



16.30 - PROYECTO FINAL

INSTITUTO TECNOLÓGICO DE BUENOS AIRES

---

## PROYECTO FINAL

Estimación de la Presión Arterial No Invasiva con  
Calidad Médica a partir de PPG mediante Redes  
Neuronales Artificiales

---

*Tutores:*

ING. SZLAIN, Darío  
DR. ING. RAMELE, Rodrigo

*Autores:*

ANANÍA, Pilar (Leg. 57226)  
TORLASCHI, Megan Ann (Leg. 56603)  
TU, Alex (Leg. 56184)

PROYECTO FINAL PRESENTADO PARA LA OBTENCIÓN DEL TÍTULO  
DE BIOINGENIERÍA

BUENOS AIRES, ARGENTINA  
OCTUBRE 2021

# Índice

<b>Resumen</b>	<b>5</b>
<b>Abstract</b>	<b>6</b>
<b>1 Glosario de Contracciones</b>	<b>7</b>
<b>2 Introducción</b>	<b>11</b>
<b>3 Objetivos</b>	<b>12</b>
<b>4 Hipótesis y Alcance</b>	<b>13</b>
<b>5 Marco Teórico</b>	<b>14</b>
5.1 Presión Arterial . . . . .	14
5.1.1 Valoración de la BP . . . . .	17
5.1.2 Métodos Ambulatorios: AMPA y MAPA . . . . .	19
5.2 Fotopletismografía . . . . .	20
5.2.1 Oxímetro de Pulso . . . . .	22
5.3 Presión Arterial y Fotopletismografía . . . . .	26
5.3.1 Presión Central y Periférica . . . . .	28
5.4 Redes Neuronales Artificiales . . . . .	29
5.4.1 Funcionamiento Biológico Neuronal . . . . .	29
5.4.2 Aspectos Generales de las ANNs . . . . .	31
5.4.3 Funciones de Costo . . . . .	33
5.4.3.1 Mean Squared Error (MSE) . . . . .	34
5.4.3.2 Mean Absolute Error (MAE) . . . . .	34
5.4.3.3 Mean Squared Logarithmic Error (MSLE) . . . . .	35
5.4.4 <i>Learning Rate</i> . . . . .	35
5.4.5 Funciones de Activación . . . . .	35
5.4.5.1 Función Lineal . . . . .	36
5.4.5.2 Función Sigmoidea . . . . .	36
5.4.5.3 ReLU . . . . .	37
5.4.5.4 ELU . . . . .	38
5.4.6 Optimizadores . . . . .	39
5.4.6.1 Gradient Descent . . . . .	39
5.4.6.2 RMSprop . . . . .	40
5.4.6.3 Adam . . . . .	40
5.4.7 Fenómeno de <i>Overfitting</i> . . . . .	41
5.4.7.1 Regularización . . . . .	42
5.4.8 Tipos de ANNs . . . . .	42
5.4.9 Calibración . . . . .	48

<b>6</b>	<b>Calidad Médica</b>	<b>48</b>
6.1	Métricas de Evaluación . . . . .	48
6.2	Estándar AAMI-ISO 81060-2:2018 . . . . .	49
6.3	Criterio Simplificado . . . . .	51
<b>7</b>	<b>Estado del Arte</b>	<b>52</b>
<b>8</b>	<b>Materiales y Métodos</b>	<b>57</b>
8.1	Materiales . . . . .	57
8.1.1	Base de Datos . . . . .	57
8.1.2	Entorno de Ejecución . . . . .	59
8.1.3	Herramienta de Análisis Estadístico . . . . .	59
8.2	Métodos . . . . .	59
8.3	Descarga de la Base de Datos . . . . .	60
8.4	Descarte de Señal por Peso de Archivo . . . . .	61
8.5	Análisis Estadístico . . . . .	62
8.6	Preprocesamiento . . . . .	64
8.6.1	Sustitución de valores NaN . . . . .	66
8.6.2	Filtrado de Señal de PPG . . . . .	66
8.6.3	Filtrado de Señal de ABP . . . . .	73
8.6.4	Clasificación de Segmentos de Señal . . . . .	73
8.6.4.1	Deformaciones Morfológicas . . . . .	74
8.6.4.2	Definición de Calidad . . . . .	76
8.6.4.3	Clasificación Manual de Segmentos . . . . .	78
8.6.4.4	Clasificadores . . . . .	80
8.6.4.4.1	<i>Features</i> de Calidad de Señal . . . . .	80
8.6.4.4.1.1	Dominio Temporal . . . . .	81
8.6.4.4.1.2	Dominio Frecuencial . . . . .	82
8.6.4.4.2	Métricas de Calidad de Clasificadores . . . . .	83
8.6.4.4.3	Implementación de Clasificadores . . . . .	85
8.6.5	Extracción de Pulsos . . . . .	86
8.6.6	Correspondencia entre Pulsos . . . . .	89
8.6.7	Obtención de Presión Arterial Sistólica y Diastólica . . . . .	90
8.6.8	Obtención de Features . . . . .	92
8.6.9	<i>Zero Padding</i> . . . . .	94
8.6.10	Normalización . . . . .	94
8.7	Procesamiento . . . . .	96
8.7.1	Regresión Lineal . . . . .	96
8.7.2	MLP . . . . .	97
8.7.3	CNN . . . . .	102
8.7.3.1	Selección de Especificaciones de las CNNs . . . . .	102
8.7.3.1.1	Red V1 . . . . .	106
8.7.3.1.2	Red V2 . . . . .	106
8.7.3.2	Evaluación de Calidad Médica en CNNs . . . . .	107

8.8	Evaluación de Modelos . . . . .	109
8.8.1	MLP . . . . .	110
8.8.2	CNN . . . . .	110
<b>9</b>	<b>Resultados</b>	<b>113</b>
9.1	Preprocesamiento . . . . .	113
9.2	Descarte de Señal por Peso de Archivo . . . . .	113
9.2.1	Sustitución de valores NaN . . . . .	113
9.2.2	Clasificadores de Segmentos de Señal . . . . .	114
9.3	Regresión Lineal . . . . .	114
9.4	MLP . . . . .	115
9.5	CNN . . . . .	118
9.5.1	Selección de Especificaciones de las CNNs . . . . .	118
9.5.2	Evaluación de Calidad Médica en CNNs . . . . .	119
<b>10</b>	<b>Discusión</b>	<b>124</b>
10.1	Preprocesamiento . . . . .	124
10.1.1	Clasificadores de Segmentos de Señal . . . . .	124
10.2	Regresión Lineal . . . . .	124
10.3	MLP . . . . .	125
10.4	CNN . . . . .	126
10.4.1	Selección de Especificaciones de las CNNs . . . . .	126
10.4.2	Evaluación de Calidad Médica en CNNs . . . . .	127
10.4.2.1	Comparación con Trabajos Previos . . . . .	128
10.5	Limitaciones y Próximos Pasos . . . . .	129
<b>11</b>	<b>Conclusiones</b>	<b>135</b>
<b>12</b>	<b>Referencias</b>	<b>137</b>



## Resumen

La alta mortalidad asociada a las enfermedades cardiovasculares, en particular a la hipertensión, es la razón por la que resulta relevante la búsqueda de métodos que permitan un registro continuo de la presión arterial de forma simple y realizable en un entorno doméstico. El presente proyecto hace foco en la utilización de curvas fotopletismográficas como método de estimación de la presión arterial no invasiva, utilizando redes neuronales artificiales. La hipótesis planteada es que esto se puede cumplir con los parámetros establecidos por la AAMI-ISO 81060-2:2018 a fin de garantizar la calidad médica en las mediciones generadas por el método propuesto. Se exploran dos tipos de redes neuronales artificiales con enfoques diferentes. Por un lado, se analiza la utilización de las *Multilayer Perceptrons* a fin de estimar la presión arterial de un paciente, habiendo sido entrenadas con datos de múltiples individuos. Al probar no ser exitoso el método anterior, se investigó el uso de las *Convolutional Neural Networks* a medida del paciente, es decir que las redes neuronales fueron entrenadas y evaluadas para un único individuo, implicando una eventual necesidad de calibración. Se concluye que la hipótesis fue debidamente demostrada, donde el camino sugerido a continuar explorando es el de las *Convolutional Neural Networks*, pudiéndose extender a aplicaciones sin necesidad de calibración.

## ***Abstract***

*Cardiovascular diseases' high mortality rate, particularly that of hypertension, has deemed the search for methods that allow for a continuous registry of arterial blood pressure in a simple and domestic-friendly manner, relevant. This study focuses on the use of photoplethysmography signals as a method for estimation of blood pressure from non-invasive data, using artificial neural networks. The tested hypothesis is that the former can be achieved under the guidelines stated by the AAMI-ISO 81060-2:2018 standard, hence guaranteeing medical quality in the measurements generated by the proposed method. Two types of artificial neural networks, with different approaches, are studied. Firstly, the use of Multilayer Perceptrons for estimation of a patient's arterial blood pressure is analyzed, having been trained with data belonging to multiple subjects. Since this proved unsuccessful, patient-tailored Convolutional Neural Networks are explored, where neural networks were trained and evaluated for single individuals, implying a future need for calibration. It is concluded that the hypothesis proves to be true and that the suggested path to explore, as it shows prospect of success, is that of Convolutional Neural Networks, which could result in applications without the need for calibration.*

## 1. Glosario de Contracciones

**AAMI:** Asociación para el Avance de Instrumentación Médica, *Association for the Advancement of Medical Instrumentation*

**AMPA:** Automedición de la Presión Arterial

**ANN:** Red Neuronal Artificial, *Artificial Neural Network*

**AUC ROC:** Área Bajo la Curva ROC

**BGD:** *Batch Gradient Descent*

**BN:** *Batch Normalization*

**BP:** Presión Arterial, *Blood Pressure*

**CCD:** Cumplimiento con Criterio Simplificado en la estimación de DBP

**CCS:** Cumplimiento con Criterio Simplificado en la estimación de SBP

**CCT:** Cumplimiento con Criterio Simplificado en SBP y DBP

**CNN:** Red Neuronal Convolutacional, *Convolutional Neural Network*

**CS:** Criterio Simplificado

**DBP:** Presión Arterial Diastólica, *Diastolic Blood Pressure*

**ECG:** Electrocardiografía

**ELU:** *Exponential Linear Unit*

**ESH:** Sociedad Europea de la Hipertensión, *European Society of Hypertension*

**FFT:** Transformada Rápida de Fourier, *Fast Fourier Transform*

**FIR:** *Finite Impulse Response*

**FPR:** Tasa de Falsos Positivos, *False Positives Rate*

**FN:** Falsos Negativos

**FP:** Falsos Positivos

**GRU:** *Gated Recurring Units*

**HIPAA:** Ley de Responsabilidad y Portabilidad del Seguro de Salud, *Health Insurance Portability and Accountability Act*

**ISO:** Organización Internacional de Estandarización, *International Organization for Standardization*

**LED:** Diodo Emisor de Luz. *Light Emitting Diode*

**MAE:** Error Medio Absoluto, *Mean Absolute Error*

**MAPA:** Medida Ambulatoria de la Presión Arterial

**MBP:** Presión Arterial Media, *Mean Blood Pressure*

**MCC:** Coeficiente de Correlación de Matthews

**MIMIC:** *Medical Information Mart for Intensive Care*

**MLP:** *Multilayer Perceptron*

**MSE:** Error Cuadrático Medio, *Mean Squared Error*

**MSLE:** Error Logarítmico Cuadrático Medio, *Mean Squared Logarithmic Error*

**NaN:** *Not a Number*

**NICE:** Instituto Nacional de Salud y Excelencia Clínica de Gran Bretaña, *National Institute for Health and Care Excellence*

**OMS:** Organización Mundial de la Salud

**PCA:** Análisis de Componentes Principales, *Principal Component Analysis*

**PPG:** Fotopletismografía, *Photoplethysmography*

**PWD:** Duración del Pulso, *Pulse Wave Duration*

**ReLU:** *Rectified Linear Unit*

**SBP:** Presión Arterial Sistólica, *Systolic Blood Pressure*

**SGD:** *Stochastic Gradient Descent*

**TPR:** Tasa de Verdaderos Positivos, *True Positives Rate*

**UTI:** Unidad de Terapia Intensiva

**VN:** Verdaderos Negativos

**VP:** Verdaderos Positivos

## 2. Introducción

Según la Organización Mundial de la Salud (OMS), las enfermedades cardiovasculares son responsables de aproximadamente 17 millones de muertes por año, casi un tercio del total, siendo esta la principal causa de muerte a nivel mundial. Las complicaciones de la hipertensión son responsables de 9,4 millones de dichas muertes [1]. Para disminuir dicha tasa de mortalidad, es conveniente que los pacientes que sufren de hipertensión lleven un registro continuo de su presión arterial (BP). Teniendo esto en cuenta, las mediciones que se toman en los consultorios médicos resultan insuficientes y, varias veces, no representan de forma fidedigna la verdadera BP del paciente a lo largo del día. Por esta razón, son cada vez más importantes los dispositivos que apuntan a permitir la medición de presión arterial doméstica [2].

Actualmente, la mayoría de los artefactos disponibles se basan en la técnica oscilométrica dada su facilidad de uso. Sin embargo, las mediciones de éstos son altamente dependientes de diversas variables. En primer lugar, el correcto posicionamiento del manguito en el brazo del paciente es un factor determinante en la medición obtenida, pudiendo resultar en una sobreestimación o subestimación de la misma [2]. Por otro lado, si la medición se realiza en un contexto de consulta médica, se debe tener en cuenta el Síndrome de Guardapolvo Blanco. El mismo consiste en una medición de BP consistente con un diagnóstico hipertensivo a causa de que el paciente se encuentra bajo estrés, originado por la propia situación de evaluación médica [2]. Adicionalmente, existen reportes de personas mayores relacionados a la incomodidad o dolor que experimentan en el momento en el que se incrementa la presión del brazalete inflable con la finalidad de colapsar las arterias [3].

Teniendo en cuenta las desventajas previamente expuestas, recientemente surge como alternativa el uso de las curvas de fotopletismografía (PPG) obtenidas a partir de los oxímetros de pulso como método para la medición de la BP [4]. Adicionalmente, permitiría realizar mediciones continuas de la BP de forma no invasiva y de sus variaciones durante el ritmo circadiano, incluso facilitando la recolección de información en el ambiente doméstico. La implementación de esta tecnología en desarrollo podría resultar beneficiosa al asistir en el diagnóstico y seguimiento de pacientes contribuyendo, fundamentalmente, a mejorar su calidad de vida.

### 3. Objetivos

Esta investigación tiene como objetivo generar un método de estimación de la BP que resuelva las desventajas antes mencionadas. Esto implicaría un método de medición no invasiva y continua, que asegure resultados de forma rápida y cómoda para el paciente, garantizando que los mismos posean calidad médica. Para los propósitos de este proyecto, se considerará calidad médica a los resultados que posean un error de acuerdo a lo establecido como aceptable por los estándares internacionales, es decir lo recomendado por la Asociación para el Avance de Instrumentación Médica (AAMI). En particular, se hará foco en el consenso AAMI-ISO 81060-2:2018. Asimismo, se busca que se pueda obtener valores de BP del paciente durante el ritmo circadiano, a fin de analizar los cambios que se dan a lo largo del día. Para esto, se busca que finalmente se logre obtener una técnica que permita lo visualizable en la Figura 1, en donde por cada pulso de PPG se extraiga una estimación de la presión arterial sistólica (SBP) y de la diastólica (DBP), mediante un preprocesamiento y procesamiento adecuado.

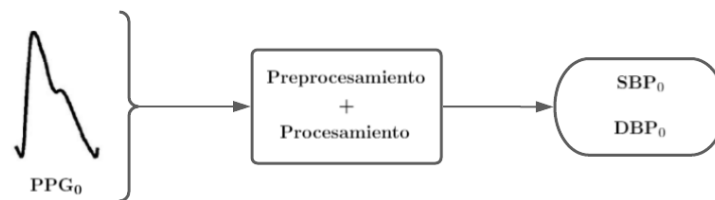


Figura 1: Diagrama de Flujo del objetivo general.

De forma adicional, se busca que todo lo desarrollado sea de acceso libre, es decir *open source*, ya que se considera relevante que sea de fácil acceso a fin de incrementar la cantidad de conocimiento disponible con respecto al tema.



## 4. Hipótesis y Alcance

La hipótesis que se busca demostrar en este proyecto de investigación es que es posible estimar la presión arterial con calidad médica a partir de las curvas de fotoplethysmografía mediante el uso de Redes Neuronales Artificiales.

A fin de poner a prueba la hipótesis se utilizará una única base de datos, cuyo tratamiento será *offline* mediante el uso del lenguaje de programación Python y las librerías de aplicación en el contexto de Inteligencia Artificial y procesamiento de señales. Se hará un análisis de preprocesamiento a modo de generar un correcto condicionamiento de las señales que serán las entradas de las Redes Neuronales Artificiales. De estas, se hará foco en las *Multilayer Perceptrons* y en las *Convolutional Neural Networks*. El enfoque a tomar es a medida del paciente, implicando una calibración.

## 5. Marco Teórico

### 5.1. Presión Arterial

La presión arterial se define como la tensión que ejerce la sangre al circular por las arterias [5]. Su medición se descompone en dos valores: presión arterial sistólica (SBP), medida durante la contracción ventricular (presión arterial máxima); y presión arterial diastólica (DBP), medida durante la relajación ventricular (presión arterial mínima). Estos valores constituyen un índice básico que refleja la hemodinamia del paciente y tienen una gran importancia a la hora del diagnóstico [6]. Es por esto que es de importancia la precisión en la medición de la presión arterial. Para esto, existen mediciones continuas o mediciones discretas. La curva estándar de la presión arterial se puede observar en la Figura 2, donde se encuentran señalados los valores de presión arterial sistólica y diastólica, además de la media.

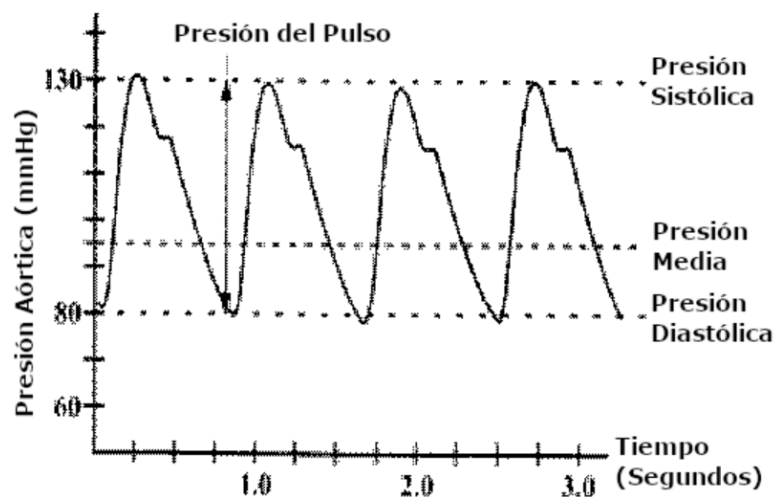


Figura 2: Curva estándar de presión sanguínea. Fuente: [7].

La medición continua de la presión arterial se suele obtener mediante un catéter intraarterial, como muestra el esquema de la Figura 3, y es usado como *gold standard*. Resulta de especial utilidad para aquellos pacientes que puedan experimentar cambios bruscos en su hemodinamia [7]. Entre las ventajas de este tipo de medición, además de su precisión, se puede mencionar su capacidad para medir presiones bajas y de presentar mediciones en tiempo real o latido-a-latido. Las desventajas que más se destacan de este método son el alto nivel de invasividad y los riesgos que este puede suponer, como lo son las infecciones o los distintos tipos de trombosis [8].

El método más preciso para la medición discreta de la presión arterial es mediante

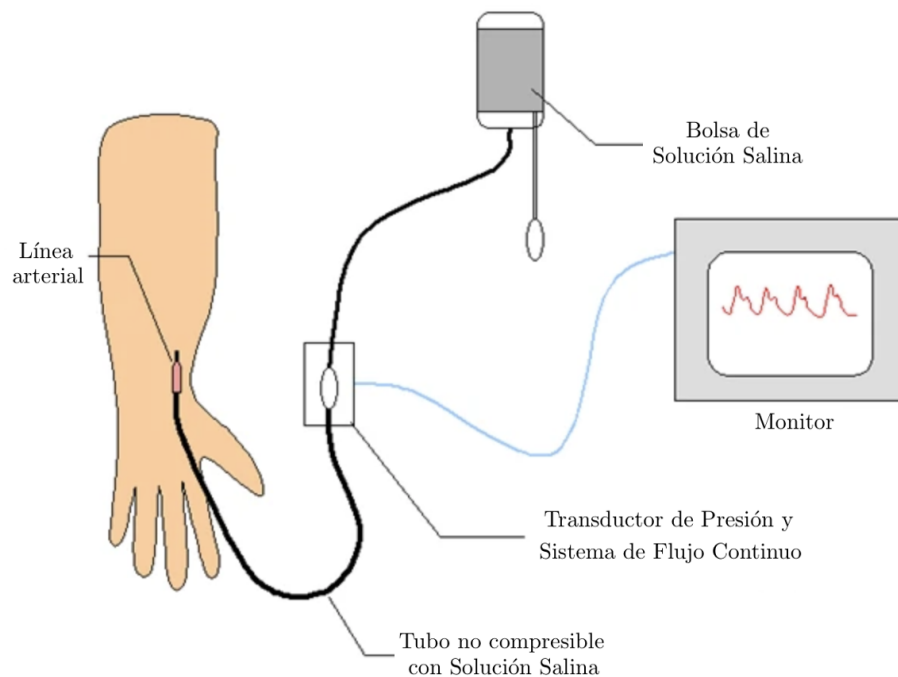


Figura 3: Esquema de Método de Medición de BP Invasiva. Fuente: [10].

el uso del esfigmomanómetro de mercurio [9]. Este método necesita de un manómetro, un brazalete inflable o manguito y un estetoscopio para auscultar de forma clara los sonidos de Korotkov (ver Figura 4). Estos son los sonidos que el personal de salud capacitado escucha durante la medición de la presión arterial. El procedimiento para obtener la BP mediante esta técnica se detalla a continuación [11]:

- El paciente debe permanecer en reposo por al menos 5 minutos previos a la toma de la medición.
- El paciente debe estar sentado en una silla, con los pies en el suelo y las piernas descruzadas.
- El brazo debe quedar apoyado de forma tal que el antebrazo esté a nivel del corazón.
- Se coloca el brazalete en el brazo izquierdo del paciente de forma tal que el borde inferior del manguito quede a 2.5 cm por encima del doblez del codo.
- Se coloca el estetoscopio sobre la arteria braquial y se infla rápidamente el manguito hasta superar la presión de oclusión en la cual no se escuchan ruidos.
- La válvula del manguito se abre ligeramente para que la presión descienda lenta-

mente.

- En cuanto se escucha la sangre pulsando por el flujo turbulento (primer ruido de Korotkov) se registra la presión sistólica.
- El flujo luego comienza a ser laminar, por lo que los ruidos van disminuyendo en intensidad. Se registra el punto en el cual los sonidos dejan de ser audibles, siendo esta la presión diastólica.

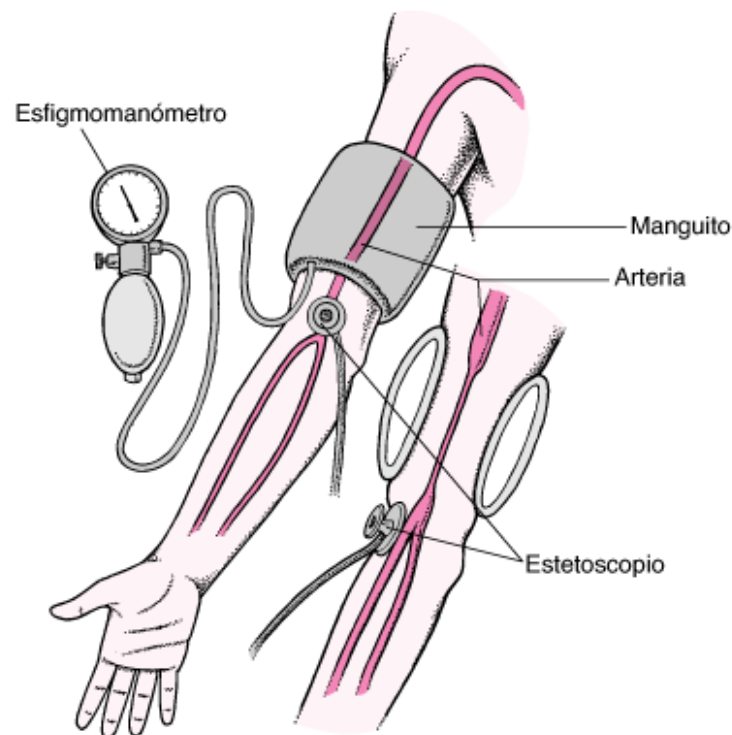


Figura 4: Esquema de medición de la presión arterial por Método Auscultatorio. Fuente: [12].

Un punto a considerar es la toxicidad del mercurio, razón por la cual este método no es el recomendado e incluso está siendo gradualmente prohibido [13]. Es por esto que, actualmente, la mayoría de los dispositivos utilizan la técnica oscilométrica como alternativa para la medición de presión arterial de forma no invasiva [14]. De forma similar al método anteriormente mencionado, este consta de un manguito inflable cuyo objetivo es colapsar la arteria para luego disminuir la presión de forma gradual. El flujo turbulento de la sangre produce oscilaciones, que son detectadas por un sensor. Al detectarse, se determina la presión sistólica. Al finalizar las oscilaciones, se determina la presión diastólica.

El método oscilométrico posee la ventaja de ser fácil de usar. Sin embargo, presenta

altos niveles de variabilidad y, por lo tanto, es difícil obtener mediciones confiables, particularmente cuando los pacientes son los encargados de manipular el dispositivo. Esto se debe a que, entre los factores que afectan a la medición, se encuentran la posición del manguito y la del paciente, dos variables difíciles de controlar en las mediciones domésticas.

### 5.1.1. Valoración de la BP

En función del valor de BP habitual o fisiológico del paciente se lo puede clasificar como hipotenso, normotenso o hipertenso. Para que un individuo sea considerado normotenso, sus valores de presión sistólica deben ser menores a 120 mmHg y los de diastólica deben ser menores a 80 mmHg. Por otro lado, para que sea considerado hipotenso, la presión sistólica debe ser menor a 90 mmHg y la diastólica menor a 60 mmHg. Los síntomas que esta condición puede acarrear son debilidad, fatiga, sudoración y náuseas, entre otros. Por último, la hipertensión es diagnosticada cuando un paciente posee una presión sistólica mayor a 130 mmHg y una diastólica mayor a 80 mmHg [15]. La Tabla 1 esquematiza los distintos niveles de hipertensión. Dependiendo de la BP del paciente, es asignado en una de las categorías correspondientes.

Categoría de la Presión Arterial	SBP (mmHg)		DBP (mmHg)
Normal	<120	y	<80
Elevada	120-129	y	<80
Hipertensión Nivel 1	130-139	o	80-89
Hipertensión Nivel 2	>140	o	>90
Crisis de Hipertensión	>180	y/o	>120

Tabla 1: Categorías de presión arterial. Fuente: [16].

La hipertensión arterial daña las arterias, lo que eventualmente puede resultar en un infarto cerebrovascular o de miocardio. Además, puede derivar en una falla cardíaca debido a que el corazón realiza un mayor trabajo. Adicionalmente, se encuentran otras complicaciones como enfermedad o falla renal, pérdida de visión, disfunción sexual, angina de pecho y arterioesclerosis [17]. En la Tabla 2 se observan los riesgos asociados a diferentes rangos de BP con distintas comorbilidades metabólicas. En la misma, se hace referencia al riesgo de morbilidad y mortalidad cardiovascular en los próximos 10 años de un individuo, análogo al sistema de puntaje de riesgo de Framingham [18]. A partir de la Tabla 2 se puede ver como, a mayor cantidad de factores de riesgo metabólicos o enfermedades y mayor la BP, mayor es el riesgo cardiovascular.

Es por estas complicaciones que, como se ha mencionado previamente, resulta importante realizar un seguimiento y poder obtener mediciones confiables de la BP del paciente. Sin embargo, en un ámbito de consultorio, donde únicamente se toma una cantidad limitada de mediciones repetidas a intervalos regulares, existe mucho lugar a error [17]. Uno anteriormente mencionado es que se puede dar el Síndrome de Guardapolvo Blanco,

Otros factores de riesgo, daño orgánico o enfermedades	BP (mmHg)				
	Normal: SBP 120-129 o DBP 80-84	Normal Alto: SBP 130-139 o DBP 85-89	Hipertensión grado 1: SBP 140-159 o DBP 90-99	Hipertensión grado 2: SBP 160-179 o DBP 100-109	Hipertensión grado 3: SBP $\geq 180$ o DBP $\geq 110$
Sin factores de riesgo	Riesgo normal	Riesgo normal	Riesgo agregado bajo	Riesgo agregado moderado	Riesgo agregado alto
1-2 factores de riesgo	Riesgo agregado bajo	Riesgo agregado bajo	Riesgo agregado moderado	Riesgo agregado moderado	Riesgo agregado muy alto
3 o más factores de riesgo	Riesgo agregado moderado	Riesgo agregado alto	Riesgo agregado alto	Riesgo agregado alto	Riesgo agregado muy alto
Enfermedad cardiovascular o renal establecida	Riesgo agregado muy alto	Riesgo agregado muy alto	Riesgo agregado muy alto	Riesgo agregado muy alto	Riesgo agregado muy alto

Tabla 2: Estratificación del riesgo vascular en 4 categorías: riesgo bajo, moderado, alto y muy alto de presencia de evento cardiovascular, el cual puede o no ser fatal. El término “agregado” refiere a que es mayor al promedio. Fuente: [19].

resultando en una BP medida más elevada. La Tabla 3 resume algunos de los errores más comunes en este ámbito, no relacionados al método sino al ambiente. Es por todas estas razones por lo que se utilizan las técnicas de medición llamadas automedición de presión arterial (AMPA) y monitoreo ambulatorio de presión arterial (MAPA). Estas técnicas ambulatorias son las aceptadas para el diagnóstico de la hipertensión según el Instituto Nacional de Salud y Excelencia Clínica (NICE) de Gran Bretaña [20].

Factor	Fuente de error	Efecto en la medición de BP
Procedimiento	Actitud del paciente	Significativa sobreestimación
	Postura del paciente	
	Colocación del brazo	
Individuo	Infantes (tamaño del manguito)	Estos grupos de pacientes pueden tener distintas particularidades que afecten en la precisión del dispositivo
	Embarazo, Pre-clampsia	
	Arritmia	
	Obesidad (tamaño del manguito)	
	Diabetes	
Dispositivo y manguito	Mercurio	Errores en la calibración del dispositivo. Si se utiliza un manguito mas chico al necesario se produce una sobreestimación y con mas grandes, se subestima.
	Oscilométrico	
	Sistema de inflado y desinflado	
Observador	Error de medición	Estimación errada en la medición de BP.
	Capacidades físicas (visión, audición)	
	Sesgos	

Tabla 3: Fuentes de error en la medición de BP en el consultorio. Fuente: [14].

### 5.1.2. Métodos Ambulatorios: AMPA y MAPA

La AMPA consiste en la medición de BP obtenida por el mismo paciente fuera del ámbito de la consulta. Esto posee numerosas ventajas al suceder en un contexto más próximo al de la vida cotidiana. De esta manera, se disminuyen los efectos del Síndrome del Guardapolvo Blanco, al complementar los valores de las mediciones obtenidas en el consultorio [21]. Si bien cuenta con ciertos requerimientos para que los valores puedan ser considerados como válidos, como el uso de dispositivos adecuados y su correcta colocación, es una técnica sencilla de aplicar.

Otro método de utilidad para disminuir el efecto del Síndrome de Guardapolvo Blanco es el del MAPA. Es un método no invasivo, que consta de un brazalete con un sensor y un dispositivo de registro que es colocado por un profesional. Este brazalete se infla de forma periódica durante un tiempo programado, generalmente cada 15 o 20 minutos durante el día y 20 o 30 minutos durante la noche [22]. Con este método se obtiene información adicional del cambio de la BP durante el ritmo circadiano.

Si bien el método AMPA presenta las ventajas de permitir obtener mediciones en el domicilio del paciente, en un contexto semejante al de su vida cotidiana, se corre el riesgo de adquirir resultados erróneos si falla la colocación del dispositivo. Asimismo, asume la necesidad de un dispositivo fuera del consultorio para la medición, ya sea en el domicilio del paciente o en un Centro de Atención Primaria. Por otro lado, el método MAPA es manejado por un profesional, por lo que el riesgo de medición incorrecta por mal posicionamiento disminuye. Asimismo, el profesional suele encargarse de brindarle instrucciones al paciente previo al inicio del estudio en cuanto al correcto posicionamiento

que debe llevar durante las mediciones. Adicionalmente, permite obtener valores de BP nocturnos que proporcionan información crítica del individuo [23]. Sin embargo, presenta la contraparte de que no contempla la obtención de medidas de la BP del paciente en su vida cotidiana en largos períodos de tiempo.

Ambos presentan una limitación en cuanto al tiempo del estudio. Cabe destacar que el método MAPA tiene una duración de un día y, por lo tanto, valores anormales que no se observen en este periodo no serán documentados. El método AMPA se puede realizar durante toda la vida del paciente pero se encuentran requerimientos tanto de obtener el dispositivo como de conocer su correcta colocación. En los casos donde los pacientes no posean un esfigmomanómetro en sus domicilios y recurran a un establecimiento, no se estaría evitando el efecto del Síndrome de Guardapolvo Blanco.

## 5.2. Fotopletismografía

La fotopletismografía es una técnica óptica que permite obtener la curva de pletismograma que refleja los cambios del volumen de sangre en los vasos sanguíneos. Esta curva se puede adquirir a partir del oxímetro de pulso y, actualmente, se utiliza principalmente para medir la saturación de oxígeno en sangre [24].

La curva obtenida a partir de PPG está compuesta por una componente pulsátil (denominada “AC”), que contiene información sobre los cambios de volúmenes generados por el ciclo cardíaco pulso a pulso, y por una componente continua que varía lentamente (denominada “DC”), modificada por la absorción de los huesos, de los tejidos y de otros factores que se mantienen constantes para un mismo paciente, durante una misma medición [25]. Cabe aclarar que las denominaciones para las componentes provienen del ámbito biomédico y, por convención, no se utiliza la misma definición de “AC” y “DC” que en el tratamiento de señales eléctricas. Además, como se puede observar en la Figura 5, la curva es periódica y cada ciclo suele tener una fase anacrótica hasta el pico primario o pico sistólico. Luego, comienza la fase catacrótica que se corresponde con la diástole del corazón y con la onda reflejada en la periferia. Debido a la correspondencia, las fases se conocen como sistólica y diastólica, respectivamente. En la fase catacrótica, se encuentra un valle conocido como *dicrotic notch*, marcado en la Figura 6, que se corresponde con el cierre de la válvula aórtica.

Se debe hacer mención al hecho de que el valor “DC” de la curva es, por convención, el valor absoluto mínimo de la señal, y no el RMS, como se observa en la Figura 5.



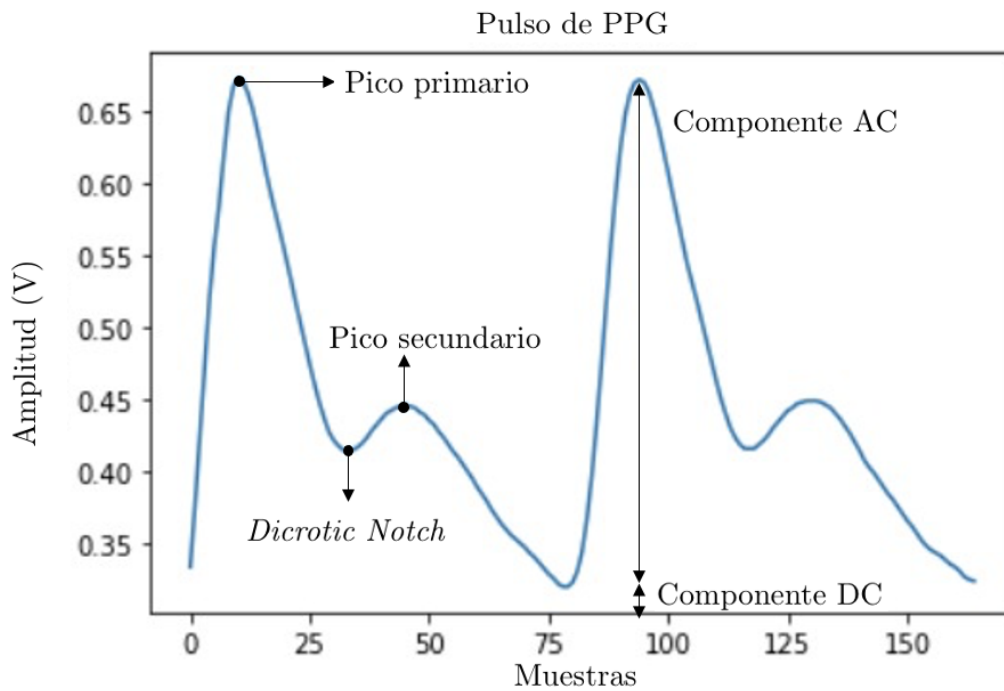


Figura 5: Curva de PPG donde se puede observar ambos componentes (DC y AC) y las fases anacrótica y catacrótica. Fuente: [26].

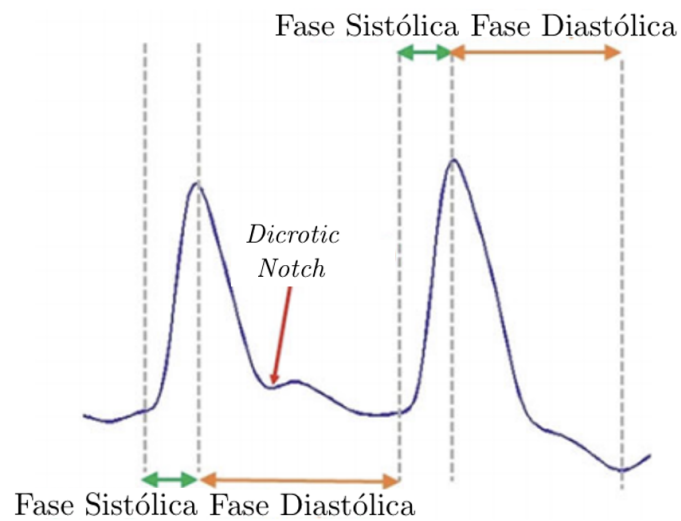


Figura 6: Curva de PPG donde se observa la fase sistólica y diastólica con el *dicrotic notch*. Fuente: [24].

### 5.2.1. Oxímetro de Pulso

Como se ha mencionado previamente, el oxímetro de pulso es utilizado para obtener la señal de PPG. Para esto, un diodo emisor de luz (LED) ilumina el tejido con dos longitudes de onda distintas y un fotodiodo capta la intensidad de la luz no absorbida del otro lado. Luego, un transductor convierte la intensidad de luz reflejada o transmitida en una señal de voltaje (PPG) [27]. La Figura 7 muestra un esquema básico del funcionamiento de un oxímetro.

