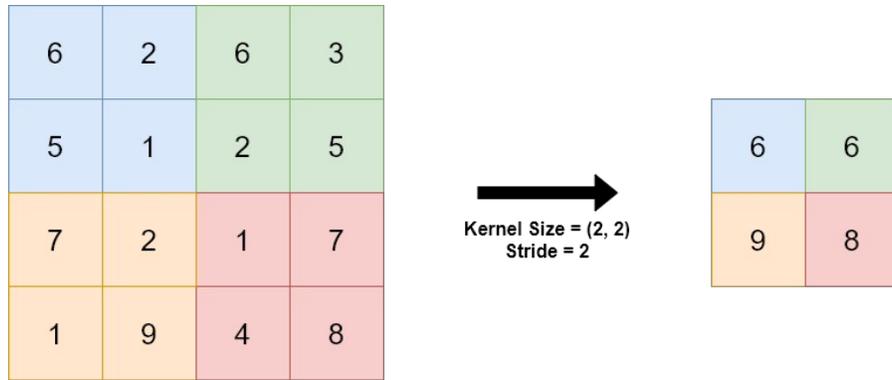


Fig. 4.13: Ejemplo de operación *max pooling*.

4.4.4 Convolución transpuesta

Cuando se trata de realizar una tarea de segmentación o localización de objetos en una imagen, no solo basta con que la red neuronal sólo haya extraído características que le son útiles aprender, sino también es necesario que conozca la localización del objeto de interés y al haber reducido la dimensión de la entrada con una serie de capas de convolución, *ReLU* y *subsampling*, esta información se va perdiendo. Afortunadamente, existen operaciones o técnicas de *upsampling*, pero una de ellas destaca por ser una técnica que aprende de los datos: la operación de convolución transpuesta.

La convolución transpuesta recupera la resolución espacial y también (con el apoyo de otras técnicas como la concatenación del mapa de características en el caso de la U-Net) la información espacial que se había perdido en la etapa de extracción de características, ayuda al modelo a conocer la localización de dónde provienen estas características y por ende a realizar de mejor manera la tarea de segmentación.

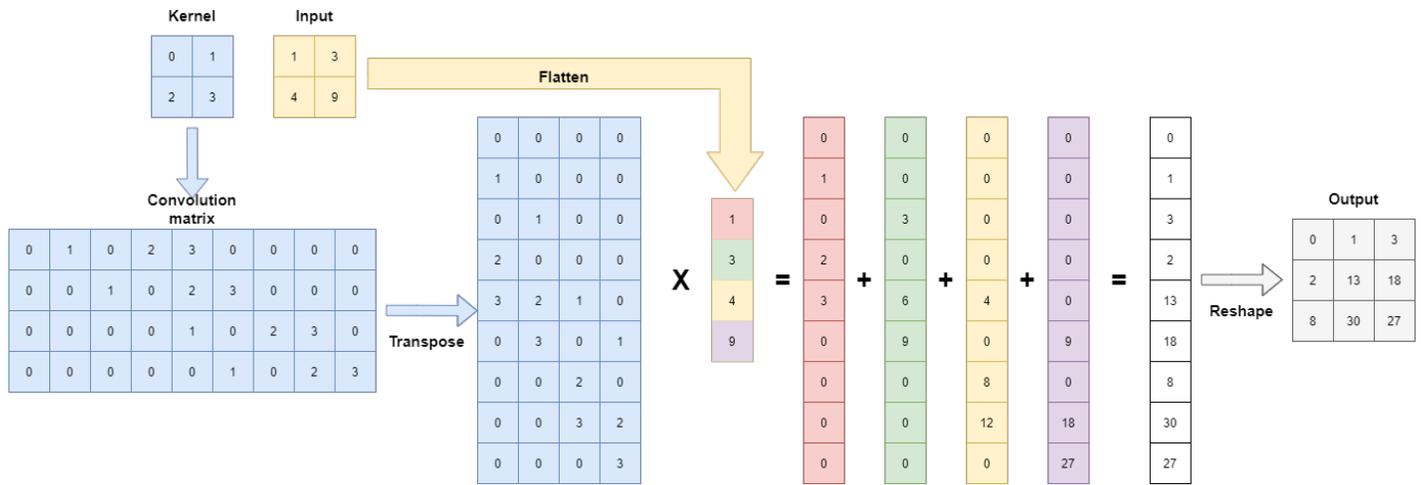


Fig. 4.14: Ejemplo de operación de convolución transpuesta.

Capítulo 5

Desarrollo y metodología

5.1 Conjunto de datos

El conjunto de datos utilizado fue proporcionado por el Centro para la Inteligencia Artificial en Medicina e Imagenología de la Universidad de Stanford, quienes pusieron a disposición de manera *online* y de forma gratuita su conjunto de datos *EchoNet-Dynamic Cardiac Ultrasound* el cual contiene más de 10 mil vídeos de ecocardiografía apical en plano de cuatro cámaras, cabe destacar que en cada vídeo se observa el ciclo cardíaco completo.

De este conjunto se extrajeron, con la ayuda de un médico especialista, *frames* correspondientes a imágenes que cumplieran el criterio que consideró el especialista. Los requisitos consistieron en seleccionar las imágenes con mejor calidad y nitidez, donde las cuatro cavidades de interés y sus delimitaciones anatómicas fueran visibles en su totalidad (tanto en sístole y diástole). Finalmente, se obtuvieron 418 imágenes con las que se construyó un conjunto final y, además, se tuvieron que crear, con la supervisión del médico especialista, las máscaras o etiquetas de las cuatro cavidades cardíacas para cada una de las imágenes del conjunto final ya que las imágenes del conjunto original no contaban con las etiquetas necesarias.

También es de mencionar que cada una de las imágenes (y de su respectiva máscara) cuenta con una resolución espacial de 112 por 112 píxeles y, que, al estar

en una escala de grises, tienen solo un canal de color (profundidad) con valores dentro de un rango de 0 a 255.

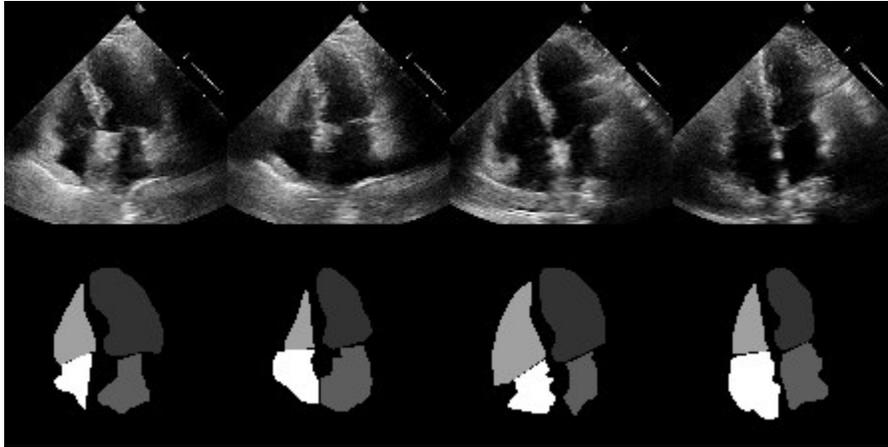


Fig. 5.1: Algunos ejemplos del conjunto de datos (fila superior) y sus respectivas máscaras (fila inferior). En las máscaras se puede visualizar las cuatro cavidades cardíacas en diferente tono de gris. En orden de tono más bajo a tono más alto: ventrículo izquierdo, aurícula izquierda, ventrículo derecho y aurícula derecha.

Por último, del total de las imágenes y de las máscaras se generaron volúmenes en el formato sin compresión *Tagged Image File Format (TIFF)* con un tamaño de 5.3 MB en cada volumen. Esto último se realizó con el propósito de agilizar su posterior lectura dentro del algoritmo.

5.2 Herramientas de desarrollo

Tomando en cuenta que la segmentación es una tarea que consiste en clasificar un grupo de píxeles de acuerdo a cierta clase, y también el hecho de que las imágenes biomédicas son generalmente residentes de objetos complejos con estructuras a veces difíciles de detectar, se habían contemplado el desarrollo de algunas de las dos técnicas muy comunes para estas tareas de segmentación, por una parte, el algoritmo de Active Shape Model (ASM) [39] y, por otro lado, el algoritmo de aprendizaje profundo llamado U-Net.

De manera preliminar se trabajó primero con la segmentación de una sola cavidad cardíaca (Ventrículo izquierdo) y posteriormente en las cuatro cavidades cardíacas en ambos algoritmos. Los resultados obtenidos del primer caso en ambos algoritmos fueron buenos, aunque los resultados numéricos del algoritmo U-Net superaron a los del algoritmo de ASM (coeficiente Dice 0.86 vs 0.78, respectivamente). El segundo caso de segmentación de las cuatro cavidades cardíacas fue decisivo para decidir trabajar con el algoritmo U-Net, debido a que la complejidad de la tarea había aumentado considerablemente y U-Net arrojó resultados preliminares que seguían siendo buenos (Coeficiente Dice de 0.89). Justo a todo lo mencionado dentro este trabajo preliminar es que se decidió seguir el desarrollo de una U-Net por sobre el algoritmo de Active Shape Model.