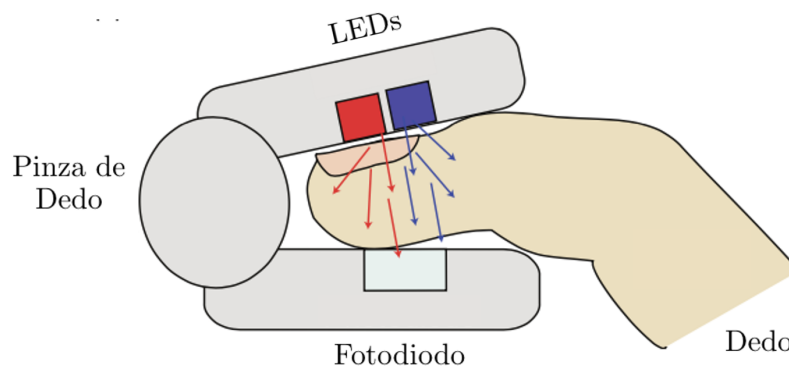


Figura 7: Oxímetro de pulso de transmisión con sus componentes. Fuente: [27].

Cuando el volumen de sangre en las arterias y arteriolas aumenta durante la sístole, la proporción de luz incidente absorbida también aumenta debido al carácter elástico de las paredes arteriales que responden cambiando el diámetro debido a variaciones en la presión transmural [27]. Similarmente, pero de forma contraria, la absorción de luz disminuye con la reducción del volumen sanguíneo en la diástole.

El inventor de este dispositivo, Takuo Aoyagi, partió de los siguientes principios para su desarrollo [27]:

- La sangre venosa no es pulsátil y, por lo tanto, no afecta en la medición.
- Si la densidad óptica de la parte pulsátil es medida mediante dos longitudes de onda, se puede obtener el valor de saturación de oxígeno ( $SaO_2$ ).

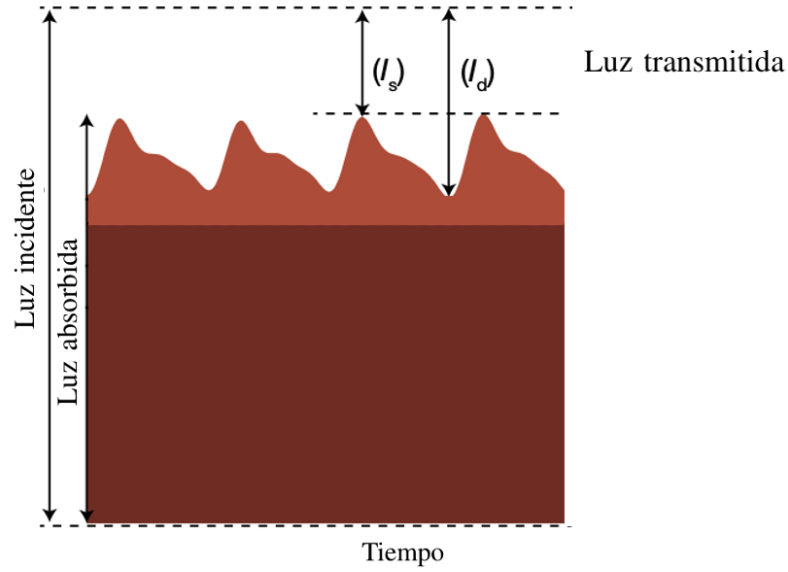


Figura 8: Esquema de la absorción de luz en la oximetría de pulso. Las intensidades  $I_s$  e  $I_d$  corresponden a la fase sistólica y diastólica, respectivamente [27].

La absorbancia de la luz por el componente pulsátil se encuentra determinada por la ley de *Lambert-Beer*, la cual establece que la absorción de la luz en un medio es proporcional a la concentración de absorbentes de luz presentes ( $C$ ), las propiedades ópticas de los mismos y la longitud del camino óptico realizado por el haz de luz ( $L$ ) [27].

$$A_\lambda = \alpha_\lambda CL \quad (1)$$

donde  $A_\lambda$  es la absorbancia de la luz y  $\alpha_\lambda$  es la absorbancia molar de los absorbentes, ambas en la longitud de onda  $\lambda$ . La ecuación puede reescribirse para incluir la transmitancia durante la sístole ( $I_s$ ) y la transmitancia durante la diástole ( $I_d$ ) de la siguiente forma:

$$A_\lambda = \alpha_\lambda CL = \ln\left(\frac{I_s}{I_d}\right) \quad (2)$$

Ya que  $AC_\lambda = I_s - I_d$  (ver Figura 8) donde  $AC_\lambda$  es la componente AC de la señal transmitida a la longitud de onda  $\lambda$ .

$$\ln\left(\frac{I_s}{I_d}\right) = \ln\left(\frac{I_d + AC_\lambda}{I_d}\right) = \ln\left(1 + \frac{AC_\lambda}{I_d}\right) \approx \frac{AC_\lambda}{I_d} \quad (3)$$

Teniendo en cuenta que la componente AC es mucho menor que la DC, se puede tomar que  $DC_\lambda$  es igual a  $I_d$  [27]. Por lo tanto,

$$A_\lambda = \frac{AC_\lambda}{DC_\lambda} \quad (4)$$

tomando el *ratio* de *ratios* de la absorbancia de la luz en dos longitudes de onda distintas

$$R = \frac{(\frac{AC}{DC})_{\lambda_1}}{(\frac{AC}{DC})_{\lambda_2}} = \frac{A_{\lambda_1}}{A_{\lambda_2}} \quad (5)$$

Luego, un microprocesador en los oxímetros de pulso usa este *ratio* para determinar la  $SaO_2$  a partir de una curva de calibración generada de forma empírica. Este proceso de calibración incluye desaturar voluntarios sanos haciéndolos respirar mezclas de gas hipóxicas con una saturación de oxígeno entre 70 y 100 %. La Ecuación 6 relaciona el *ratio*  $R$  y la  $SaO_2$ .

$$SaO_2 = \frac{k_1 - k_2 R}{k_3 - k_4 R} \quad (6)$$

De esta manera, se obtienen los valores de las constantes  $k_i$  con  $i = 1, 2, 3, 4$  mediante un análisis de mejor ajuste [27]. A continuación, se presenta una curva de la relación empírica entre la saturación de oxígeno y el *ratio* mencionado.

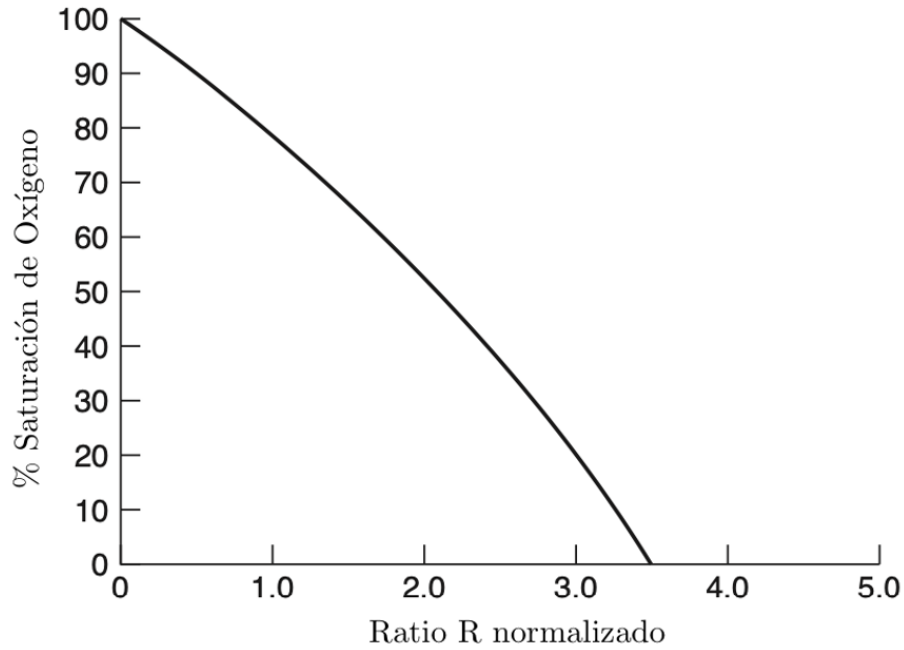


Figura 9: Curva de calibración del oxímetro Fuente: [28].

Debido a que la calibración se realiza hasta una saturación de 70 %, la precisión de la medición deteriora luego de este punto. El porcentaje de saturación de oxígeno medido por un oxímetro de pulso en pacientes sanos se encuentra entre el 95 y 100 %. Bajo condiciones normales de operación, la precisión del oxímetro de pulso es aceptable para la mayoría de las aplicaciones clínicas siendo de entre el 2 y el 3 % [29].

En cuanto a la elección de las longitudes de onda a utilizar se deben considerar los principales absorbentes de la sangre. La hemoglobina es la proteína responsable del transporte de oxígeno en el cuerpo. Se encuentra en dos estados: en su forma oxigenada, conocida como oxihemoglobina ( $HbO_2$ ), y en su forma reducida, desoxihemoglobina ( $Hb$ ). Es por esto que la  $SaO_2$  se define según la siguiente Ecuación:

$$SaO_2 = \frac{[HbO_2]}{[HbO_2] + [Hb]} \quad (7)$$

Donde  $[ ]$  simboliza la concentración. Para la selección de las longitudes de las ondas a utilizar, se tiene en cuenta la sensibilidad de la hemoglobina en sus dos estados y se busca maximizar la diferencia entre la absorción de la  $HbO_2$  y  $Hb$  [27].

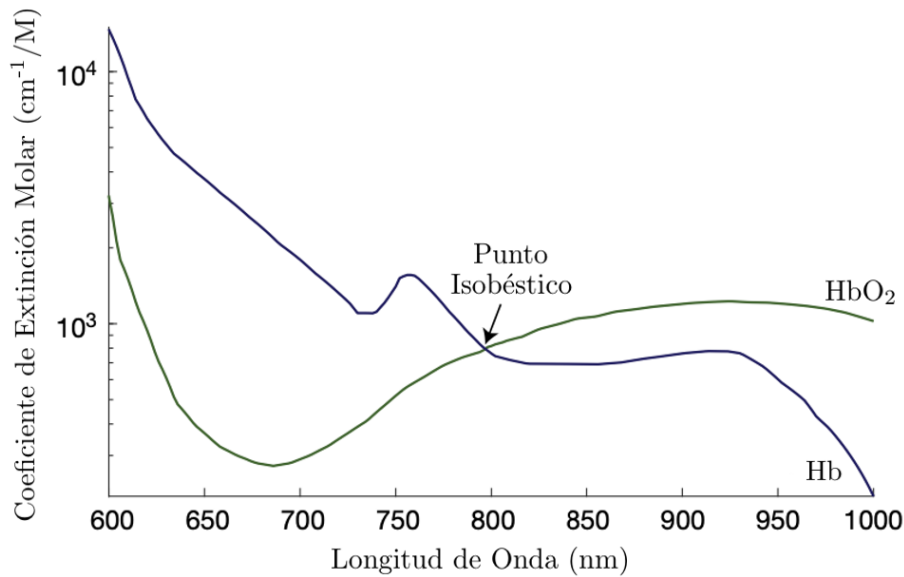


Figura 10: Espectro de absorción de  $Hb$  y  $HbO_2$  en el rango de 600 nm a 1000 nm. La hemoglobina tiene una absorción característica al estado en el que se encuentra y es distinta, excepto en el punto isobéptico. Fuente: [27].

Como se puede ver en la Figura 10, la absorción de  $Hb$  y  $HbO_2$  es dependiente de la longitud de onda y es diferente para ambos estados de la hemoglobina, excepto en el punto isobéptico. En la región anterior a este punto, la absorción se debe principalmente a  $Hb$  y, luego de este punto, a  $HbO_2$ . Por otro lado, por debajo de los 600 nm la melanina

Fuentes de Error	Efectos en $SaO_2$
Hipotensión	Posible pérdida de señal
Vasoconstricción	Posible pérdida de señal y reducción de $SaO_2$
Hipotermia	Posible pérdida de señal y reducción de $SaO_2$
Carboxihemoglobinemia	Valores de $SaO_2$ falsamente altos
Metahemoglobinemia	Valores de $SaO_2$ falsamente bajos, cercanos a 85 %
Pulsaciones venosas	Valores de $SaO_2$ falsamente bajos
Uso del manguito para la medición de presión arterial en simultaneo	Pérdida de la señal

Tabla 4: Fuentes de error y sus efectos en  $SaO_2$ . Fuente: [27].

absorbe la mayoría de la luz y, por arriba de los 950 nm, el agua en los tejidos. Por todo esto, y para maximizar la diferencia de absorción del  $Hb$  y  $HbO_2$ , las dos longitudes de onda típicamente elegidas son 660 nm y 940 nm [27].

Es importante destacar que la curva en cuestión presenta altos niveles de variabilidad y, en base a la investigación actual, no se conocen con exactitud todos los factores que generan estas variaciones. De todas formas, se sabe que el volumen sanguíneo, el movimiento de las paredes de los vasos y la orientación de los glóbulos rojos influye en la cantidad de luz que es detectada en el fotodetector. Además, pueden influir en la medición los cambios en la intensidad de la señal, los movimientos, interferencias de color y la bilirrubina. En la Tabla 4 se pueden encontrar algunas situaciones que interfieren en la medición de  $SaO_2$ .

De la Tabla 4 se puede destacar el uso de métodos oscilométricos, como el manguito, que colapsa las arterias y, por lo tanto, se pierde la señal. También, es importante destacar la complejidad de la interacción de la luz con el tejido biológico en donde se producen los procesos de absorción, dispersión, reflexión y transmisión, entre otros.

Mas allá de estas limitaciones y debido a su carácter portátil, de bajo costo y facilidad de uso es que se comenzó a analizar su uso con otras finalidades [24].

### 5.3. Presión Arterial y Fotopletismografía

La curva de PPG comenzó a cobrar mayor importancia en los últimos años y, por lo tanto, la información sobre cómo se relaciona esta con otras variables fisiológicas es escasa. Sin embargo, un estudio conducido por Martínez *et al.* realizado en el 2018 demostró

que existe una correlación morfológica entre las curvas de BP y PPG para pacientes normotensos, prehipertensos e hipertensos, como se puede observar en la Figura 11 [30]. Además, la investigación demostró que esta correlación se da tanto en el dominio temporal como frecuencial.

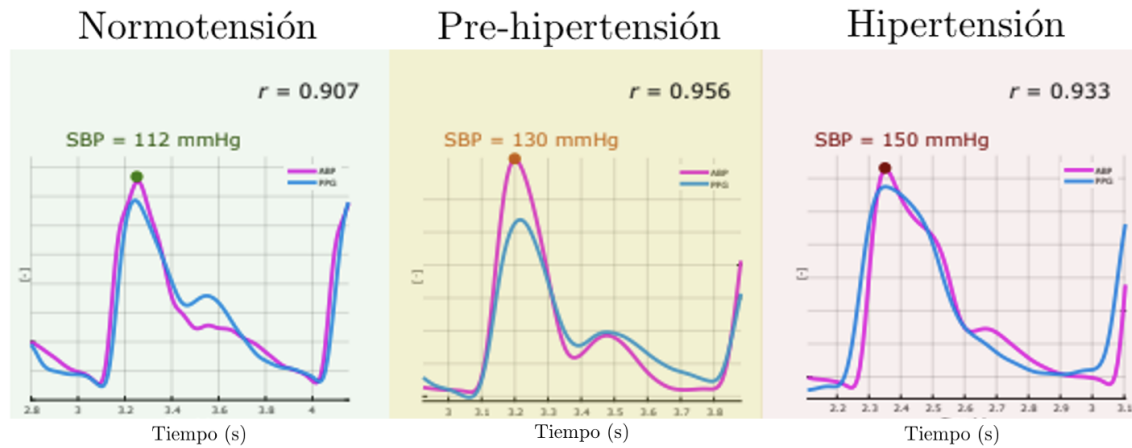


Figura 11: Comparación entre curvas de presión arterial (roja) y PPG (azul) en fase, donde  $r$  representa la correlación morfológica de ambas curvas. Fuente: [30].

En la Figura 12 se pueden observar las curvas de BP y PPG superpuestas y en fase. De esta forma, se puede apreciar visualmente la alta similitud que existe entre las mismas. De todas formas, se debe tener en cuenta que a medida que aumenta la BP, el desfase entre las señales comienza a modificarse [30].

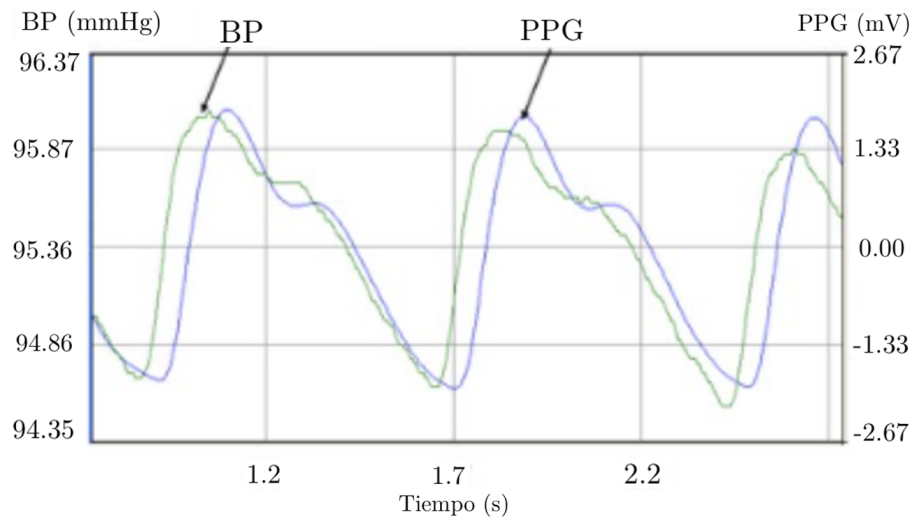


Figura 12: Comparación entre curvas de presión arterial y PPG en fase. Fuente: [31].

Ya demostrada la correlación morfológica entre las curvas de BP y PPG, se deben

tener en cuenta las ventajas que presenta la medición de BP a través de PPG. Este método permite la medición continua y no invasiva eliminando errores sistemáticos y obteniendo una medición de forma rápida. Esto último es de gran importancia en situaciones donde se necesita una respuesta inmediata, como lo es en los servicios de emergencia. Esta técnica podría ser de utilidad durante el *Triage*, el cual representa un sistema de selección y clasificación de pacientes basado en sus necesidades terapéuticas y los recursos disponibles, agilizando los procesos involucrados y ayudando a reducir la demanda sobre el sistema sanitario.

Por último, resulta importante mencionar que se demostró la existencia de una correlación entre la curva de PPG radial, es decir la medida en las manos, con la de BP en el dominio temporal y frecuencial [30]. Este hallazgo sustenta la hipótesis planteada en el presente trabajo de investigación.

### 5.3.1. Presión Central y Periférica

Es pertinente destacar que la medición de presión arterial mediante el uso de PPG es periférica. Por lo tanto, se deben mencionar las diferencias entre la medición de presión arterial central y periférica.

Hoy en día, la presión arterial medida con manguito y esfigmomanómetro en la arteria braquial representa un predictor ampliamente aceptado del riesgo cardiovascular. Sin embargo, la presión sistólica varía a lo largo del árbol arterial como se puede ver en la Figura 13. De hecho, la presión aórtica es menor que la braquial, aunque hay cierta variabilidad entre pacientes, teniendo una diferencia de aproximadamente 40 mmHg entre ambas [32].

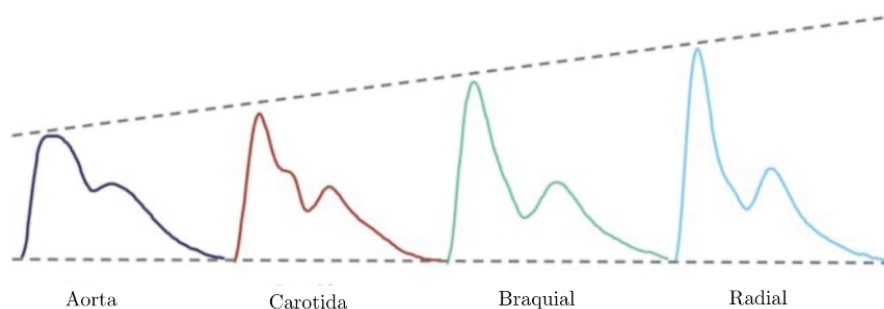


Figura 13: Comparación entre curvas de presión arterial en las distintas partes del árbol arterial. Fuente: [32].

No obstante, la presión diastólica y la BP media (MBP) se mantienen relativamente constantes [32]. Este fenómeno de amplificación de la BP surge por el componente elástico arterial y el aumento en la rigidez a medida que se avanza por el árbol arterial, observándose-



se un pico que se angosta, haciéndose más prominente. Uno de los paradigmas utilizados para explicar este cambio en la morfología asume que la señal se compone de una onda transmitida que es reflejada en los sitios de diferentes impedancias, como las bifurcaciones. A medida que se avanza se generan más ondas reflejadas, las cuales se suman y aumentan la presión sistólica [32].

Los métodos oscilométricos generalmente utilizan la arteria braquial o radial aunque el oxímetro, en su colocación habitual, puede tomar mediciones en las arterias de las falanges. Como se expuso previamente, hay una diferencia entre las presiones en los distintos puntos del árbol arterial y, por lo tanto, este dispositivo podría generar mediciones de presión arterial más elevadas. A los fines prácticos de este proyecto de investigación, es invariante ya que se espera que mediante el uso de redes neuronales artificiales se realice una estimación adecuada.

## **5.4. Redes Neuronales Artificiales**

Las redes neuronales artificiales (ANNs) son modelos computacionales inspirados en las neuronas y las redes neuronales biológicas. Para entender más acerca de ellas es relevante comprender su comportamiento.

### **5.4.1. Funcionamiento Biológico Neuronal**

Las neuronas son las unidades fundamentales del sistema nervioso, siendo éstas las células responsables de recibir estímulos sensitivos del mundo externo, de enviar comandos motores a los músculos y de transformar y relevar las señales eléctricas en cada paso intermedio [33]. Las partes principales son el soma, prolongaciones del mismo llamadas dendritas y una prolongación larga llamada axón (ver Figura 14). Las dendritas reciben impulsos eléctricos o estímulos de los axones terminales de neuronas adyacentes y los transmiten al soma. Estas entradas se procesan en el núcleo, donde se integra la información y se decide si desencadenar un estímulo hacia otra neurona como respuesta. De ser así, se propaga el potencial de acción por el axón hacia las terminaciones del mismo. Esto provoca la liberación de un neurotransmisor hacia las dendritas de otra neurona, produciéndose una sinapsis. De esta manera, la salida de esta neurona es la entrada de otras miles. El espacio entre las terminaciones del axón de la primera neurona y el soma de la segunda se conoce como hendidura sináptica [33].

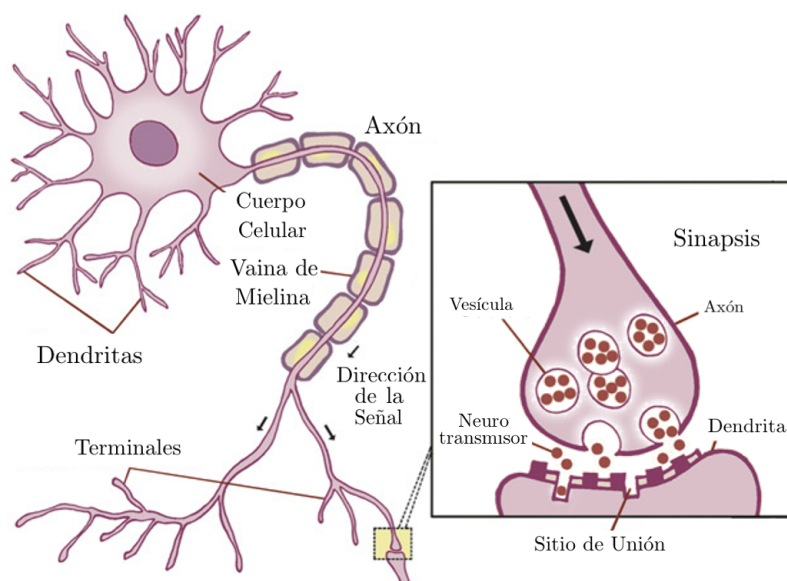


Figura 14: Esquema de una neurona y sinapsis. Fuente: [34]

A continuación, en la Figura 15, se presenta una de las primeras imágenes de neuronas dibujadas por Santiago Ramón y Cajal, considerado el padre de la neurociencia moderna, en 1899.

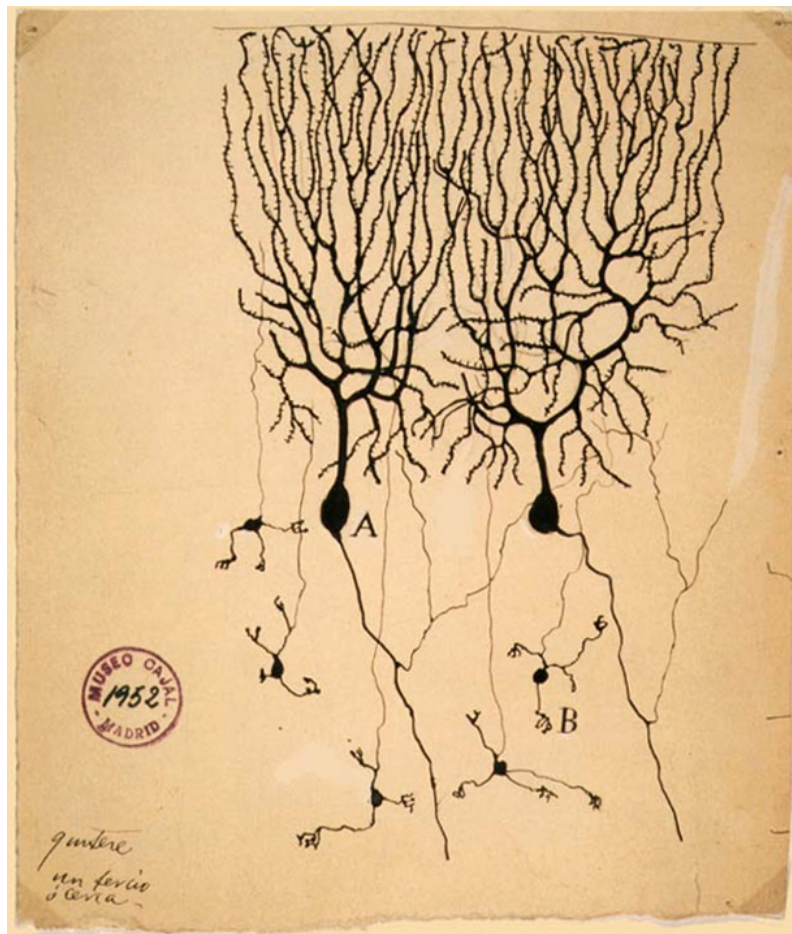


Figura 15: Dibujo de Santiago Ramón y Cajal de las neuronas de un cerebelo de paloma.  
Fuente: [35].

Se puede observar en la Figura 15 que las neuronas destacadas con la letra “A” son de mayor tamaño y se conectan con otras neuronas. Debido a que se encuentran con una gran cantidad de estímulos, las neuronas deciden cómo actuar frente a ellos dado que no pueden reaccionar a todos al mismo tiempo. Entonces, funcionan integrando todos los estímulos (que pueden ser excitatorios e inhibitorios) recibidos para así determinar si se produce, o no, una respuesta.

#### 5.4.2. Aspectos Generales de las ANNs

De forma similar al dibujo de Ramón y Cajal de las redes neuronales (Figura 15), las neuronas de las ANNs poseen dendritas, soma y axón ya que se encuentran inspiradas en estas. En la Figura 16 se observa el paralelismo entre el modelo biológico y el modelo

matemático donde la salida del axón está mediada por una función de activación.

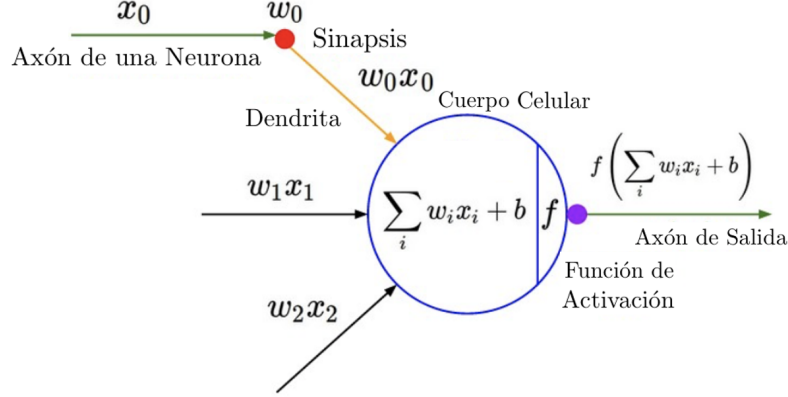


Figura 16: Esquema del modelo matemático de las redes neuronales con referencia al modelo biológico. Fuente: [35].

Las redes neuronales artificiales consisten de distintas capas, donde la primera recibe las entradas y se la conoce como *input layer*. La información llega a estas neuronas y luego se transmite a las neuronas de la siguiente capa. La capa final, que provee la salida, es conocida como *output layer* [36]. Todas las capas ubicadas entre estas dos son denominadas *hidden layers*. A este movimiento hacia adelante de la información se lo llama *forward propagation*, donde esta se mueve desde la *input layer* hacia la *output layer*. Este proceso consiste en el cálculo de ciertos valores ( $z$ ) según la Ecuación 8, donde si se tiene un vector de observaciones de entrada con una dimensionalidad  $d$ , es decir,  $x \in \mathbb{R}^d$ :

$$z = W_1 x + b \quad (8)$$

donde  $W_1 \in \mathbb{R}^{h \times d}$  representa los pesos de las neuronas de la *hidden layer* y  $b \in \mathbb{R}^h$  es el *bias*, un parámetro intrínseco a la red que ajusta la salida [36]. Luego, una función de activación le asigna un valor que será la entrada de la siguiente capa. En el caso en que la función sea una regresión lineal, la salida es  $z$ . Existen otras funciones de activación, tales como la función sigmoidea, que le asigna un valor en un rango de 0 a 1. En la Figura 17 se resume lo comentado en un modelo simplificado con 1 neurona, donde se le debe aplicar una función de activación a la salida de la neurona en blanco. El *bias* se encuentra representado con la letra **b**.

Uno de los problemas que puede llegar a presentarse debido a esta forma de entrenamiento es el *internal covariate shift* [36]. El mismo consiste en que, debido al cambio constante de pesos durante el procesamiento, puede generarse una gran diferencia entre los datos entrantes de las primeras capas y los datos entrantes de las últimas capas. El *Batch*

*Normalization* (BN) hace referencia a un método que busca contrarrestar este problema agregando capas de normalización para disminuir la variabilidad de los datos entrantes en las diferentes capas [36]. Las capas de normalización se encargan de reparametrizar los datos de forma tal que posean una media igual a 0 y un desvío estándar de 1. Esta estandarización evita cambios drásticos de los datos entrantes de cada capa y contrarresta el problema mencionado.

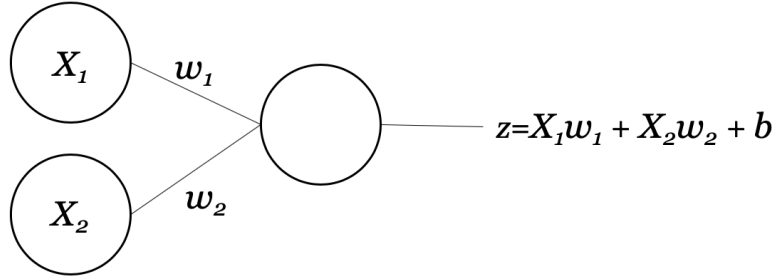


Figura 17: Esquema simplificado de *forward propagation* en un perceptrón lineal.

Por otro lado, *backpropagation* refiere al método en donde se calcula el gradiente de la función de costo (concepto que se desarrolla en la Sección 5.4.3) con respecto a los parámetros de la red neuronal para actualizar el vector de pesos [36]. Este proceso se hace en cada iteración dentro del entrenamiento de una ANN y genera que algunas entradas sean más relevantes sobre otras. Esto se realiza de forma inversa, desde la *output layer* a la *input layer*, utilizando la regla de la cadena para obtener el gradiente. La actualización de pesos para cada iteración del entrenamiento (*epoch*) se calcula según:

$$\vec{W}^{t+1} = \vec{W}^t - \eta \nabla J(W) \quad (9)$$

Donde  $\vec{W}^t$  es el vector de pesos de la  $t$ -ésima *epoch*,  $\eta$  es el *learning rate* y  $J$  es la función de costo.

#### 5.4.3. Funciones de Costo

Como se mencionó anteriormente, el algoritmo de aprendizaje consiste en un problema de inferencia de los parámetros libres de la red dado por las conexiones entre las neuronas [37]. Con el fin de optimizar estos parámetros, es necesario poder determinar el error entre el valor estimado y el valor real. La función de costo es una función matemática que caracteriza dicho error y por lo tanto se trata de minimizar. En otras palabras, las ANNs son herramientas que buscan resolver un problema de optimización, donde lo que se

busca minimizar es el error entre el valor estimado y el valor real, descrito por la función de costo. A continuación se explican brevemente las funciones de costo utilizadas.

#### 5.4.3.1. Mean Squared Error (MSE)

El Error Cuadrático Medio o *Mean Squared Error* es la función de costo más utilizada para problemas de regresión [38]. Su función matemática es (Ecuación 10):

$$MSE = \frac{\sum_{i=1}^N (x_i - y_i)^2}{N} \quad (10)$$

Donde  $N$  representa la cantidad total de muestras analizadas y  $(x_i - y_i)$  es la diferencia entre el valor estimado y el valor real. El MSE al ser convexo, simétrico y diferenciable, es fácilmente optimizable. Sus gradientes y matrices Hessianas son fáciles de computar y por ende tiene un bajo costo computacional. Sin embargo, un aspecto importante a tener en cuenta es que el MSE es muy sensible a los *outliers*, ya que eleva el error al cuadrado [38].

#### 5.4.3.2. Mean Absolute Error (MAE)

El *Mean Absolute Error* es la suma de los valores absolutos de los errores. Matemáticamente viene representado por (Ecuación 11):

$$MAE = \frac{\sum_{i=1}^N |x_i - y_i|}{N} \quad (11)$$

Esta función, comparada con el MSE, es más robusta frente a los *outliers*, lo cual representa una ventaja si se sabe que la base de datos con la cual se trabaja contiene *outliers* no relevantes para el análisis del problema.

Un problema que puede llegar a presentar el MAE es que el gradiente se mantiene constante. Esto podría afectar el proceso de aprendizaje de la red, especialmente cuando los valores del error ya son pequeños. Una posible solución a este problema es trabajar con valores dinámicos de *learning rate*, un concepto que será desarrollado en la Sección 5.4.4. Por otro lado, debido a que la derivada de la función no es continua, se podrían generar ineficiencias a la hora de encontrar una solución [39].

#### 5.4.3.3. Mean Squared Logarithmic Error (MSLE)

El *Mean Squared Logarithmic Error* es una variante del MSE, que se explicó anteriormente. La función matemática es (Ecuación 12):

$$MSLE = \frac{\sum_{i=1}^N (\log(x_i + 1) - \log(y_i + 1))^2}{N} \quad (12)$$

Una característica importante del MSLE es que trabaja con las diferencias porcentuales entre los valores reales y predichos. Otra característica que tiene es que penaliza más los valores subestimados que los sobreestimados, lo cual introduce una asimetría en la curva del error [40].

#### 5.4.4. Learning Rate

El *learning rate* es un parámetro que regula la tasa de aprendizaje de la ANN. Como se puede observar en la Ecuación 9, la función de esta variable es modular hasta qué punto la información recientemente adquirida modifica a la anterior. Este parámetro puede ser modificado dinámicamente con el objetivo de que la convergencia se obtenga más eficientemente y de evitar caer en un mínimo local en la función de costo [36].

Un método de amplia utilización para la modificación dinámica del *learning rate* es la de la reducción del mismo, en un factor determinado, al detectarse un posible mínimo local en la función de costo. Por lo general, se especifica un valor inicial a partir del cual se parte, un valor mínimo al cual se permite llegar, y un factor de reducción. Asimismo, se suele monitorear la función de costo por una cantidad determinada de *epochs* y, si no se detecta un decremento significativo, se procede a una reducción del *learning rate*.

#### 5.4.5. Funciones de Activación

La función de activación es una función matemática que define la salida de una neurona dada una serie de datos de entrada [41]. A continuación se explican las funciones de activación consideradas en este trabajo.

#### 5.4.5.1. Función Lineal

La función lineal genera una salida directamente proporcional al valor de entrada. La misma viene definida por la Ecuación 13:

$$f(x) = ax \quad (13)$$

donde  $a$  es una constante elegida por el usuario. Como se puede ver en la Figura 18, donde  $a$  toma el valor de 2, el gradiente de esta función es un valor constante. Esto hace que no presente grandes beneficios en la reducción del error al trabajar con el método de *backpropagation* y que impida la detección de patrones complejos en los datos de entrada. Es por esto que la función de activación lineal no es recomendada para resolver problemas no lineales [42].

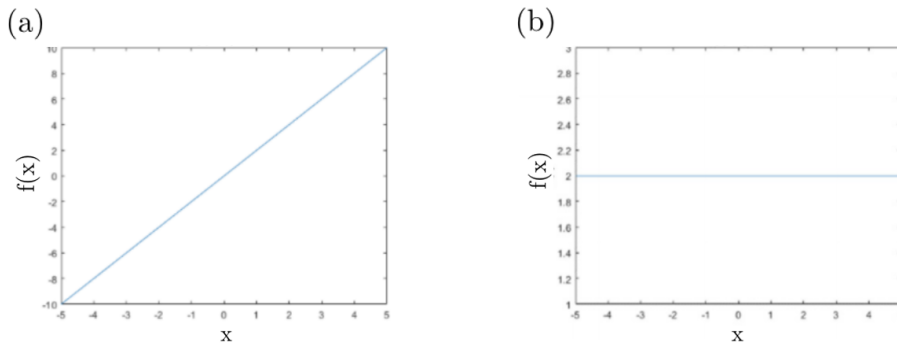


Figura 18: Gráficos de la función lineal con constante de 2 (a) y su derivada (b). Fuente:[42]

#### 5.4.5.2. Función Sigmoidea

La función sigmoidea se define de forma matemática de la siguiente manera (Ecuación 14):

$$\sigma(x) = \frac{1}{1 + e^{-x}} \quad (14)$$

Esta es una función no lineal cuyo valor de salida se encuentra entre 0 y 1. En la Figura 19 se presenta el gráfico de la función y su derivada.



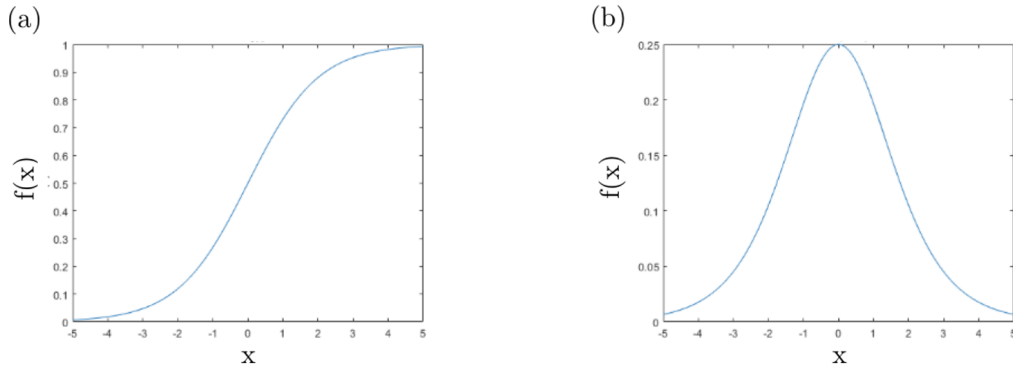


Figura 19: Gráficos de la función sigmoidea (a) y su derivada (b). Fuente: [42]

Una de las desventajas de esta función es que cuando esta se encuentra cerca del 0 o del 1, el gradiente en ambas regiones se acerca a 0. Esto impide que la red tenga un nivel adecuado de entrenamiento [42], ya que no se actualizan los pesos, finalizando el proceso de aprendizaje aunque no se haya llegado a la convergencia. Este fenómeno es conocido como *Vanishing Gradients*. Adicionalmente, como se puede observar en la Figura 19, la función sigmoidea no se encuentra centrada en 0 (con respecto al *eje y*). Esto hace que las salidas de estas neuronas tengan siempre el mismo signo, lo cual restringe la capacidad de aprendizaje de la red. Es por esto que se suele escalar la función con distintos valores para mejorar el rendimiento de la misma [42].

#### 5.4.5.3. ReLU

La función de activación ReLU (*Rectified Linear Unit*) es una función no lineal que se define de la siguiente manera (Ecuación 15):

$$ReLU(x) = \begin{cases} 0, & x < 0 \\ x, & x > 0 \end{cases} \quad (15)$$

En la Figura 20, se puede observar que es una función no lineal ya que para todos los valores de entrada menores a 0, la salida es 0. Debido a esto, esta función de activación tiene la característica de que no activa a todas las neuronas al mismo tiempo. Esto se traduce en un menor costo computacional o una convergencia más rápida en comparación con la función sigmoidea [35].

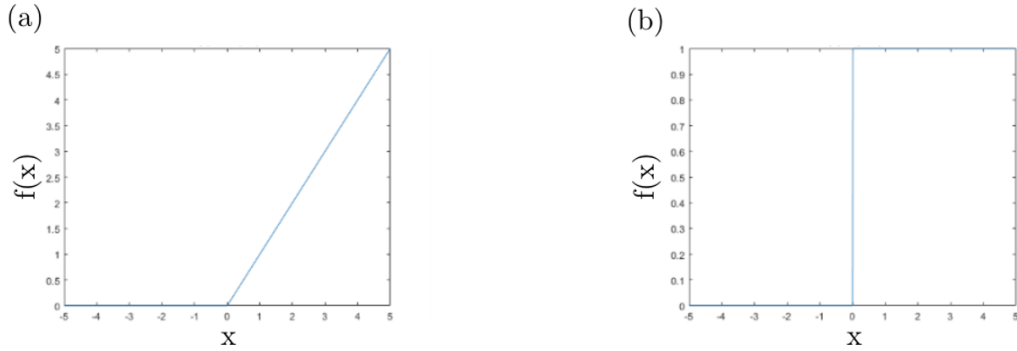


Figura 20: Función de activación ReLu (a) y su derivada (b). Fuente: [42]

Como se puede observar en la derivada de la función ReLu en la Figura 20, el gradiente de la función es 1 para las entradas mayores a 0. Esto rectifica el problema de *Vanishing Gradients* que ocurre con la función de activación sigmoidea.

Una de las desventajas que tiene esta función de activación es el fenómeno conocido como *Dying ReLu*. Dado que el gradiente es 0 para todas las entradas con valores menores a 0, los pesos no van a poder ser ajustados y las neuronas entran en un estado de inactividad lo cual termina el entrenamiento de la red [35].

#### 5.4.5.4. ELU

La función de activación ELU (*Exponential Linear Unit*) es una variante de la ReLu explicada anteriormente. Esta variación se incorpora para solucionar el problema *Dying ReLu*. Para esto, se introduce una pendiente para los valores negativos de la función lo cual no sólo disminuye el problema de *Vanishing Gradients*, sino que además acelera el proceso de aprendizaje de la red [36]. La función se puede observar en la Figura 21 y se define de la siguiente manera (Ecuación 16):

$$ELU(x) = \begin{cases} a(e^x - 1), & x < 0 \\ x, & x > 0 \end{cases} \quad (16)$$

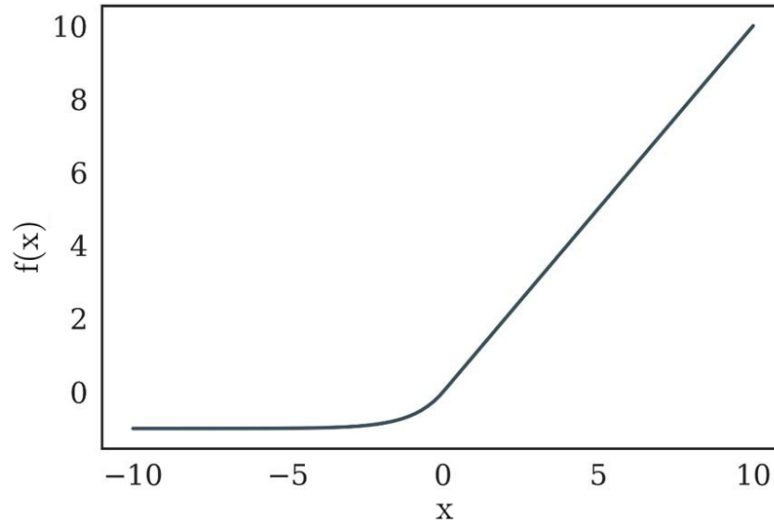


Figura 21: Función de Activación ELU. Fuente: [43]

#### 5.4.6. Optimizadores

Los optimizadores hacen referencia a las distintas técnicas o algoritmos que implementan las redes de forma intrínseca para disminuir la función de costo. Existen distintos optimizadores y a continuación se presentan algunos.

##### 5.4.6.1. Gradient Descent

El método *Gradient Descent* es uno de los más populares a la hora de optimizar redes neuronales. Para esto, se analiza el gradiente de la función de costo para actualizar los valores de los parámetros. En primer lugar, se busca obtener la dirección de máxima variación de la función. Luego, se modifican los parámetros hacia la dirección opuesta para minimizar la función de costo. Esto se realiza de forma iterativa hasta que la red llegue a un punto de convergencia [36].

En el caso del *Batch Gradient Descent* (BGD), se computa el gradiente de la función de costo a los parámetros de toda la serie de datos. Dado que es necesario calcular los gradientes de toda el conjunto de datos para generar una actualización en los parámetros, es un método lento y caro computacionalmente [36]. La siguiente ecuación define como se adaptan los pesos [44].

$$\theta = \theta - \eta \Delta_{\theta} J(\theta) \quad (17)$$

Donde  $\theta$  son los parámetros,  $\eta$  es el *learning rate* y  $J$  es la función de costo.

Por el contrario, en el *Stochastic Gradient Descent* (SGD) se actualizan los parámetros por cada dato. Al trabajar con un dato a la vez y evitar cálculos innecesarios de gradientes, es más rápido que el BGD. Las actualizaciones que se generan utilizando este método suelen tener altos niveles de varianza y por eso la función de costo a analizar puede presentar fuertes fluctuaciones [44]. La siguiente ecuación define cómo se adaptan los pesos con SGD [44].

$$\theta = \theta - \eta \Delta_{\theta} J(\theta, x(i), y(i)) \quad (18)$$

#### 5.4.6.2. RMSprop

A diferencia del SGD, el RMSprop es un método que trabaja variando los *learning rates* de los distintos parámetros. Esto es útil porque los parámetros que tienen derivadas parciales grandes, suelen generar fuertes oscilaciones en la función de costo [36]. Al penalizar las actualizaciones de este tipo, se puede llegar a un punto de convergencia de forma más rápida. Para lograr esto, en primer lugar se calcula para cada iteración una constante  $A_i$  que representa el agregado de la derivada parcial respecto de cada parámetro. Para cada iteración,  $A_i$  se calcula de la siguiente forma (Ecuación 19) [36]:

$$A_{i+1} = \rho A_i + (1 - \rho) \left( \frac{dL}{dw_i} \right)^2 \quad (19)$$

donde  $\frac{dL}{dw}$  es la derivada parcial de la función de costo respecto al parámetro analizado y  $\rho$  es un factor de decaimiento que va entre 0 y 1. La siguiente ecuación define cómo se actualizan los pesos (Ecuación 20):

$$w_{i+1} = w_i - \frac{\alpha}{\sqrt{A_i}} \left( \frac{dL}{dw_i} \right) \quad (20)$$

#### 5.4.6.3. Adam

El optimizador Adam, al igual que el RMSprop, trabaja con la constante  $A_i$  ya definida previamente. Además se incorpora un nuevo valor  $F_i$  a la hora de actualizar los pesos. El mismo se define de la siguiente manera:

$$F_{i+1} = \rho_f F_i + (1 - \rho_f) \left( \frac{dL}{dw_i} \right) \quad (21)$$

La constante  $F_i$  representa una suavización exponencial al gradiente de la función de costo. Este término se añade para tener en cuenta la variación del momento o *momentum* de la función al actualizar los pesos [36]. Por otro lado, el *learning rate* aplicado en este optimizador es distinto al de RMSprop. La misma viene definida de la siguiente forma (Ecuación 22):

$$\alpha_t = \alpha \left( \frac{\sqrt{1 - \rho^t}}{1 - \rho_f^t} \right) \quad (22)$$

Esto es porque el *learning rate* del algoritmo RMSprop introduce *bias* en las primeras iteraciones con determinadas inicializaciones. El *learning rate* en Adam corrige esto mediante el nuevo término que se le agrega. Si se analiza detalladamente, se puede observar que el nuevo término sólo genera cambios en las primeras iteraciones ya que cuando  $t$  es grande, el término converge a 1 y  $\alpha_t$  converge a  $\alpha$  [36].

Teniendo todo esto en cuenta, los pesos se actualizan siguiendo la siguiente expresión (Ecuación 23):

$$W_{i+1} = W_i - \frac{\alpha}{\sqrt{A_i}} F_i \quad (23)$$

#### 5.4.7. Fenómeno de *Overfitting*

Un aspecto importante a tener en cuenta es que las ANNs, al igual que cualquier modelo de *Deep Learning*, son entrenadas con un conjunto de datos determinado, denominado conjunto de entrenamiento o *set de training*. Sin embargo, a fin de evaluar su desempeño, se suelen utilizar conjuntos separados denominados de validación (o *validation*) que la ANN no recibió como entrada durante el entrenamiento. Esto se debe a que es posible que se obtengan modelos que no sean del todo útiles. Por ejemplo, puede ocurrir que la ANN entrenada presente un buen desempeño frente al *set de training*, mientras que al realizar una evaluación con datos que no haya previamente recibido la red, los resultados no sean óptimos. En otras palabras, la ANN podría “memorizar” el conjunto de entrenamiento sin habilidad de generar buenas predicciones para datos nuevos [36]. Es por esto que resulta de importancia mantener un conjunto de información separado (el *set de validación*) que no sea utilizado durante el entrenamiento de la ANN. Este conjunto luego servirá para cuantificar la diferencia del desempeño para los conjuntos de validación y de entrenamiento. Grandes diferencias podrían indicar la existencia de *overfitting* [36].

El *overfitting* es un problema intrínseco en el diseño de las ANNs resultando en que la red tenga un buen desempeño en el conjunto de *training* pero uno pobre en el de

validación, es decir, en la información que no ha visto previamente [36]. En estos casos, las redes fallan en identificar patrones estadísticamente significativos para la clasificación o estimación de datos de entrada desconocidos. El incremento en la cantidad de nodos de una ANN suele promover la aparición del *overfitting*. Existen investigaciones recientes cuyo foco es la arquitectura de la ANN, además de los cálculos realizados dentro de cada nodo a fin de minimizar la problemática del *overfitting*. Adicionalmente, se debe mencionar que la forma en que la ANN es entrenada tiene un impacto en la calidad del modelo final [36].

Se debe aclarar que, en ciertas situaciones, se puede necesitar una división adicional del conjunto de validación. En definitiva, esto resultaría en tres conjuntos finales: el de entrenamiento, el de validación y el de *test*. La razón por la cual se puede realizar esto es porque el *set* de validación suele estar involucrado en el proceso de selección de la arquitectura y los componentes de la ANN en cuestión. Una vez seleccionados, se aplica el *set* de *test* para la verificación del desempeño de la ANN final, siendo este un conjunto que el modelo no ha visto anteriormente.

#### 5.4.7.1. Regularización

La regularización hace referencia a los métodos que se implementan en la ANN para reducir la cantidad de parámetros y evitar el *overfitting*. Para lograr esto se puede restringir la cantidad de parámetros utilizados o disminuir el valor absoluto de los mismos mediante penalizaciones. Uno de los métodos más populares para regularizar es el *Dropout*. Este último consiste en seleccionar un conjunto de neuronas e ignorar sus valores de salida de manera que provoquen un efecto probabilístico en el accionar de la red y eviten que ciertas neuronas aprendan el error en vez del dato en sí mismo. Usualmente las probabilidades que se usan para eliminar neuronas están entre el 20 % y el 50 % [36].

Otro método de regularización conocido es el *Early Stopping*. Este consiste en detener el proceso de entrenamiento luego de una determinada cantidad de iteraciones. Existen varias formas de implementar el *Early Stopping* pero las más conocidas son determinando la cantidad de iteraciones o frenando el proceso de entrenamiento de la red cuando empieza a decaer el rendimiento de la misma. Dado que dicho método finaliza de forma prematura el entrenamiento de la red, la cantidad de parámetros intervinientes se ve restringido y por ende se lo considera una forma de regularización [36].

#### 5.4.8. Tipos de ANNs

Las Redes Neuronales Artificiales son procesadores poderosos de señales no lineales [45]. Dependiendo de las características y de los componentes que las constituyen, existen

distintos tipos de redes. En este proyecto se estudiaron las *Multilayer Perceptron* (MLP) y las *Convolutional Neural Networks* (CNN).

Las MLP son las redes más antiguas y, por mucho tiempo, fueron las más implementadas [45]. Las mismas consisten en redes con capas de entrada, de salida y, entre ellas, capas ocultas. Estas redes también son conocidas como redes *feed-forward* ya que la transmisión de datos sigue el esquema de *forward propagation* explicado anteriormente. La red más simple es el perceptrón, el cual consiste únicamente de una capa de entrada y un nodo de salida. A medida que se van agregando *hidden layers* con funciones de activación no lineales, la red adquiere capacidad para trabajar con funciones cada vez más complejas [36].

Las redes convolucionales 2D, por su parte, obtuvieron más popularidad a partir de la competencia *ImageNet 2012* debido a que demostraron ser herramientas de gran utilidad a la hora de tratar con problemas que involucran grandes cantidades de datos y características a analizar [46]. Esto es debido a que las CNNs son capaces de realizar la obtención y el procesamiento de los *features* o características de la entrada en la misma red. Dado que la técnica fue diseñada para entradas bidimensionales, la obtención de *features* se realiza mediante la convolución bidimensional entre los datos de entrada y filtros en las capas convolucionales. Los filtros son arreglos bidimensionales de menor tamaño que la entrada y con pesos fijos. Estos filtros recorren toda la entrada realizando una convolución 2D entre sus pesos y los correspondientes valores de los datos crudos. Dado que en cada instancia el resultado es un valor único, se obtiene una matriz de menor tamaño denominada *feature map* (Figura 22), a la que luego se le aplicarán las distintas funciones de activación. La aplicación de filtros de forma sistemática permite la detección de *features* específicos en cualquier ubicación de la entrada [36]. Por otro lado, a diferencia de las redes *fully connected*, cada *feature* puede ser codificado en un solo conjunto de pesos, en lugar de ser codificado en forma redundante, ya que aparece en distintas posiciones en el proceso de entrenamiento (fenómeno que surge por la naturaleza de las redes *fully connected*). Esto último permite trabajar con una cantidad menor de parámetros ajustables en la red, lo cual contribuye a reducir el riesgo de incurrir en el fenómeno de *overfitting* [47].

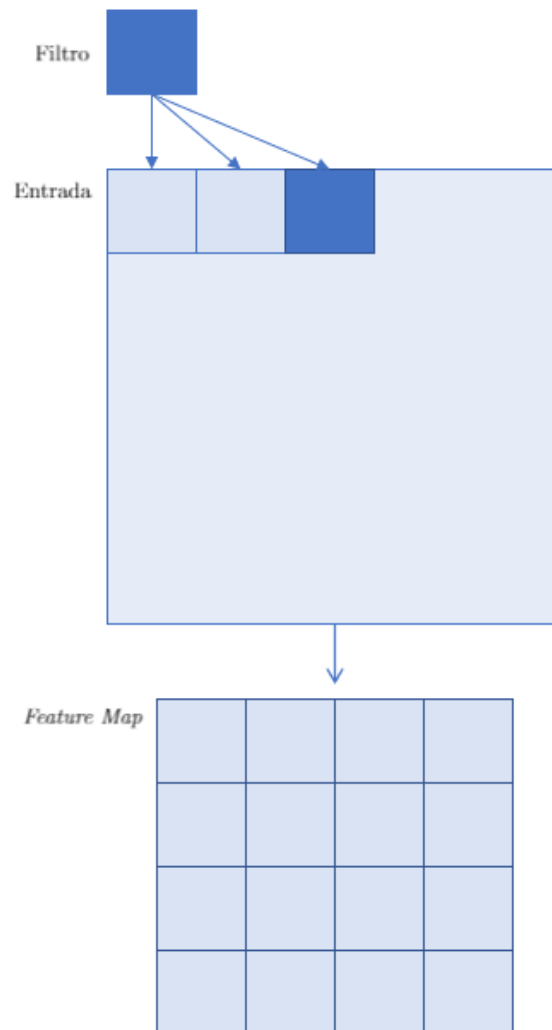


Figura 22: Convolución bidimensional entre la entrada y el filtro para la obtención del *feature map*.

Otra herramienta importante de las redes CNN son las capas *pooling*. Una de las desventajas de las capas convolucionales es que guardan la posición específica de los *features* detectados haciendo que las redes sean muy sensibles a pequeños cambios posicionales. Las capas *pooling* reducen en gran medida la resolución espacial pero mantienen la información más relevante. Esto permite que las redes convolucionales no sean tan sensibles a las deformaciones, distorsiones o traslaciones menores en las entradas [48]. Para lograr esto, las capas *pooling* trabajan con matrices o *kernels* de tamaño fijo que recorren el *feature map* con un determinado paso o *stride*. A medida que el *kernel* se desplaza por la operación de la convolución, este realiza un submuestreo de la zona afectada para permitir una mayor capacidad de generalización. Por esta razón, se suelen incorporar capas *pooling*



luego de las convolucionales [48].

Dependiendo del tipo de *pooling* que se realice, el criterio que se toma para generar la salida varía. Las capas *pooling* más utilizadas son las *Max Pooling* [48]. En este caso, los *kernels* de tamaño fijo dividen la entrada en subregiones y de cada subregión se toma solo el máximo valor. En la Figura 23 se muestra un ejemplo de *Max Pooling* con *kernels* de tamaño  $2 \times 2$  y paso 2.

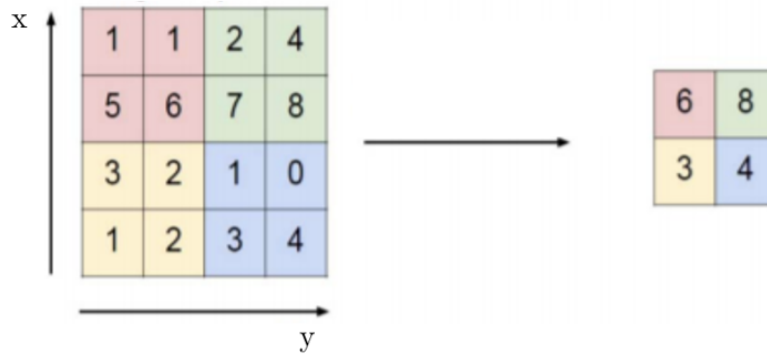


Figura 23: Demostración de *Max Pooling* con *kernel* de tamaño  $2 \times 2$  con paso 2. Fuente: [48].

Por último, para complementar el procesamiento realizado en las CNNs, se suelen introducir capas densas o *fully connected* al final de las arquitecturas. Estas capas funcionan de la misma manera que las capas de las MLPs. Se basan en los conceptos de *Backpropagation* y *Forwardpropagation* previamente explicados.

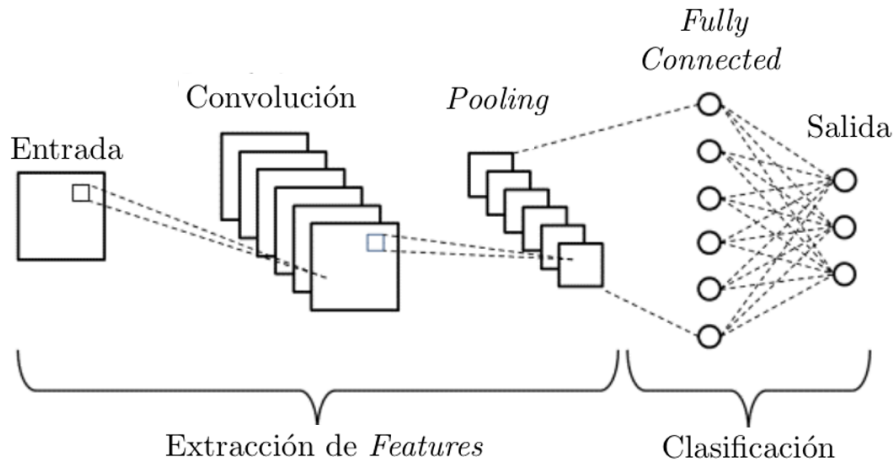


Figura 24: Arquitectura general de las redes CNN. Fuente: [49]

A modo de resumen, se puede decir que las CNNs tienen 3 tipos de capas principales [36],

que se pueden ver en la Figura 24:

- **Convolutacional:** En el caso de las convolucionales 2D la entrada tiene un formato de  $H \times W \times C$  donde  $H$  y  $W$  son las dimensiones de la imagen y  $C$  representa la cantidad de canales. Como ya se mencionó anteriormente, se trabaja con un *kernel* o detector de características con dimensión  $X \times Y \times C$ , donde  $X$  e  $Y$  son las dimensiones del *kernel*. La salida de la capa consiste en la convolución bidimensional de la entrada y del *kernel*. Es importante destacar que la convolución es una operación que genera pérdida de resolución espacial, que depende del tamaño del *kernel* y del paso con el que el mismo recorre la entrada.
- **Pooling:** Esta capa está dedicada a reducir los parámetros para evitar el fenómeno de *overfitting*. Para esto, se reduce en gran medida la resolución espacial pero se mantiene la información más relevante, lo cual resulta en una mayor generalización a la hora de analizar las entradas. Esta generalización permite a las redes convolucionales no ser tan sensibles a las deformaciones, distorsiones o traslaciones menores en las entradas.
- **Fully Connected:** Se implementan después de las capas convolucionales y las capas de *pooling*. En cuanto a la última capa, se debe tener en cuenta que la selección de la cantidad de neuronas se realiza en función de la cuestión a resolver; en un problema clasificación la cantidad de neuronas coincide con la cantidad de clases o categorías en las cuales se clasifican a los datos de entrada, mientras que en un problema en donde se busca obtener un dato numérico, es decir por ejemplo una regresión, se utiliza una neurona por parámetro a estimar. Teniendo esto en cuenta, la función de activación utilizada en la última capa dependerá también del problema, siendo la lineal la opción más utilizada para situaciones de estimación de parámetros.

Generalmente, la información entra en una capa convolucional donde una operación convolucional se encuentra definida y se aplica entre los datos de entrada y los filtros tridimensionales. Subsiguientemente, se aplica la función de activación, para luego ingresar en la siguiente capa. Esta puede tratarse de una capa de *Pooling*. Este proceso se repite a lo largo de la arquitectura de la CNN y, finalmente, la salida de la porción convolucional de la red ingresa a una sección de capas densas o *fully connected* que generan el valor de salida de la red [36].

Hasta este punto se han presentado las características principales de las redes CNN bidimensionales (CNNs 2D). Este tipo de ANNs se encuentran diseñadas para trabajar con datos bidimensionales, como pueden ser las imágenes. Dada la evidente discordancia de los requerimientos de esta herramienta para su uso en señales, se introdujeron las CNNs unidimensionales (CNNs 1D) [50]. Se deben tener en cuenta ciertas diferencias significativas en cuanto costo computacional. Al trabajar con convoluciones bidimensionales, las CNNs 2D presentan complejidades computacionales del orden  $O(N^2 K^2)$ , mientras que las

CNNs 1D tienen una complejidad del orden  $O(NK)$ . Esto hace que las CNNs 1D puedan utilizarse para realizar análisis en tiempo real sin la necesidad de emplear *hardware* complejo. Por otro lado, estudios recientes muestran que las CNNs 1D son superiores en casos donde los datos presentan altos niveles de variación como, por ejemplo, señales de electrocardiografía (ECG) [50].

### 5.4.9. Calibración

Debe hacerse una especial mención al término calibración cuando se utiliza en el presente proyecto, referido a las ANNs. En el trabajo se proponen arquitecturas y configuraciones de entrenamiento, no modelos ajustados. Cuando se toman enfoques a medida del paciente, se refiere a estos como calibrados o con necesidad de calibración, o términos derivados. Esto se debe a que, al entrenamiento por sujeto en aplicaciones de señales fisiológicas se lo suele llamar calibración, por lo que se consideró buena práctica utilizar este término [51]. Adicionalmente, parte de la bibliografía referenciada en la Sección 7 hace uso de esta terminología, por lo que resulta consistente. Sin embargo, no debe confundirse con el uso del término en contextos de estadística, cuando se hace referencia al ajuste de parámetros de modelos paramétricos.

## 6. Calidad Médica

### 6.1. Métricas de Evaluación

El objetivo de esta sección radica en desarrollar las métricas aplicadas para el análisis de las estimaciones de BP obtenidas con las ANNs en la etapa de Procesamiento (Sección 8.7). En particular, considerando que se decidió evaluar si las ANNs proporcionan valores de BP con calidad médica (Sección 6.2), se decidió comparar los resultados arrojados por las redes, es decir las estimaciones, con los valores de presión arterial reales. En consecuencia, se definen la media del error ( $\mu_e$ ), el Error Absoluto Medio (MAE) y el desvío de este ( $\sigma_e$ ) para cada red:

$$e_i = x_{pred,i} - x_{gs,i} \quad (24)$$

$$\mu_e = \frac{\sum_{i=1}^N e_i}{N} \quad (25)$$

$$MAE = \frac{\sum_{i=1}^N |e_i|}{N} \quad (26)$$

$$\sigma_e = \sqrt{\frac{\sum_{i=1}^N |e_i - \mu_e|^2}{N}} \quad (27)$$

Donde  $N$  es la cantidad de pulsos en el conjunto de validación,  $x_{pred,i}$  es el valor de BP estimado por la ANN entrenada para el  $i$ -ésimo pulso dentro del conjunto de validación,  $x_{gs,i}$  es el valor real de BP correspondiente al  $i$ -ésimo pulso dentro del conjunto de validación, y  $e_i$  es el error absoluto de estimación de la ANN para el  $i$ -ésimo pulso dentro del conjunto de validación.

## 6.2. Estándar AAMI-ISO 81060-2:2018

En los últimos 30 años numerosas organizaciones han desarrollado protocolos para la validación de dispositivos que miden la BP [52]. Las organizaciones que se destacan en este área son la AAMI, la Organización Internacional de Estandarización (ISO) y la Sociedad Europea de Hipertensión (ESH). Fundada en 1967 en Estados Unidos, la AAMI es una organización sin fines de lucro dedicada al desarrollo y uso seguro de tecnologías en el área de la salud [53] [54]. Desde el año 1987, han publicado protocolos de validación que fueron revisados en los años 1992 y 2002 [55]. La ISO es la organización encargada en el desarrollo de estándares y normas técnicas internacionales cuya publicación del 2009 incluyó varios aspectos de los propuestos por la AAMI [56]. La ESH se enfoca en la creación de protocolos referidos a la presión arterial particularmente. En el año 2002, publicaron un protocolo con un tamaño de muestra menor al sugerido por la AAMI y la ISO [57]. Luego, en el año 2018, la AAMI y la ISO publicaron una versión revisada en conjunto [58].

Debido a la vasta cantidad de protocolos con diferentes métodos pese al objetivo común, en el año 2018 estas organizaciones decidieron realizar un estándar en conjunto que sirva de referencia mundial, eliminando discordancias. El estándar AAMI-ISO 81060-2 especifica los métodos y requisitos para el desarrollo de los equipos utilizados para la estimación no invasiva de presión arterial mediante el uso de un manguito. Si bien este estándar no aplica de forma directa a este trabajo, resulta razonable utilizarlo por encontrarse ampliamente aceptado, además de la falta de un estándar que sea idóneo al método propuesto debido a la novedad del mismo.

El procedimiento de validación de dispositivos de medición de BP propuesto implica la toma de 4 mediciones por parte de los observadores, las cuales representan el *ground truth* y se las conoce como referencia, alternadas con 3 mediciones en el dispositivo a validar. Se compara cada una de estas con el promedio entre dos mediciones consecutivas por los observadores tal como muestra la Figura 25. Se deben excluir los pacientes que tengan una diferencia de SBP mayor a 12 mmHg y 8 mmHg en DBP en 2 de las 4 mediciones de referencia.

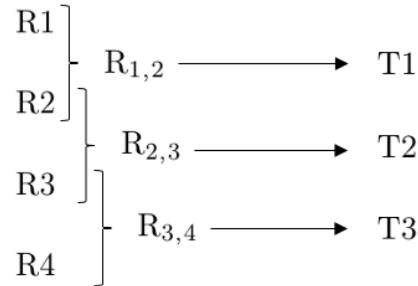


Figura 25: Procedimiento para la validación de dispositivos de medición de BP donde se toman 4 mediciones de referencia ( $R_1$ - $R_4$ ) alternadas por 3 mediciones del dispositivo ( $T_1$ - $T_3$ ). Estas se comparan con el promedio entre las mediciones de referencia consecutivas ( $R_{1,2}$ ,  $R_{2,3}$  y  $R_{3,4}$ ).

Se establece que un dispositivo posee calidad médica si cumple con los dos particulares criterios. El **Criterio 1** consiste en tomar las 3 mediciones resultantes de, por lo menos, 85 pacientes según el procedimiento comentado y calcular las diferencias entre las mediciones de referencia y la del dispositivo. De esta forma, se obtienen, al menos, 255 errores con los cuales se procede a calcular la media y el desvío estándar, que deben tener un valor menor a 5 y 8 mmHg respectivamente. El **Criterio 2** postula que la media de los errores medios por paciente debe ser menor a 5 mmHg y que el desvío estándar sea tal que el error medio en al menos 85 % de los pacientes sea menor a 10 mmHg. Por último, se deben publicar los gráficos de Bland-Altman para realizar un análisis del acuerdo entre ambos métodos de medición. En la Figura 26 se observa un ejemplo presentado por la AAMI-ISO a modo de referencia sobre este tipo de gráficos [52].

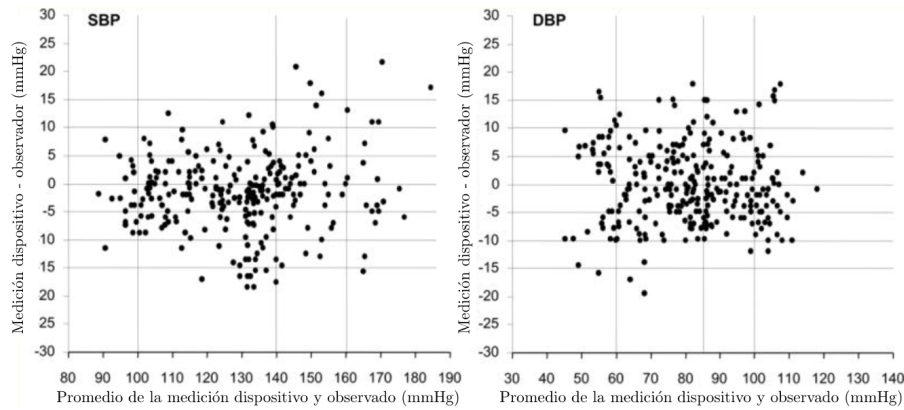


Figura 26: Gráficos de Bland-Altman para SBP y DBP. En el eje  $x$  se encuentra la media de la BP medida y la estimada. Se deben marcar con una línea vertical de referencia en 80, 100, 140 y 160 mmHg para SBP y 60, 85, 100 mmHg para DBP. En el eje  $y$  se encuentra el error entre ambas mediciones. Se colocan líneas horizontales de referencia en un intervalo de 5 mmHg desde -15 mmHg a 15 mmHg. Los errores mayores a 30 mmHg deben ser graficados a 30 mmHg. Fuente: [52].

Por último, se requiere presentar la cantidad de mediciones cuyo error, en valor absoluto es menor a 5, 10 y 15 mmHg.

### 6.3. Criterio Simplificado

Considerando que la totalidad de la bibliografía consultada (ver Sección 7) utiliza como parámetro de calidad médica el cumplimiento de un MAE menor o igual a 5 mmHg y  $\sigma_e$  menor o igual a 8 mmHg, se decidió definir este mismo como un **Criterio Simplificado (CS)**, inspirado a partir del estándar impuesto por la AAMI-ISO (Sección 6.2). El mismo se utilizó para las ANNs implementadas, es decir tanto las MLPs como las CNNs. Esto permitirá, además, una comparación de los resultados obtenidos con los de las investigaciones publicadas. Se debe notar que este método surge de una simplificación de los estándares de calidad médica de la AAMI-ISO detallados en la Sección 6.2.

Se debe tener en cuenta que el CS se utilizará para evaluar el rendimiento de las ANNs en sus resultados de desempeño finales, y no se aplicará de forma individual para analizar el rendimiento de una configuración para un paciente particular.

## 7. Estado del Arte

Investigaciones recientes buscan estimar la BP a partir de la curva de PPG. La relación precisa entre ambas curvas no se encuentra demostrada ni determinada con precisión, pero sí se encuentra establecida una correlación morfológica (ver Sección 5.3) [30]. La mayoría de las investigaciones referidas al tema procuran estimar mediciones de BP a partir de curvas de PPG mediante el uso de ANNs. En conjunto con dicha herramienta, utilizan bases de datos que contienen señales fisiológicas de pacientes.

En un principio, los estudios fueron realizados con redes MLP sencillas que tenían como entrada pocos *features*. A modo de ejemplo, se pueden mencionar los estudios realizados por Xing *et al.* (2016) y el de Priyanka *et al.* (2018) [59][60]. En el primer caso, se utilizó la base de datos MIMIC-II, a partir de la cual se obtuvo información de 69 pacientes. De allí se extrajeron *features* frecuenciales que luego fueron utilizados para el entrenamiento, la validación y la evaluación de su ANN. Para esto último, el conjunto de datos se particionó de forma tal que el 70 % fue utilizado para el entrenamiento, el 15 % para la validación y el último 15 % para la evaluación de la ANN. La red propuesta fue una MLP con una sola capa oculta y dos neuronas en la capa de salida. Los errores obtenidos para la SBP y la DBP fueron de 0.06 y 0.01 mmHg mientras que los desvíos estándares fueron de 7.08 y 4.66 mmHg respectivamente [59]. En el segundo caso se estudió el rendimiento de una red MLP con dos capas ocultas entrenada con cuatro *features* temporales provenientes de señales de PPG: la fase sistólica, la fase diastólica, el período cardíaco y el *reflective pulse time transit* (RPTT). Este último hace referencia al tiempo que ocurre entre el pico sistólico y su pico consecutivo. El conjunto de datos que se utilizó en este caso fue una base de datos propia donde el 90 % fue utilizado para el entrenamiento y el 10 % restante para la validación. Los MAEs obtenidos en este caso fueron de 0.17 y 0.3 mmHg para la SBP y DBP respectivamente con unos desvíos estándares de 3.90 y 1.80 mmHg.

Si bien en ambos casos los resultados logran cumplir con el Criterio Simplificado o CS, es decir un MAE menor o igual a 5 mmHg y un desvío estándar del error menor o igual a 8 mmHg (concepto desarrollado en Sección 6.1), se debe hacer notar que existe una falta de detalle en la información publicada, lo cual dificulta o incluso imposibilita la reproducibilidad de los trabajos para así poder verificar la veracidad de sus resultados. Por un lado, en el estudio de Xing *et al.* no se encuentran detallados los *features* específicos utilizados ni los componentes de la ANN propuesta. Entre los datos de relevancia faltantes se encuentran la cantidad de neuronas en cada capa, las funciones de activación y los optimizadores. Por otro lado, la investigación de Priyanka *et al.* detalla en mayor profundidad sus métodos pero esto no logra ser suficiente para su reproducibilidad, ya que no se encuentra información como la función de costo analizada o los optimizadores implementados. Adicionalmente, al trabajar con una base de datos propia que, además, no se encuentra publicada, generar comparaciones y extraer información reproducible de dicho trabajo es complejo.



A medida que el tema en discusión fue ganando notoriedad, el uso de *features* frecuenciales comenzó a ser analizado en mayor detalle, como en los estudios de Slapnicar *et al.* (2019) y Lin *et al.* (2021) [61] [62]. En ambos casos se logró probar que el análisis conjunto de *features* temporales y frecuenciales mejoran el rendimiento de las redes. Dentro de estos *features* temporales, se incluyeron algunos obtenidos de la primera y segunda derivada de la señal de PPG. Esto último se implementó porque estudios como el de Yousef *et al.* (2012) demostraron previamente la relevancia de estos puntos a la hora de analizar curvas PPG [63]. Parte de la explicación para lo postulado anteriormente es que ciertos puntos característicos en las derivadas ayudan a calcular de forma más precisa los *features* temporales de la curva de PPG [63].

El estudio de Slapnicar *et al.* implementó el método *Random Forest* de *scikit-learn* para generar la regresión lineal múltiple. Los mejores resultados obtenidos fueron dados por un entrenamiento con calibración donde el MAE del SBP fue de 13.62 mmHg y el del DBP fue de 11.73 mmHg. Por otro lado, Lin *et al.* utilizaron un total de 65 *features* temporales y frecuenciales e implementaron una regresión lineal múltiple. Los resultados reportados en este estudio cumplen con condición del Criterio Simplificado ya que con una base de datos de 109 pacientes lograron un MAE y un desvío estándar de 4.59 y 1.78 mmHg para la SBP y 2.47 y 1.09 mmHg para la DBP. Dicho estudio sólo fue consultado a la hora de determinar *features* para las redes MLP ya que parte del objetivo de la presente investigación es el uso de las ANNs.