Por otra parte, dentro del campo del aprendizaje profundo existen diversas herramientas de alto nivel para facilitar el desarrollo de aplicaciones de algoritmos de dicha índole. Entre estas herramientas sin duda destacan las siguientes: PyTorch, Tensorflow, Keras y Theano [52].

Se eligió a PyTorch por ser una biblioteca de código abierto que proporciona funciones de alto nivel e intuitivo para el cálculo y operaciones con tensores, y además de brindar aceleración por GPU, PyTorch también es capaz de procesar los datos de manera paralela dentro de la unidad de procesamiento en la que se esté trabajando (CPU o GPU).

Finalmente, es de mencionar que se utilizó Python como el lenguaje de programación de desarrollo junto con otras bibliotecas exclusivas de Python, tales como: NumPy, Matplotlib, Python Imaging Library y Scikit-image con la finalidad de realizar operaciones auxiliares.

5.3 Arquitectura propuesta

Se eligió la red neuronal completamente convolucional *U-Net* para desarrollar el modelo propuesto, gracias a su capacidad de trabajar con conjuntos de datos limitados en cantidad, un caso que suele producirse de manera frecuente en la

mayoría de veces cuando de aplicaciones biomédicas se refiere. Al contar con un conjunto de 418 imágenes para una aplicación biomédica, además de tomar en cuenta el éxito que tienen las redes convolucionales en el área de visión computacional, esta red neuronal fue sin duda la primera opción contemplada.

La arquitectura propuesta para este trabajo consiste de los siguientes elementos:

- Cuatro bloques de contracción.
- Un bloque de *bottleneck*.
- Cuatro bloques de expansión.
- Un bloque final.

Cabe señalar que la arquitectura con la que se trabajó se basa en la arquitectura original propuesta por los autores en su publicación: *U-Net: Convolutional Networks for Biomedical Image Segmentation* [7].

Por otra parte, en cada uno de los bloques mencionados anteriormente se cuentan con capas de convolución, y a su vez, cada una de ellas cuenta con normalización por lote y activación *ReLU*. En la Figura 5.2 se puede observar la estructura de algunos de estos bloques.

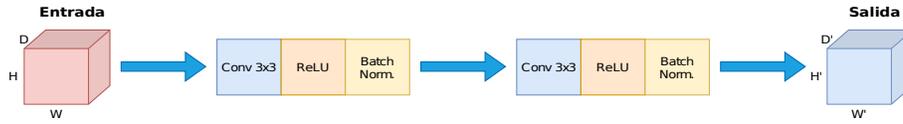


Fig. 5.2: Estructura del bloque de contracción, *bottleneck* y expansión. Se puede observar que existen tres diferentes capas: convolución, *ReLU* y de normalización por lote.

Es de considerar la importancia de la normalización por lote, ya que esta ayuda a que el modelo entrene de manera más rápida, la inicialización de parámetros sea más fácil y reduzca la posibilidad de que se presente el famoso problema de *overfitting*. Esta regularización consiste en estandarizar el lote en media 0 y varianza 1.

Dado un lote $B = \{x_1, x_2, \dots, x_m\}$ y parámetros entrenables γ y β . Se calcula la media y varianza del lote de acuerdo a las ecuaciones 5.1 y 5.2, respectivamente:

$$\mu_B = \frac{1}{m} \sum_{i=1}^m x_i \quad (5.1)$$

$$\sigma_B^2 = \frac{1}{m} \sum_{i=1}^m (x_i - \mu_B)^2 \quad (5.2)$$

Donde B es el lote a normalizar con m elementos, μ_B la media del lote y σ_B la desviación estándar del lote.

Se aplica la normalización a cada elemento x_i del lote después de calcular sus estadísticas (Ecuación 5.3):

$$\hat{x}_i = \frac{x_i - \mu_B}{\sqrt{\sigma_B^2 + \epsilon}} \quad (5.3)$$

Donde ϵ es un valor que se agrega al denominador para evitar divisiones por cero.

Por último, se aplica la escala y desplazamiento a cada elemento \hat{x}_i (Ecuación 5.4):

$$y_i = \gamma \hat{x}_i + \beta \quad (5.4)$$

Donde γ y β son parámetros entrenables para escalar y desplazar cada uno de los datos estandarizados del lote.

Estos últimos parámetros se ajustan lo suficiente durante el entrenamiento para mantener la expresividad de la capa a la vez que se acelera el tiempo entrenamiento de este, es decir, al mismo tiempo que se acelera la convergencia.

El lote por normalización puede ser usado tanto antes o después de la función de activación.

Por otra parte, en la capa *ReLU* radica esta función de activación que es descrita como en la Ecuación 5.5:

$$y_i = \begin{cases} x_i & \text{si } x_i \geq 0 \\ 0 & \text{si } x_i < 0 \end{cases} \quad (5.5)$$

Es de las funciones que comúnmente se utilizan en redes convolucionales, se caracteriza por ser una función de activación que no sufre saturación y evita el desvanecimiento del gradiente, algo que podría pasar con otras funciones de activación como la sigmoide o la tangente hiperbólica. La *ReLU* evita valores negativos que resulten después de las capas convoluciones como una forma de reducir costo computacional y acelerar el proceso de aprendizaje.

Al final de la red neuronal, se aplica un bloque extra denominado como bloque final (Figura 5.3).

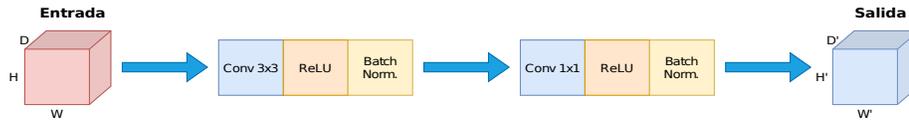


Fig. 5.3: Bloque final, a diferencia del diagrama presentado en la Figura 5.2, en el bloque final se cambia la segunda capa de convolución 3×3 por una de 1×1 .

Se puede observar que existe una última convolución de 1×1 , esta sirve para que cada mapa de características que entra sea mapeado al número de clases deseadas en la salida final.

Por otro lado, el esquema completo de la arquitectura se presenta en la Figura 5.4, mientras que en la tabla 5.1 se presenta de manera más estructurada la misma arquitectura junto al número de parámetros.

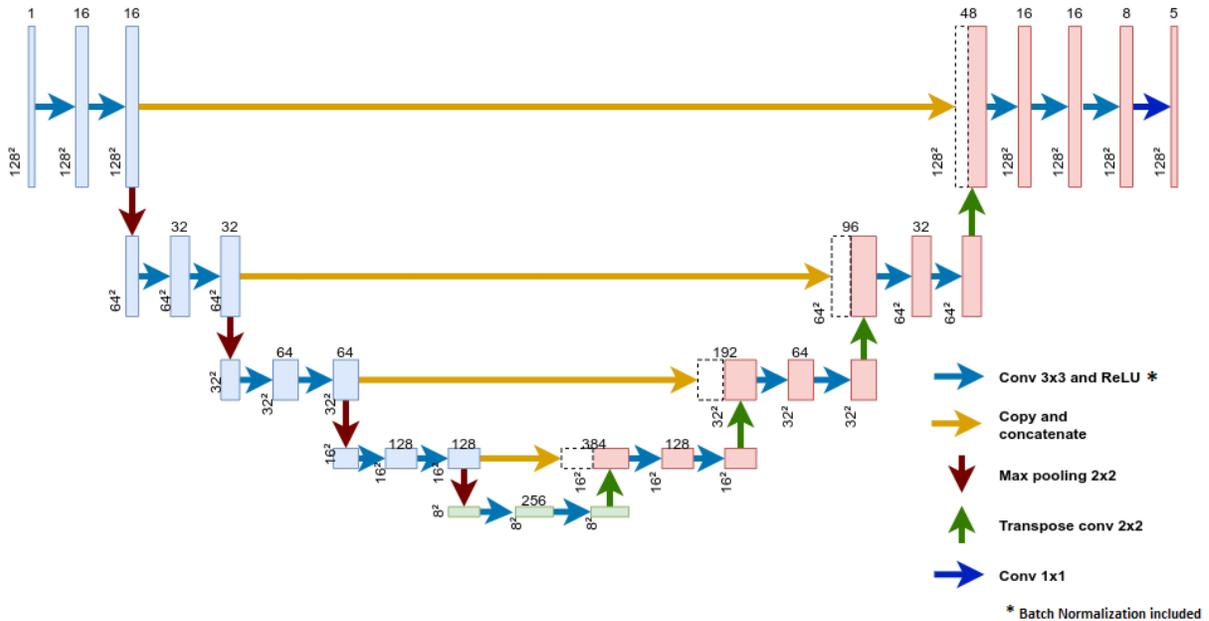


Fig. 5.4: Esquema de arquitectura U-Net utilizada.