En lo que refiere a bibliografía de base para la implementación de CNNs, se debe mencionar que al momento de comenzar con dicha etapa las publicaciones eran escasas. Particularmente, se podía encontrar el estudio de Slapnicar *et al.* que, además de aplicar *Random Forest*, estudiaron el uso de CNNs combinadas con *Gated Recurrent Units* (GRU) para comparar el rendimiento de ambos métodos. El mejor resultado se obtuvo calibrando la red previamente con información de los distintos sujetos analizados. Para este caso, el MAE del SBP fue de 9.43 mmHg y el del DBP fue de 6.88 mmHg. Como se puede observar, los resultados obtenidos con la ANN se acercan más a cumplir con el CS que aquellos obtenidos con *Random Forest*. Dicho estudio fue consultado principalmente para la etapa de preprocesamiento ya que la base de datos utilizada es la misma. Para la etapa de procesamiento y el diseño de la ANN, no se pudo tomar en cuenta debido a la falta de información publicada. Algunos de los detalles faltantes son la cantidad de capas convolucionales y el tamaño de los filtros en las mismas, el *learning rate* y las funciones de activación implementadas. Además, a la hora de presentar resultados, el trabajo publica las medias de los errores obtenidos pero no el desvío estándar de los mismos. Esto último dificulta el análisis y resulta cuestionable ya que representa una falta de información provista por parte de los autores.

En los últimos dos años, esta combinación de CNNs con GRUs fue estudiada en profundidad y se lograron resultados muy positivos que alcanzaron el Criterio Simplificado. Ejemplos de estos son los estudios realizados por Wang *et al.* (2020), Esmalpoor *et al.* (2020) o el de Jeong *et al.* (2021) [64] [65] [66]. Estos estudios pudieron ser consultados

en una etapa ya avanzada del proyecto, por lo cual sirven a modo de comparación de resultados pero no aportaron al diseño experimental.

La investigación realizada por Wang *et al.* utilizó la base de datos MIMIC que tiene información sobre 90 pacientes. En primer lugar, separaron las señales de PPG y ECG en pulsos individuales y luego los normalizaron. Estos pulsos normalizados se utilizaron como entrada a una ANN con aproximadamente 8 *hidden layers* incluyendo 7 capas de convolución y una densa. En una primera etapa, se analizó el rendimiento de la red con 37600 pulsos de un paciente. El 65 % de los datos se usaron para entrenar el modelo, el 20 % se utilizó para validarlo y finalmente se lo evaluó con el 15 % restante. En una segunda instancia, se estudió el rendimiento de la red frente al *transfer learning*. Para esto, se separaron los datos de los pacientes en *train* y *test*. De forma preliminar, se entrenó la red con el conjunto *train*. Luego, el conjunto *test* se volvió a dividir en dos subconjuntos: uno para complementar el entrenamiento de la red y otro para evaluarla. Los MAEs con sus desvíos estándares obtenidos fueron de 3.95 y 4.38 mmHg para SBP y 2.14 y 2.40 mmHg para DBP, lo cual cumple con los requisitos de los estándares del CS. Como desventaja se puede mencionar el uso de señales ECG ya que la adquisición de las mismas implica el uso de equipamiento complejo y costoso. También se puede mencionar la falta de información necesaria para su reproducibilidad. Algunos de los datos faltantes son las funciones de activación, las cantidad de neuronas por capa densa y las funciones de activación.

El trabajo de Esmalpoor *et al.* propone el uso de dos redes. Una red para la estimación de la SBP y otra para la DBP. Cada red consta de dos etapas: La primera etapa, constituida por una CNN, cumple la función de extraer *features* y de generar una primera estimación de los valores de SBP y DBP. La segunda etapa, constituida por una GRU, sirve para mejorar la estimación realizada en la primera etapa. Para el trabajo se obtuvieron de la base de datos MIMIC II 51884 pulsos de 200 pacientes diferentes. El 70 % de estos pulsos fue utilizado para entrenar las redes, un 10 % se usó para validar y se evaluó el rendimiento final con el 20 % remanente. Se logró alcanzar el CS ya que los MAEs para SBP y DBP fueron de 1,91 y 5,55 mmHg respectivamente y los desvíos estándares obtenidos fueron de 0,67 y 2,84 mmHg. Si bien en el estudio se publican esquemas de las arquitecturas de las ANNs utilizadas, siguen faltando detalles necesarios para la reproducibilidad de las mismas. Algunos de los datos faltantes son las funciones de activación, los optimizadores y los *learning rate*.

Jeong *et al.* proponen el uso de una red entrenada con señales de ECG en conjunto con las de PPG para la estimación de presión sistólica y diastólica. A diferencia de los otros estudios mencionados previamente, en esta investigación se usa la diferencia entre la señal de ECG y la señal de PPG como entrada a la red. Para dicho estudio se utilizó la base de datos MIMIC de la cual obtuvieron 48s de información de 48 pacientes diferentes. El 80 % de los datos fue utilizado para entrenar la red, un 10 % se utilizó para la validación del modelo y el 10 % restante fue usado para la evaluación. Los MAEs obtenidos en este trabajo fueron de 0.02 y 0.16 mmHg con desvíos estándares de 1.62 y 1.30 mmHg para la SBP y DBP respectivamente. Se pudo observar, al igual que los otros estudios analizados,

que faltaban datos para poder reproducir la red propuesta. Algunos de estos datos faltantes son los pasos que toman los *kernels* en las capas CNNs o la cantidad de neuronas en sus dos capas LSTM unidireccionales.

Cabe destacar que en los últimos años, han surgido estudios que no sólo intentan estimar los valores de presión sistólica y diastólica, sino que buscan obtener la curva de presión arterial completa. Se puede mencionar a modo de ejemplo el estudio realizado por Athaya *et al.* (2021) [4]. Este último trabaja con la base de datos MIMIC-III y utiliza como conjunto de entrada las señales de PPG de los distintos pacientes. Para el estudio se obtuvieron, en promedio, 3.4 horas de datos de 100 pacientes distintos. El 70 % de los datos disponibles se utilizó para entrenar la red, un 15 % se usó para la validación y finalmente se evaluó el modelo con el 15 % restante.

Para lograr su objetivo, proponen una red basada en la *U-Net* propuesta por Ronnenberger *et al.* (2015) [67]. Dado que esta última fue creada para procesar imágenes bidimensionales, la red tuvo que ser modificada para recibir señales unidimensionales.

Para analizar la similitud entre la curva predicha y la curva de referencia, se calculó el coeficiente de *Person*. La media de esta última fue de 0.993, lo cual indica que la red fue capaz de predecir de forma precisa gran parte de la curva de presión arterial. Por otro lado, también se analizaron los valores de SBP y DBP obtenidos y sus errores. Los MAEs para la SBP y la DBP fueron de 3.68 y 4.42 mmHg y los desvíos estándares fueron de 1.97 y 2.92 mmHg. Hasta la fecha, dicho estudio es el más completo en cuanto información publicada. El único dato que genera complicaciones es la cantidad de filtros que tiene capa convolucional. Si bien se presenta un diagrama general de la red propuesta con algunos valores numéricos, no queda en claro la cantidad de filtros de cada capa convolucional.

En la Tabla 5 se muestra un resumen de la bibliografía consultada a modo de panorama general de los estudios que se tomaron en consideración para el desarrollo de este trabajo.

En conclusión, de toda la bibliografía consultada, muchos de los estudios utilizan la señal de ECG en conjunto con la señal de PPG. Si bien la señal de ECG prueba ser útil, la misma requiere de equipamiento complejo y costoso para su adquisición [4]. Adicionalmente, se debe mencionar que esto complejiza el procesamiento y el problema, donde una de las ventajas de trabajar únicamente con señales de PPG es la simplificación del proceso. Por otro lado, se puede observar que todas las publicaciones que hacen uso de las CNNs trabajan en conjunto con GRUs. Los modelos basados en esta combinación tienden a ser costosos computacionalmente ya que las GRUs presentan baja eficiencia en el aprendizaje y por lo tanto tardan más en converger [4].

Un punto importante a destacar es que a lo largo de toda la bibliografía analizada, tanto para redes MLP como para redes CNN, se puede observar que la estimación de DBP es más precisa que la de SBP. Esto puede observarse claramente en la Tabla 5, que muestra

Autores	Año	Base de Datos	Método	Calibración	SBP   DBP	
					MAE (mmHg)	$\sigma_e^*$ (mmHg)
Xin <i>et al.</i> [59]	2016	MIMIC-II	MLP con features frecuenciales de PPG	No	0.060   0.010	7.080   4.660
Priyanka <i>et al.</i> [60]	2018	Propia	MLP con features temporales de PPG	No	0.170   0.300	3.900   1.800
Slapnicar <i>et al.</i> [61]	2019	MIMIC-III	Random Forest	Sí	13.620   11.730	-   -
Slapnicar <i>et al.</i> [61]	2019	MIMIC-III	CNN + GRU con features temporales y frecuenciales de PPG	Sí	9.430   6.880	-   -
Wang <i>et al.</i> [64]	2020	MIMIC	CNN+GRU con pulsos de ECG y PPG	Sí	3.950   2.140	4.380   2.400
Esmaelpoor <i>et al.</i> [65]	2020	MIMIC-II	CNN + GRU con pulsos de PPG	Sí	1.910   0.670	5.550   2.840
Jeong <i>et al.</i> [66]	2021	MIMIC	CNN+GRU con pulsos de ECG y PPG	Sí	0.020   0.160	1.620   1.300
Lin <i>et al.</i> [62]	2021	MIMIC-III	Regresión Lineal con features temporales y frecuenciales de ECG y PPG	Sí	4.590   2.470	1.780   1.090
Athaya <i>et al.</i> [4]	2021	MIMIC-III	CNN U-NET con pulsos de ECG y PPG	Sí	3.680   1.970	4.420   2.920

\* $\sigma_e$ : Hace referencia al desvío estándar del error.

Tabla 5: Resumen de la bibliografía consultada.

los resultados obtenidos en SBP y DBP en los diversos estudios mencionados a lo largo de esta sección.

Debe mencionarse que ninguna de las investigaciones hace uso de los datos clínicos de los pacientes en el entrenamiento de sus ANNs. Tampoco se ha podido encontrar bibliografía en la cual se hayan reproducido los trabajos mencionados. Esto puede deberse a la falta de información publicada. Otro aspecto importante a señalar es que todos los estudios mencionados afirman cumplir con calidad médica según la AAMI-ISO. Sin embargo, el análisis de los resultados obtenidos que realizan no es el correcto ya que no siguen el proceso de validación explicado en la Sección 6.2.

Finalmente, cabe destacar que los estudios publicados son sumamente recientes y, por lo tanto, se puede decir que el tema se encuentra en boga. Se intuye que una posible causa del reciente interés en la temática es la pandemia de COVID-19, que puso en el foco de interés al oxímetro de pulso y, en consecuencia, a sus funcionalidades y potenciales usos adicionales.

## 8. Materiales y Métodos

### 8.1. Materiales

Los principales componentes utilizados en el desarrollo del presente trabajo son una base de datos, un entorno de ejecución y una herramienta de análisis estadístico. En las siguientes subsecciones se presentan de forma más detallada las herramientas utilizadas.

Todos los materiales utilizados en este trabajo son de carácter *open source* y gratuitos. Esto consiste en dar libre acceso al código fuente de un programa, lo que permite que el mismo sea de libre uso, ejecución, distribución y modificación. De esta forma, se aumenta la reproducibilidad, facilitando futuros trabajos e innovaciones basados en el presente proyecto.

#### 8.1.1. Base de Datos

La base de datos utilizada es MIMIC-III (*Medical Information Mart for Intensive Care*) [68]. La misma es de uso libre y se puede encontrar en PhysioNet [69]. Esta base de datos es de acceso público pero para utilizarla, se debe completar un curso en protección de datos de los participantes de investigación que incluye información sobre la Ley de Transferencia y Responsabilidad de Seguro Médico o *Health Insurance Portability and Accountability Act* (HIPAA) [70]. La base se encuentra dividida en dos partes fundamentales: una que contiene las curvas fisiológicas de los pacientes, MIMIC-III *Waveform Database*<sup>1</sup>, y la que contiene los datos clínicos, MIMIC-III *Clinical Database*<sup>2</sup>. Es importante destacar que no todos los pacientes que se encuentran en MIMIC-III *Clinical Database* tienen registros de curvas fisiológicas. Los datos se encuentran asociados a más de 40000 pacientes que se encontraron en la unidad de terapia intensiva (UTI) del Beth Israel Deaconess Medical Center durante el 2001 al 2012 [68].

Uno de los intereses del proyecto es trabajar con las curvas de PPG y de BP, obtenidas mediante oxímetros de pulso y de forma invasiva respectivamente. Adicionalmente, otro interés es realizar un análisis estadístico y utilizar variables como la edad y el sexo de los pacientes, para lo que se utilizó un subconjunto de MIMIC-III: *MIMIC-III Waveform Database Matched Subset*<sup>3</sup>. Este último contiene información clínica de algunos de los pacientes que se poseen registros en MIMIC-III *Waveform Database*. Cabe aclarar que se encuentran otros datos clínicos como el peso, la altura y el diagnóstico asociado, entre otros. La obtención de estos datos no es trivial, en el sentido de que no se encuentra una

<sup>1</sup><https://physionet.org/content/mimic3wdb/1.0/>

<sup>2</sup><https://physionet.org/content/mimiciii/1.4/>

<sup>3</sup><https://physionet.org/content/mimic3wdb-matched/1.0/>

forma estandarizada para reportarlos. En el caso del diagnóstico vinculado a cada paciente se pueden encontrar distintas nomenclaturas para las mismas enfermedades generando que sea dificultoso el procesamiento de esta información. Por otro lado, los valores de peso y altura de cada paciente deben accederse en una tabla separada donde se ubican estas variables con un número identificador. Asimismo, estos valores se encuentran expresados en distintas unidades, por ejemplo, el peso se lo puede encontrar tanto en Kg como en lb. Además, en algunos casos estos valores no se encuentran documentados o son NaN. Los registros de los pacientes en ambas bases de datos se encuentran relacionados mediante un SUBJECT\_ID en común.

Todos los datos de cada paciente se encuentran guardados en diez directorios, denominados de p00 a p09. Por lo tanto, el nombre de cada archivo tiene la forma:

pXX/pXXNNNN/pXXNNNN-YYYY-MM-DD-hh-mm

donde pXX corresponde al directorio, pXXNNNN es el subdirectorio (con NNNN el SUBJECT\_ID) y YYYY-MM-DD-hh-mm son el año, el mes, el día, la hora y los minutos en el que se tomaron las mediciones.

Se debe aclarar que, además de poseer un SUBJECT\_ID, los individuos cuentan con un ICUSTAY\_ID, que representa un código de internación. Esto implica que un SUBJECT\_ID puede tener un ICUSTAY\_ID asociado, o más. Por este motivo, se optó por asumir independencia entre cada ICUSTAY\_ID de un mismo SUBJECT\_ID, considerándolos como pacientes independientes. Es pertinente mencionar que, de utilizar datos clínicos de los pacientes, cada ICUSTAY\_ID posee asociado los datos correspondientes del individuo al momento de ingreso a la UTI, lo cual presenta una ventaja.

Se utilizaron dos sistemas para la recolección de datos:

- *Philips CareVue Clinical Information System* (modelos M2331A y M1215A; Philips Health-care)
- *iMDsoft MetaVision ICU* (iMDsoft)

En el caso de las variables de interés para este proyecto de investigación, los sistemas son similares dado que la forma de recolección de datos se encuentra estandarizada, por lo que datos provenientes de ambos sistemas fueron incorporados en una única tabla. Adicionalmente, cabe aclarar que la frecuencia de muestreo para las señales fue de 125 Hz.

Es relevante remarcar que antes de ser volcados en la base, los datos fueron deidentificados en concordancia con la ley HIPAA, que contempla la limpieza de la información personal de los pacientes y el corrimiento de las fechas. Lo que se hizo fue una traslación temporal hacia el futuro para cada paciente de forma aleatoria, resultando en fechas de internación comprendidas entre los años 2100 y 2200. En el caso particular de los pacientes mayores a 89 años, sus fechas de nacimiento se desplazaron con el fin de encubrir su verda-

dera edad y, de esta forma, cumplir con las regulaciones HIPAA. Esto resultó en pacientes en la base de datos con edades mayores a 300 años. Para tratar con este problema, lo que se hizo fue reemplazar la edad de dichos individuos con la mediana de estos, que es igual a 91.4 años, proporcionada por los investigadores detrás de la confección de la base de datos [68].

### 8.1.2. Entorno de Ejecución

El entorno de ejecución utilizado fue *Google Colaboratory*, que es un entorno de servicio de cloud (o computación en la nube) basado en las notebooks de Jupyter [71]. Este permite utilizar software, datos y servicios a través de la red y desde cualquier ubicación mediante un dispositivo. Esta herramienta presenta ventajas como la posibilidad de acceder a entornos de ejecución gratuitos con GPUs y RAMs de distintos tamaños, facilitando así el procesamiento de grandes cantidades de datos y de algoritmos de *machine learning* [72].

Las *notebooks* creadas en Google Colaboratory, al momento de ejecución del proyecto, trabajan con el lenguaje de programación *Python 3*. Este es un lenguaje de alto nivel ya que contiene diversas librerías, como *TensorFlow* y *Keras*, que permiten realizar algunas tareas complejas en pocas líneas de código y de manera legible [73].

### 8.1.3. Herramienta de Análisis Estadístico

Se utilizó *R Studio*, un entorno para el lenguaje de programación *R* como herramienta de análisis estadístico. Es *open source* y fue diseñado con esta finalidad, además de permitir el diseño de gráficos de alta complejidad. Además, permite incorporar distintos paquetes para análisis y visualización de datos [74].

## 8.2. Métodos

Haciendo uso de los materiales mencionados (Sección 8.1), se debe tener en consideración que se implementaron ANNs de tipo MLP y CNN (Sección 8.7). En las primeras etapas del proyecto, se evaluó la utilización de las MLPs con la finalidad de entrenarlas con una gran cantidad de sujetos y así lograr reproducir, o incluso mejorar de cierta forma, los trabajos de investigación publicados en ese entonces. Como se demostrará más adelante (Secciones 9.4 y 10.3), esto no se consideró exitoso, por lo que se optó por aplicar las CNNs 1D en un enfoque individualizado, es decir con necesidad de calibración previa para cada paciente.

Se debe tener en cuenta que si bien ciertas particularidades en el proceso aplicado para las MLPs y las CNNs difieren, las mismas se aclaran a lo largo del desarrollo subsecuente y el procedimiento aplicado para la obtención de resultados se dividió en dos grandes secciones: el Preprocesamiento (Sección 8.6) y el Procesamiento (Sección 8.7). A grandes rasgos, el Preprocesamiento hace referencia a todo proceso realizado sobre la señal previo a su ingreso en la ANN, mientras que el Procesamiento refiere al diseño y entrenamiento de las ANNs para, así, obtener resultados. La Figura 27 muestra un esquema simplificado de lo descrito. Adicionalmente, a lo largo del trabajo se hará una distinción entre estos tipos de ANNs.

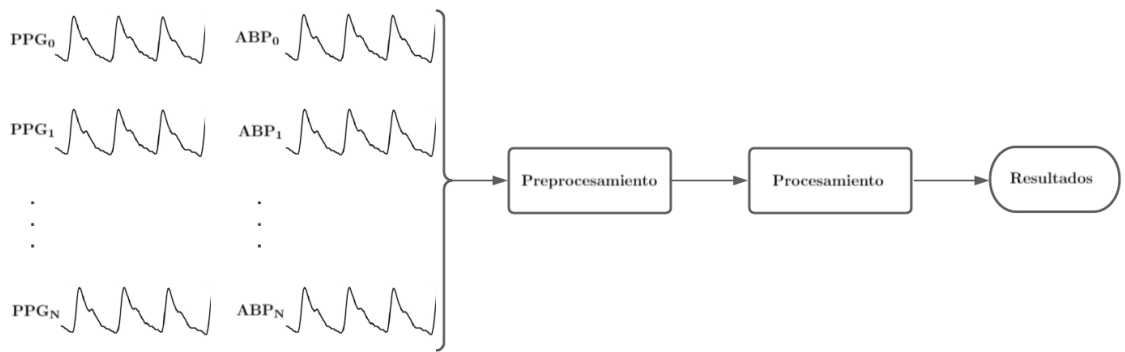


Figura 27: Diagrama de flujo del proceso general.

### 8.3. Descarga de la Base de Datos

Con el objetivo de descargar la base de datos, se confeccionó un algoritmo o *script* que recorre la base de datos MIMIC-III *Waveform Database Matched Subset*, analiza qué señales posee cada archivo y sólo se descargan los que posean la señales de interés, PPG y ABP.

Cada paciente posee una cantidad  $N$  de archivos donde se encuentran las señales digitalizadas y, por cada uno de ellos, se encuentra un archivo de extensión *.hea*. Este es un encabezado o *header* con información descriptiva del archivo que contiene las señales. El valor de  $N$  es variable para cada paciente.

En la Figura 28 se observa un caso donde el archivo que contiene las señales, en este caso el 3770665\_0004.dat, posee las señales ABP y PLETH, que corresponde a PPG. Por lo tanto, el *script* identifica que el archivo con el mismo nombre pero con extensión *.dat* debe ser descargado. Para esto, se utilizó el paquete *Python waveform-database* (WFDB) el cual fue diseñado específicamente para manejar bases de datos de *PhysioBank*. Se descargaron archivos pertenecientes a 3012 pacientes.



```

3770665_0004 6 125 5500 18:56:19.616
3770665_0004.dat 16 4093(2)/pm 12 2048 1023 -9596 0 RESP
3770665_0004.dat 16 2048(1023)/mV 12 2048 0 0 0 II
3770665_0004.dat 16 2046(1024)/mV 12 2048 0 0 0 V
3770665_0004.dat 16 2048(1023)/mV 12 2048 0 0 0 AVR
3770665_0004.dat 16 20.4733(307)/mmHg 12 2048 1987 -28396 0 ABP
3770665_0004.dat 16 1023(0)/NU 12 2048 -32768 12132 0 PLETH

```

Figura 28: Ejemplo de *header* donde se observa que se encuentran las curvas de interés denominadas como PLETH y ABP. PLETH es una abreviación de *plethysmography*, o pletismografía, que corresponde a PPG.

A continuación, se presenta, en la Figura 29, un esquema del flujo usado en la descarga de los archivos de la base MIMIC-III. Como se observa, los archivos descargados son los que poseen ambas señales de interés.

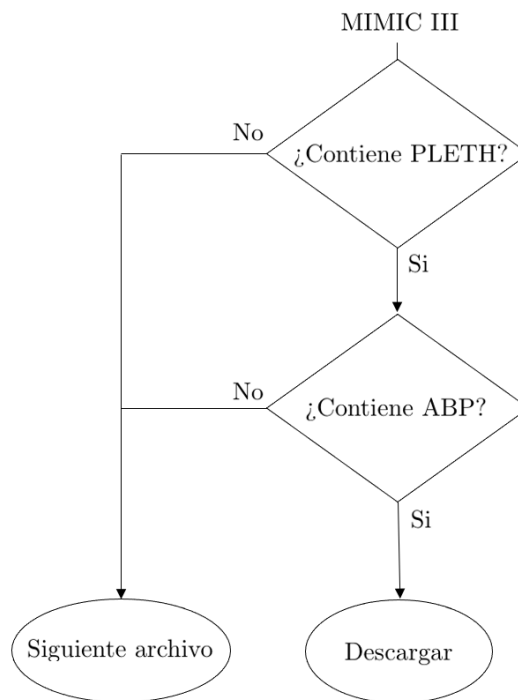


Figura 29: Esquema de flujo de la descarga de archivos según si contiene PLETH y ABP.

#### 8.4. Descarte de Señal por Peso de Archivo

Se pudo observar que la base de datos cruda contiene archivos que tienen fragmentos cortos de señal o se encuentran vacíos. Estas señales son indeseables ya que, debido a su

alta inestabilidad, no aportan información confiable para la red [61]. Es por esto que se decidió trabajar únicamente con archivos que contienen al menos 75 segundos de señal. Teniendo en cuenta que las señales poseen una frecuencia de muestreo de 125 Hz, un archivo con 75 segundos de señal, o más, contiene al menos 9375 muestras.

En el cálculo del peso de un archivo *.csv*, se debe tener en cuenta tanto la cantidad de muestras como el peso que aporte el encabezado (los títulos de las columnas). Por lo tanto, se obtuvo que si el archivo contiene 75 s de señal, esto equivale a un peso de 468.762 kb. En consecuencia, todos los archivos con menor peso fueron descartados.

## 8.5. Análisis Estadístico

Al aplicar lo descripto en la Sección 8.3, luego del *script* de descarga de los archivos de PhysioBank, del descarte por peso de archivo, y de la clasificación de señales, la cantidad de pacientes resultantes considerando SUBJECT\_IDs fue de 2362. Sin embargo, como se consideraron los diferentes ingresos en la UTI como individuos independientes (ICUSTAY\_IDs), la cantidad de pacientes estudiados fue de 2509.

Se procedió a realizar un análisis de la distribución de la edad en la población en estudio, para lo cual se generaron un histograma (Figura 30) y un diagrama de Caja y Bigote (Figura 31).

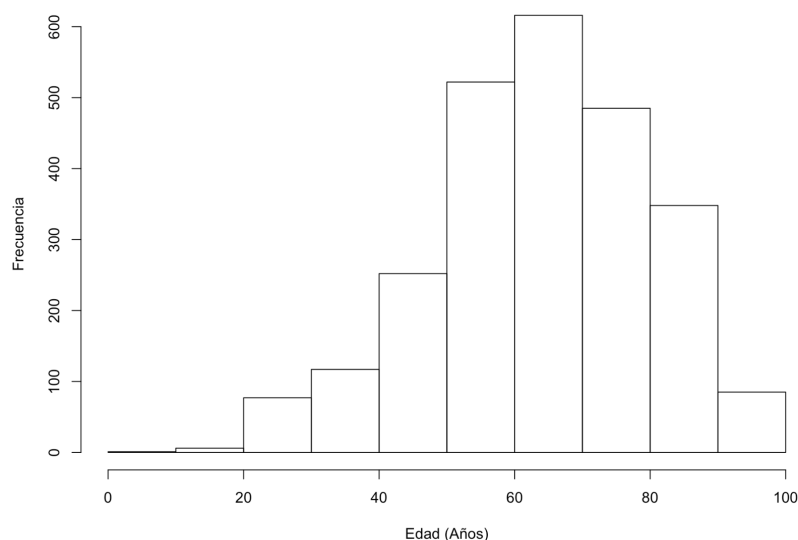


Figura 30: Histograma de la edad de los pacientes.

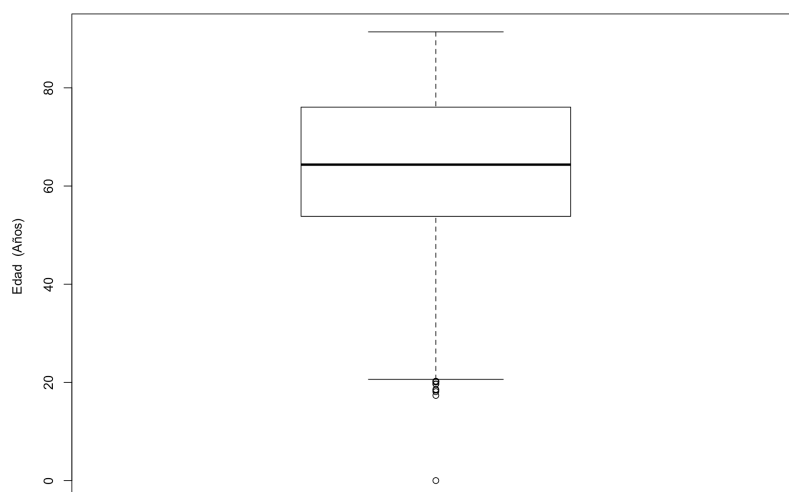


Figura 31: Diagrama de Caja y Bigote de la edad de los pacientes.

Como se puede observar mediante los métodos analíticos presentados anteriormente, la distribución de la edad de la población en estudio es normal. Por lo tanto, la edad puede ser descripta mediante la media y el desvío estándar. A continuación la Tabla 6 resume lo dicho, separado por sexo:

Sexo	Cantidad de Pacientes	Media de Edad (Años)	Desvío Estándar de Edad (Años)
F	1076	64.287	16.236
M	1433	63.313	15.642

Tabla 6: Descripción de la base de datos según sexo, cantidad de pacientes y edad.

Por último, debido a que se encuentra un único paciente cuya edad es 0 años, el cual representa un valor *outlier*, se decidió excluirlo de la base de datos.

## 8.6. Preprocesamiento

Se entiende por preprocesamiento a todos los tratamientos aplicados sobre la base de datos, lo que representa una parte fundamental de cualquier proceso de aprendizaje automático, a fin de realizar: un descarte de señales que no aportan información relevante a los fines de la investigación y un acondicionamiento de las de relevancia. En algunos casos, se busca corregir deficiencias en los datos que puedan dañar el aprendizaje, como omisiones, ruido y valores extremos. En otros, el objetivo es adaptar los datos al modelo que se pretende entrenar para optimizar el proceso [75].

Una de las ventajas que se tiene al trabajar con una base de datos extensa como MIMIC-III es que, al momento de preprocesar los datos, se puede optar por un umbral de calidad mayor en el proceso de selección de la información. En otras palabras, debido al gran volumen de información con el que se cuenta, se puede ser más estricto a la hora de fijar los parámetros para eliminar datos que puedan dañar el aprendizaje de la ANN [61].

El preprocesamiento fue llevado a cabo mediante un *script* implementado para recorrer cada uno de los pacientes y sus respectivos archivos, descargados por contener las señales fisiológicas de interés (ver Sección 8.3). Para cada uno de los pacientes, se analizó la totalidad de sus archivos aplicando las técnicas e instancias de preprocesamiento que se detallarán en las siguientes subsecciones. Una vez finalizada esta etapa se generaron, para cada individuo, dos archivos de tipo *.csv*: uno con los 12 *features* y la SBP y DBP correspondiente a cada uno de los pulsos extraídos para el paciente, y otro con los pulsos extraídos para ese mismo paciente.

Se decidió por trabajar con segmentos de 30 s de señal de PPG, lo que equivale a 3750 muestras con una frecuencia de muestreo de 125 Hz, por dos motivos: establecer un tamaño determinado y acotado del vector a analizar facilita las operaciones a realizar y se ha demostrado que la BP no varía significativamente en intervalos menores o iguales a 30s [24]. En otras palabras, se realizó el preprocesamiento de a 30 s de longitud de señal de PPG. Es necesario destacar que se analizó en simultáneo la señal de ABP. Como se evidenciará en el desarrollo subsiguiente (ver Secciones 8.6.4.3 y 8.6.6), fue necesario trabajar con 35 s de señal de presión arterial invasiva, lo que corresponde a 4376 muestras. La Figura 32 a continuación muestra un esquema de los distintos niveles de *dataset* manejados a lo largo del proyecto (señal, segmento y pulso).

A continuación, en la Figura 33 se presenta un diagrama de flujo para resumir el preprocesamiento realizado.

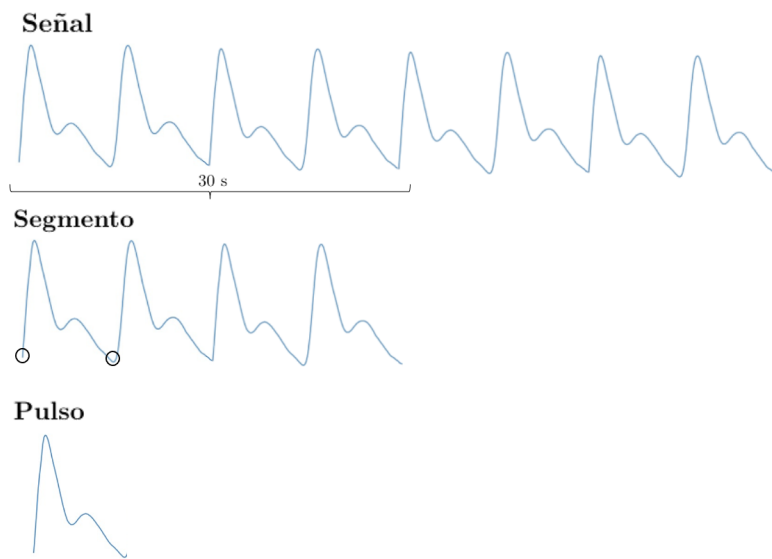


Figura 32: Esquema demostrativo de los niveles de *dataset* manejados: señal, segmento y pulso.

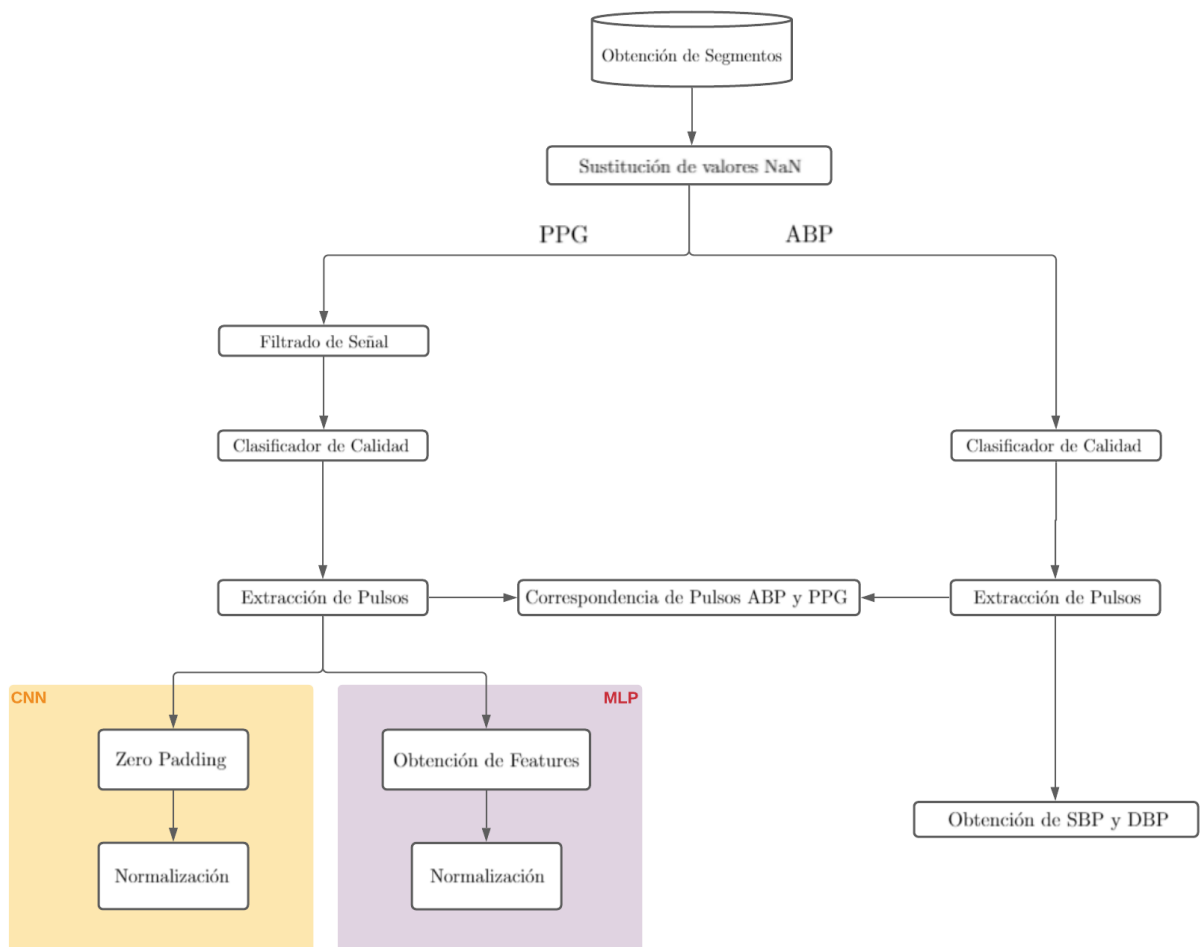


Figura 33: Diagrama de flujo del Preprocesamiento.

### 8.6.1. Sustitución de valores NaN

Un problema detectado en las etapas iniciales de prueba con señales, tanto de PPG como de ABP, es la presencia de valores NaN (*Not a Number*). Esto representa un inconveniente a la hora de procesar las señales, por lo que se decidió detectar y reemplazar estos valores NaN por 0. Se optó por esta solución debido a que se buscó preservar las señales originales en la mayor medida posible, sin introducir modificaciones o cambios. Es por esta razón, además, que se determinó no utilizar métodos predictores de tendencia, como por ejemplo el promedio móvil. Adicionalmente, debe tenerse en cuenta que, estos valores NaN se encontraron de a intervalos, no de forma aislada. Teniendo esto en cuenta y que, en pasos posteriores se realizará una clasificación de segmentos de señal, lo que se buscó fue generar una deformación morfológica de tipo *flat line* y, de esta forma, descartar estos segmentos de señal (ver Sección 8.6.4.1).

### 8.6.2. Filtrado de Señal de PPG

Una de las principales desventajas que posee la señal de PPG es su limitación debido a distintos artefactos y su alto grado de variabilidad [24]. Para corregir esto, se pueden aplicar diversos filtros a la señal. En primer lugar, se pudo observar que las señales de PPG se encuentran montadas sobre una componente continua, lo cual dificulta la interpretación de los valores de la señal. Adicionalmente, se observa un movimiento de línea de base de baja frecuencia a causa de distintos artefactos, como por ejemplo el movimiento del paciente.

A fin de eliminar el corrimiento de línea de base, se empleó la librería *heartpy*, especialmente diseñada para el análisis de señales cardíacas en *Python*, pudiendo también ser utilizada en señales de PPG [76]. Una de las cosas que permite obtener la función *process* de la librería es la señal de corrimiento de base. En la Figura 34 se puede ver la señal original con su movimiento de línea de base correspondiente, obtenido con *heartpy*. Una vez adquirido el corrimiento, se le restó a la señal original. En la Figura 35 se puede observar la eliminación del corrimiento para un segmento de PPG dado.