Tamaño de entrada	Bloque	Operación	Número de parámetros
128×128×1	Contracción 1	Conv2d 3×3, ReLU, BatchNorm2d	192
128×128×16		Conv2d 3×3, ReLU, BatchNorm2d	2,352
128×128×16		MaxPool2d	
64×64×16	Contracción 2	Conv2d 3×3, ReLU, BatchNorm2d	4,704
64×64×32		Conv2d 3×3, ReLU, BatchNorm2d	9,312
64×64×32		MaxPool2d	
32×32×32	Contracción 3	Conv2d 3×3, ReLU, BatchNorm2d	18,624
32×32×64		Conv2d 3×3, ReLU, BatchNorm2d	37,056
32×32×64		MaxPool2d	
16×16×64	Contracción 4	Conv2d 3×3, ReLU, BatchNorm2d	74,112
16×16×128		Conv2d 3×3, ReLU, BatchNorm2d	147,840
16×16×128		MaxPool2d	
8×8×128	Bottleneck	Conv2d 3×3, ReLU, BatchNorm2d	295,680
8×8×256		Conv2d 3×3, ReLU, BatchNorm2d	590,592
8×8×256		ConvTranspose2d	262,400
16×16×256		Concat	

$16 \times 16 \times 384$	Expansión 4	Conv2d 3×3 , ReLU, BatchNorm2d	442,752
$16 \times 16 \times 128$		Conv2d 3×3 , ReLU, BatchNorm2d	147,840
$16 \times 16 \times 128$		ConvTranspose2d	65,664
$32 \times 32 \times 128$		Concat	
$32 \times 32 \times 192$	Expansión 3	Conv2d 3×3 , ReLU, BatchNorm2d	110,784
$32 \times 32 \times 64$		Conv2d 3×3 , ReLU, BatchNorm2d	37,056
$32 \times 32 \times 64$		ConvTranspose2d	16,448
$64 \times 64 \times 64$		Concat	
$64 \times 64 \times 96$	Expansión 2	Conv2d 3×3 , ReLU, BatchNorm2d	27,744
$64 \times 64 \times 32$		Conv2d 3×3 , ReLU, BatchNorm2d	9,312
$64 \times 64 \times 32$		ConvTranspose2d	4,128
$128 \times 128 \times 32$		Concat	
$128 \times 128 \times 48$	Expansión 1	Conv2d 3×3 , ReLU, BatchNorm2d	6,960
$128 \times 128 \times 16$		Conv2d 3×3 , ReLU, BatchNorm2d	2,352
$128 \times 128 \times 16$	Final	Conv2d 3×3 , ReLU, BatchNorm2d	1,176
$128 \times 128 \times 8$		Conv2d 1×1 , ReLU, BatchNorm2d	55
Total de parámetros			2,315,135

Tabla 5.1: Estructura de la arquitectura U-Net y el total de parámetros.

5.4 Entrenamiento

Para el proceso de entrenamiento se utilizó el siguiente algoritmo a alto nivel:

Algoritmo *U-Net*

1. Se carga conjunto de datos.
2. Realizar un *split* al conjunto de datos (conjunto de entrenamiento, validación y prueba).
3. Crear un objeto tipo *dataset* donde se realicen algunas transformaciones morfológicas o numéricas del conjunto de datos.
4. Crear los *dataloaders* en donde se define el tamaño de lote, número de procesos paralelos y, si se desea, desordenar los datos.
5. Definir la arquitectura *U-Net*.
6. Definir hiperparámetros.
7. Iniciar entrenamiento.
8. **Hacer:**
 - Seleccionar un lote de entrenamiento.
 - Ingresar dicho lote al modelo.
 - Calcular la función de pérdida.
 - Actualizar pesos.
9. **Hasta:** Cumplir el criterio definido de número de épocas

5.4.1 Datos

El conjunto de datos fue dividido en tres subconjuntos:

- Entrenamiento: Subconjunto exclusivo para el entrenamiento del modelo de aprendizaje. Este conjunto ayuda encontrar los pesos óptimos.
- Validación: Especialmente para medir el desempeño que está teniendo el modelo conforme cada etapa de entrenamiento.
- Prueba: Son ejemplos que el modelo no ve durante el entrenamiento, usados para replicar una situación real para asegurar el correcto funcionamiento del modelo.

De las 418 imágenes con las que se cuenta, el 80% corresponde a imágenes de entrenamiento y 10% tanto para validación como pruebas, lo que resulta en 334, 41 y 43 imágenes respectivamente. Existe una importancia que se considera al momento de separar el conjunto de datos, y es el hecho que es útil para detectar si el modelo cae en un caso de *overfitting*, donde el modelo tiene buen desempeño con los ejemplos que observó, pero en cambio, uno pobre con nuevos ejemplos, o en un caso de *underfitting* donde el modelo tendrá un pobre desempeño tanto con los ejemplos que observó como con los nuevos (Figura 5.5).

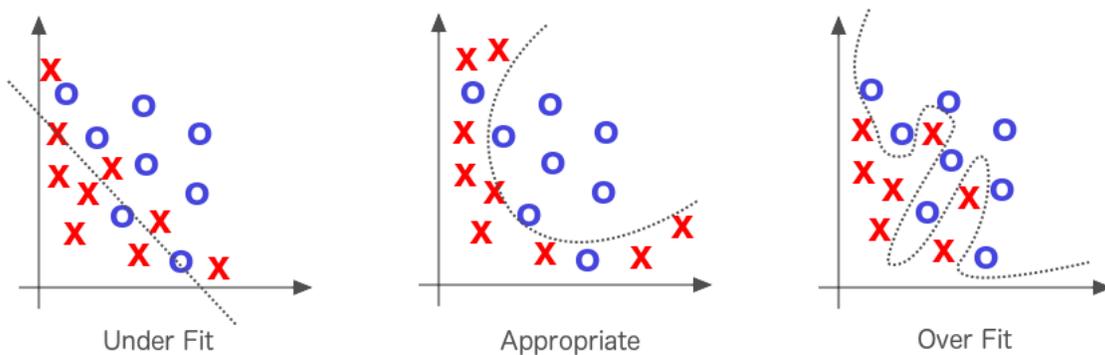


Fig. 5.5 Casos de *underfitting* (izquierda), un *fit* apropiado (centro) y *overfitting* (derecha). Figura tomada de: <https://www.kaggle.com>

Cabe destacar que se tomaron las siguientes medidas con el conjunto de datos:

- Se redimensionaron todas las imágenes de 112×112 píxeles a 128×128 píxeles, con la intención de trabajar con una dimensión de potencia dos para llevar un tipo de orden y consistencia al momento de reducir la resolución espacial durante la primera etapa de contracción de la U-Net.
- Se realizó una normalización de valores de píxeles de las imágenes a valores entre 0 y 1 para que la red neuronal aprendiera de forma óptima los parámetros de entrenamiento, ya que en caso contrario, los pesos llegarían a tornarse en valores grandes conforme se van actualizando y provocando una inestabilidad numérica que, a su vez, conduce a un desempeño pobre.
- Al decidir utilizar lotes de entrenamiento, con la ayuda de la experimentación se debe encontrar un tamaño para estos con el que el modelo tenga un buen desempeño y no sólo inclinarse a un tamaño muy pequeño. Esto condujo a la decisión de dividir el conjunto de entrenamiento en lotes de tamaño 8 para disminuir el tiempo de entrenamiento y la carga computacional que se pueda generar en comparación con entrenar con todo el conjunto de imágenes.
- Y finalmente, también mencionar que durante el entrenamiento se desordenaron los ejemplos al momento de alimentar el modelo para evitar que este aprendiera del orden de los ejemplos y para darle un comportamiento diverso durante el entrenamiento.

5.4.2 Función de pérdida

Al tratarse de un problema multiclase se utilizó la función de pérdida de entropía cruzada (Ecuación 5.6) que fue aplicada a una función *softmax* con entrada n -dimensional x :

$$\text{loss}(x, \text{clase}) = -\log \left(\frac{\exp(x[\text{clase}])}{\sum_j \exp(x[j])} \right) \quad (5.6)$$

Donde $x[\text{clase}]$ es la probabilidad predicha calculada de la clase positiva clase y N_c es el número de clases.

5.4.3 Optimizador

La importancia de un optimizador recae en la generación de pesos cada vez mejores de manera eficiente, por lo que elegir uno es sumamente relevante en el desarrollo de modelo de aprendizaje.

Adam (*ADAPtative Moment estimation*), propuesto en 2015 es un método de optimización estocástica que solo requiere gradientes de primer orden y calcula tasas de aprendizaje adaptativos para diferentes parámetros a partir de estimaciones de primer y segundo momento de los gradientes [41]. Este también es un método sencillo de implementar, computacionalmente eficiente, que requiere pocos requerimientos de memoria y que es apto para problemas que usan gran cantidad de datos y/o de parámetros [41].

5.4.4 Hiperparámetros

Las redes neuronales cuentan con una serie de hiperparámetros configurables que sirven para ajustar el algoritmo de una u otra determinada manera, según hayan sido estos configurados.

Los hiperparámetros de entrenamiento para este modelo fueron definidos de la siguiente manera:

- Tamaño de mini lote: 8.

- Número de épocas: 50.
- Tasa de aprendizaje: $1e-3$.

Es importante destacar que este último hiperparámetro debe ser lo suficientemente bueno para que los cambios de pesos que se van actualizando por cada época de entrenamiento permitan al modelo adaptarse correctamente a la tarea para la que se está entrenando.

Capítulo 6

Resultados

El análisis exhaustivo de resultados en un modelo de aprendizaje como el desarrollado en este trabajo es de suma importancia ya que permite conocer el desempeño general del modelo, no solo el que pueda tener al final del entrenamiento, sino que también es importante conocer su desempeño durante el transcurso de este.

Existen diferentes maneras de evaluar un modelo de aprendizaje, especialmente de manera cuantitativa a través de números o cifras, y de manera cualitativa por medio de resultados visuales. Todos los análisis correspondientes a los resultados obtenidos y al desempeño del modelo se estarán presentando en este apartado.

6.1 Métricas utilizadas

Las métricas que se utilizaron para evaluar el modelo son tres y que generalmente se utilizan en tareas de segmentación.

- Coeficiente de Sorensen-Dice: Métrica que está en rango $[0,1]$, descrita por la siguiente expresión (Ecuación 6.1).

$$dicecoeff = 2 \frac{NoP(target \cap prediction)}{NoP(target) + NoP(prediction)} \quad (6.1)$$

Donde **NoP** es el número de píxeles.

- *Intersection Over Union (IoU)*: También conocido como el índice Jaccard calcula la proporción de solapamiento (Ecuación 6.2).

$$IoU = \frac{NoP(target \cap prediction)}{NoP(target \cup prediction)} \quad (6.2)$$

Donde **NoP** es el número de píxeles.

- *Mean Pixel Accuracy*: Mide la precisión media de píxeles correctamente clasificados.

$$PA = \frac{NoP(correctlyclassified)}{ToP(class)} \quad (6.3)$$

$$MeanPA = \frac{1}{m} \sum_{i=1}^m PA_i \quad (6.4)$$

Donde **PA** es *Pixel Accuracy*, **NoP** es número de píxeles, **ToP** es total de píxeles y **m** es el número de clases.

Las métricas obtenidas en la etapa de validación se presentan en la Tabla 6.1

Métrica	Valor obtenido
Coefficiente de Sorensen-Dice	0.9196
Intersection over Union	0.8517
Mean Pixel Accuracy	0.9352

Tabla 6.1: Métricas obtenidas en la etapa de validación.

De acuerdo a los resultados presentados en la Tabla 6.1, se presentó una mejor cifra en *Mean Pixel Accuracy* con un 93.52% de precisión, enseguida fue el coeficiente de Dice con un 91.96% y, por último, *IoU* con un 85.17%. De este último puede existir una controversia por ser la más baja por una diferencia de 6.79% con respecto al coeficiente de Dice, sin embargo, es necesario también analizar resultados cualitativos que serán presentados más adelante en este capítulo.

6.2 Gráficas de pérdida y precisión

Se produjeron dos gráficas de históricos, una para la pérdida donde se presenta tanto para el caso de entrenamiento como el de validación y otra para las métricas utilizadas. El generar estas gráficas es importante para revisar si el modelo está presentando algún problema de entrenamiento como lo puede ser el *overfitting* o si es que está teniendo un comportamiento poco usual que pueda comprometer a la tarea de clasificación que está realizando el modelo.

En la Figura 6.1 se puede apreciar que el *overfitting* no se presentó durante el entrenamiento del modelo, ya que prácticamente las dos gráficas, tanto de entrenamiento como de validación, son descendentes a lo largo de las épocas de entrenamiento. Aunque en la validación existió un repunte alrededor de la época 20, este no es significativo y no afectó al comportamiento descendente de la gráfica de validación que posteriormente fue teniendo.

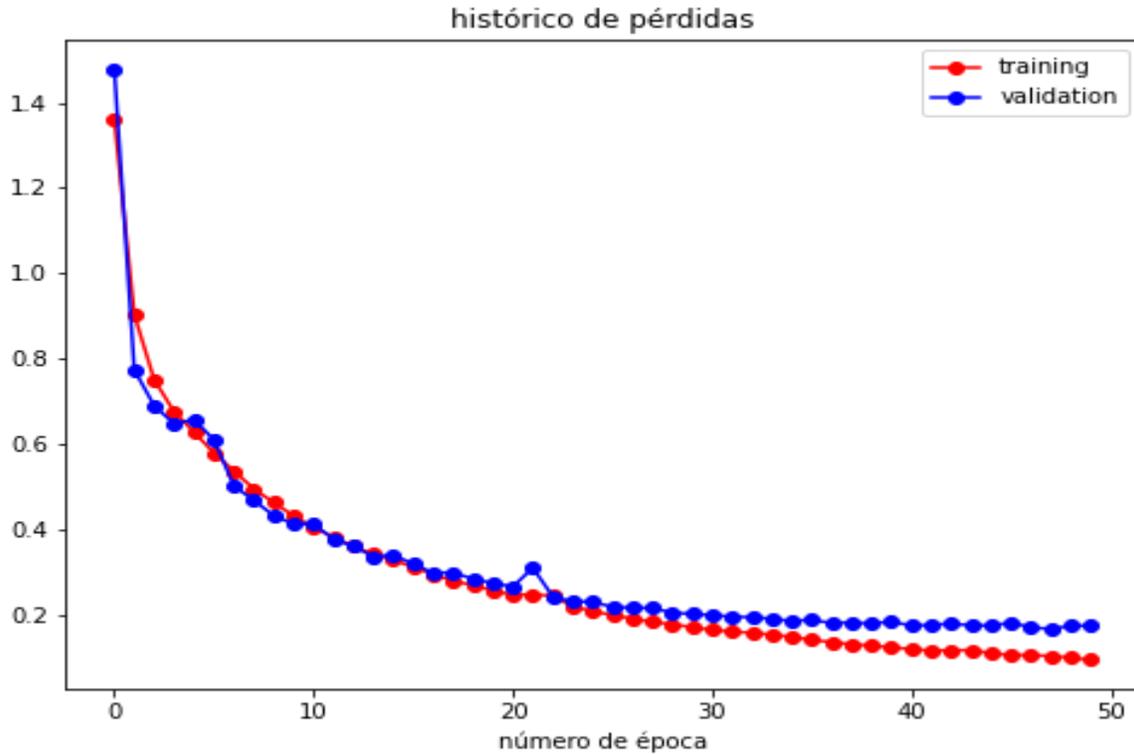


Fig. 6.1: Histórico de pérdidas. Una de las señales más claras de que no existe overfitting es que ambas pérdidas (entrenamiento y validación) desciendan a lo largo del entrenamiento.

Por otro lado, para la gráfica de historico de métricas, se puede observar en la Figura 6.2 su evolución a lo largo del proceso de validación del modelo de aprendizaje. Las métricas tuvieron un ascenso muy significativo en las primeras épocas, con una bajada notable alrededor de la época 5 y un repunte hacia números que persistieron en las métricas con un comportamiento regularmente estable a lo largo de las épocas de entrenamiento. Finalmente, es de destacar que las métricas terminaron por reflejar un buen desempeño del modelo de aprendizaje.

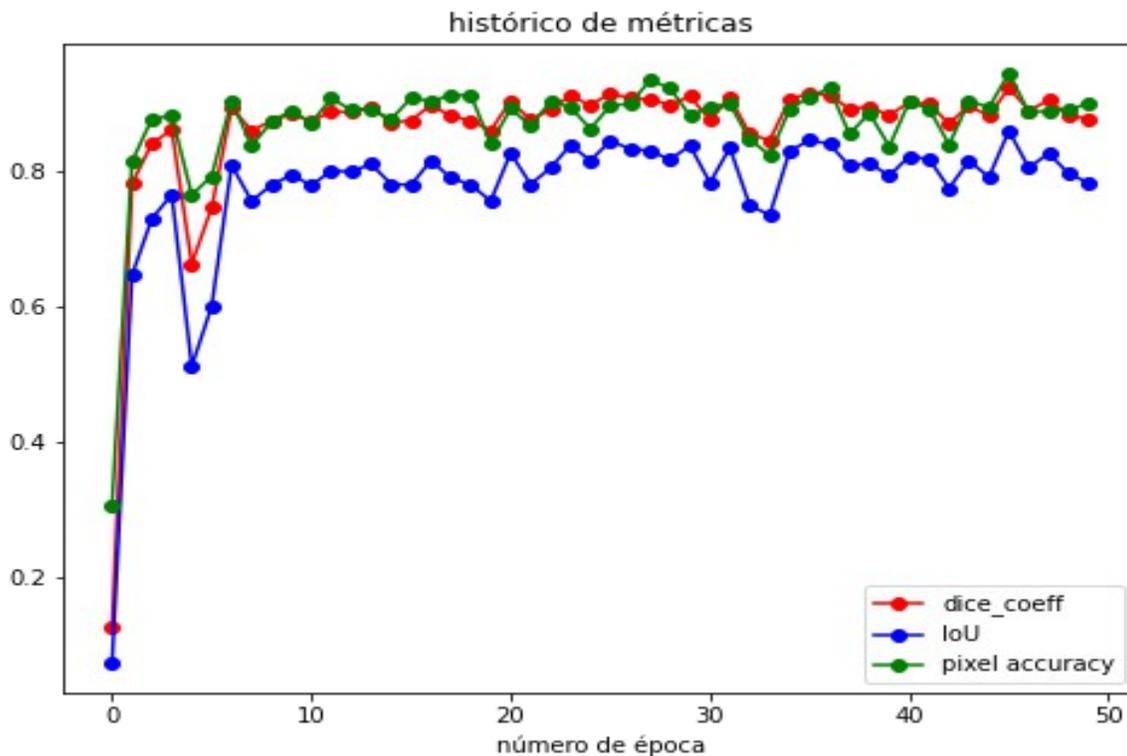


Fig. 6.2: Histórico de métricas a lo largo del entrenamiento. Todas las métricas toman una tendencia ascendente para posteriormente mantenerse dentro de un rango de valores óptimos.