En segundo lugar, si se observa el espectro de frecuencias de la señal de PPG (Figura 36 y Figura 37), se puede observar que la información está contenida en dos picos principales de información. El primer pico se encuentra siempre por debajo de los 0.5 Hz, generalmente alrededor de los 0.1 Hz. Este contiene información relacionada a la actividad del sistema nervioso simpático, la respiración y la termoregulación del paciente [77]. El segundo pico, generalmente ubicado entre 0.5 y 2 Hz, contiene información sobre el ritmo cardíaco del paciente. De acuerdo con la bibliografía, cualquier otro pico que se encuentre por encima de los 2 Hz, se considera información clínicamente no relevante [78]. Sin embargo, en la práctica, los filtros aplicados suelen tener una frecuencia de corte de 8 Hz

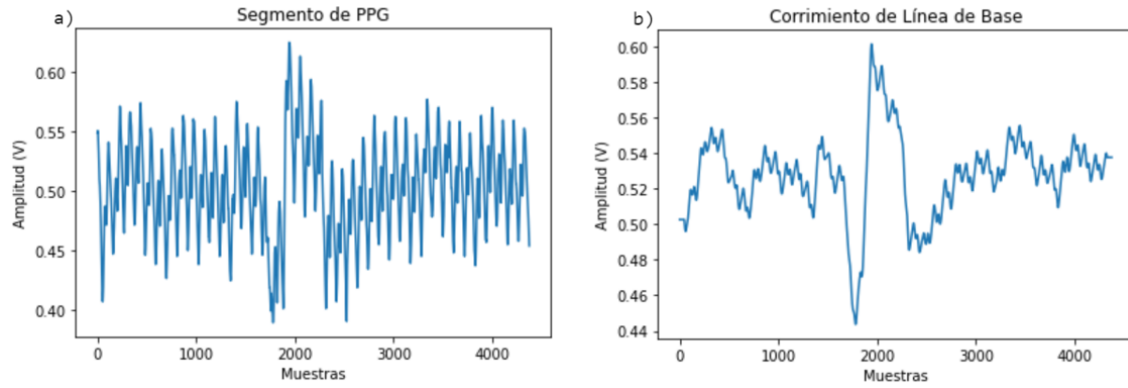


Figura 34: Segmento de PPG (a) con corrimiento de línea de base y (b) su movimiento de línea de base.

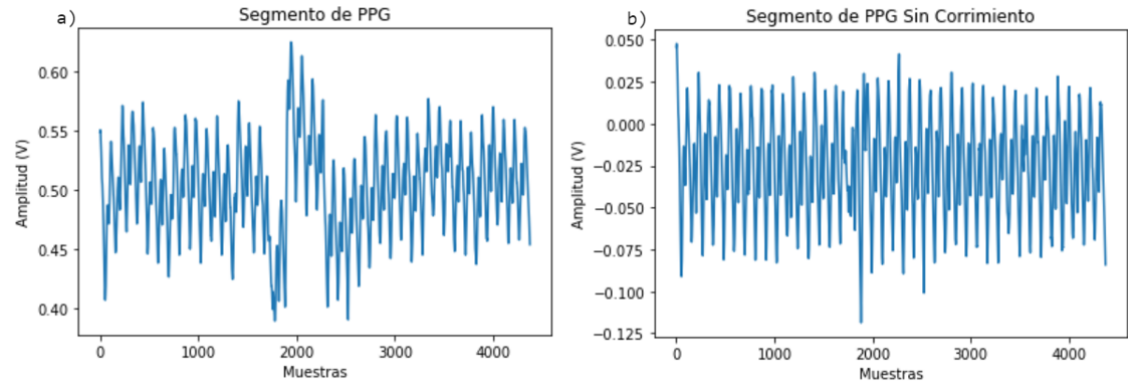


Figura 35: Comparación entre un segmento de PPG (a) con corrimiento de línea de base y (b) sin corrimiento.

o mayor. Esto se debe a que se intentan incorporar todos los armónicos para evitar la pérdida de información sobre la morfología de la señal.

Considerando el análisis frecuencial previo, se concluye la necesidad de aplicar filtros que sean capaces de conservar la integridad de la información relevante de la señal, atenuando el ruido presente en el rango de altas frecuencias. En consecuencia, en la bibliografía se reporta el uso de diversos formatos de filtros pasabajos, donde el filtro Butterworth de orden 4 se considera el *Gold Standard* para el filtrado de señales de PPG [80]. Sin embargo, Liang *et al.* estudiaron en 2018 la aplicación de filtros de Chebyshev II de orden 4 [80]. Por otro lado, también existen trabajos que analizaron el uso de filtros *Finite Impulse Response* (FIR) de orden alto para remover el ruido [82].

Estas diferencias en las técnicas aplicadas en el preprocesamiento podrían deberse a la implementación de distintas bases de datos y métodos de adquisición, que introducen

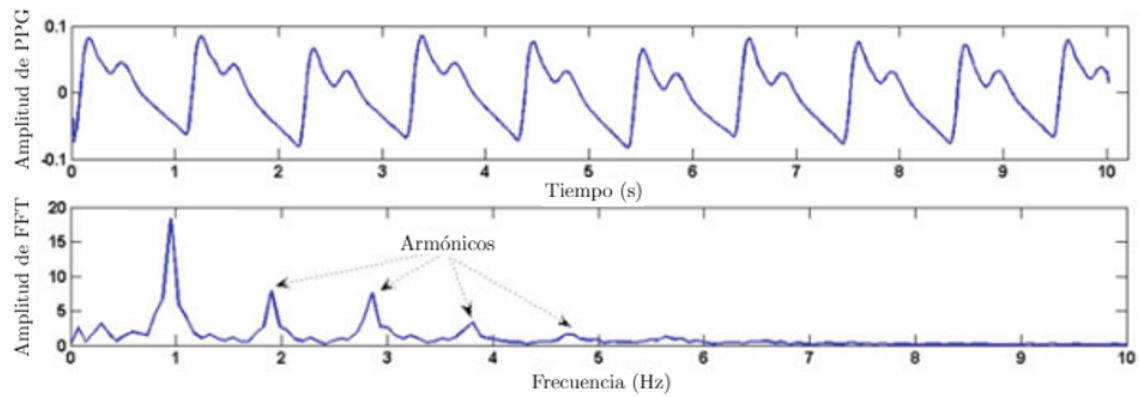


Figura 36: Señal de PPG representada en el dominio temporal y frecuencial. Fuente: [24].

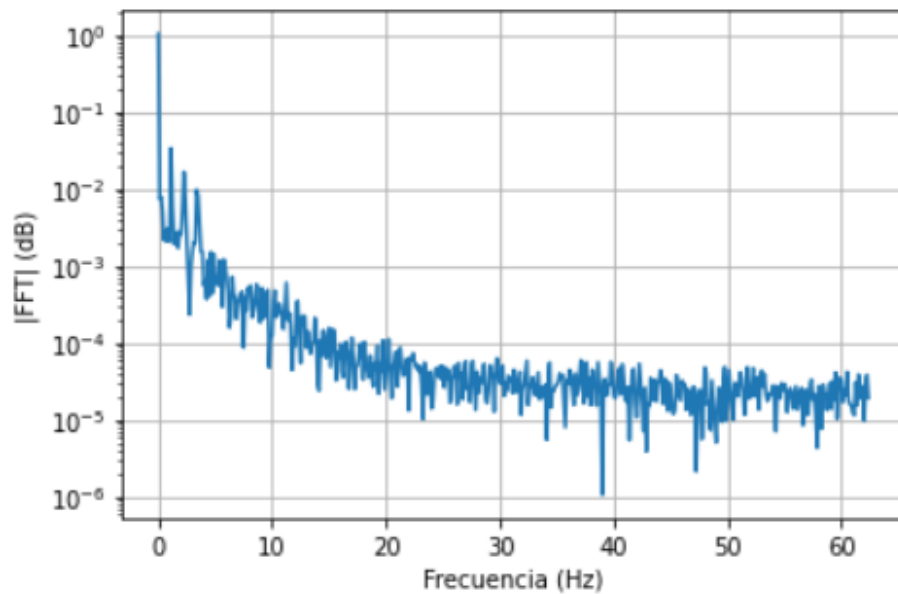


Figura 37: Espectro en amplitud de PPG en dB

sus artefactos particulares. Es por esto que se decidió comparar el uso de estos tres filtros en señales de MIMIC-III para ver cuál arroja mejores resultados.

En las Figuras 38, 39 y 40 se pueden observar los diagramas de Bode de los filtros Butterworth, Chebyshev II y FIR respectivamente.



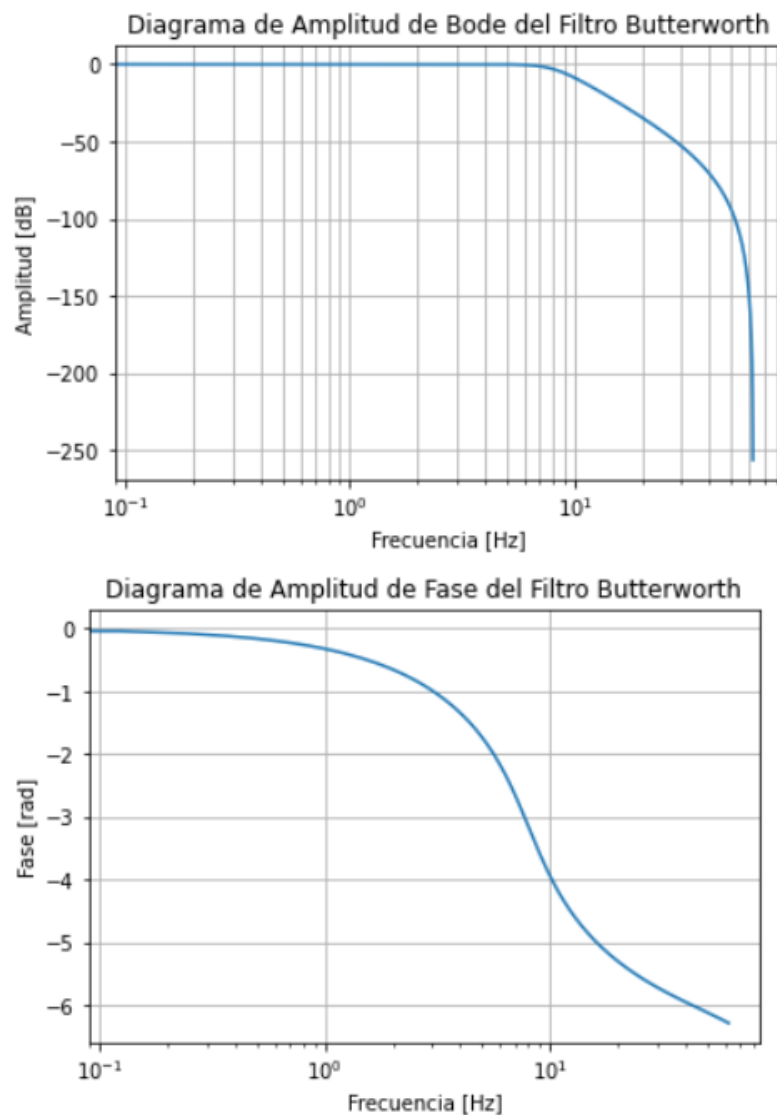


Figura 38: Diagrama de Bode de filtro Butterworth

Para el diseño del filtro Butterworth, se usaron como parámetros la frecuencia de corte, la frecuencia de muestreo y el orden. La frecuencia de corte y el orden del filtro fueron determinados por el *Gold Standard* con valores de 8 Hz y 4, respectivamente. La frecuencia de muestreo, como se mencionó en secciones anteriores, es de 125Hz. Una característica de los filtros Butterworth es la ausencia de *ripple*, tanto en la banda de rechazo como en la banda de paso. Todo esto se puede observar en la Figura 38.

A diferencia del filtro Butterworth, los filtros Chebyshev II sí tienen *ripple*. Estos filtros presentan una relación de compromiso o trade-off entre el *ripple* y la pendiente del

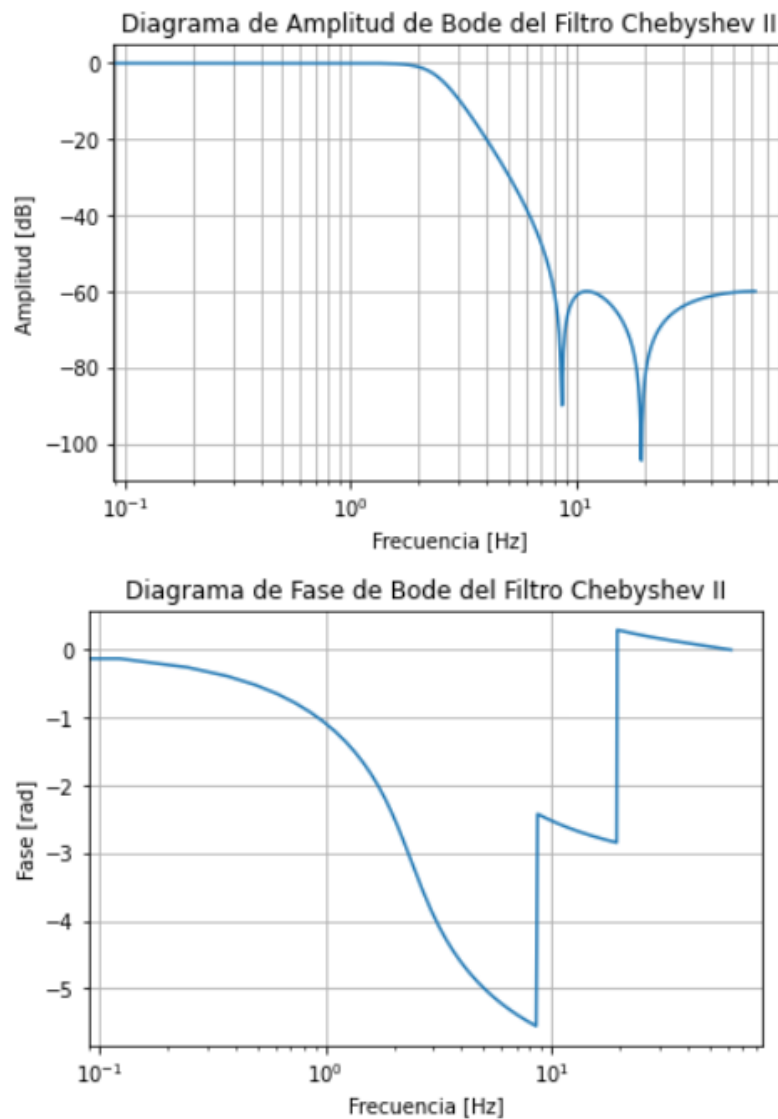


Figura 39: Diagrama de Bode de filtro Chebyshev II

filtro (*roll-off*), lo cual significa que, a mayor amplitud de *ripple* permitido, el *roll-off* del filtro puede ser mejor, acortando así la banda de transición [79]. Para el diseño del filtro se usaron como parámetros: el orden, la atenuación mínima requerida en la banda de rechazo y la frecuencia crítica a la cual la ganancia del filtro alcanza el valor de atenuación mínimo. Estos valores fueron elegidos siguiendo el estudio realizado por Liang *et al.* [80].

Por último, los filtros FIR de orden alto se caracterizan por ser más precisos y tener mejor desempeño en general a cambio de un alto costo computacional. Se decidió diseñar un filtro FIR teniendo en cuenta los parámetros observados en el estudio de Krishna *et*

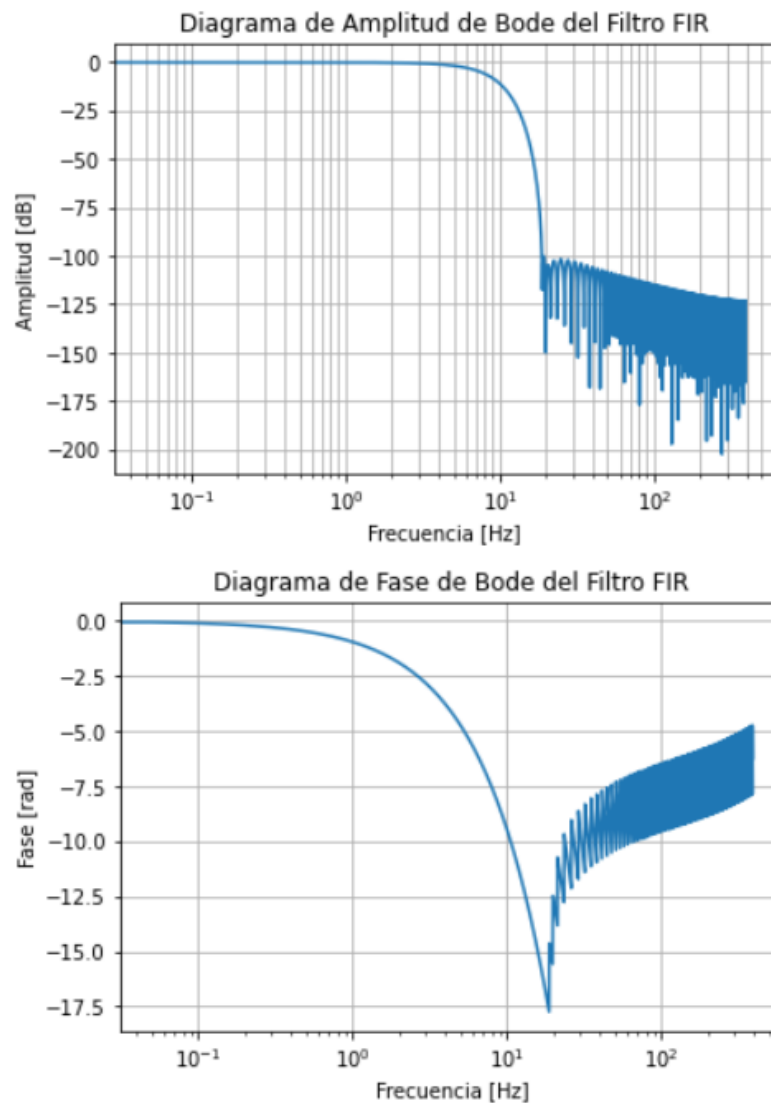


Figura 40: Diagrama de Bode de filtro FIR

*al.* [82]. Si bien en ese trabajo se propone el uso de filtros pasabandas, debido al preprocesamiento que se llevó a cabo, se decidió diseñar un filtro pasabajos. Para el diseño se utilizó una ventana *Kaiser* cuyo orden fue determinado por la atenuación y el ancho de banda de transición. Luego se incorporaron parámetros como la frecuencia de muestreo y la frecuencia de corte deseada.

Teniendo todo esto en consideración, se decidió utilizar el filtro Butterworth. En la Figura 41 se puede observar la comparación del funcionamiento de los distintos filtros. En primera instancia, la señal filtrada por el filtro Chebyshev II muestra claras deformaciones introducidas en las señales que pueden ser debido al *trade-off* entre el *ripple* y el *roll-*

*off* que se mencionó anteriormente. Por otro lado, se puede observar que tanto el filtro Butterworth como el filtro FIR cumplen con su función. Finalmente se decidió utilizar el filtro Butterworth debido a que es considerado *Gold Standard* y el filtro FIR es más costoso computacionalmente.

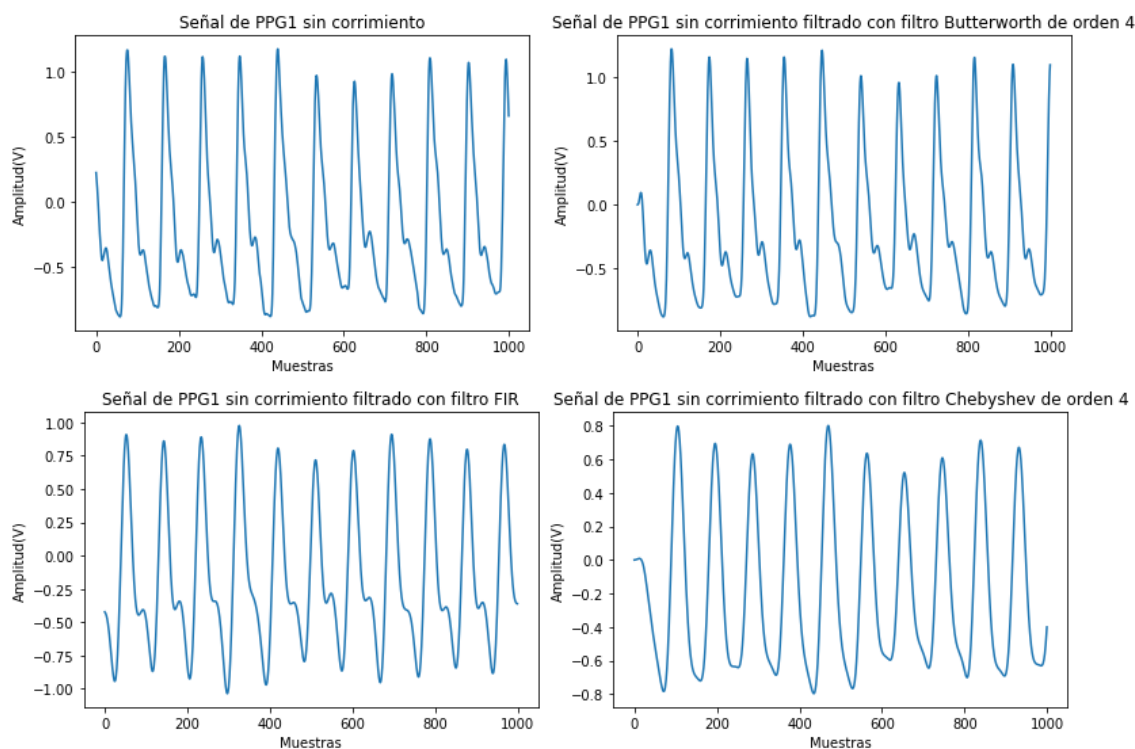


Figura 41: Señal de PPG filtrada con los distintos filtros.

### 8.6.3. Filtrado de Señal de ABP

La señal de presión arterial que se encuentra registrada en la base de datos proviene de una medición continua invasiva mediante catéter, lo cual es considerado como *Gold Standard* a nivel mundial. Con el fin de evitar alterar dichas mediciones, considerando que el método es de alta precisión, se decidió no aplicarle ningún tipo de filtro a esta señal [61].

La Figura 42 muestra un ejemplo de una señal de ABP extraída de un paciente, sin filtros aplicados.

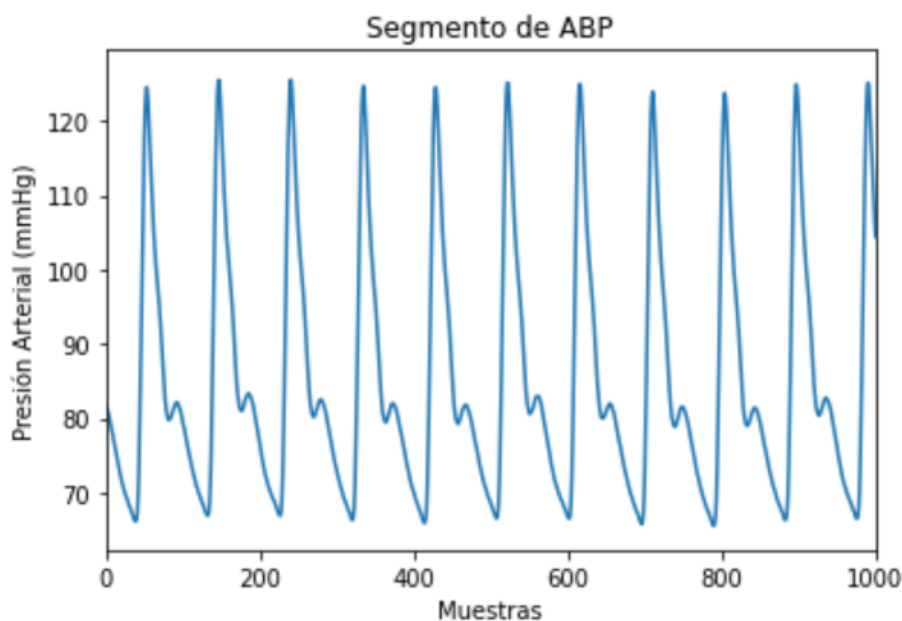


Figura 42: Segmento de señal de ABP sin filtros aplicados.

### 8.6.4. Clasificación de Segmentos de Señal

Dada la diversidad de la calidad de las señales encontradas en la base de datos empleada, además del alto volumen de información con el que se trabajó, se identificó la necesidad de elaborar clasificadores de señal, tanto para PPG como para ABP, a fin de descartar curvas que no aporten información relevante y, de esta forma, evitar el procesamiento de señales inadecuadas y reducir el tiempo de cómputo. Para esto, se reconocieron los principales tipos de deformaciones morfológicas presentes en las señales, se confeccionó un *ground truth* y se realizaron los clasificadores correspondientes.

#### 8.6.4.1. Deformaciones Morfológicas

Luego de realizar un análisis preliminar de la base de datos, se pudo observar que varios segmentos de señal presentan deformaciones morfológicas. Estas pueden agruparse en tres grupos principales:

- *Flat line*: Las deformaciones de este tipo se observan principalmente en las señales de ABP. En la Figura 43 se muestra un ejemplo de la misma y se puede observar que el segmento tiene un valor constante igual a 0 a lo largo de una cantidad de muestras mayor a  $n$ , donde  $n$  se tomó como 3 [61]. Una posible explicación para este fenómeno es la desconexión del cable o transductor durante la medición.

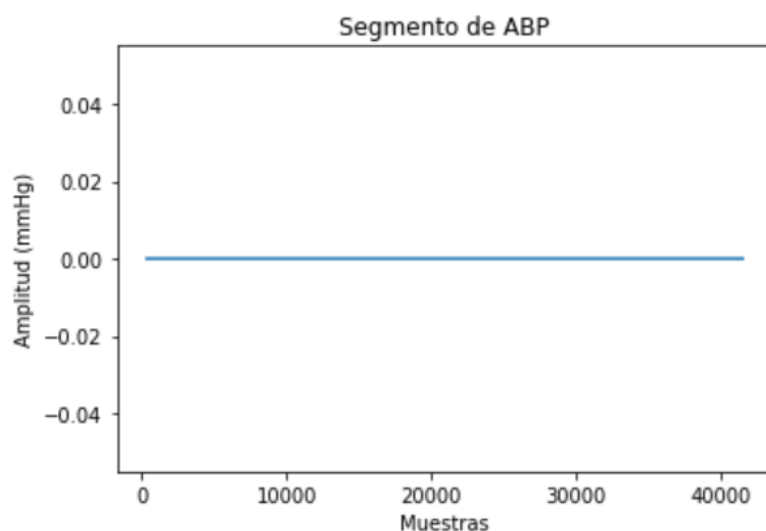


Figura 43: Ejemplo de Desvío Morfológico tipo *flat line*.

- Picos Saturados: Estos se encuentran predominantemente en las señales de ABP. Como se puede observar en la Figura 44, la señal se encuentra saturada, impidiendo la correcta detección de los picos sistólicos.

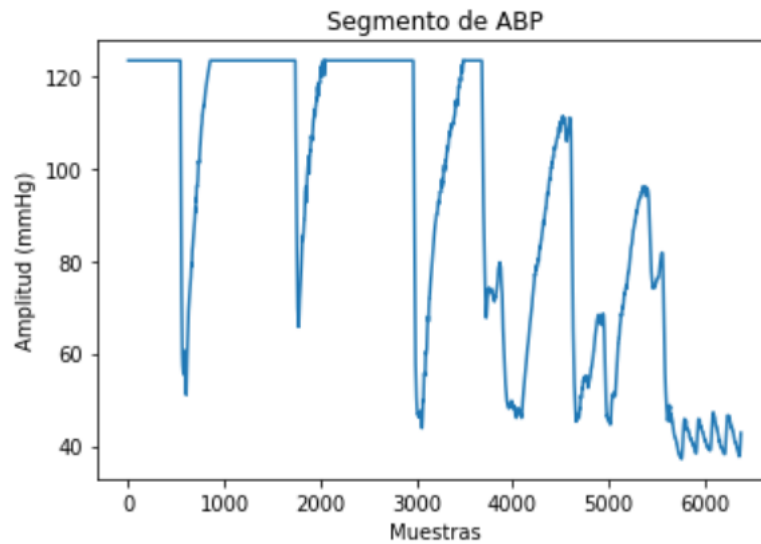


Figura 44: Ejemplo de Picos Saturados en señal de ABP.

- Desvíos Morfológicos: refiere a aquellas alteraciones en la señal visualmente evidenciables que resultan en una diferencia notable con respecto a la forma de onda teórica, no siendo atribuibles a factores fisiológicos. La Figura 45 muestra un ejemplo de dichas desviaciones.

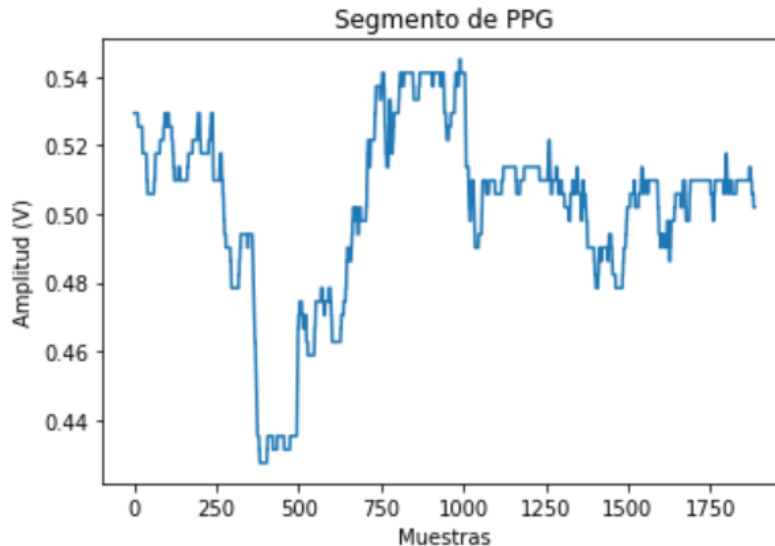


Figura 45: Segmento de PPG con Desvíos Morfológicos.

En función de los problemas explicitados, se decidió confeccionar clasificadores que determinen si la calidad de un segmento de señal dado es admisible. Esto se realizó tanto para señales de PPG como de ABP de manera separada.

#### 8.6.4.2. Definición de Calidad

Previo al armado de los clasificadores, es necesario definir qué se considera como una señal con calidad tolerable para ambos tipos de curva. Teniendo en cuenta que no se encontraron precedentes de este proceso en la bibliografía consultada, se decidió clasificar manualmente segmentos obtenidos aleatoriamente teniendo en consideración las deformaciones morfológicas explicadas en la Sección 8.6.4.1. De esta forma, se generó un *ground truth* que sirvió como base de construcción para los clasificadores implementados.

A continuación se encuentran 2 señales de PPG con su respectiva señal de ABP. En la Figura 46, ambas señales fueron clasificadas de buena calidad debido a su forma de onda, ya que no se evidencian las alteraciones definidas previamente.



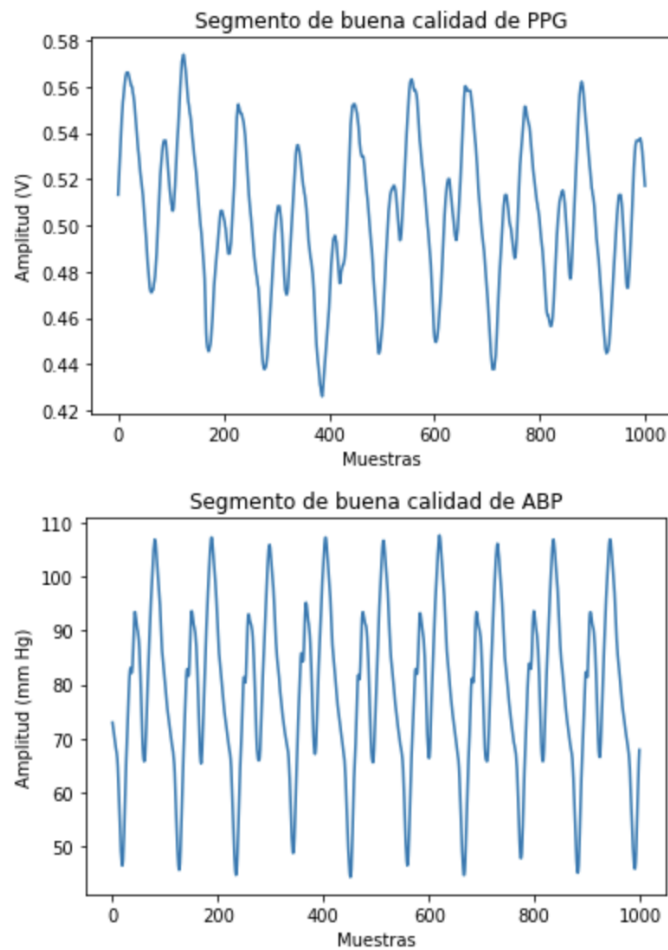


Figura 46: Señal de PPG con su respectiva señal de ABP donde se observa una buena calidad de las señales

La Figura 47 muestra una curva de PPG que se clasificó de mala calidad ya que presenta características compatibles con las desvíos morfológicos. Sin embargo, la curva de ABP que le corresponde es de buena calidad. Esto deberá ser tenido en cuenta a futuro, ya que, para que la información obtenida sea aprovechable, ambas curvas deben ser de buena calidad.

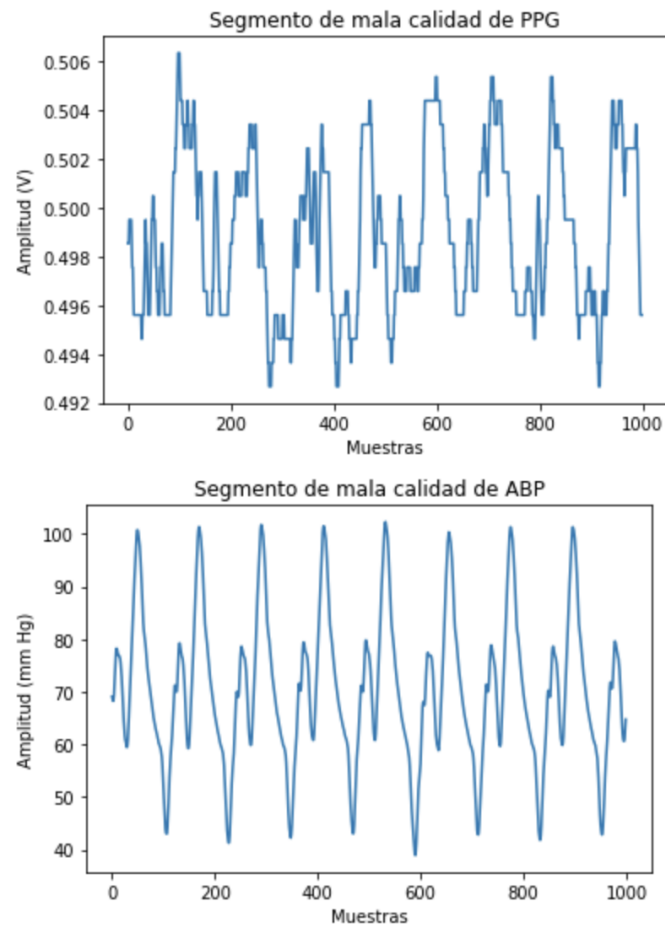


Figura 47: Señal de PPG de baja calidad con su respectiva señal de ABP

#### 8.6.4.3. Clasificación Manual de Segmentos

La primera etapa de la construcción de los clasificadores consistió en la clasificación manual de los segmentos de PPG y BP para, de esta forma, obtener un *ground truth*.

Para esto, se implementó un algoritmo que toma una cantidad  $n$  de segmentos de señal y las grafica, de forma tal que permite al usuario inspeccionar visualmente cada una de ellas y decidir si se trata de un segmento de buena o mala calidad, en base a los criterios anteriormente explicados. Más específicamente, se siguieron los siguientes pasos:

- Se obtuvo el número de parte de la base de datos de forma aleatoria (el mismo puede variar entre 0 y 9).
- Ya obtenida la parte, se accedió al directorio correspondiente en Google Drive y se

obtuvo la lista de directorios dentro del mismo. En otras palabras, se extrajeron los nombres de las carpetas de los correspondientes pacientes. A modo de ejemplo, los items obtenidos son del tipo: \p079617

- Se obtuvieron  $n$  posiciones aleatorias dentro de la lista de nombres de pacientes obtenidas en el paso anterior, resultando en  $n$  pacientes a analizar por vez.
- Para cada uno de los pacientes aleatoriamente seleccionados, se abrió el correspondiente directorio y se obtuvo la lista de directorios, es decir la lista con los nombres de los archivos para ese paciente en particular.
- Para un archivo aleatorio dentro de la lista de posibles archivos para un paciente, se obtuvo, también aleatoriamente, una posición de inicio del segmento de señal a analizar. Las particularidades dependieron de si se trataba de una señal de PPG o de ABP:
  - Señales de PPG: la posición posible de inicio de segmento se encuentra entre 0 y  $L - 3750$ , siendo  $L$  la cantidad total de muestras del archivo. A partir de esta posición aleatoria, se tomaron 3750 muestras para analizar por inspección visual. Dado que la frecuencia de muestreo es de 125 Hz, 3750 muestras equivalen a 30 s de señal.
  - Señales de ABP: la posición posible de inicio de segmento se encuentra entre 0 y  $L - 4376$ . En análisis posteriores, a un segmento de PPG dado se le deberá asignar su correspondiente segmento de ABP, por lo que, hay que considerar que existe un desplazamiento temporal entre la señal de PPG y ABP. Es por esto que se toman 4736 muestras en ABP, que corresponden a considerar 2.5 s antes y después del segmento de PPG.
- Finalmente, cada segmento graficado fue categorizado manualmente según su calidad. La clasificación fue realizada con las recomendaciones y la información encontrada en la bibliografía.

Los pasos anteriores se repitieron hasta obtener un total de 250 segmentos clasificados por inspección visual, tanto para PPG como para ABP. Esto se realizó en dos etapas, ya que primero se obtuvieron 125 segmentos y luego otros 125 adicionales, llegando al total de 250, para así posteriormente evaluar cambios en la sensibilidad de los clasificadores obtenidos a mayor  $N$  de entrenamiento.

	Buena Calidad	Mala Calidad	Proporción*
Cantidad de Segmentos	52	73	0.416

\*Proporción: Cantidad de Segmentos de Buena Calidad/Cantidad Total de Segmentos.

Tabla 7: Cantidad de segmentos con buena y mala calidad en señales de PPG para  $N:125$

	Buena Calidad	Mala Calidad	Proporción*
Cantidad de Segmentos	119	131	0.476

\*Proporción: Cantidad de Segmentos de Buena Calidad/Cantidad Total de Segmentos.

Tabla 8: Cantidad de segmentos con buena y mala calidad en señales de PPG para N:250

	Buena Calidad	Mala Calidad	Proporción*
Cantidad de Segmentos	73	52	0.584

\*Proporción: Cantidad de Segmentos de Buena Calidad/Cantidad Total de Segmentos.

Tabla 9: Cantidad de segmentos con buena y mala calidad en señales de ABP para N:125

	Buena Calidad	Mala Calidad	Proporción*
Cantidad de Segmentos	133	117	0.532

\*Proporción: Cantidad de Segmentos de Buena Calidad/Cantidad Total de Segmentos.

Tabla 10: Cantidad de segmentos con buena y mala calidad en señales de ABP para N:250

#### 8.6.4.4. Clasificadores

##### 8.6.4.4.1. *Features* de Calidad de Señal

Se utilizaron distintos *features* definidos en la bibliografía para poder valorizar la calidad de los segmentos [24]. El objetivo general fue obtener el valor de los *features*, armar clasificadores y así poder determinar cuáles de ellos presentan los mejores resultados para los objetivos buscados.

<i>Feature</i>	Fórmula	Descripción
Skewness ( $\hat{S}$ )	$\frac{1}{N} \sum_{i=1}^N (x_i - \hat{x})^3$	Mide la simetría de la distribución.
Kurtosis ( $\hat{K}$ )	$\frac{1}{N} \sum_{i=1}^N (x_i - \hat{x})^4$ $\hat{\sigma}^4$	Mide qué tan definido es el pico de la distribución.
Entropía (E)	$-\sum_i x_i^2 \ln(x_i^2)$	Es una medida del desorden de los datos.

Tabla 11: *Features* de calidad para clasificadores. Fuente: [24].

Estos fueron aplicados a las señales de PPG y ABP tanto en el dominio del tiempo como el de la frecuencia teniendo en cuenta que el análisis de los resultados es dependiente del espacio.

A continuación se presenta el análisis contemplado al momento de comparar los *features*.

#### 8.6.4.4.1.1. Dominio Temporal

Una distribución con una gran cantidad de *outliers* tendrá un valor de *Skewness* y de *Kurtosis* bajo debido a que su distribución será más asimétrica, acercándose al valor 0 de forma más lenta. En el caso contrario, tendrá mayor simetría, con un pico más definido y se acercará al 0 más rápidamente. Debido al carácter periódico de la señal de PPG, se espera que un pulso con buena calidad tenga un valor de *Skewness* asociado alto.

A continuación, se muestra una imagen donde se pueden comparar y analizar lo comentado sobre los *features* y su comportamiento con las distintas calidades de señal.

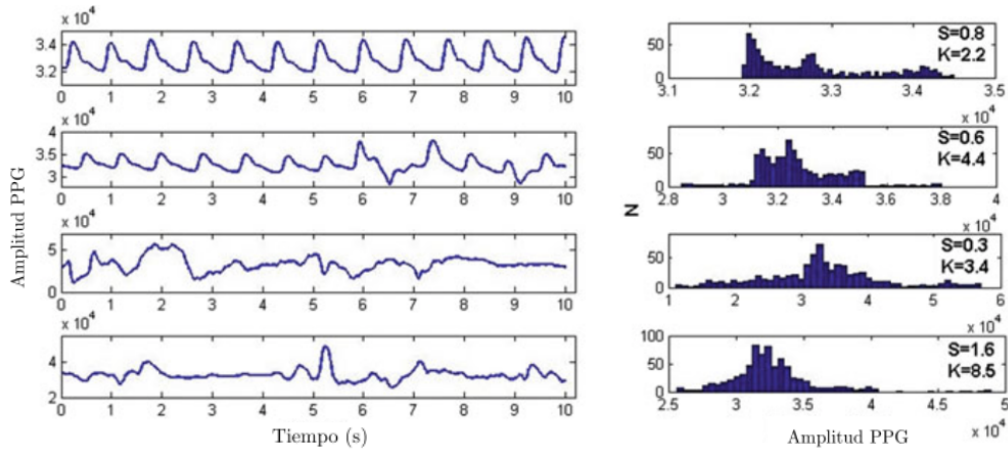


Figura 48: Señales de PPG de distintas calidades con sus medidas de Skewness (S) y Kurtosis (K). Fuente: [24].

En los primeros 3 casos de la Figura 48 se observa que el *Skewness* aumenta y la *Kurtosis* disminuye a medida que la señal tiene mayor calidad. En el último caso, se observa que el valor de *Skewness* es alto, lo que indicaría una señal de buena calidad, aunque se observa que no lo es. Esto se puede deber a la presencia de ruido, por lo que en estos casos sirve utilizar de forma combinada ambos *features* para poder determinar la calidad.

#### 8.6.4.4.1.2. Dominio Frecuencial

Adicionalmente, se decidió incorporar medidas del espectro en frecuencia. Para esto, se utilizó la *Kurtosis* de la Transformada Rápida de Fourier (FFT) de la señal.

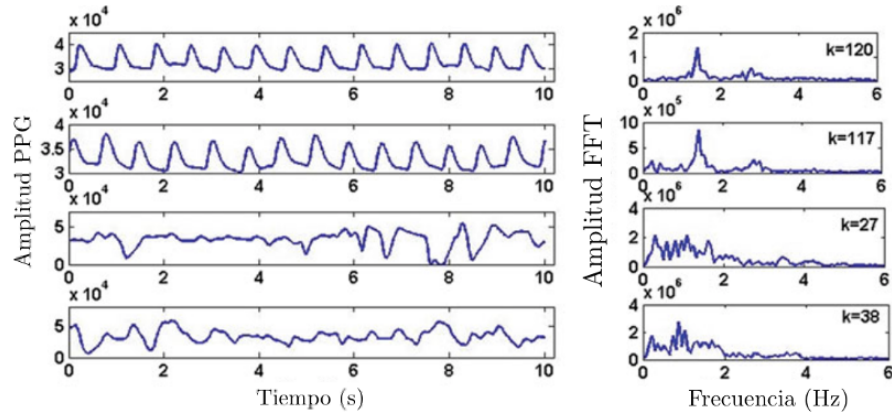


Figura 49: Señales de PPG de distintas calidad con su FFT y Kurtosis asociado. Fuente: [24].

Tal como se muestra en la Figura 49, ante la presencia de *outliers*, la distribución se vuelve más ancha y plana, resultando en valores de *Kurtosis* bajos. En consecuencia, una FFT con baja presencia de ruido tendrá un pico más definido, al tener una cantidad finita de componentes frecuenciales significativos. Si la señal posee artefactos, la FFT tendrá componentes frecuenciales asociados a los mismos [24].

#### 8.6.4.4.2. Métricas de Calidad de Clasificadores

A fin de poder comparar los diferentes clasificadores de señal generados, se utilizaron diferentes métricas. A continuación se presentan las definiciones de aquellos indicadores utilizados:

- Área Bajo la Curva ROC (AUC ROC): Para entender esta métrica es necesario definir la curva ROC. La curva ROC viene dada por la tasa de verdaderos positivos (TPR) y la tasa de falsos positivos (FPR) de un clasificador, como se puede observar en la Figura 50. El área bajo esta curva proporciona una medición del rendimiento del clasificador en todos los umbrales posibles. Para un clasificador binario, el valor oscila entre 0.5 y 1, siendo 1 cuando el clasificador logra clasificar correctamente todos los datos de entrada.

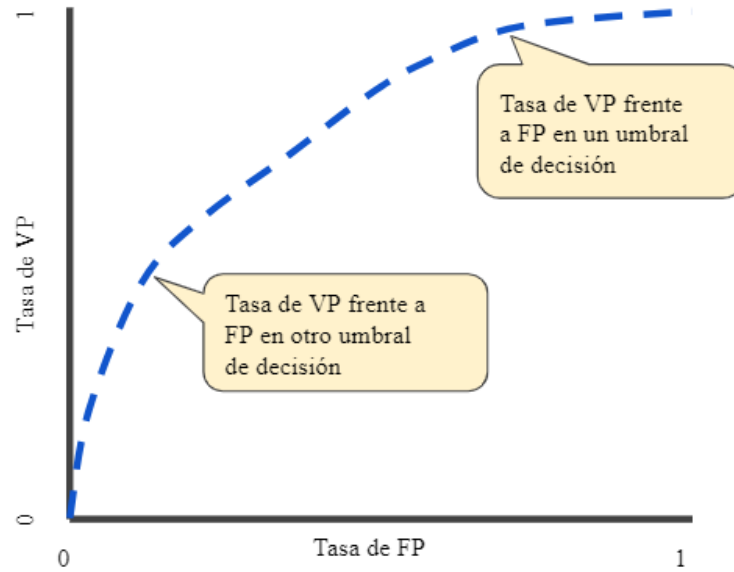


Figura 50: Curva ROC. Fuente: [81].

- Falsos Positivos (FP): Se consideran falsos positivos a los casos en los que la prueba realizada indica la presencia de cierta condición mientras que el *Ground Truth* indica lo contrario. En este contexto, serán los casos en los cuales el clasificador señala que la calidad del segmento es buena, aunque no sea así.
- Falsos Negativos (FN): Se consideran falsos negativos a los casos en los que la prueba realizada no detecta la presencia de cierta condición y el *Ground Truth* expresa la presencia de la misma. En este contexto, serán los casos en los cuales el clasificador señala que la calidad del segmento es mala, aunque no sea así.
- Sensibilidad: La sensibilidad se interpreta como la proporción de señales correctamente clasificadas. También se la conoce como la tasa de verdaderos positivos y se define de la siguiente forma:

$$TPR = \frac{VP}{VP + FN} \quad (28)$$

- Especificidad: La especificidad se interpreta como la probabilidad de que el pulso quede clasificado como malo cuando no cumple con los requisitos necesarios. También se la conoce como la tasa de verdaderos negativos. Se define de la siguiente forma:



$$Especificidad = \frac{VN}{VN + FP} \quad (29)$$

- Precisión: La precisión es una métrica que establece la proporción de predicciones positivas correctas. La misma se define como:

$$Precisión = \frac{VP}{VP + FP} \quad (30)$$

- Coeficiente de Correlación de Matthews (MCC): El coeficiente de Correlación de Matthews es una métrica que varía entre -1 y 1. El mismo es -1 cuando el clasificador binario clasifica de forma incorrecta todos los valores de entrada y 1 cuando lo hace de forma correcta. El MCC se define de la siguiente forma:

$$MCC = \frac{VP \times VN - FP \times FN}{[(VP + FP)(FN + VN)(FP + VN)(VP + FN)]^{1/2}} \quad (31)$$

- F1 Score: El F1 Score es una métrica que combina la precisión y la sensibilidad en un sólo valor. El mismo puede valer entre 0 y 1. Se obtiene el valor máximo cuando la precisión y la sensibilidad son perfectas, mientras que el valor mínimo se obtiene cuando alguno de los valores de la precisión y la sensibilidad da 0. La fórmula para calcular el F1 Score se presenta a continuación:

$$F1 = \frac{2 \times \text{precisión} \times \text{sensibilidad}}{\text{precisión} + \text{sensibilidad}} \quad (32)$$

#### 8.6.4.4.3. Implementación de Clasificadores

Teniendo en cuenta los *features* de calidad de señal descriptos en la Sección 8.6.4.4.1, se generaron distintos regresores logísticos usando diversas combinaciones de *features* y cantidad de datos de entrada. En un principio se probaron clasificadores entrenados con 125 muestras y luego se aumentó la cantidad a 250.

En primer lugar, resulta de importancia destacar que una de las prioridades en el proceso de selección de clasificadores fue obtener la menor cantidad de falsos positivos posibles, de forma tal que el clasificador sea capaz de actuar como un primer filtro que descarte señal considerada de mala calidad y, de esta manera, evitar preprocesar un gran volumen de información innecesaria. Asimismo, otro parámetro a tener en cuenta es el de falsos negativos para no perder señal potencialmente valiosa. Como existe un *trade off* entre el monitoreo de los falsos positivos y el de los falsos negativos, se decidió priorizar

la disminución de los falsos positivos sin dejar de lado el efecto que esto tiene sobre los falsos negativos.

En lo referido a la cantidad de muestras, el objetivo fue ir aumentándola e ir comparando el impacto de dicho incremento en el rendimiento de los clasificadores. La idea fue encontrar el punto en el cual un aumento en la cantidad de muestras de entrada no generaba un cambio significativo en el desempeño del clasificador. Como se verá posteriormente, no se decidió continuar aumentando la cantidad de datos de entrada dado que el proceso de clasificación manual (Sección 8.6.4.3) resultó dificultoso y llevó mucho tiempo. Es por esto que, al obtener resultados satisfactorios con un  $N$  de 250, se decidió frenar el aumento.

Para señales de PPG, se definieron el *Clasificador PPG 1* y el *Clasificador PPG 2*. El primero utiliza los *features* de kurtosis y kurtosis de la FFT. El segundo utiliza como *features* de entrada a la kurtosis, el skewness y la kurtosis de la FFT. Adicionalmente, se analizó la diferencia en rendimiento al modificar la cantidad de segmentos de señal utilizados para su entrenamiento.

Por otro lado, para las señales de ABP se definió el *Clasificador ABP 1*. El mismo tiene como *features* de calidad la kurtosis y la kurtosis de la FFT. Se analizaron los efectos del aumento de la cantidad de datos de entrada (segmentos de señal) en su rendimiento.

### 8.6.5. Extracción de Pulsos

Una vez clasificados los segmentos de señal, se procedió a separar los pulsos de cada segmento de señal. Por un lado, esto es necesario ya que, según lo consultado con la bibliografía, es posible obtener información relevante a partir de la morfología de cada pulso mediante el análisis de características o *features* [83]. Por otro lado, al trabajar con pulsos individuales es posible establecer una correspondencia o sincronía entre el pulso de PPG y el pulso de ABP. Esto último es de gran importancia ya que, sin esta sincronía temporal, se complejizaría el entrenamiento la red neuronal artificial. Para cortar los pulsos de un segmento de PPG o ABP, en primer lugar se identificaron tanto los máximos como los mínimos locales del segmento a analizar mediante la función *process* de la librería *heartpy*. Una vez hecho esto, se pudo observar que el algoritmo clasificaba varios puntos como si fuesen máximos locales cuando no lo eran. Por este motivo, se decidió incorporar más condiciones para la detección correcta de los mismos. La primera condición implementada fue que las muestras del pulso podían considerarse máximos si, y solo si, su valor era mayor al 85 % del valor máximo del segmento. De la misma forma, una muestra podía considerarse mínimo si y solo si al reflejar la señal sobre el eje  $x$ , el valor de la muestra superaba el 85 % del valor máximo. Esto último se hizo dado que la librería implementada cuenta con un detector de picos máximos fiable y no con un detector de mínimos locales. A continuación la Figura 51 muestra un ejemplo de un pulso de BP en donde se detecta

un máximo local que debería ser descartado, razón por la cual se incluye el umbral.

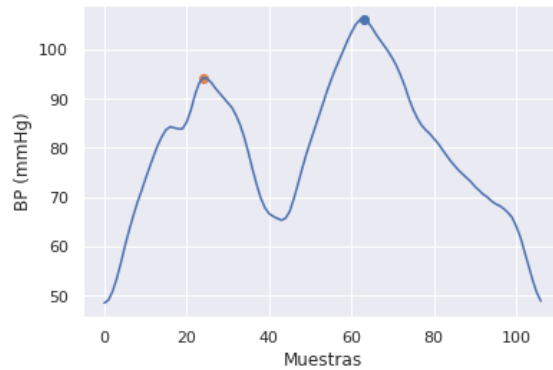


Figura 51: Pulso de ABP con detección de dos picos: el sistólico en azul (máximo absoluto) y el diastólico en naranja (máximo local).

Por último, se debe tener en cuenta que para sujetos sanos que no se encuentran ejercitando el valor de la frecuencia cardíaca debe estar entre los 30 y los 120 latidos por minuto [24]. Sin embargo, considerando que se trata de pacientes internados en la UTI que presentan diversas patologías y comorbilidades, se debe tener en cuenta que la frecuencia cardíaca máxima podría ser de hasta 180 latidos por minuto [84]. Adicionalmente, como las muestras fueron obtenidas a una frecuencia de muestreo de 125Hz, se concluye que los mínimos deben estar separados entre sí por al menos 42 muestras. Esto representa lo mínimo que puede durar un latido. La correcta detección de los extremos locales de un segmento de ABP se puede observar en la Figura 52. Cada pulso se comprende entre dos mínimos locales, marcados en naranja en la Figura. El pulso recortado mediante esa detección de máximos y mínimos locales se puede observar en la Figura 53.

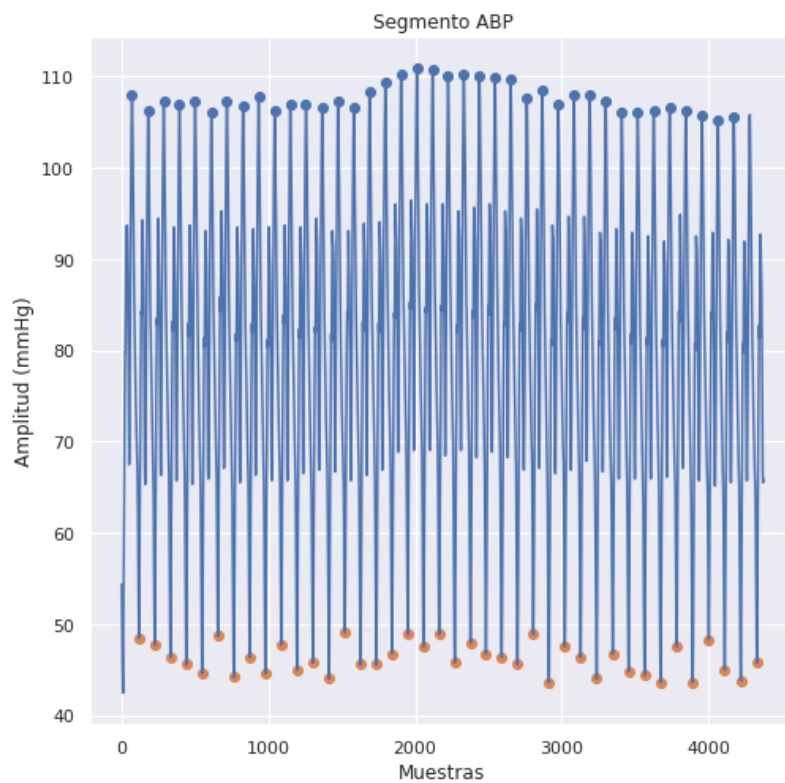


Figura 52: Detección de máximos y mínimos locales en un segmento de señal ABP.

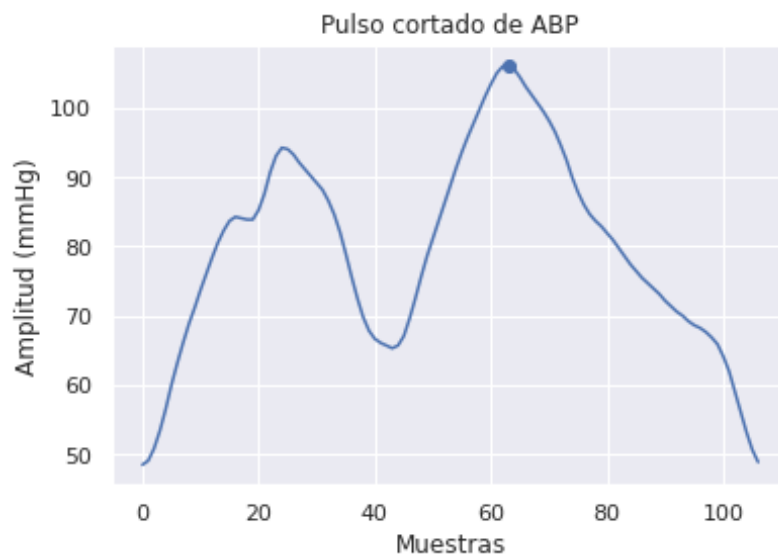


Figura 53: Recorte de pulso de ABP

Se debe hacer una breve mención al funcionamiento de la detección de los picos de la función *process* de la librería *heartpy*. La función *hp.process* contiene una función para *peak detection* en señales de origen cardíaco. La naturaleza de esta detección utiliza un umbral adaptativo, seguido de pasos de identificación de *outliers*. Para la identificación de los pulsos, utiliza un filtro de media móvil con una ventana de 0.75 s a ambos lados de la señal en cada punto de los datos[86]. Los primeros y últimos 0.75 s de la señal son poblados con la media de la señal, no se utiliza la media móvil en estas secciones. Regiones de Interés (RI) son identificadas y marcadas entre dos puntos de intersección donde la amplitud de la señal es superior a la media móvil, lo cual es una manera estándar de detección de picos. Los máximos se marcan en el máximo de cada RI [86].

Durante la etapa de la detección de picos, el algoritmo ajusta la amplitud del umbral paso a paso. Para encontrar el mejor ajuste, el desvío estándar entre las diferencias sucesivas se minimiza y la frecuencia cardíaca se verifica. Esto representa un método rápido para aproximar el umbral óptimo, ya que se aprovecha la regularidad relativa de la señal de la frecuencia cardíaca [86].

#### 8.6.6. Correspondencia entre Pulsos

Una vez cortados los pulsos, se procedió a determinar la correspondencia entre cada pulso de PPG con su respectivo pulso de ABP. Esto se realiza porque entre la señal de PPG y la de ABP existe un desfase desconocido [61]. Sin embargo, debido a que los valores de presión arterial no suelen presentar grandes variaciones en períodos cortos de tiempo, se atribuyó a cada pulso de PPG un valor de presión arterial que resulta ser el promedio de dos pulsos de la señal de ABP: el del pulso correspondiente y el del pulso anterior. En caso de que no se tenga un pulso anterior porque se esté trabajando con el primer pulso de la señal, sólo se toma el valor de ese pulso y no se realiza el promedio.

En la Figura 54 se puede observar cómo quedan cortados y sincronizados temporalmente los pulsos de PPG y ABP.

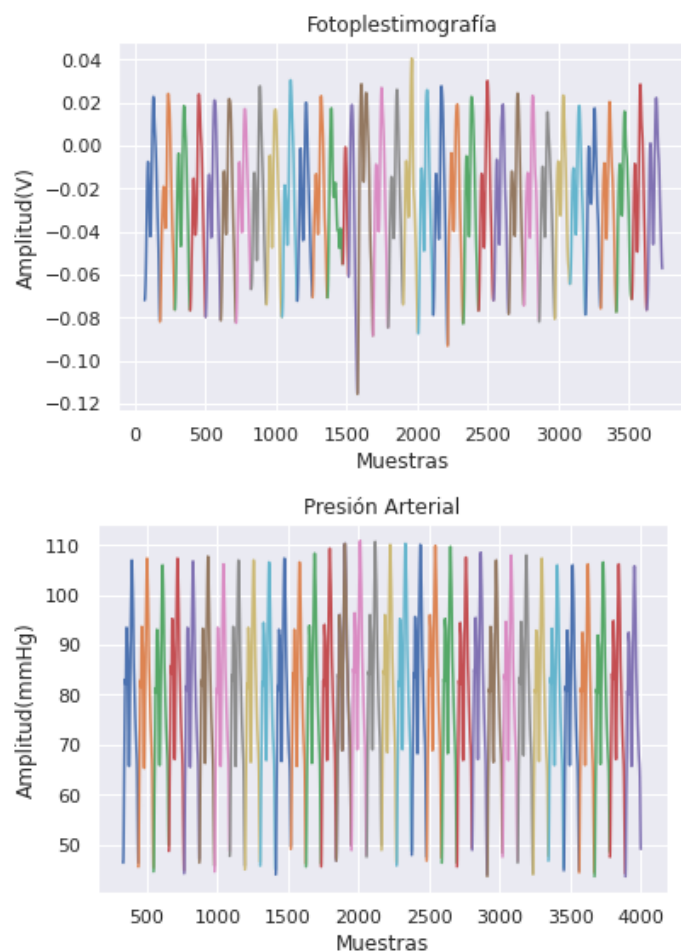


Figura 54: Pulsos cortados en señal de PPG y ABP. La correspondencia de pulsos entre ambas señales se observa con los distintos colores.

#### 8.6.7. Obtención de Presión Arterial Sistólica y Diastólica

La obtención de los valores de SBP y DBP se realizó para cada pulso individual de la señal de ABP. Es decir que este es un paso que necesita que los pulsos de ABP sean extraídos previamente. La Figura 55 muestra cómo se obtienen estos valores a partir de la curva de ABP invasiva.

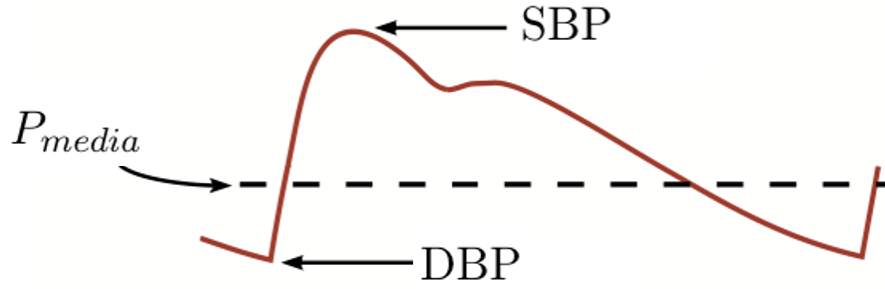


Figura 55: Pulso de ABP con la SBP y DBP señaladas. Fuente: [85].

En función de lo anterior, se desarrolló un algoritmo que recibe un pulso de ABP y devuelve los valores de DBP, SBP y de MBP. Como se indica en la Figura 55, el valor de SBP corresponde al valor máximo del pulso, por lo que se utilizó la función *max* de la librería *numpy* para obtenerlo. Además, el valor inicial del pulso representa la DBP. Una vez obtenidos estos valores, el valor de MBP se obtuvo mediante la Ecuación 33:

$$MBP = \frac{1}{3}SBP + \frac{2}{3}DBP \quad (33)$$

Se decidió no tomar en cuenta el valor de MBP debido a que es una combinación lineal de SBP y DBP y, por lo tanto, sería redundante incorporar esta información.

En la Figura 56 se visualiza un único pulso de ABP obtenido aleatoriamente, con el punto del cual se obtiene la SBP marcado en rojo y una línea punteada verde que señala la DBP.



Figura 56: Pulso cortado de señal de ABP, donde el punto rojo es del cual se obtiene la SBP y la línea punteada verde señala la DBP.

Un punto importante a destacar es que se descartaron los valores de BP, y por ende los pulsos de PPG y/o vectores de *features* asociados a ellos, considerados *outliers* de acuerdo a los rangos fisiológicos normales. En cuanto a la SBP se tomó un rango aceptable de entre 60 mmHg y 180 mmHg, mientras que para la DBP se tomó entre 45 mmHg y 110 mmHg [24].

#### 8.6.8. Obtención de Features

Como se mencionó anteriormente, parte del trabajo consiste en estudiar el funcionamiento de las redes neuronales de tipo MLP para la estimación de la presión arterial en base a la curva de fotopletomografía. Las redes MLP, a diferencia de las redes CNN 1-D, tienen como entrada características o *features* de las señales. Cabe destacar que la cantidad y la correcta selección de los mismos influyen en gran medida al rendimiento de la red. En función de la bibliografía consultada y de las investigaciones previas, se decidió trabajar con *features* temporales, frecuenciales y clínicos [24]. Los *features* elegidos son:

1. Duración del pulso (*Pulse Wave Duration*, PWD): Este *feature* hace referencia a la longitud del pulso en segundos.
2. Fase sistólica (*Systolic Phase*): Se conoce como fase sistólica a la duración temporal desde el comienzo de la curva de PPG hasta llegar al pico sistólico.



3. Fase diastólica (*Diastolic Phase*): La fase diastólica es el tiempo desde el pico sistólico hasta el final del pulso de PPG.
4. Amplitud sistólica (*Systolic Amplitude*): Es la amplitud que tiene el pulso de PPG cuando se da el pico sistólico.
5. Energía FFT: Corresponde a la energía de la transformada rápida de Fourier del pulso de PPG.
6. *Kurtosis* FFT: Es una característica basada en el cálculo de la *Kurtosis* de la transformada rápida de Fourier del pulso de PPG.
7. Entropía FFT: Es la entropía de la transformada rápida de Fourier del pulso de PPG.
8. Ratio SPDDP: Es la relación entre la fase sistólica y la fase diastólica.
9. Continua (DC) : Es el valor de la componente continua del pulso de PPG.
10. Edad: Es la edad del paciente a la hora de tomarle las curvas de PPG y ABP.
11. Sexo: Es el sexo del paciente, pudiendo ser masculino (M) o femenino (F).
12. Ratio BA: Es un *feature* calculado en base a la segunda derivada de la curva de PPG. Esta tiene varios puntos relevantes dentro de los cuales se ubican el punto “a” y el punto “b” como se muestra en la Figura 57. El Ratio BA se calcula como la relación entre estos dos puntos y es un *feature* que se encuentra estrechamente vinculado con la hemodinamia de las personas. Takazawa *et al.* demostraron que el Ratio BA refleja un aumento en la rigidez arterial y por ende aumenta con la edad de los pacientes [87]. Imanaga *et al.* lograron demostrar que el Ratio BA presenta una correlación directa con la distensión de las arterias perisféricas y por lo tanto su valor podría ser un indicador no invasivo de la arteriosclerosis [88]. Por último Otsuka *et al.* descubrieron la correlación positiva entre dicho *feature* y el *Framingham score*, utilizado para estimar el riesgo cardiovascular de los pacientes [89]. Teniendo todo lo anterior en cuenta, el Ratio BA fue utilizado como *feature* para la estimación de presiones arteriales en estudios como el de Mousavi *et al.* [90].

Se debe notar que el *feature* Ratio SPDDP, al no ser una combinación lineal de los *features* que involucra, podría aportar información adicional.

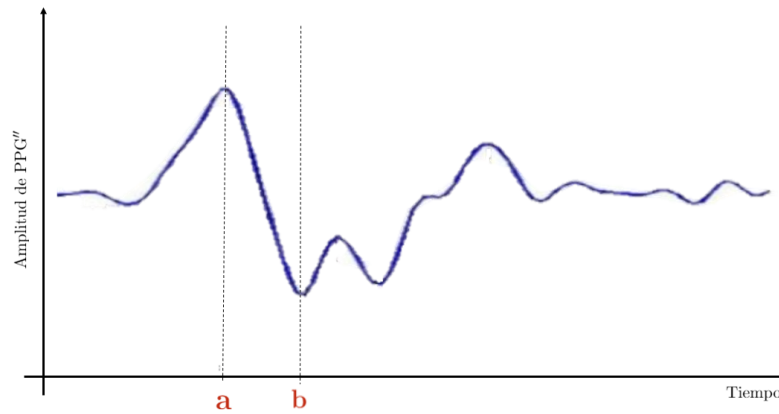


Figura 57: Segunda derivada de un pulso de PPG con los puntos de interés marcados en la misma. Fuente: [90].

#### 8.6.9. Zero Padding

Teniendo en cuenta que las redes CNN tienen como entrada los pulsos, es conveniente homogeneizar las longitudes de estos.

Para esto, se tomaron las señales y se les concatenó una cantidad de ceros  $N$  para que todas las señales queden de un largo de 300 muestras (*zero padding*). Se tomó este valor ya que representa el máximo de longitud de los pulsos obtenidos previamente de la base de datos. Se analizó utilizar un largo mayor pero la idea fue descartada debido a que esto aumentaría la dimensionalidad del problema, por lo tanto, resultando en un detrimento del rendimiento y desempeño de las ANNs.

#### 8.6.10. Normalización

Generalmente, un paso de vital importancia a realizar antes de introducir los datos en las ANNs es la normalización de los datos de entrada, ya sea en una red de tipo MLP o CNN. Este paso suele asegurar un mejor rendimiento por parte de la red ya que es frecuente que los valores relativos de los *features* varíen por más de un orden de magnitud. En estos casos, el aprendizaje de parámetros se enfrenta al problema de encontrarse mal condicionado, por lo que la función de costo tiende a ser más sensible a ciertos parámetros antes que a otros. Particularmente, esto afecta el rendimiento de los optimizadores de tipo *gradient descent* [36].

En lo que refiere a los datos de entrada de las redes MLP, se optó por aplicar una normalización de tipo estandarización *Z-Score*. Se obtuvieron la media ( $\mu$ ) y el desvío ( $\sigma$ ) de cada *feature* en el conjunto de entrenamiento y se normalizó cada valor según (Ecuación

34):

$$x_{norm} = \frac{x - \mu}{\sigma} \quad (34)$$

Los segmentos que ingresaron a las redes CNN también fueron estandarizados mediante el método *Z-Score*. En este caso, al tratarse de segmentos, se normalizó de a pulso: para cada uno se calculó la media y el desvío dentro del pulso y, luego, a cada muestra se le aplicó la estandarización según la Ecuación 34. Debe aclararse que, como se verá en la Sección 8.7.3, también se probó la alternativa de entrenar CNNs con pulsos de entrada sin normalizar.

## 8.7. Procesamiento

Esta etapa consiste principalmente en el diseño y entrenamiento de las redes neuronales artificiales. Como se mencionó anteriormente, se probaron dos tipos de redes neuronales artificiales diferentes: las MLPs y las CNNs unidimensionales. Esta etapa es de crítica importancia dado que la bibliografía existente no suele entrar en detalles específicos acerca de los componentes de las redes, proporcionando datos genéricos acerca de la arquitectura que no permiten la reproducibilidad de las ANNs. Es por esto que en esta instancia se intentó generar la mayor cantidad de información posible. Para esto, se analizaron los siguientes aspectos de las redes neuronales artificiales, donde corresponda:

- Cantidad de capas ocultas
- Tipo y cantidad de filtros
- Conjunto de datos de entrada:

Cantidad de *features* por vector de entrada (MLP)

Cantidad de pulsos en el conjunto de entrenamiento (CNN)

- Optimizadores
- Funciones de activación
- Funciones de costo

Independientemente de los distintos diseños de red, se decidió que la capa de salida tenga siempre dos neuronas: una para estimar la presión sistólica y otra para la diastólica. Si bien se consideró utilizar las ANNs para realizar una clasificación en grupos separando los pacientes en hipotensos, normotensos e hipertensos y comparar el grado de acuerdo mediante el método de Bland-Altman, se decidió que la salida sea un valor de presión y así poder estimar el valor de BP ya que posee mayor relevancia clínica [91].

### 8.7.1. Regresión Lineal

Antes de adentrarse en el estudio de las ANNs, se analizó el rendimiento de una regresión lineal. Esto se realizó para tener un dimensionamiento de la complejidad del problema a tratar y justificar el uso de herramientas más complejas como el *Deep Learning*.

Para esto se utilizó la librería *sklearn*. La misma permite importar un modelo de regresión lineal y métricas para la interpretación del mismo. Se analizaron el Error Cuadrático Medio usando la función *Mean Squared Error* y el MAE.

En este caso, se intentó ajustar el modelo de regresión lineal con 160 pulsos de un paciente y realizar la validación con 40 pulsos del mismo paciente. En otras palabras, se utilizó un 20 % del conjunto de datos para realizar la validación. Al ser un proceso con fines demostrativos, se estudiaron únicamente dos pacientes seleccionados aleatoriamente. Esto implicó que, para cada paciente, se entrenó un modelo con *features* de cada uno de los 160 pulsos. Adicionalmente, se aclara que cada *feature* de entrada a la regresión lineal constó de una muestra de un pulso de PPG.

Como se evidencia en la Sección 9.3, los resultados obtenidos no son satisfactorios, por lo cual se justifica el uso de las herramientas subsiguientes. El resultado de MAE obtenido será utilizado como valor referencia, donde el objetivo es adquirir mejores desempeños mediante el uso de ANNs.

### 8.7.2. MLP

En primer lugar, se estudió el rendimiento de las redes MLP. Mediante un proceso iterativo respaldado en la bibliografía consultada (Sección 7), se logró obtener una arquitectura genérica en cuanto a cantidad de capas, cantidad de neuronas por capa, funciones de activación y utilización de *Batch Normalization* y *Dropout*. Esta misma estructura general de las MLPs diseñadas puede visualizarse en la Figura 58. En todas las redes estudiadas se mantuvo la misma estructura donde lo que se varió fueron la función de costo y el optimizador. Dado que se trabajó con 12 features (mencionados en la Sección 8.6.8), los datos de entrada al momento del entrenamiento tenían un formato de  $43908 \times 12$ . La salida de la red consistió en un vector de  $1 \times 2$  correspondiente a los valores predichos de presión sistólica y presión diastólica.

Como se desarrolla en profundidad en la Sección 8.8.1, se utilizó un 15 % de los datos para la validación y otro 15 % para la etapa de *test*. Las Figuras 59, 60 y 61 muestran las distribuciones de la SBP y la DBP en los conjuntos de entrenamiento, validación y *test*, respectivamente. Se puede observar que los conjuntos se encuentran balanceados, considerando que las distribuciones se mantienen similares entre los diferentes *sets*. A continuación, en la Tabla 12 se presenta la media y el desvío estándar para cada conjunto de datos en SBP y DBP.

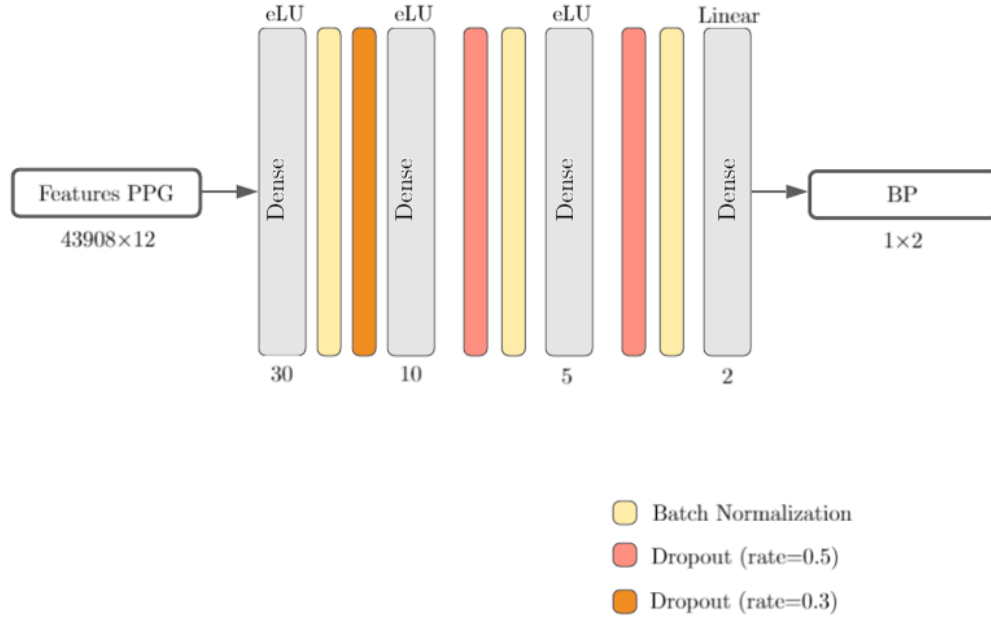


Figura 58: Esquema de la red MLP utilizada donde las capas en amarillo son de *Batch Normalization* y las de color naranja y rojo de *Dropout*.

	SBP		DBP	
	$\mu_e$ (mmHg)	$\sigma_e$ (mmHg)	$\mu_e$ (mmHg)	$\sigma_e$ (mmHg)
Entrenamiento	128.178	23.429	65.006	12.496
Validación	128.542	23.454	65.452	12.622
Verificación	128.098	23.424	65.083	12.467

Tabla 12: Media y desvío de los conjuntos de datos utilizados para entrenamiento, validación y verificación en MLP para SBP y DBP

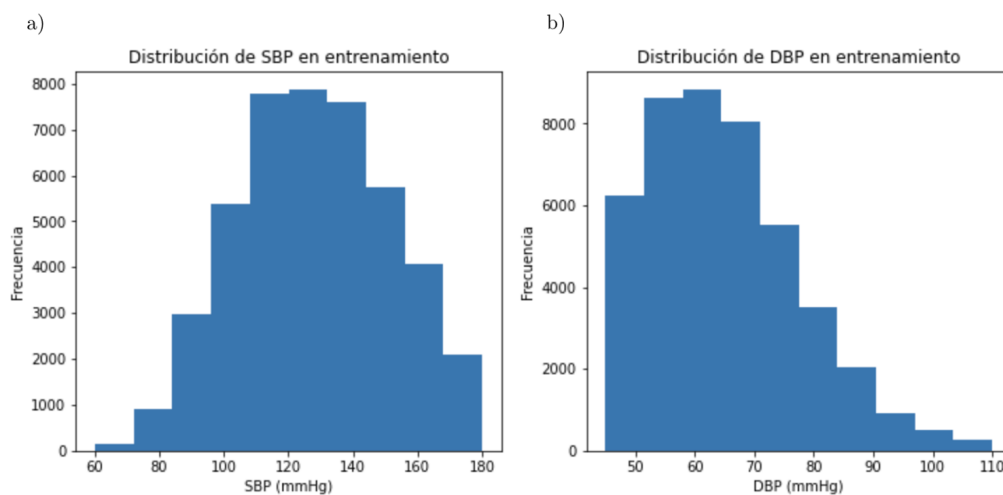


Figura 59: Distribución de (a) SBP y (b) DBP en el conjunto de entrenamiento para las MLPs.