6.3 Matriz de confusión

La matriz de confusión es otra forma de visualizar el desempeño del modelo a través de una cuadrícula. Usualmente las columnas representan el *ground truth* o el *target*, mientras que las filas las predicciones hechas por el algoritmo de aprendizaje.

Existen matrices de confusión donde son “todas las clases contra todas las clases”, es decir, interclase y que consiste en presentar el total de instancias correctamente clasificadas a su respectiva clase y la cantidad de instancias clasificadas en alguna otra clase que no les correspondía. Sin embargo, debido a que se realizó una clasificación a un gran número de píxeles, para esta ocasión, nos interesaría más presentar una fracción o porcentaje de instancias correctamente clasificadas en su respectiva clase ya que esto sería mucho más representativo como resultado y observar qué tan bien se clasificaron los píxeles.

Se construyó la matriz de confusión interclase (Figura 6.3) utilizando cinco clases: el fondo, el ventrículo izquierdo, aurícula izquierda, ventrículo derecho y aurícula derecha. Los valores están en un rango de 0 a 1. Dentro de esta matriz se puede observar diferentes tasas: para el fondo se obtuvo una tasa de 98.4% de píxeles que correspondían a fondo y que fueron correctamente clasificados, mientras que 1% de estos píxeles se mal clasificaron como ventrículo izquierdo, 0.3% como aurícula izquierda, 0.2% como ventrículo derecho y 0.1% como aurícula derecha. Esto quiere decir que el modelo prácticamente no tiene problema alguno para clasificar píxeles del fondo.

Para el ventrículo izquierdo se presenta una tasa de 90.9% de píxeles correctamente clasificados, mientras que el 86.5% fueron correctamente clasificados como aurícula izquierda, 86.4% ventrículo derecho y 90.4% aurícula derecha, siendo el ventrículo izquierdo y la aurícula derecha las cavidades con mayor tasa y mejores clasificados, mientras que la cavidad con menor tasa la obtuvo el ventrículo derecho.

Por otra parte, cabe resaltar que la mayor tasa de píxeles erróneamente clasificados recae en el fondo con tasas de 7.4%, 7.2% y 6.8% en los casos de ventrículo izquierdo, aurícula izquierda y ventrículo derecho, respectivamente, lo que quiere decir que es más probable que el modelo mal clasifique como fondo a un píxel de alguna de estas tres cavidades. Cabe destacar que son tasas aparentemente bajas.

En el caso de la aurícula derecha, es más probable que el modelo mal clasifique (con una tasa de 3.5%) como ventrículo derecho a un píxel de esta cavidad esto se debe a que la pared que los divide es muy estrecha o poco visible en algunas de las imágenes por lo que puede ser algo difícil encontrar un patrón (por ejemplo, bordes) que ayude a dividirlos de mejor manera, aunque aparentemente la tasa de esta incidencia es baja (35 píxeles de cada 1000) por lo que no parece representar un problema del cual preocuparse, sin embargo se estarán analizando en el siguiente apartado resultados visuales para evaluar de manera cualitativa el desempeño del modelo.

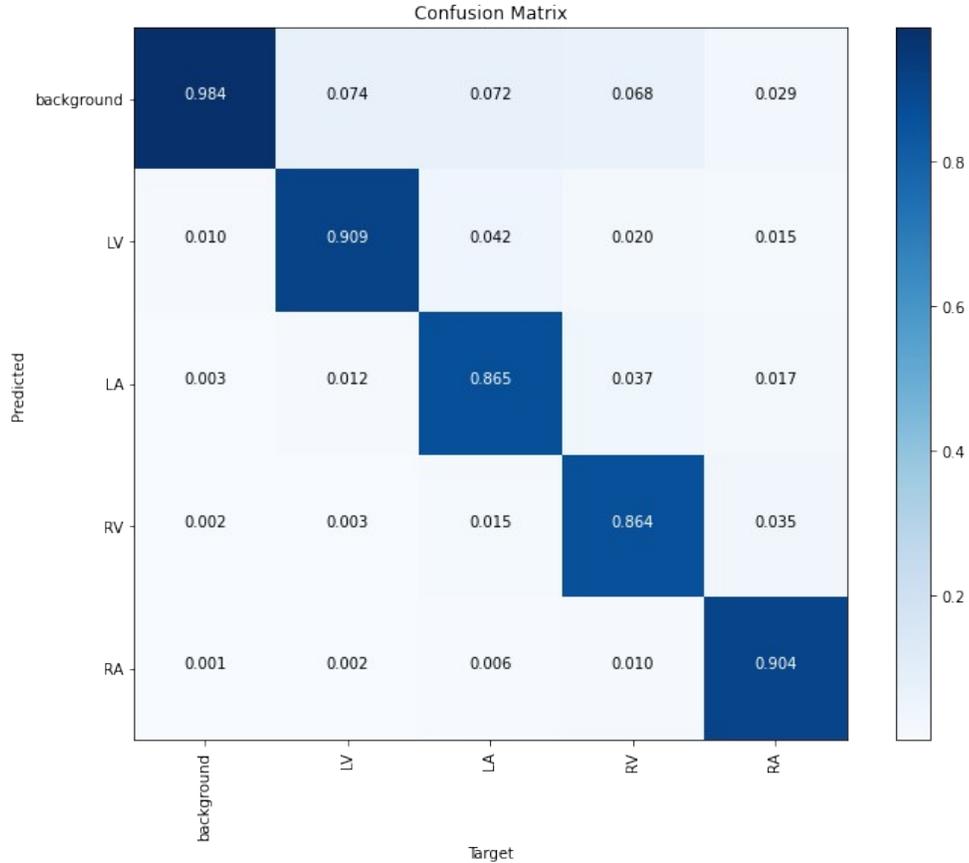


Fig. 6.3: Matriz de confusión del modelo de aprendizaje desarrollado en este trabajo. Es de destacar que las tasas para instancias incorrectamente clasificadas son bajas.

6.4 Resultados visuales

Para generar estos resultados se utilizaron ejemplos del conjunto de prueba mencionado en el apartado 5.4.1. Hay que recordar que en este conjunto existen 43 imágenes a las que se les realizó inferencia utilizando la U-Net ya entrenada. De la misma manera, cabe destacar que son imágenes que el modelo no vio durante el entrenamiento. También es importante dejar en claro que, tanto las segmentaciones como las anotaciones realizadas por el especialista, se presentan a manera de contornos dentro de las imágenes.

Dentro de los resultados obtenidos se detectaron diferentes casos que son interesantes de analizar.

Uno de estos casos se presenta en la Figura 6.4, donde las cavidades cardíacas a segmentar presentan formas muy regulares y, visualmente, el modelo se aproxima muy bien al *ground truth* definido por las anotaciones realizadas.

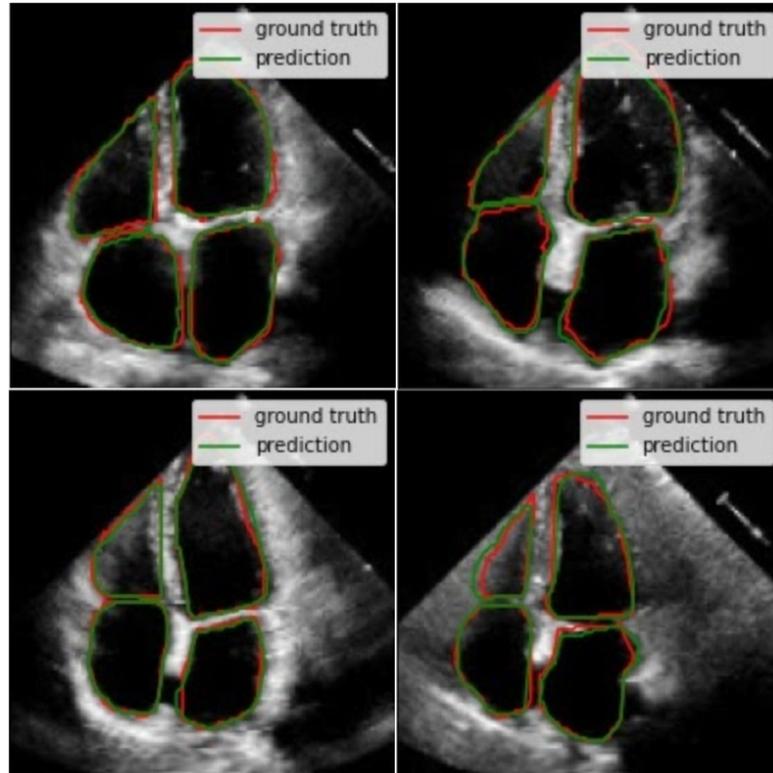


Fig. 6.4: Instancias detectadas como un primer caso de formas regulares. El modelo realiza muy buenas segmentaciones.

Para el siguiente caso, el cual se presenta en la Figura 6.5, dentro de las imágenes existen una o más cavidades con una forma más irregular respecto al caso anterior (generalmente en la aurícula izquierda). Sin embargo, el modelo, a pesar de ello, también logró realizar muy bien la segmentación, incluso presentándose traslapes casi perfectos entre la predicción y el *ground truth*. También se puede observar que el modelo utiliza los bordes para ajustarse bien a las formas de estas cavidades irregulares.