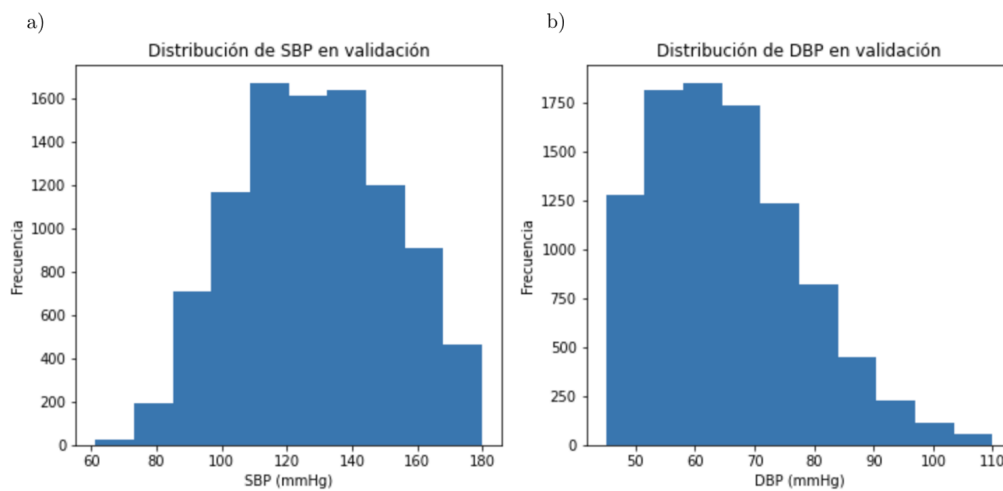


Figura 60: Distribución de (a) SBP y (b) DBP en el conjunto de validación para las MLPs.

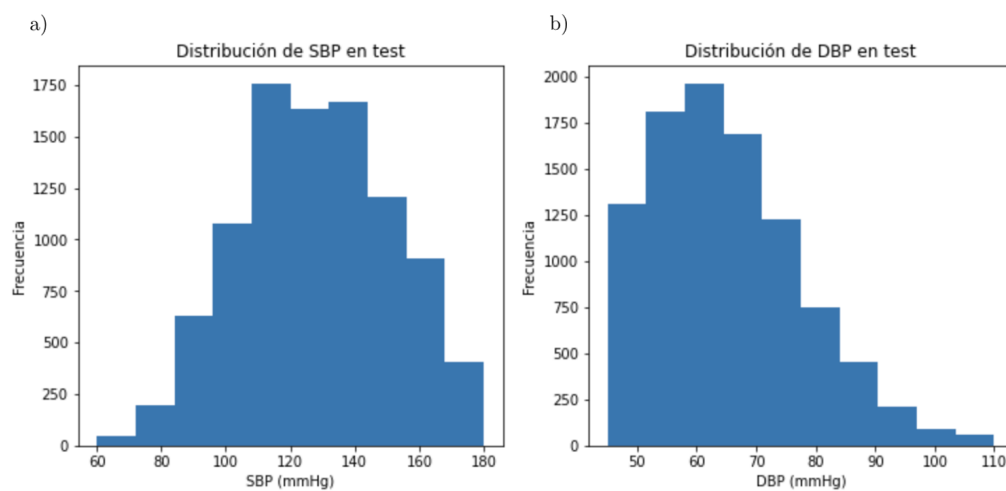


Figura 61: Distribución de (a) SBP y (b) DBP en el conjunto de *test* para las MLPs.



Cabe destacar que se utilizó el método *EarlyStopping* de la librería Keras, donde se definió monitorear el valor de *loss* en validación con una paciencia de 200 *epochs*. Además, se utilizó un *learning rate* variable, para lo cual se seleccionó el método implementado en *Keras.callbacks* llamado *ReduceLROnPlateau* con un valor inicial de  $10^{-3}$  y un valor mínimo de  $10^{-7}$ , con una paciencia de 5 *epochs* monitoreando la *loss* en validación. El factor seleccionado para reducir el *learning rate* fue de 0.1, por lo que cada vez que se necesitó disminuirlo se multiplicó al mismo por dicho factor.

Las funciones de costo analizadas en esta instancia se listan a continuación:

- MAE
- MSE
- MSLE

Adicionalmente, los optimizadores puestos a prueba se detallan en la siguiente lista:

- Adam
- RMSprop
- SGD

Una vez obtenida la mejor combinación, se procedió a realizar la prueba sobre el conjunto de *test* para dicha arquitectura. Además, como se mencionó en la Sección 8.8.1, se realizó un análisis de selección de *features* teniendo en cuenta a la SBP y a la DBP por separado, a modo de determinar aquellos *features* con mayor relevancia a la hora de estimar la BP. Una vez obtenido este análisis, se contrastó el resultado en *test* de entrenar a la red con los 12 *features* con el obtenido entrenándola con aquellos de mayor importancia.

### 8.7.3. CNN

Como ya se mencionó, las redes CNN suelen estar conformadas, en primer lugar, por capas que funcionan como filtros (capas convolucionales y capas *pooling*) seguidas de capas densas o *fully connected*.

La determinación de parámetros para las capas de filtro depende de los datos de entrada y, por eso, suele ser específica para cada proyecto. Si bien es un proceso iterativo, se tomaron en cuenta algunos criterios fundamentales. En primer lugar, se tuvo en cuenta que la cantidad de parámetros utilizados no sea excesivamente alta para la cantidad de datos de entrada para, de esta forma, evitar el *overfitting*. Asimismo, tampoco debe ser muy reducida para garantizar un buen rendimiento de la red. Por lo tanto, se debe asegurar que el nivel de parámetros sea adecuado en función de los datos de entrada, logrando la convergencia y un entrenamiento correcto.

Los parámetros de las capas densas fueron elegidos en base a derivaciones realizadas de la arquitectura genérica de las redes MLP. Por ejemplo, para ambas configuraciones de CNN armadas (Red V1 y Red V2), se utilizó la misma cantidad de capas densas que en MLP. Además, para la Red V1 se utilizó la función de activación ELU en todas las capas (menos en la de salida), además de utilizar el mismo esquema de *Dropout*.

Las redes fueron entrenadas con distintos objetivos:

- **Selección de Especificaciones de las CNNs:** Se analizó el rendimiento de una arquitectura de red neuronal entrenada, validada y verificada con datos de un solo paciente, realizando este proceso para 50 pacientes. Esto implica que la red sería capaz de estimar la presión arterial de un sujeto si y sólo si es calibrada previamente con datos de ese mismo paciente. Además, se variaron algunos parámetros de entrenamiento.
- **Evaluación de Calidad Médica en CNNs:** Se seleccionó la arquitectura y configuración de entrenamiento con mejor desempeño dentro de las evaluadas en el **Selección de Especificaciones de las CNNs** y se buscó demostrar la calidad médica según la AAMI-ISO, realizando el entrenamiento de la red a medida de cada paciente, para 150 pacientes.

#### 8.7.3.1. Selección de Especificaciones de las CNNs

En vista de alcanzar el objetivo de la calidad médica no alcanzado utilizando MLPs (ver Secciones 9.4 y 10.3), se optó por evaluar el desempeño de un enfoque particular utilizando CNNs para luego, en un futuro, progresar hacia uno general. Es por esto que se

partió por estudiar el rendimiento de CNNs calibradas para pacientes de forma individual. Esto se realizó tomando  $N$  pulsos de igual longitud (debido al *zero padding*) para un mismo paciente y entrenando con estos datos para luego validarlos con otro grupo de datos del mismo individuo, para un total de 50 pacientes. Los valores de  $N$  estudiados fueron de 120, 200 y 400. La elección de estos valores tuvo en cuenta el eventual tiempo de calibración a emplearse si se deseara utilizar en un paciente, resultando en la conclusión de que no puede ser un valor de excesiva magnitud. Dado que se utilizó el 10 % de los datos para la validación en un proceso de *Cross Validation*, considerando que la media obtenida de la duración de los pulsos de PPG es de 96 muestras y la frecuencia de muestreo es 125 Hz, esto resultaría en tiempos de calibración de 1.434 s para el  $N$  de 120, 2.304 s para el  $N$  de 200 y 4.608 s para el  $N$  de 400. Adicionalmente, se consideró que la cantidad de datos de entrada debe ser lo suficientemente grande para que la ANN pueda ser entrenada de forma adecuada. Como se mencionó en la Sección de Preprocesamiento (Sección 8.6.10), se optó por estudiar los resultados obtenidos a partir del entrenamiento de las redes con y sin normalizar.

Las Figuras 62, 63 y 64 muestran las distribuciones de la SBP y la DBP para valores de  $N$  de 120, 200 y 400, respectivamente. Se puede observar que, al variar  $N$ , las distribuciones se mantienen similares. Se debe notar que no se presentan las distribuciones de los conjuntos de entrenamiento y validación por separado ya que, siendo que se realizó *Cross Validation*, no aportaría información relevante. En la Tabla 13 se encuentra la media y el desvío estándar para los conjuntos usados en los valores de  $N$  de 120, 200 y 400.

Se utilizaron dos arquitecturas diferentes a fin de poder realizar una comparación en cuanto a su desempeño y su capacidad para alcanzar la calidad médica buscada. A fines prácticos, se denominarán Red V1 y Red V2. Ambas arquitecturas presentan aspectos en común: se utilizó la función de costo MSE dado su buen rendimiento en modelos de regresión y los antecedentes mencionados en Estado del Arte (Sección 7), el optimizador de elección fue Adam con un *learning rate* variable y la métrica monitoreada fue el MAE dado que es de utilidad para el problema en cuestión.

En lo que refiere al *learning rate*, se utilizó el implementado en *keras.callbacks* llamado *ReduceLROnPlateau*, reduciendo a la mitad el *learning rate* cada vez que detecta por más de 20 *epochs* que la función de costo en validación se encuentra en un mínimo local. Se debe mencionar que el valor inicial de *learning rate* fue de  $10^{-3}$  y que el mínimo permitido fue de  $10^{-7}$ . Adicionalmente, se aplicó *EarlyStopping* de la misma librería con una paciencia de 150 *epochs*, también monitoreando el valor de la función de costo o *loss* en validación.

N	SBP		DBP	
	$\mu_e$ (mmHg)	$\sigma_e$ (mmHg)	$\mu_e$ (mmHg)	$\sigma_e$ (mmHg)
120	132.090	28.123	63.264	14.608
200	132.068	28.022	62.853	14.278
400	130.744	27.718	62.269	13.733

Tabla 13: Media y desvío de los conjuntos de datos utilizados para los N de 120, 200 y 400 en la Selección de Especificaciones de las CNNs para SBP y DBP

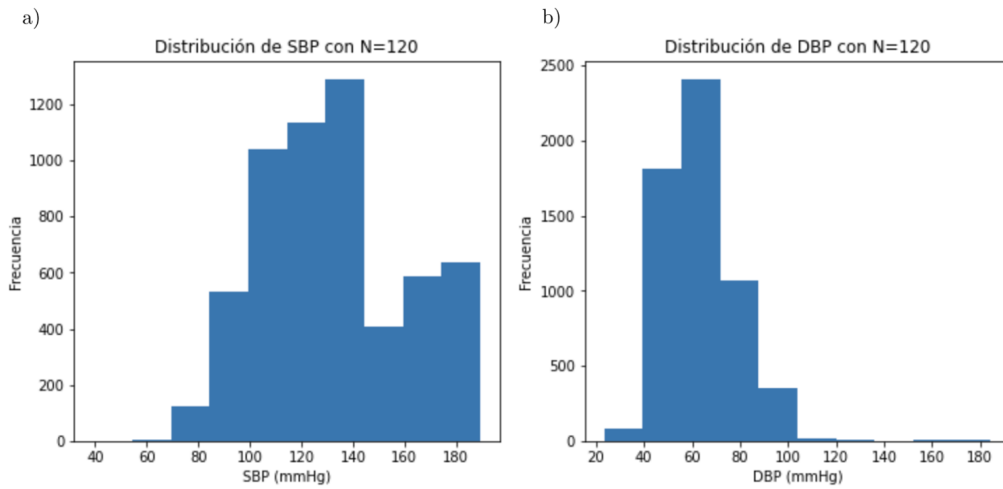


Figura 62: Distribución de (a) SBP y (b) DBP para N = 120 en la Selección de Especificaciones de las CNNs.

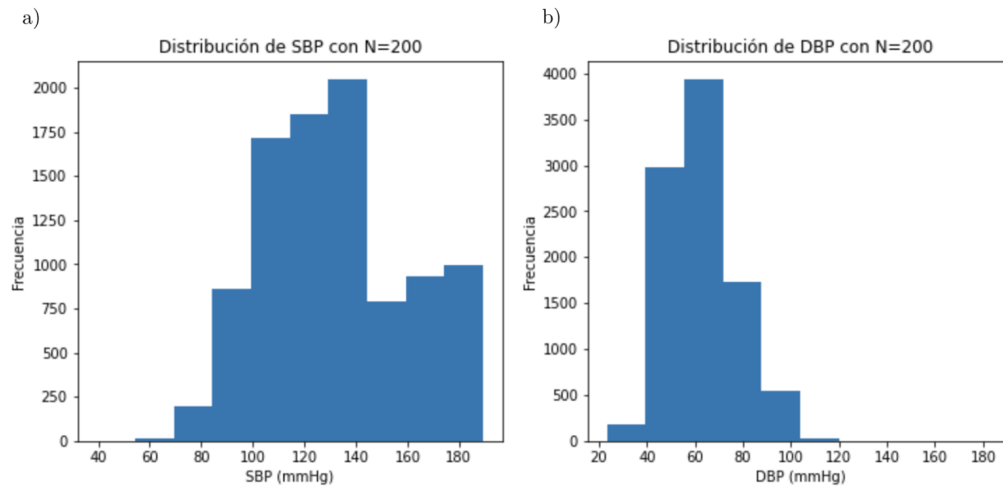


Figura 63: Distribución de (a) SBP y (b) DBP para N = 200 en la Selección de Especificaciones de las CNNs.

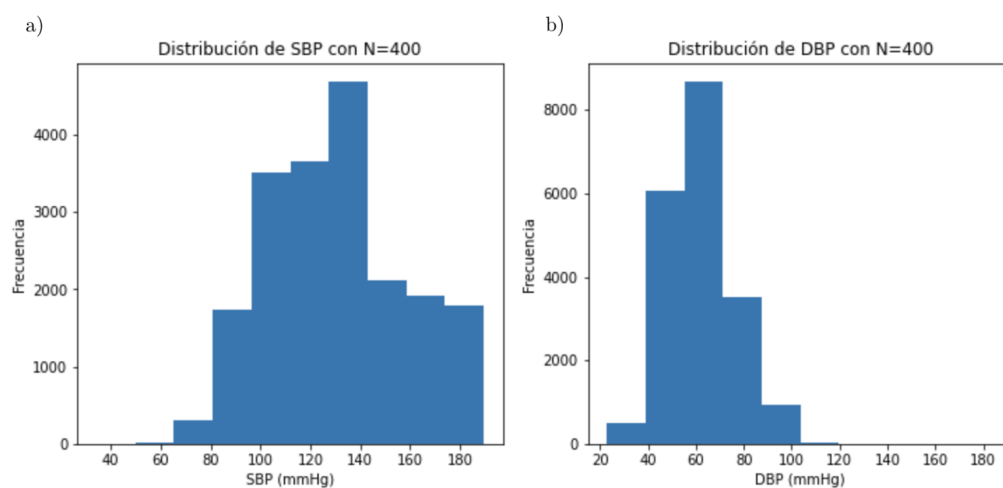


Figura 64: Distribución de (a) SBP y (b) DBP para  $N = 400$  en la Selección de Especificaciones de las CNNs.

### 8.7.3.1.1. Red V1

La Red V1 utilizada es una CNN-1D de 8 capas ocultas cuyos parámetros se pueden encontrar en la Figura 65.

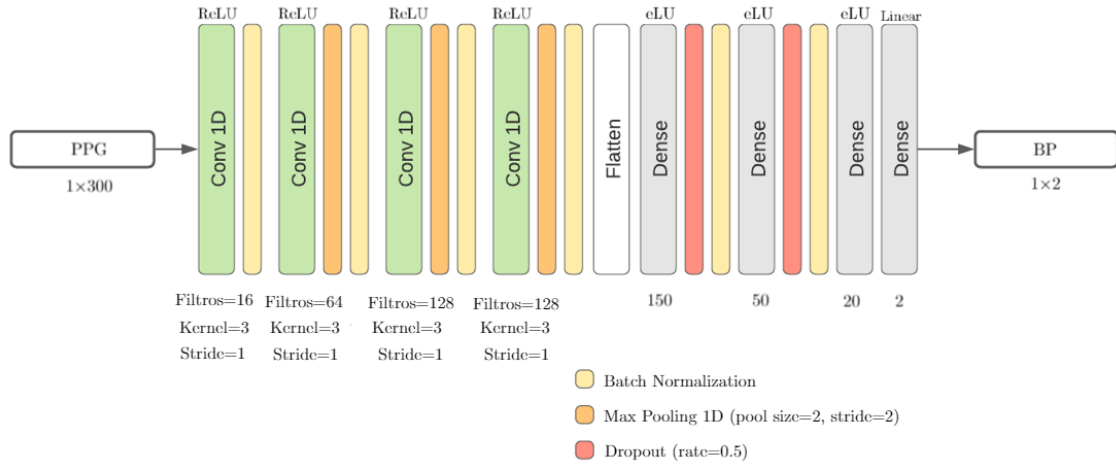


Figura 65: Esquema de la CNN Red V1 utilizada donde las capas en amarillo son de *Batch Normalization*, las de color naranja de *Max Pooling* y las rojas de *Dropout*.

La arquitectura de la ANN fue inspirada en la de la VGG-16<sup>4</sup>, en el sentido de que la cantidad de filtros en las capas convolucionales se van duplicando respecto de la anterior [92]. En el caso particular de las últimas dos, se mantiene la cantidad de filtros.

### 8.7.3.1.2. Red V2

La Red V2 es una modificación de la Red V1, donde se quitó el *Dropout* y el *Batch Normalization* en las capas densas. Adicionalmente, se incluyó 'ReLU' como la función de activación en las capas densas en donde antes se utilizaba 'eLU'.

<sup>4</sup>La VGG-16 es un modelo de CNN propuesto por K. Simonyan y A. Zisserman en la publicación de la Universidad de Oxford "Very Deep Convolutional Networks for Large-Scale Image Recognition" [93]. Fue uno de los famosos modelos submitidos a la competencia *Large Scale Visual Recognition Challenge* en 2014.

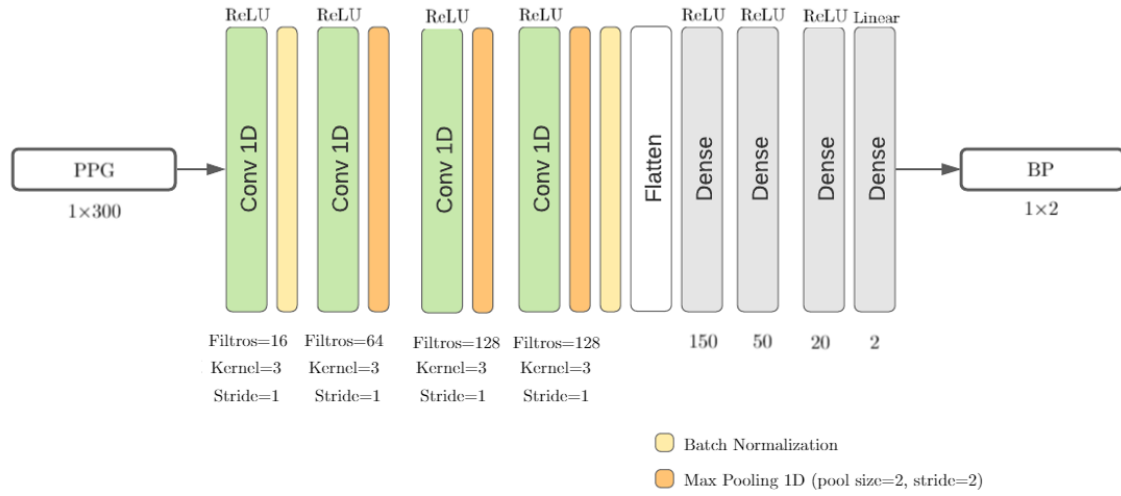


Figura 66: Esquema de la CNN Red V2 utilizada donde las capas en amarillo son capas de *Batch Normalization* y las de color naranja de *Max Pooling*

#### 8.7.3.2. Evaluación de Calidad Médica en CNNs

El objetivo principal de la Evaluación de Calidad Médica en CNNs es el de poder verificar el cumplimiento con los requisitos de la AAMI-ISO para la calidad médica (Sección 6.2), analizando el mejor modelo obtenido a partir de la Selección de Especificaciones de las CNNs (Sección 8.7.3.1) en cuanto a desempeño. Los resultados obtenidos para la Selección de Especificaciones de las CNNs pueden verse en la Sección 9.5.1, donde la red seleccionada fue la Red V2 normalizando con 400 pulsos, de ahora en más, denominada ATT-Net por ser la abreviación del término *Artificial Tailored Tensiometer*. La justificación puede encontrarse en la Sección 10.4.

A fin de cumplir con el objetivo planteado, se seleccionaron 150 pacientes (diferentes a los de la Selección de Especificaciones de las CNNs) de forma aleatoria y se entrenó, para cada uno de ellos, una ANN de acuerdo al modelo seleccionado por su rendimiento (ATT-Net). En este caso, no se aplicó el método de Validación Cruzada, sino que se obtuvo un único modelo entrenado para cada individuo. Se debe hacer notar que se utilizó un 10 % del conjunto de pulsos de cada paciente para el proceso de validación. Debido a que se utilizaron 400 pulsos en total, el conjunto de validación se compone de 40 pulsos por pacientes, los cuales son consecutivos a los usados en entrenamiento. A continuación, en la Figura 68 se presenta la distribución de SBP y DBP en este conjunto y en la Figura 67 del conjunto de entrenamiento. En la Tabla 14 se encuentra la media y desvío estándar de los conjuntos usados para entrenamiento y validación. Se puede observar que, al presentar distribuciones similares, los conjuntos de datos se encuentran balanceados. Adicionalmente, se destaca el parecido entre las morfologías de estas distribuciones y aquellas correspondientes a la Selección de Especificaciones de las CNNs (ver Sección

8.7.3.1).

	SBP		DBP	
	$\mu_e$	$\sigma_e$	$\mu_e$	$\sigma_e$
	(mmHg)	(mmHg)	(mmHg)	(mmHg)
Entrenamiento	127.516	26.527	65.65048	15.63601
Validación	128.375	25.669	65.035	13.046

Tabla 14: Media y desvío de los conjuntos de datos utilizados para entrenamiento y validación para la Evaluación de Calidad Médica en CNNs para SBP y DBP

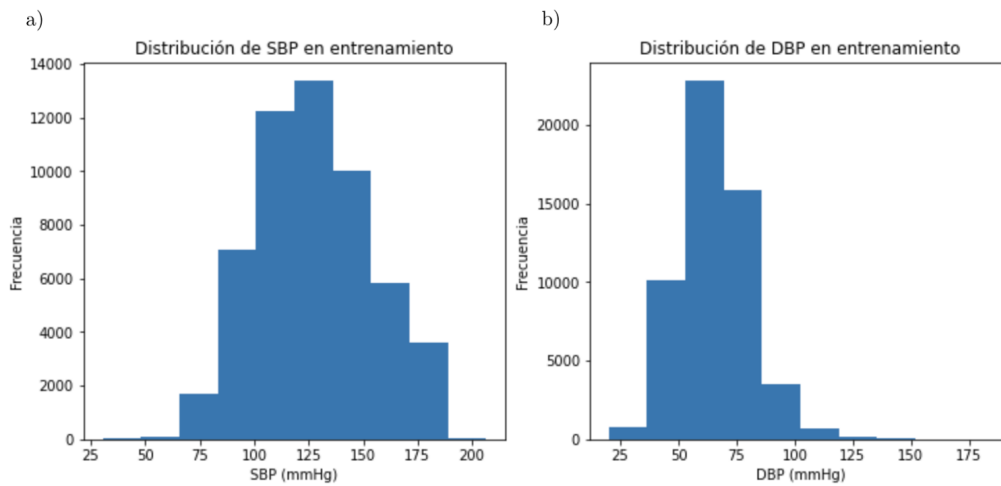


Figura 67: Distribución de la (a) SBP y (b) DBP en el conjunto utilizado de entrenamiento para la Evaluación de Calidad Médica en CNNs.

La media en SBP es de 127.516 mmHg y el desvío estándar 26.525 mmHg. En DBP, la media es de 65.650 mmHg y el desvío estándar de 15.635 mmHg. La población utilizada es de 150 pacientes mayores a 12 años tal como lo exige la AAMI-ISO.

Debido a lo propuesto por la AAMI-ISO, se tomaron 3 valores por paciente, consecutivos y al azar, para obtener las metricas requeridas. Además, se debieron excluir 11 pacientes debido a una diferencia contra las mediciones de referencia mayor a 12 mmHg en SBP y a 8 mmHg en DBP en cumplimiento por lo requerido en el estándar AAMI-ISO, tal como se comentó en la Sección 6.2. La distribución final de los datos de referencia se presenta en los histogramas de la Figura 69 para este conjunto en SBP y DBP.

Como se observa en la Figura 69 que muestra la distribución en los conjuntos finales, la media es de 128.375 mmHg y el desvío estándar 25.638 mmHg en SBP y de 65.035 mmHg y 13.031 mmHg, respectivamente, en DBP.



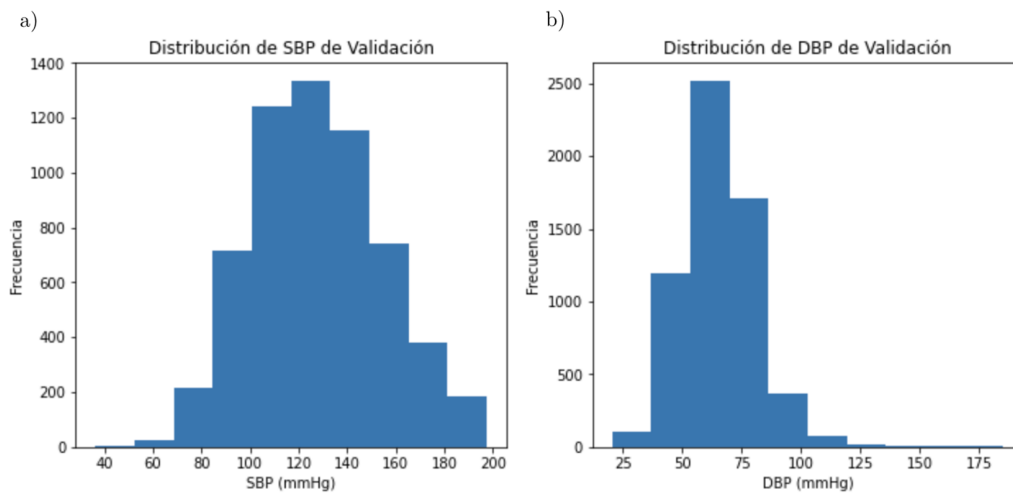


Figura 68: Distribución de la (a) SBP y (b) DBP en el conjunto utilizado de validación para la Evaluación de Calidad Médica en CNNs.

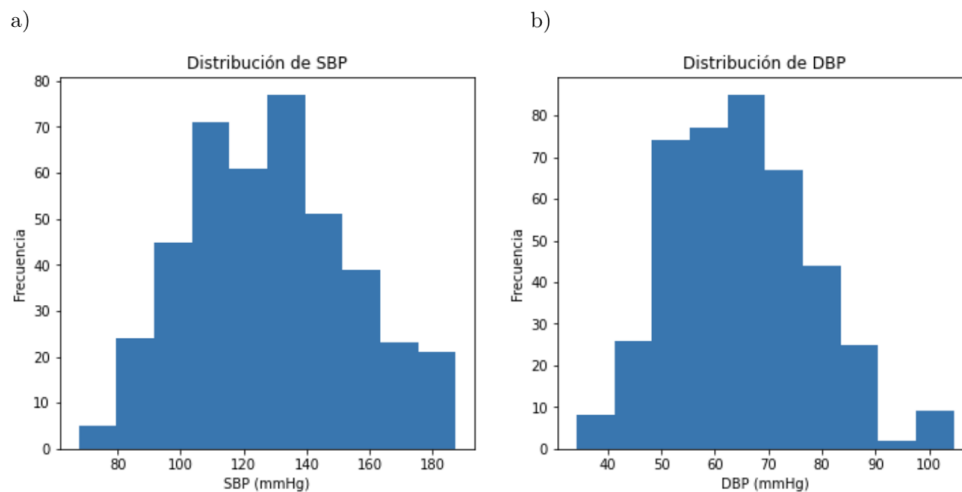


Figura 69: Distribución de la (a) SBP y (b) DBP en el conjunto utilizado de validación luego de la aplicación de los criterios propuestos por la AAMI-ISO.

## 8.8. Evaluación de Modelos

El objetivo de esta sección es dilucidar el procedimiento de evaluación llevado a cabo en el presente proyecto de investigación. De esta forma, se busca aportar un panorama del flujo de trabajo implementado. A continuación se desarrollan los diferentes procedimientos empleados para el caso de las redes MLP y CNN por separado.

### 8.8.1. MLP

En el análisis de las MLPs se decidió evaluar las diferencias en el desempeño presentadas utilizando diferentes combinaciones de funciones de costo (Sección 5.4.3) y optimizadores (Sección 5.4.6). Para esto, se determinó la cantidad de capas y su correspondiente cantidad de neuronas, así como también la adición de *Batch Normalization* y de regularización mediante el uso de *Dropout*, a través de un proceso iterativo apoyado en la bibliografía consultada (ver Sección 7). Dicho proceso resultó en la determinación de un esquema genérico de la arquitectura de la red a utilizar (ver Sección 8.7.2), donde luego se puso foco en encontrar la combinación óptima de función de costo y optimizador.

Se decidió trabajar con 25 pulsos por paciente ya que se intentó replicar la capacidad de predicción de BP de las MLPs reportadas (Sección 7) sin necesidad de calibración previa. Esto implica que se utilizaron 25 pulsos de cada uno de los 2509 pacientes resultantes (ver Sección 8.5), obteniéndose 12 *features* por cada pulso (Sección 8.6.8) a partir del preprocesamiento (Sección 8.6). Esto resultó en un arreglo de datos de 62725x12, de los cuales el 15 % se utilizaron para *test* y otro 15 % como conjunto de validación. Se debe hacer mención, sin embargo, de que la partición contempló que no hubiera información dividida de los pacientes entre los *sets* de entrenamiento, validación y *test*. Para cada ANN entrenada se obtuvo el MAE para el conjunto de validación. Una vez obtenidos los resultados, se seleccionó la red con el mejor desempeño considerando el valor de MAE en validación y se analizó el error en el conjunto de *test*, obteniéndose el error de predicción (Ecuación 24), su MAE (Ecuación 26) y desvío estándar (Ecuación 27) tanto para la presión sistólica como para la diastólica.

Adicionalmente, se realizó un análisis de selección de *features* utilizando la función *feature\_selection* de la librería *sklearn*. Dicha función básicamente calcula el valor F o *F-score* entre el valor de BP y los *features* y, a mayor *F-score*, mayor la influencia del *feature* sobre la BP. Esto se aplicó por separado para la SBP y para la DBP, obteniéndose resultados de relevancia de cada *feature* para cada tipo de BP. Una vez obtenidos, se comparó el resultado en *test* de entrenar a la red con los 12 *features* o con aquellos de mayor importancia.

### 8.8.2. CNN

En todo lo referido a las CNNs, el preprocesamiento presenta leves disimilitudes con el necesario para las redes MLP, las cuales son destacadas a lo largo de la Sección 8.6. Adicionalmente, el procesamiento fue dividido en dos etapas, la Selección de Especificaciones de las CNNs (Sección 8.7.3.1) y la Evaluación de Calidad Médica en CNNs (Sección 8.7.3.2). Es necesario mencionar que la implementación de este tipo de ANNs surgió como alternativa luego de que los resultados obtenidos con las MLPs no solo no cumplieron con la hipótesis planteada, sino que además demostraron no poder reproducir los resultados

publicados en otras investigaciones (referirse a la Sección 9.4 y Sección 10.3 para mayor detalle). Por estas razones, se optó por explorar el desempeño de las CNNs partiendo de un enfoque individual, es decir que las ANNs entrenadas en esta etapa fueron particulares para cada sujeto evaluado, implicando entonces una necesidad de calibración previa a la estimación de la BP.

En cuanto la Selección de Especificaciones de las CNNs, la misma consistió en analizar los resultados obtenidos para dos redes, denominadas Red V1 (Sección 8.7.3.1.1) y Red V2 (Sección 8.7.3.1.2), variando la cantidad de pulsos con la que se trabajó para 50 pacientes diferentes. Esta cantidad se varió en 120, 200 y 400. Los detalles de este procesamiento se encuentran en la Sección 8.7.3.1. Se decidió tomar 50 pacientes en vez de utilizar los 2509 disponibles para optimizar el tiempo de cómputo.

Para la Selección de Especificaciones de las CNNs, se decidió estudiar el desempeño de las ANNs utilizando Validación Cruzada (*Cross Validation*). Este método implica la división del conjunto de datos en  $k$  segmentos de igual dimensión, de los cuales uno será utilizado como conjunto de *test* y los restantes  $k-1$  segmentos conforman el conjunto de *train*. Esto se repite  $k$  veces, donde se alterna el subconjunto de *test* y, por lo tanto, el de *train* [36].

Las ventajas de esta técnica radican en su efectividad para medir el desempeño del modelo considerando que los resultados son independientes de la partición del conjunto de datos. En otras palabras, es un método que permite analizar la robustez del modelo entrenado. Sin embargo, se debe destacar que la desventaja principal de aplicarlo es la eficiencia por el gasto de cómputo, sobretudo en los casos en los que se trabaja con conjuntos de datos extensos. Es por esto que los valores típicos de  $k$  son 5 y 10 [36].

Se optó por utilizar un valor de  $k$  igual a 10 a fin de analizar la robustez de los modelos implementados, teniendo en cuenta las desventajas mencionadas previamente. Esto implicó que, en cada instancia, el 10 % de los datos se utilizaron como conjunto de *validation*, como se ve ilustrado en la Figura 71. Luego, para cada iteración de la Validación Cruzada se calcularon el error de predicción (Ecuación 24), su MAE (Ecuación 26) y desvío estándar (Ecuación 27) tanto para la presión sistólica como para la diastólica. Entonces, esto resulta en 10 valores de MAE y  $\sigma_e$  los cuales fueron resumidos obteniendo sus respectivas medianas ( $M_{MAE}$  y  $M_{\sigma}$ ) para no asumir una distribución normal. Este proceso se realizó para 50 pacientes, por lo que se optó por obtener la mediana de estas últimas. Se consideraron otras medidas de tendencia central, como la media, la cual se desestimó ya que no se puede asumir que se cumplan los supuestos de distribución normal en un conjunto de 10 datos por paciente. En la Figura 70 se puede observar este procedimiento para el paciente **n** para la SBP. El procedimiento es análogo para DBP.

Luego, se puede obtener una medida de tendencia central para todos los pacientes, para lo cual se utilizó la mediana. Esto implica que para la SBP, de las 50 medianas obtenidas previamente para el MAE, se obtuvo una mediana final  $M_{MAE}$  y lo mismo se realizó

SBP	
MAE (mmHg)	$\sigma$ (mmHg)
$MAE_{s0}$	$\sigma_{s0}$
$MAE_{s1}$	$\sigma_{s1}$
$MAE_{s2}$	$\sigma_{s2}$
$MAE_{s3}$	$\sigma_{s3}$
$MAE_{s4}$	$\sigma_{s4}$
$MAE_{s5}$	$\sigma_{s5}$
$MAE_{s6}$	$\sigma_{s6}$
$MAE_{s7}$	$\sigma_{s7}$
$MAE_{s8}$	$\sigma_{s8}$
$MAE_{s9}$	$\sigma_{s9}$
<div style="display: flex; justify-content: space-around; align-items: center;"> <div style="text-align: center;"> <b>Paciente n</b>  <math>\left. \begin{matrix} MAE_{s0} \\ MAE_{s1} \\ MAE_{s2} \\ MAE_{s3} \\ MAE_{s4} \\ MAE_{s5} \\ MAE_{s6} \\ MAE_{s7} \\ MAE_{s8} \\ MAE_{s9} \end{matrix} \right\} M_{nMAEs}</math> </div> <div style="text-align: center;"> <math>\left. \begin{matrix} \sigma_{s0} \\ \sigma_{s1} \\ \sigma_{s2} \\ \sigma_{s3} \\ \sigma_{s4} \\ \sigma_{s5} \\ \sigma_{s6} \\ \sigma_{s7} \\ \sigma_{s8} \\ \sigma_{s9} \end{matrix} \right\} M_{n\sigma s}</math> </div> </div>	

Figura 70: Esquema de la obtención de métricas por paciente. Luego de *Cross Validation*, se utiliza la mediana de las 10 MAEs y 10 desvíos estándar obtenidos para conseguir  $M_{nMAEs}$  y  $M_{n\sigma s}$  respectivamente.

para las medianas de los desvíos, consiguiendo su mediana final  $M_{\sigma}$ . El procedimiento análogo fue llevado a cabo para DBP.

La Evaluación de Calidad Médica en CNNs consistió en, habiendo seleccionado la arquitectura y configuración de entrenamiento con mayor desempeño al final de la Selección de Especificaciones de las CNNs, la verificación del cumplimiento con la calidad médica (Sección 6.2) en 150 pacientes aleatoriamente seleccionados, distintos a aquellos utilizados en la Selección de Especificaciones de las CNNs. A diferencia de la Selección de Especificaciones de las CNNs, no se aplicó el método de Validación Cruzada, sino que se obtuvo un único modelo entrenado para cada paciente. Adicionalmente, es pertinente mencionar que se utilizó el 10% del conjunto de pulsos de cada paciente para el proceso de validación.