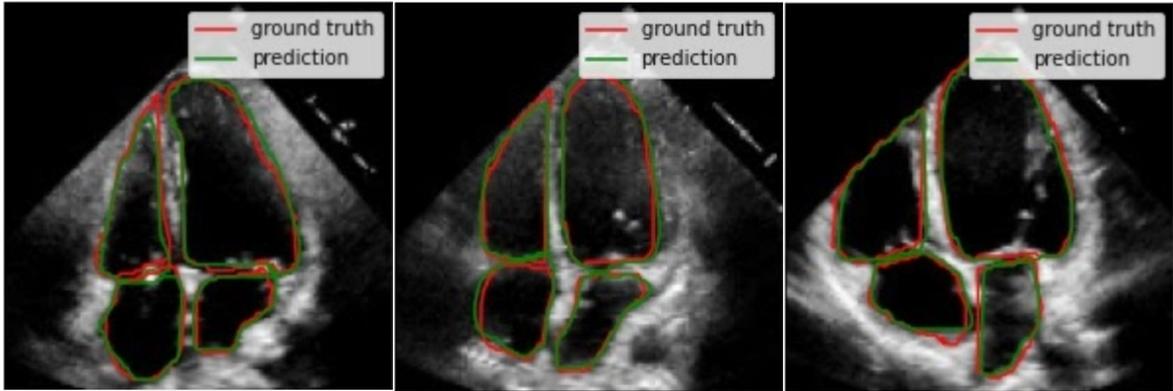


Fig. 6.5: Instancias detectadas como un segundo caso. Formas ligeramente irregulares en la aurícula izquierda, a pesar de ello, el modelo también logra realizar buenas segmentaciones.

Así como existieron casos buenos y muy buenos de segmentación por parte del modelo, existieron algunos casos donde el modelo no logra tener una segmentación igual de efectivas que los dos casos anteriores (Figura 6.6). En resumen, estas segmentaciones realizadas por el modelo, resultaron ser de menor calidad, sin embargo, tampoco son malas segmentaciones.

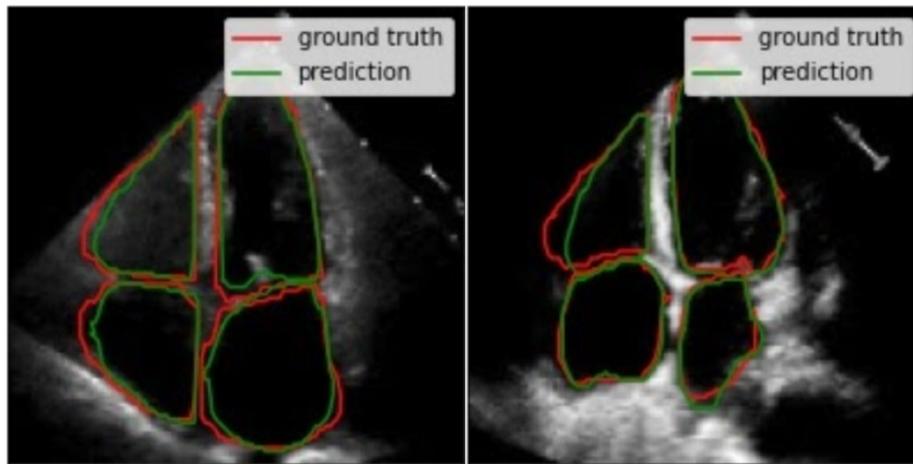


Fig. 6.6: Instancias detectadas como un tercer caso. En esta ocasión el modelo realiza segmentaciones más débiles que en los dos anteriores casos.

Para finalizar, existen resultados de este caso que se consideran interesantes de analizar debido a que se puede observar que las segmentaciones realizadas por el modelo distan, en cierto grado, de las anotaciones del *ground truth* (Figura 6.7). El modelo, a priori, parece ser capaz de ayudar a corregir anotaciones erróneas que

puedan surgir durante el análisis de estas imágenes por parte del especialista. Es interesante cómo es que en situaciones donde una anotación tenga formas muy irregulares, imposibles para una cavidad cardíaca, el modelo es capaz incluso de realizar la segmentación intentando lo más posible mantener formas regulares propias de las cavidades cardíacas.

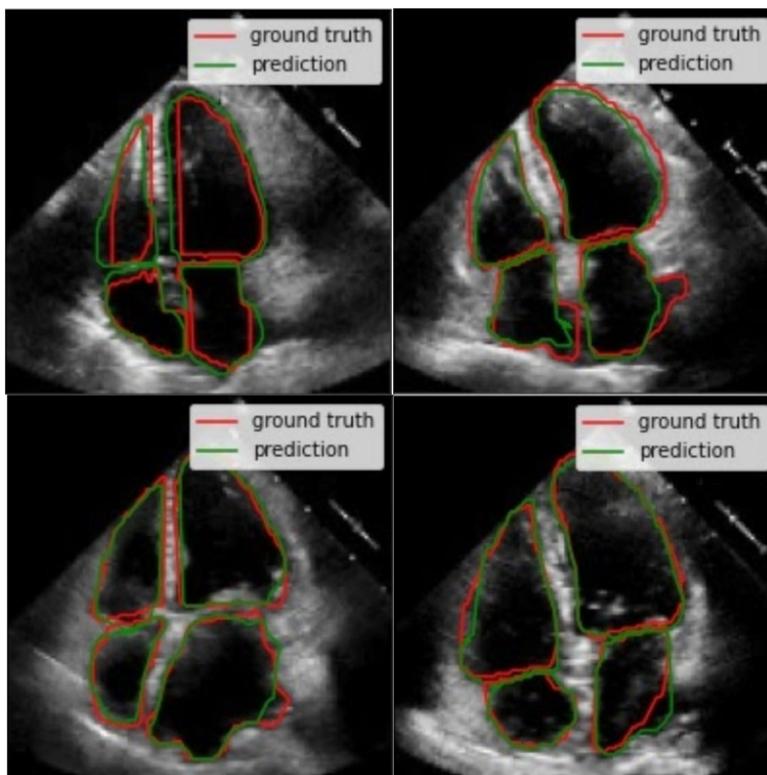


Fig. 6.7: Instancias detectadas como un cuarto caso. El modelo logra realizar buenas segmentaciones a comparación de algunas de las anotaciones.

6.5 Resultados con datos clínicos

Además de los resultados presentados anteriormente, se realizaron pruebas con 204 nuevas imágenes de ecocardiografía apical de cuatro cámaras, obtenidas de una base de datos ajena (Figura 6.8) a la base de datos descrita en el apartado 5.1.



Fig. 6.8: Una de las 204 nuevas imágenes de ecocardiografía apical de cuatro cámaras. Se observa información del estudio dentro de la imagen que posteriormente fue procesada.

Antes de alimentar a la red neuronal entrenada con estas nuevas imágenes, fue necesario crear un módulo extra para acondicionar las imágenes. Este módulo extra consistió en realizar un centrado de cada una de las ecocardiografías y lograr en estas una resolución espacial de 128×128 píxeles a través de un *cropping* (Figura 6.9).

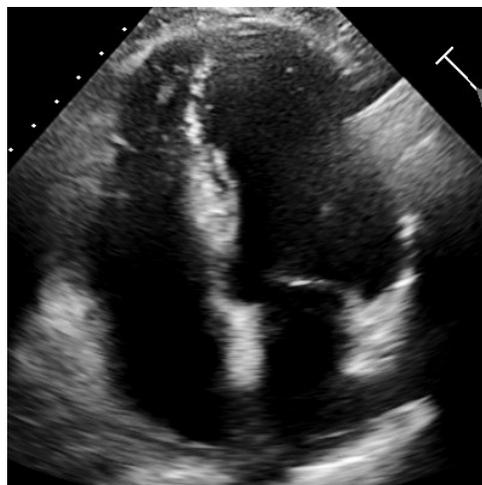


Fig. 6.9: Imagen después del proceso del módulo extra de centrado y *cropping*.

Los resultados cualitativos obtenidos en estas nuevas imágenes lograron conservar segmentaciones buenas y satisfactorias en las cuatro cavidades cardíacas. Lo más notable de los resultados es que la red neuronal artificial es capaz de mantener las formas de las cavidades cardíacas y también de buscar ajustarse a los bordes que estén presentes para realizar cada una de las segmentaciones (Figura 6.10 y Figura 6.11).

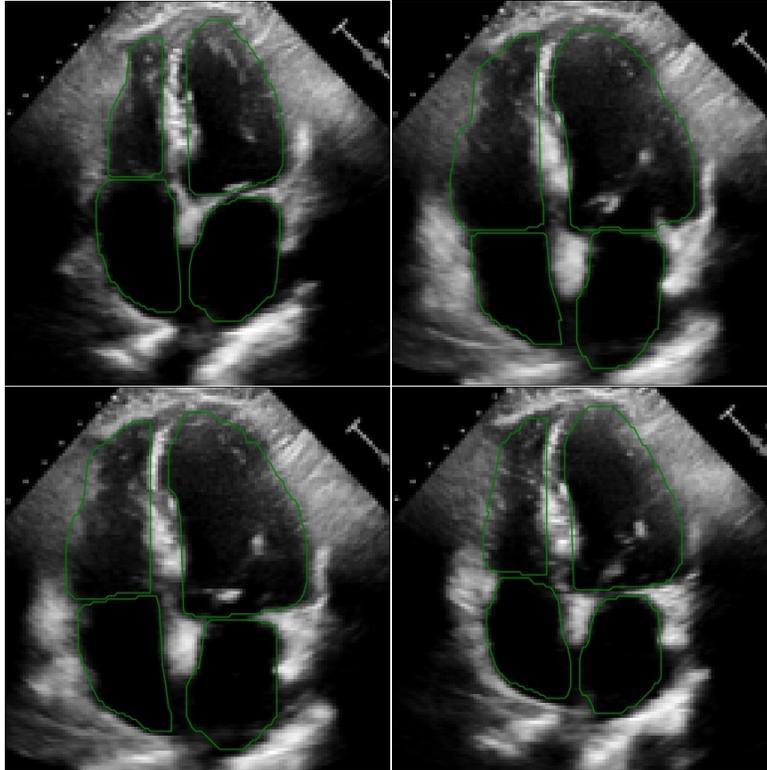


Fig. 6.10: Se puede observar que la red neuronal busca mantener la forma de cada cavidad además de buscar ajustarse a los bordes, siempre y cuando estos estén presentes dentro de la imagen.

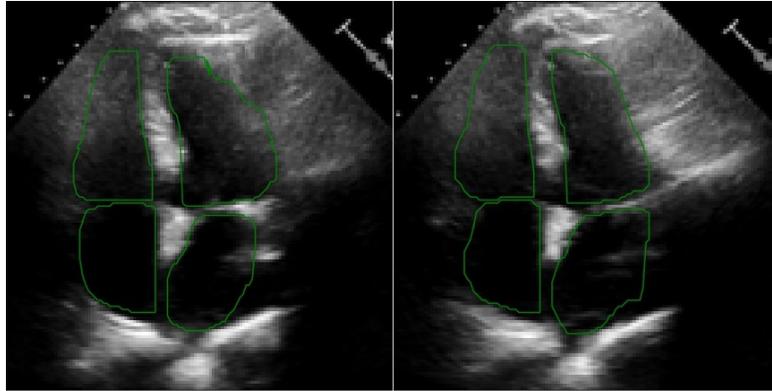


Fig. 6.11: En estos casos de mayor complicación, se observa que a pesar de que existan bordes muy difuminados (Como los bordes exteriores de los dos ventrículos), la red neuronal artificial aún es capaz de buscar mantener las formas de estas cavidades y aprovechar al máximo los bordes que puedan existir para ajustarse a la hora de segmentar.

Capítulo 7

Discusión y conclusiones

7.1 Conclusiones

Una vez más, queda demostrada la confianza que brinda una red neuronal convolucional como *U-Net* en tareas de segmentación de imágenes biomédicas, incluso si se trata de imágenes cardíacas de ultrasonido, que suelen ser complicadas de trabajar, por lo que no es de extrañarse la popularidad que estos algoritmos de aprendizaje van ganando actualmente. Es altamente recomendable utilizar esta tecnología siempre y cuando se cuente con un conjunto de datos lo suficientemente extenso y representativo del problema a resolver.

Así fue, teniendo en cuenta lo anterior, que en este trabajo se presentó un modelo de aprendizaje profundo *U-Net* en una tarea de segmentación de cavidades cardíacas que no había sido abordada de la manera que se realizó en el presente trabajo. Los resultados obtenidos son satisfactorios, se analizaron tanto numéricamente como visualmente y el algoritmo mostró responder muy bien segmentando cada una de las cavidades cardíacas de interés, incluso con nuevos datos clínicos, se replican resultados satisfactorios. Por otro lado, dentro de lo visual, el algoritmo responde muy bien a características dentro de la imagen, tales como los bordes.

Por último, también hay que hacer énfasis en que a pesar de que un algoritmo de aprendizaje profundo como el desarrollado en este trabajo presente buenos resultados, nunca debe de tomarse como verdad absoluta y tampoco busca

suplantar el trabajo realizado por los especialistas, sino que tiene un propósito como herramienta computacional para mejorar la calidad de los diagnósticos médicos, así como disminuir el tiempo de estos.

7.2 Trabajo a futuro

En este trabajo se presentó un modelo de aprendizaje profundo que realiza una segmentación automática de cada una de las cavidades cardíacas del corazón en imágenes de ecocardiografía apical de cuatro cámaras, tanto en sístole como en diástole. Al haber observado ya los buenos resultados que arrojó en la segmentación y tomando en cuenta que este trabajo forma parte de un sistema mucho más completo, se tiene como trabajo a futuro varios puntos:

- Obtener parámetros médicos de manera automática a partir de las segmentaciones realizadas por el modelo para poder clasificar entre un corazón sano y un corazón con alguna patología (que presente anomalías).
- Con la ayuda de estas segmentaciones, dar más prioridad al estudio del ventrículo derecho, una cavidad con menos relevancia clínica que el ventrículo izquierdo pero que, sin embargo, es igual de importante estudiarlo para diagnosticar enfermedades como la hipertensión pulmonar.
- Diseñar algoritmos que sean de apoyo para este modelo.
- Extrapolar el algoritmo a ecocardiografías en 3D.
- Someter al modelo y sus resultados a una revisión y/o evaluación por especialistas.

Bibliografía

- [1] Organización Mundial de la Salud, “Cardiovascular diseases (CVDs)”, 2017. [En línea]. Disponible en: [https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-\(cvds\)](https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)) [Accedido: 10-Agosto-2020].
- [2] “¿Qué causa la mayoría de las muertes”, *Institute for Health Metrics and Evaluation: México*, 2020. [En línea]. Disponible en: <http://www.healthdata.org/mexico?language=149> [Accedido: 13-Agosto-2020].
- [3] Secretaría de Salud, “Programa Nacional para la Reducción de la Mortalidad por Infarto Agudo al Miocardio IAM-MX (PREMIA)”, 2018. [En línea]. Disponible en: <http://www.calidad.salud.gob.mx/site/iam/> [Accedido: 10-Agosto-2020].
- [4] R. J. McDonald *et al.*, “The Effects of Changes in Utilization and Technological Advancements of Cross-Sectional Imaging on Radiologist Workload”, *Academic Radiology*, vol. 22, núm. 9, pp. 1191–1198, 2015, doi: 10.1016/j.acra.2015.05.007.
- [5] Imaging Technology News, “Global Medical Imaging Trends”, 2017. [En línea]. Disponible en: <https://www.itnonline.com/article/global-medical-imaging-trends> [Accedido: 14-Septiembre-2020].
- [6] C. Chen *et al.*, “Deep Learning for Cardiac Image Segmentation: A Review”, *Frontiers in Cardiovascular Medicine*, vol. 7, 2020, doi:10.3389/fcvm.2020.00025.
- [7] O. Ronneberger, P. Fischer y T. Brox, “U-Net: Convolutional Networks for Biomedical Image Segmentation”, *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 2015, vol. 9351, pp. 234–241, 2015. doi: 10.1007/978-3-319-24574-4_28
- [8] V. Zyuzin *et al.*, “Identification of the left ventricle endocardial border on two-dimensional ultrasound images using the convolutional neural network Unet”, *2018 Ural Symposium on Biomedical Engineering, Radioelectronics and Information Technology (USBREIT)*, 2018, doi: 10.1109/usbereit.2018.8384554.
- [9] J. Borrajo-Sánchez y F. Cabrero-Fraile, "Tomografía por emisión de positrones (PET): fundamentos y limitaciones tecnológicas.", *Sociedad Española de Oftalmología*, vol. 85, no. 4, pp. 129-130, 2010. [En línea]. Disponible en: https://scielo.isciii.es/scielo.php?script=sci_arttext&pid=S0365-66912010000400001. [Accedido: 21-Abril-2021].

- [10] CancerQuest, "La Resonancia Magnética (RM)", 2021. [En línea]. Disponible en: <https://bit.ly/3mIHbTv> [Accedido 21-Abril-2021].
- [11] M. Verdugo P., "Tomografía Computada Multicorte", *Revista Chilena de Cirugía*, vol. 56, no. 2, pp. 185-190, 2004. [En línea]. Disponible en: <https://bit.ly/3nViCCh>. [Accedido 26-Marzo-2021].
- [12] "World Population Ageing (Highlights)", United Nations, Department of Economic and Social Affairs, Population Division, Nueva York, EE. UU., pp. 1-9, 2017. [En línea]. Disponible en: https://www.un.org/en/development/desa/population/publications/pdf/ageing/WPA2017_Highlights.pdf [Accedido: 30-Marzo-2021].
- [13] The Cancer Atlas, "The Burden of Cancer", 2019. [En línea]. Disponible en: <https://bit.ly/3wg2UFN> [Accedido: 15-Abril-2021].
- [14] R. E. Berg, "Sound", *Britannica*, 2021, [En línea]. Disponible en: <https://www.britannica.com/science/sound-physics>
- [15] M. García-Fernández, J. Zamorano y J. García-Robles, "Conceptos generales de la ecocardiografía", en *Manual de ecocardiografía: Indicaciones e interpretación en la práctica clínica*, M.A. García Fernández y J.L. Zamorano, Ed. Madrid, España: EDIMED, pp. 1-17.
- [16] H. Feigenbaum, "Echocardiography: An overview", *Journal of the American College of Cardiology*, vol. 1, no. 1, pp. 216-224, 1983.
- [17] I. Edler y K. Lindström, "The history of echocardiography", *Ultrasound in Medicine & Biology*, vol. 30, no. 12, pp. 1565-1644, 2004. doi: 10.1016/s0301-5629(99)00056-3.
- [18] M. Maleki y M. Esmailzadeh, "The Evolutionary Development of Echocardiography", *Iranian Journal of Medical Sciences*, vol. 37, no. 4, pp. 222-232, 2012. [En línea]. Disponible en: <https://bit.ly/3odF4H9> [Accedido: 24-Abril-2021].
- [19] J. Ángeles-Valdés, "Ecocardiografía: Principios y aplicaciones", *Acta Médica*, no. 4, pp. 225-229, 2003.
- [20] I. Edler y C. Hertz, "The Use of Ultrasonic Reflectoscope for the Continuous Recording of the Movements of Heart Walls.", *Clinical Physiology and Functional Imaging*, vol. 24, no. 3, pp. 118-136, 1954. doi: 10.1111/j.1475-097x.2004.00539.x
- [21] H. Feigenbaum, "Ultrasound Measurements of the Left Ventricle", *Archives of Internal Medicine*, vol. 129, no. 3, pp. 461-467, 1972. doi: 10.1001/archinte.1972.00320030081009
- [22] R. Gowda, I. Khan, B. Vasavada, T. Sacchi y R. Patel, "History of the evolution of echocardiography", *International Journal of Cardiology*, vol. 97, no. 1, pp. 1-6, 2004. doi: 10.1016/j.ijcard.2003.07.018

- [23] J. Chasco Ronda, "El ecocardiograma", *Imagen Diagnóstica*, vol. 1, no. 1, pp. 14-18, 2010. doi: 10.1016/s2171-3669(10)70004-4
- [24] A. Eden, *The Search for Christian Doppler*. Vienna: Springer Wien, 1992.
- [25] R. Nishimura, F. Miller, M. Callahan, R. Benassi, J. Seward y A. Tajik, "Doppler Echocardiography: Theory, Instrumentation, Technique, and Application", *Mayo Clinic Proceedings*, vol. 60, no. 5, pp. 321-343, 1985. doi: 10.1016/s0025-6196(12)60540-0
- [26] T. Shiota, "3D echocardiography: The present and the future", *Journal of Cardiology*, vol. 52, no. 3, pp. 169-185, 2008. doi: 10.1016/j.jjcc.2008.09.004
- [27] N. Espínola-Zavaleta, F. Roldán, P. Chávez, A. Romero-Cárdenas y J. Vargas-Barrón, "Ecocardiografía tridimensional: Técnica, Aplicaciones clínicas y perspectivas", *Archivos de cardiología de México*, vol. 71, no. 1, pp. 88-95, 2001.
- [28] W. McCulloch y W. Pitts, "A logical calculus of the ideas immanent in nervous activity", *The Bulletin of Mathematical Biophysics*, vol. 5, no. 4, pp. 115-133, 1943. doi: 10.1007/bf02478259
- [29] F. Rosenblatt, "The perceptron: A probabilistic model for information storage and organization in the brain.", *Psychological Review*, vol. 65, no. 6, pp. 386-408, 1958. doi: 10.1037/h0042519
- [30] M. Minsky y S. Papert, *Perceptrons*. Cambridge, USA: M.I.T. Press, 1969.
- [31] P. Werbos, "Beyond regression : new tools for prediction and analysis in the behavioral sciences", Tesis Doctoral, Harvard University, 1974.
- [32] A. Géron, *Hands-on machine learning with Scikit-Learn and TensorFlow*, 1ra edición, California, USA: O'Reilly Media, 2017.
- [33] J. Long, E. Shelhamer y T. Darrell, "Fully convolutional networks for semantic segmentation", *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 1, pp. 3431-3440, 2014. doi: 10.1109/cvpr.2015.7298965
- [34] A. Otte et al., "PET and SPECT in whiplash syndrome: a new approach to a forgotten brain?", *Journal of Neurology, Neurosurgery & Psychiatry*, vol. 63, no. 3, pp. 368-372, 1997. doi: 10.1136/jnnp.63.3.368
- [35] M. Serna-Gandia, *Plano ecocardiográficos*. Valencia, España: Ecocritic, 2019, pp. 6-32.
- [36] W. Ertel, *Introduction to Artificial Intelligence*, 2da edición, Alemania: Springer, 2017, p. 246.
- [37] Y. Ma, K. Liu, Z. Guan, X. Xu, X. Qian y H. Bao, "Background Augmentation Generative Adversarial Networks (BAGANs): Effective Data Generation Based on GAN-

- Augmented 3D Synthesizing", *Symmetry*, vol. 10, no. 12, p. 734, 2018. doi: 10.3390/sym10120734
- [38] A. Zhang, Z. Lipton, M. Li y A. Smola, *Dive into Deep Learning*, 1ra edición. 2021.
- [39] T.F. Cootes, C.J. Taylor, D.H. Cooper, y J. Graham. "Active Shape Models Their Training and Application", *Computer Vision and Image Understanding*, 1995.
- [40] M. García-Fernández, J. Zamorano y J. García-Robles, "Técnicas ecocardiográficas especiales", en *Manual de ecocardiografía: Indicaciones e interpretación en la práctica clínica*, M. Orejas, L. Pérez de Isla y J.L. Zamorano, Ed. Madrid, España: EDIMED, pp. 239-245.
- [41] D.P.Kingma y J.L. Ba: "Adam: A Method for Stochastic Optimization". *3rd International Conference for Learning Representations*, 2015.
- [42] C. Jenkins, K. Bricknell, J. Chan, L. Hanekom y T. H. Marwick, "Comparison of two- and three-dimensional echocardiography with sequential magnetic resonance imaging for evaluating left ventricular volumen and ejection fraction over time in patients with healed myocardial infarction", *The American Journal of Cardiology*, vol. 99, no. 3, pp. 300-306, 2007. doi: /10.1016/j.amjcard.2006.08.026
- [43] C. Santos-Gallego *et al.*, "3D-echocardiography demonstrates excellent correlation with cardiac magnetic resonance for assessment of left ventricular function and volumes in a model of myocardial infarction", *Journal of the American College of Cardiology*, vol. 59, no. 13, p. E1564, 2012. doi: 10.1016/s0735-1097(12)61565-4
- [44] L.Joskowicz, "Accelerating Deep Learning Medical Image Analysis in Radiology", en *1er Seminario Virtual de Computación del CECAV UNAM*, Ciudad de México, 2020.
- [45] A. Schneider y H. Feussner, "Diagnosis Procedures", en *Biomedical Engineering in Gastrointestinal Surgery*, Academic Press, 2017, pp. 97-99.
- [46] A. M. Zador, "A critique of pure learning and what artificial neural networks can learn from animal brains", *Nature Communications*, vol. 10, no. 3770, pp. 1-7, 2019. doi: 10.1038/s41467-019-11786-6
- [47] X. S. Tomas, "Tomografía de perfusión pulmonar (SPECT) en Neumología". *Neumología y Salud*, vol. 2, no. 2, pp. 15-24, 2009.
- [48] "How does MRI work?", *National Institute of Biomedical Imaging and Bioengineering*, 2021. [En línea]. Disponible en: <https://www.nibib.nih.gov/science-education/science-topics/magnetic-resonance-imaging-mri#pid-951> [Accedido: 5-Noviembre-2021].
- [49] N. Singh, "Artificial Neural Networks Applications and Algorithms", *Xenonstack*, 2021. [En línea]. Disponible en: <https://www.xenonstack.com/blog/artificial-neural-network-applications>

- [50] O. Davydova, “10 Applications of Artificial Neural Networks in Natural Language Processing”, *Medium*, 2017. [En línea]. Disponible en: <https://medium.com/@datamonsters/artificial-neural-networks-in-natural-language-processing-bcf62aa9151a>
- [51] V. Kaushik, “8 Applications of Neural Networks”, *Analytics Steps*, 2021. [En línea]. Disponible en: <https://www.analyticssteps.com/blogs/8-applications-neural-networks>
- [52] I. Vyas, “Best Python Libraries for Machine Learning and AI”, *Citrus Bug*, 2021. [En línea]. Disponible en: <https://citrusbug.com/blog/python-libraries-for-machine-learning>
- [53] H. L. Rufiner, “¿Puede la inteligencia artificial dar una “mente” a las máquinas”, *Universidad Austral*, 2018. [En línea]. Disponible en: <https://www.austral.edu.ar/cerebroypersona/es/2018/10/puede-la-inteligencia-artificial-dar-una-mente-a-las-maquinas/>
- [54] E. Roberts, “Comparison between conventional computers and neural networks”, Stanford: *Neural Networks*, 2021, [En línea]. Disponible en: <https://stanford.io/3HGZ7X2>
- [55] Omnia Health, “Report: Medical Imaging Trends 2020”, 2020. [En línea]. Disponible en: <https://insights.omnia-health.com/reports/report-medical-imaging-trends-2020>