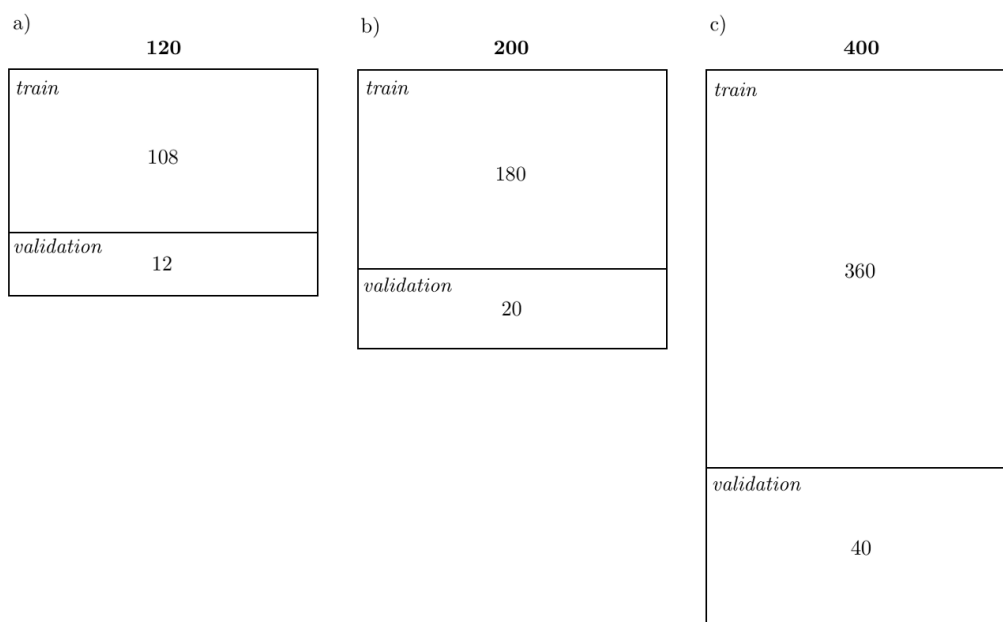


Figura 71: Esquema de las particiones realizadas para las CNNs en a)  $N = 120$ , b)  $N = 200$  y c)  $N = 400$ , con  $N$  siendo la cantidad de pulsos de PPG.

## 9. Resultados

### 9.1. Preprocesamiento

### 9.2. Descarte de Señal por Peso de Archivo

En esta etapa, se descartaron 113 pacientes resultando en una cantidad de 2899 pacientes.

#### 9.2.1. Sustitución de valores NaN

	Señales PPG	Señales ABP
Porcentaje de muestras NaN (%)	0.069	0.294

Tabla 15: Resultados de porcentaje de muestras NaN para las señales PPG y señales ABP.

### 9.2.2. Clasificadores de Segmentos de Señal

En esta etapa, se descartaron 537 pacientes resultando en una cantidad de 2362 pacientes (PATIENT\_IDS).

En la Tabla 16 se muestran los clasificadores para segmentos de señal de PPG que presentan el mejor desempeño.

Clasificadores de Señales PPG									
Nombre de Archivo	N	AUC ROC	FP	FN	Sensibilidad	Especificidad	Precisión	MCC	F1 Score
Clasificador PPG 1	125	0.67	12	8	0.61	0.55	0.58	0.17	0.79
Clasificador PPG 1	250	0.81	11	6	0.81	0.71	0.77	0.55	0.78
Clasificador PPG 2	250	0.81	6	3	0.75	0.79	0.88	0.76	0.88

Tabla 16: Métricas de desempeño del *Clasificador PPG 1* y del *Clasificador PPG 2*.

Respecto a los clasificadores implementados para las señales de ABP, la Tabla 17 muestra los resultados obtenidos para el *Clasificador ABP 1*.

Clasificadores de Señales ABP									
Nombre de Archivo	N	AUC ROC	FP	FN	Sensibilidad	Especificidad	Precisión	MCC	F1 Score
Clasificador ABP 1	125	0.81	4	4	0.86	0.75	0.79	0.57	0.82
Clasificador ABP 1	250	0.81	7	4	0.91	0.75	0.85	0.70	0.88

Tabla 17: Métricas de desempeño del *Clasificador ABP 1* frente a su entrenamiento con diferente cantidad de segmentos de señal (125 y 250).

### 9.3. Regresión Lineal

La Tabla 18 a continuación muestra los resultados obtenidos de MSE y MAE para los pacientes evaluados en la regresión lineal.

Nº Paciente	MSE (mmHg <sup>2</sup> )	MAE (mmHg)	$r^2$
1	29297.600	30.110	0.000
2	36073.610	31.220	0.000

Tabla 18: MSE, MAE y  $r^2$  obtenidos para los pacientes a partir de la Regresión Lineal.

## 9.4. MLP

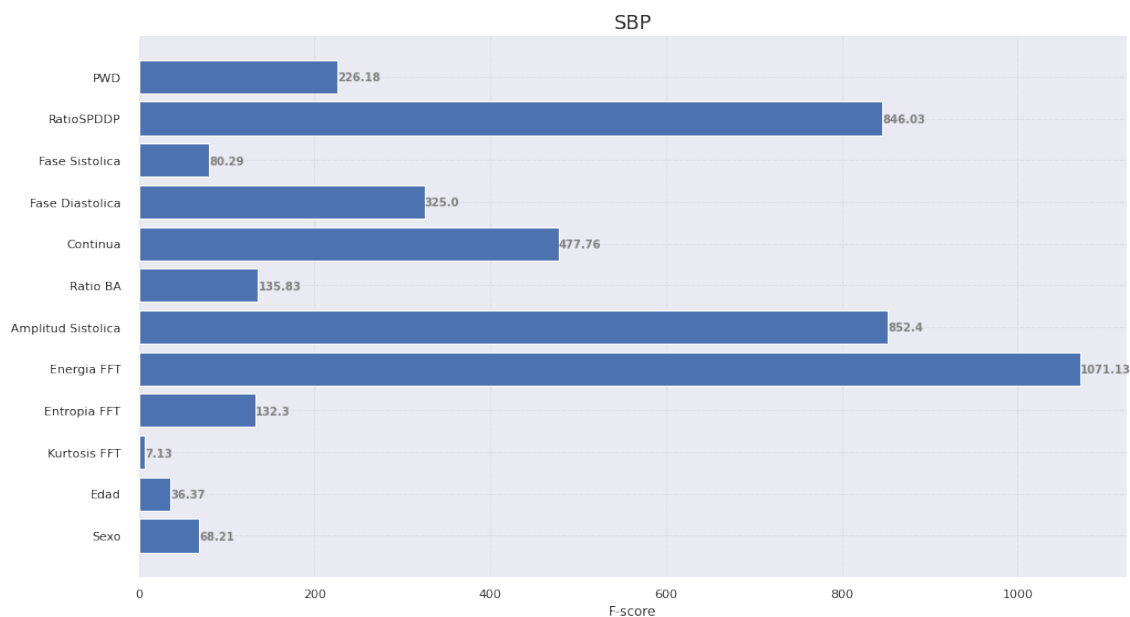
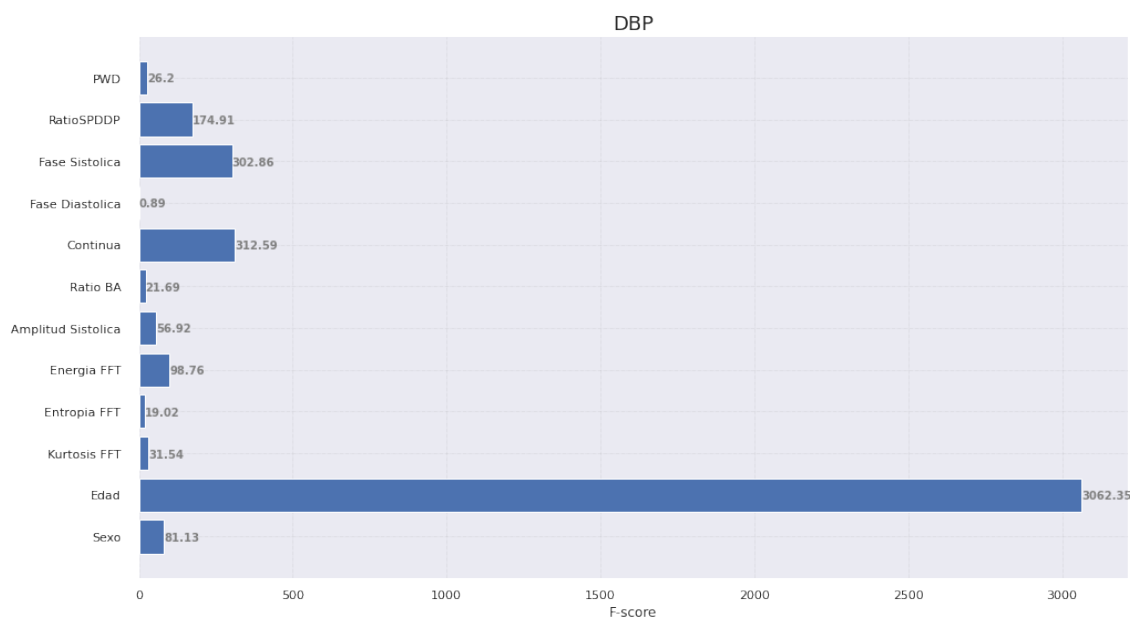
Como se explicó en la Secciones 8.8.1 y 8.7.2, se probaron distintas combinaciones de funciones de costo y optimizadores. Para encontrar la combinación óptima, se estudiaron 9 redes comparando los MAEs de cada una en el conjunto de validación, cuyos resultados se presentan en la Tabla 19. A partir de lo obtenido, se optó por la función de costo MAE junto con el optimizador Adam como combinación con mejor desempeño, siendo que el rendimiento que presenta es el menor (13.869 mmHg).

		Optimizadores		
		Adam	RMSprop	SGD
Funciones de Costo	MSE	14.063	14.112	13.930
	MAE	13.869	13.945	14.033
	MSLE	14.176	14.256	14.507
Desempeño: MAE (mmHg)				

Tabla 19: Valores de MAE de las distintas redes analizadas

Adicionalmente, se realizó una selección de features (ver Secciones 8.8.1 y 8.7.2). La Figura 72 muestra el *F-score* obtenido para cada uno de los *features* para la SBP. Análogamente, la Figura 73 demuestra lo obtenido para la DBP. A partir de esto, se identificaron los siguientes *features* de mayor importancia:

- Edad
- Energía FFT
- Ratio SPDDP
- Amplitud sistólica
- Fase diastólica
- Continua

Figura 72: F-score de cada *feature* evaluado contra la SBPFigura 73: F-score de cada *feature* evaluado contra la DBP

Teniendo en cuenta la selección de *features*, en la Tabla 20 se presentan el MAE del error de estimación y el desvío estándar de este ( $\sigma_e$ ) en el conjunto de *test* para la MLP con función de costo MAE y optimizador Adam, para la red entrenada con 12 y 6



*features*.

Función de Costo	Optimizador	Cantidad de <i>features</i>	Desempeño: MAE   $\sigma_e$ (mmHg)			
			SBP		DBP	
MAE	Adam	12	<b>19.083</b>	13.134	9.753	7.532
MAE	Adam	6	19.300	<b>13.114</b>	<b>9.752</b>	<b>7.484</b>

Tabla 20: Resultados obtenidos en el conjunto test para la MLP con función de costo MAE y optimizador Adam utilizando 12 *features* o únicamente los 6 más relevantes.

## 9.5. CNN

### 9.5.1. Selección de Especificaciones de las CNNs

A continuación en la Tabla 21 se pueden observar los resultados obtenidos con la Red V1, donde se pueden comparar según la cantidad de pulsos totales y, si se realizó o no, la normalización de los datos.

N	Normalización	Desempeño: $M_{MAE}$   $M_{\sigma}$ (mmHg)		Cumplimiento CS (%)		
		SBP	DBP	CCS	CCD	CCT
120	No	64.474   3.707	26.274   2.793	27.660	29.787	25.532
120	Si	7.558   3.826	4.090   2.405	44.681	55.319	44.681
200	No	<b>4.067</b>   3.185	3.184   2.503	61.702	65.957	59.574
200	Si	4.063   <b>2.898</b>	<b>2.642</b>   <b>2.025</b>	65.957	78.723	65.957
400	No	4.232   3.266	3.075   2.879	68.085	74.468	65.957
400	Si	4.115   3.490	2.833   2.711	<b>72.340</b>	<b>85.106</b>	<b>68.085</b>

Tabla 21: Resultados obtenidos para la Red V1 según el N de pulsos de PPG y si se aplicó la normalización de los datos. Se exhibe el desempeño en forma de mediana de los MAEs ( $M_{MAE}$ ) y mediana de los desvíos ( $M_{\sigma}$ ), y la proporción de pacientes del total en Cumplimiento con el Criterio Simplificado para la SBP (CCS), la DBP (CCD) y ambas (CCT).

En la columna de Cumplimiento CS se puede observar el porcentaje de pacientes del total en los que el error de medición cumple con el Criterio Simplificado establecido en la estimación de la SBP (CCS), en la DBP (CCD) y en ambas (CCT). Se puede observar que en todos los casos el CCD es el que alcanza los valores más altos. Esto es consistente con la bibliografía consultada que reportan mejor rendimiento en la estimación de la BP diastólica (Sección 7). Se puede apreciar cómo ambas redes con un N de 200 y de 400 cumplen con el Criterio Simplificado para el desempeño, mientras que esto no sucede en el caso de las redes con N de 120.

La Tabla 22 a continuación muestra los resultados obtenidos en el caso de la Red V2.

		Desempeño: $M_{MAE}$   $M_\sigma$ (mmHg)		Cumplimiento CS (%)		
N	Normalización	SBP	DBP	CCS	CCD	CCT
120	No	<b>3.392</b>   <b>2.675</b>	2.787   2.109	68.085	87.234	65.957
120	Si	3.534   2.729	2.730   2.290	65.960	85.106	65.960
200	No	3.560   2.816	<b>2.572</b>   2.188	68.085	85.106	65.957
200	Si	3.728   2.871	2.743   <b>2.038</b>	63.830	85.106	63.830
400	No	3.731   3.221	2.643   2.552	<b>74.468</b>	87.234	<b>74.468</b>
400	Si	3.788   3.312	2.604   2.968	<b>74.468</b>	<b>89.362</b>	<b>74.468</b>

Tabla 22: Resultados obtenidos para la Red V2 según el N de pulsos de PPG y si se aplicó la normalización de los datos. Se exhibe el desempeño en forma de mediana de las medias ( $M_{MAE}$ ) y mediana de los desvíos ( $M_\sigma$ ), y la proporción de pacientes del total en Cumplimiento con el Criterio Simplificado para la SBP (CCS), la DBP (CCD) y ambas (CCT).

Al igual que para la Red V1, se puede apreciar que los resultados obtenidos en CCD siempre son mayores que en CCS y CCT dentro de una misma red. En este caso, todas las redes cumplen con el Criterio Simplificado en el desempeño.

### 9.5.2. Evaluación de Calidad Médica en CNNs

A partir de los resultados obtenidos en la Selección de Especificaciones de las CNNs (Sección 9.5.1), como se mencionó en la Sección 8.7.3.2, se determinó que la arquitectura y configuración de entrenamiento con mejor rendimiento resultó ser la Red V2 con 400 pulsos, normalizando los datos de entrada (ATT-Net). Esto surge de comparar los resultados obtenidos en las Tablas 21 y 22.

Los resultados obtenidos en la Evaluación de Calidad Médica en CNNs con la ATT-Net pueden visualizarse en las Tablas 23 y 24. Se comparan los resultados de este trabajo con los requeridos por el estándar AAMI-ISO 81060-2:2018 para la validación de equipos médicos.

	Criterio 1			
	SBP		DBP	
	$\mu_e$ (mmHg)	$\sigma_e$ (mmHg)	$\mu_e$ (mmHg)	$\sigma_e$ (mmHg)
AAMI-ISO 81060-2:2018	$\leq 5$	$\leq 8$	$\leq 5$	$\leq 8$
ATT-Net	-0.552	5.226	0.037	3.609

Tabla 23: Resultados obtenidos en la Evaluación de Calidad Médica en CNNs comparados con la AAMI-ISO 81060-2 en el criterio 1 utilizando la media ( $\sigma_e$ ) y el desvío estándar ( $\mu_e$ ) [52].

	<b>Criterio 2</b>			
	SBP		DBP	
	$\mu_e$ (mmHg)	Probabilidad (%)	$\mu_e$ (mmHg)	Probabilidad (%)
AAMI-ISO 81060-2:2018	$\leq 5$	$\geq 85$	$\leq 5$	$\geq 85$
ATT-Net	-0.552	99.572	0.037	99.997

Tabla 24: Resultados obtenidos en la Evaluación de Calidad Médica en CNNs comparados con la AAMI-ISO 81060-2 en el criterio 2 donde  $\sigma_e$  hace referencia a la media y “Probabilidad ” al porcentaje de pacientes con error medio menor a 10 mmHg [52].

A continuación, en la Figura 74, se presenta la distribución del error en SBP y DBP en el conjunto de validación.

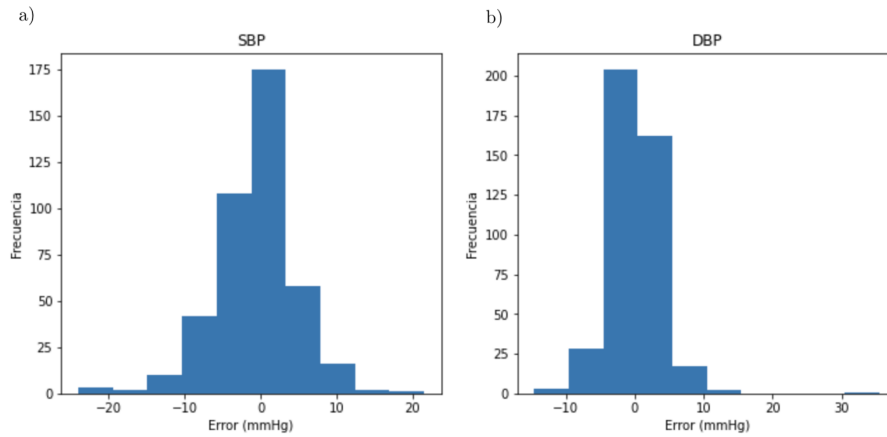


Figura 74: Distribución del error en (a) SBP y (b) DBP en el conjunto utilizado de validación para los 139 pacientes que cumplen los requisitos definidos por el estándar de aplicación.

Se observa que el error, en ambos casos, se distribuye alrededor del 0 donde el desvío estándar en SBP es menor al doble que en DBP (5.226 mmHg en SBP y 3.609 mmHg en DBP). A continuación, en la Figura 75, se encuentran los gráficos de Bland-Altman según lo propuesto por la AAMI.

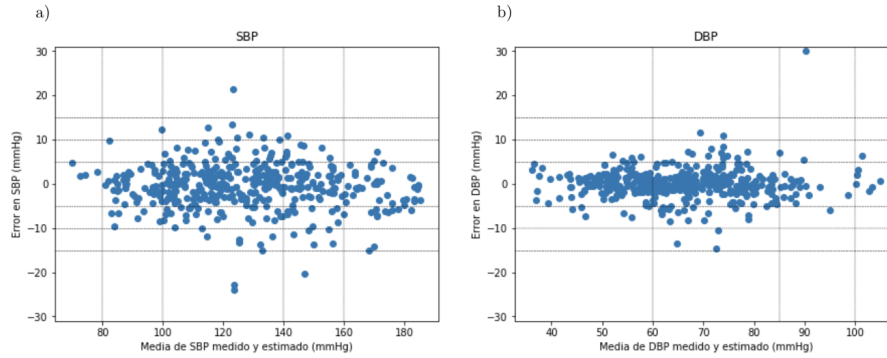


Figura 75: Gráficos de Bland-Altman para (a) SBP y (b) DBP. En el eje  $x$  se encuentra la media de la BP medida obtenida por el *ground truth* y la estimada. En el eje  $y$  se encuentra el error entre ambas mediciones. Se colocan líneas horizontales y verticales de referencia según lo establecido por el estándar [52].

Los gráficos de Bland-Altman demuestran que el método a validar estima, en promedio, 0.552 mmHg menos que el método de referencia en SBP aunque en DBP sobreestima 0.037 mmHg, en promedio. Además, se puede observar que la mayoría de los puntos en DBP se encuentran contenidos en el rango de error de -5 a 5 mmHg con una mayor concentración de puntos cercanos a 0 mmHg. Además, todos quedan contenidos en el rango de -15 a 15 mmHg excepto por un valor *outlier*, el cual tiene un error mayor a 30 mmHg. Por otro lado, la mayoría de los puntos en SBP se encuentran en el rango de -10 a 10 mmHg con una mayor densidad de estos en el rango de -5 a 5 mmHg. Se encuentran algunos pocos valores fuera del espacio de -15 a 15 mmHg. Esto mismo se puede observar en la Tabla 25, donde se toman los valores absolutos de los errores asociados a las mediciones y se calcula la cantidad de estos que se encuentran en cada rango.

	Valor absoluto del error			
	$\leq 5$ mmHg	$\leq 10$ mmHg	$\leq 15$ mmHg	Total
SBP	304	392	411	417
DBP	367	411	416	417

Tabla 25: Cantidad de mediciones cuyo error en valor absoluto es menor a 5 mmHg, 10 mmHg y 15 mmHg.

La cantidad de mediciones cuyo error en valor absoluto es menor a 5 mmHg es mayor para DBP y, como se comentó de la Figura 75, la mayoría tienen un error menor a 15mmHg. En ambas presiones se observa que la cantidad de mediciones mayores a 15 mmHg es reducida, donde se encuentran 6 mediciones para SBP y 1 para DBP.

A continuación se presentan las siguientes matrices de confusión donde se analiza la precisión del modelo utilizado en los distintos rangos de SBP y DBP (Figura 76). Si bien estas no son requeridas por el estándar en cuestión, son de interés e importancia a nivel

clínico ya que permiten analizar la estimación entre categorías. Además, en el caso de que el modelo no estime un valor dentro de su propia categoría, permite estudiar en cual otra categoría puede encontrarse.

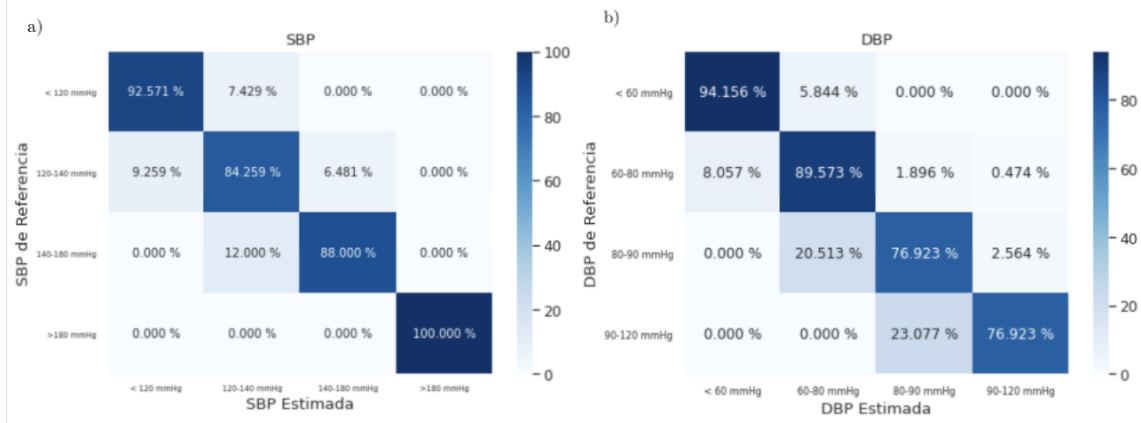


Figura 76: Matriz de confusión para los distintos rangos de BP en (a) SBP y (b) DBP. En el eje  $x$  se encuentran la SBP y DBP estimada por el modelo y en el eje  $y$  los de referencia obtenidos por el *ground truth*. Los valores son expresados como porcentaje (%).

En los valores de SBP normales, es decir, menores a 120 mmHg, la precisión de la predicción es de 92.571 %, como se observa en la Figura 76. Entre los valores considerados como SBP elevada e hipertensión de Grado 1, la precisión es de 84.259 %, donde 9.259 % de los valores son subestimados como menores a 120 mmHg mientras que 6.481 % son sobreestimados entre 140 y 180 mmHg. En cuanto a la precisión de la predicción para este último rango, consistente con hipertensión de Grado 2, el 88.000 % de las mediciones se estima en el mismo rango mientras que el 12.000 % restante se lo estima entre 120 y 140 mmHg. Por último, los valores correspondientes a una crisis hipertensiva, mayores a 180 mmHg, son estimados con un 100 % de precisión.

Por otro lado, en los valores de DBP normales se observa que la precisión es de 94.156 % y los que se encuentran entre 60 y 80 mmHg de 89.573 %. Para los valores correspondientes a hipertensión de Grado 1, la precisión es del 76.923 %, mientras que al 20.513 % se los estima entre 60 y 80 mmHg. En el rango asociado a la hipertensión de Grado 2, el 76.923 % es estimado en ese mismo rango mientras que el resto como hipertensión de Grado 1.

A fin de poder contrastar los resultados del presente Capítulo con el Criterio Simplificado, se obtuvieron los valores de MAE y  $\sigma_e$ . Los resultados alcanzados se pueden visualizar en la Tabla 26, donde tanto para la SBP como para la DBP el CS se ve cumplido.

	<b>SBP</b>		<b>DBP</b>	
	MAE (mmHg)	$\sigma_e$ (mmHg)	MAE (mmHg)	$\sigma_e$ (mmHg)
Criterio Simplificado	$\leq 5$	$\leq 8$	$\leq 5$	$\leq 8$
ATT-Net	3.843	5.226	2.347	3.609

Tabla 26: Resultados obtenidos en la Evaluación de Calidad Médica en CNNs comparados con el Criterio Simplificado utilizando el MAE y el desvío estándar ( $\mu_e$ ).

## 10. Discusión

### 10.1. Preprocesamiento

#### 10.1.1. Clasificadores de Segmentos de Señal

En cuanto a los clasificadores analizados para las señales de PPG, inicialmente se observó que a medida que se aumenta la cantidad de datos en la entrada, el rendimiento de los clasificadores mejora. Esto puede afirmarse al comparar los resultados obtenidos con  $N$  de 125 y 250. Por esto, los mejores clasificadores obtenidos tienen 250 datos de entrada. Si se analizan los resultados obtenidos para el *Clasificador PPG 1*, se puede ver que se obtienen mejores resultados utilizando 250 segmentos de señal para el entrenamiento. Por otro lado, al comparar el *Clasificador PPG 1* con el *Clasificador PPG 2*, se puede observar que el segundo posee un mejor rendimiento global.

Para los clasificadores de señal de ABP, si bien a simple vista el *Clasificador ABP 1* entrenado con 125 segmentos de señal posee un mejor rendimiento, se puede ver que al aumentar la cantidad de muestras de entrada (*Clasificador ABP 1* con 250 segmentos) los resultados se ven significativamente modificados. Esto implica que el *Clasificador ABP 1* no es capaz de generalizar correctamente con 125 segmentos de entrenamiento. Adicionalmente, se observó que los resultados obtenidos con un  $N$  de 250 fueron satisfactorios, por lo que el *Clasificador ABP 1* entrenado con 250 segmentos fue seleccionado como el óptimo.

### 10.2. Regresión Lineal

Como se preveía, el uso de una regresión lineal no resultó de utilidad para la finalidad buscada en el proyecto, es decir la de estimar la BP a partir de la curva de PPG. Esto puede evidenciarse en la Tabla 18 ya que los valores de MSE y MAE obtenidos se encuentran muy lejanos al cero, el valor ideal. Adicionalmente, el MAE se encuentra lejos de ser menor a 5 mmHg, es decir que no cumple la condición del Criterio Simplificado. En consecuencia, se justifica la subsiguiente implementación de diferentes herramientas a fin de lograr el objetivo planteado. Esta decisión se ve reforzada, además, por los valores de  $r^2$  obtenidos, que claramente indican que el método propuesto no logra resolver la problemática en cuestión.



### 10.3. MLP

Como se explicó en la Sección 8.7.2, luego de establecer la arquitectura general de la red, se hizo foco en encontrar la combinación óptima de función de costo y optimizador. En la Tabla 19 se puede observar que de las 9 combinaciones analizadas, el valor de MAE más bajo (13.869 mmHg) se obtuvo cuando se utilizó el MAE como la función de costo y *Adam* como optimizador. Por este motivo, se puede considerar que esta combinación genera el mejor rendimiento para la arquitectura de MLP diseñada. Adicionalmente, se debe hacer mención al hecho de que al utilizar como criterio de selección al menor MAE en validación, el modelo que utiliza dicha métrica como función de costo será el que mejor desempeño presente.

Una vez que se determinó la mejor red MLP, se prosiguió con la misma a la etapa de verificación. Los resultados obtenidos para la red entrenada con 12 *features*, presentados en la Tabla 20, no logran cumplir con el estándar de Criterio Simplificado, definido en la Sección 6.1 como un MAE menor a 5 mmHg y  $\sigma_e$  menor a 8 mmHg. Sin embargo, se destaca que los valores de DBP fueron estimados de forma más fidedigna. Esto se corresponde con lo estudiado en la bibliografía ya que, en todos los estudios similares (ver Sección 7), la presión arterial diastólica suele poder estimarse de forma más precisa.

Por otro lado, si se comparan los resultados obtenidos para la red entrenada con 12 y 6 *features*, se hace notable la utilidad de la selección de *features*. Si bien la segunda red fue entrenada con la mitad de los *features*, se puede observar que la diferencia entre los resultados obtenidos no es remarcable. Esto último refuerza que los 6 *features* seleccionados efectivamente son los que aportan la mayor cantidad de información a la red, o la de superior relevancia. Es por esto último que se logró, entonces, obtener resultados similares utilizando una menor cantidad de datos, conllevando un menor costo computacional. En otras palabras se logró demostrar que con vectores de entrada de longitud equivalente al 50 % de los originales, se obtienen resultados en *test* prácticamente iguales.

Se debe hacer mención de un punto importante: los resultados obtenidos para las redes en *test* no solo no cumplen con el Criterio Simplificado sino que, además, el MAE en SBP es aproximadamente 4 veces el límite superior del Criterio, con el MAE en DBP siendo casi el doble. En otras palabras, las redes MLP analizadas en *test* se encuentran lejos de cumplir con lo postulado en la hipótesis del presente proyecto de investigación.

## 10.4. CNN

### 10.4.1. Selección de Especificaciones de las CNNs

En cuanto a los resultados obtenidos con la Red V1 (Sección 9.5.1), se ve que son consistentes con lo reportado en la literatura en el sentido en que, en cada ANN, la estimación de la DBP fue más precisa, resultando en un mayor CCD. Además, como se mencionó anteriormente, tanto las redes con N de 200 como las redes con N de 400 cumplieron con el Criterio Simplificado determinado. Esto entonces sugiere que, para la Red V1, a mayor N, mejor el desempeño obtenido. Si bien las diferencias entre ANNs pueden ser sutiles si se analiza únicamente el desempeño, si se tiene en cuenta lo obtenido en Cumplimiento CS se exponen las diferencias en el rendimiento obtenido de forma más clara. Entonces, se evidencia que ambos parámetros son de importancia a la hora de la elección de la ANN.

Los resultados de la Red V1 también sugieren que las redes normalizadas suelen arrojar mejores rendimientos que las que no (para un mismo N), ya que esto sucedió para N de 120, 200 y 400. Este fenómeno podría explicarse por el hecho de que, al normalizar los datos de entrada, se consigue que la función de costo no sea más sensible a algunas muestras por sobre otras [36]. Adicionalmente, se observa que el uso de un N de 120 en PPG no es suficiente para que la red pueda generalizar y, por lo tanto, la medida de tendencia central del error ( $M_{MAE}$ ) es mayor a 5 mmHg, especialmente en el caso en que no se normalizan los datos. Sin embargo, utilizar un N de 200 o 400 sí resulta suficiente, sugiriendo que a mayores valores de N la red generaliza. Esto se debe a que la red cuenta con una mayor cantidad de información para ser entrenada, resultando en una mejor capacidad de generalización.

Teniendo en cuenta el desempeño, tanto para la presión sistólica como para la diastólica, así como también los valores de CCS, CCD y CCT, se puede ver que la configuración con mejor rendimiento global dentro del grupo de las redes de tipo V1 es la red con un N de 400 y con normalización de los datos de entrada. Al poseer valores de Cumplimiento CS mayores que la misma variante pero sin normalizar los datos de entrada, representa una clara ventaja ya que implica una mayor cantidad de individuos que cumplen con el Criterio Simplificado, implicando, entonces, que se podrían obtener mediciones fidedignas para una mayor cantidad de pacientes. Adicionalmente, se debe destacar que tanto el N de 200 como el de 400 demuestran que es posible obtener redes calibradas de tipo V1 que cumplan con el Criterio Simplificado.

En lo que refiere a la Red V2, se observa que todas las combinaciones cumplen con el Criterio Simplificado, a diferencia de lo que ocurre con V1. Además, el rendimiento en DBP es mayor al de SBP, tal como se esperaba. Por un lado, no se observa que la diferencia entre las redes sea significativa para las medidas de desempeño aunque los desvíos estándar para las redes con 400 pulsos son mayores que los correspondientes a 120 y 200 pulsos. Por

otro lado, son estas las que poseen mayor CCT alcanzando el valor de 74.468 % en ambas redes, con y sin normalización. Al comparar estas redes, se observa que la CCD es mayor cuando se normaliza los datos en la entrada.

Al compararse los resultados obtenidos para la Red V1 y para la Red V2, se puede observar que, en general, los de la Red V2 presentan un mejor rendimiento. En ambos casos se puede notar la tendencia de mejora a medida que se aumenta el N y al normalizar. Sin embargo, debe destacarse la capacidad de la Red V2 de generalizar con N de 120 frente al hecho de que la Red V1 no es capaz. Estas diferencias en rendimiento, se deben a la combinación de los efectos asociados al *Dropout*, *Batch Normalization*, la normalización y cantidad de datos de entrada y la arquitectura. Es decir, no es atribuible a un único factor. A partir de los resultados expuestos en la Sección 9.5.1, se concluye que la red que presenta el mejor rendimiento es la Red V2 con un N de 400 pulsos, normalizando los datos de entrada. Por lo tanto, se decide utilizar esta arquitectura para realizar el análisis propuesto en la Sección 8.7.3.2, la cual es renombrada a ATT-Net por ser la abreviación del término *Artificial Tailored Tensiometer*.

#### 10.4.2. Evaluación de Calidad Médica en CNNs

Como se puede observar, el desempeño de la ATT-Net es tal que permite cumplir con lo establecido por el estándar internacional, en ambos criterios (1 y 2), detallados en la Sección 6.2. En otras palabras, se puede decir que se cumplió el objetivo de obtener calidad médica para la ATT-Net.

En cuanto a los resultados obtenidos para el Criterio 1, que se muestran en la Tabla 23, se puede ver que para la SBP la ATT-Net posee una media del error negativa, lo que indica que, en promedio, se está subestimando la SBP. Por otro lado, la media del error para la DBP es positiva, por lo que en promedio se está sobreestimando el valor de DBP. Si bien esto es así, se debe tener en consideración que los valores absolutos de dichas medias son pequeños, por lo que no representan un inconveniente. Asimismo, como se puede evidenciar por los resultados obtenidos en las matrices de confusión de la Figura 76, la ATT-Net logra categorizar correctamente el valor de SBP y DBP en el rango correspondiente. Esto último implica que, además de tener un desempeño satisfactorio a nivel general, la red posee un rendimiento adecuado en la categorización. Adicionalmente, la precisión en todos los casos estudiados supera el 75 % siendo máxima en valores normales de SBP y DBP donde la precisión es de 92.571 % y 94.156 % respectivamente. Por último, se concluye que hay un alto grado de concordancia entre el método de medición de referencia (*Ground Truth*) y el propuesto, a partir de los gráficos de Bland-Altman (ver Figura 75). Esto es de gran importancia debido a que en el caso de que se haga una sobreestimación o subestimación se podría generar un diagnóstico y tratamiento inadecuado. Un ejemplo de esto es el caso de la subestimación donde se podría llevar a la falta de uso de procedimientos para disminuir la BP del paciente.

Del análisis de la Tabla 25 se pueden obtener observaciones complementarias a las obtenidas por el método de Bland-Altman. En una primera instancia, se desprende que en el caso de la SBP se encuentran 6 valores atípicos ubicados fuera del rango de -15 a 15 mmHg. Por otro lado, en el caso de la DBP, se debe notar que se encuentra un único *outlier* fuera de este rango definido. En ambos casos, se puede afirmar que el error, en la mayoría de los casos, queda contenido entre -15 y 15 mmHg. Adicionalmente, si se observa la cantidad de mediciones obtenidas en cada intervalo propuesto, se puede ver que la mayor densidad de puntos se encuentra dentro del intervalo de -5 a 5 mmHg. Esto se considera como un resultado positivo ya que indica que el error, en general, tiende a ser bajo. Como se mencionó en la Sección 9.5.2, la cantidad de puntos en este rango para DBP es mayor que en SBP, lo cual es consistente con la bibliografía consultada que siempre indica mejores resultados en cuanto a la estimación de la DBP (ver Sección 7).

En lo referido al Criterio 2, los resultados en la Tabla 24 exhiben que para la SBP la ATT-Net cumple con los requisitos necesarios, siendo que su media es menor a 5 mmHg y la probabilidad porcentual de que el error sea menor a 10 mmHg, utilizando una distribución normal, es mayor a 85 %. Estas mismas afirmaciones aplican en el caso de la DBP.

#### 10.4.2.1. Comparación con Trabajos Previos

En general, resulta dificultosa la comparación de publicaciones relacionadas debido a la utilización de diferentes métricas de evaluación y bases de datos. Sin embargo, para poder realizar un contraste, se obtuvieron los valores de MAE y  $\sigma_e$  para poder aplicar el Criterio Simplificado a la Evaluación de Calidad Médica en CNNs, considerando que ninguno de los estudios actualmente publicados analiza la calidad médica como se describe en la Sección 6.2. La Tabla 27 muestra los resultados obtenidos por los principales estudios analizados en la Sección 7 (Estado del Arte) que utilizan calibración, así como también los obtenidos en el presente proyecto en la Evaluación de Calidad Médica en CNNs. Teniendo en cuenta que en la investigación actual no se utilizan señales adicionales como el ECG y que la CNN utilizada es un modelo de simple construcción, se puede decir que los resultados recabados son positivos frente a aquellos encontrados en la literatura.

Autor	Base de Datos	Método	SBP   DBP	
			MAE (mmHg)	$\sigma_e$ (mmHg)
Slapnicar <i>et al.</i>	MIMIC-III	Random Forest	13.620   11.730	-   -
Slapnicar <i>et al.</i>	MIMIC-III	CNN+GRU con features temporales y frecuenciales de PPG	9.430   6.880	-   -
Wang <i>et al.</i>	MIMIC	CNN+GRU con pulsos de ECG y PPG	3.950   2.140	4.380   2.400
Esmaelpoor <i>et al.</i>	MIMIC-II	CNN+GRU con pulsos PPG	1.910   0.670	5.550   2.840
Jeong <i>et al.</i>	MIMIC	CNN+GRU con pulsos de ECG y PPG	0.020   0.160	1.620   1.300
Lin <i>et al.</i>	MIMIC-III	Regresión Lineal con features temporales y frecuenciales de ECG y PPG	4.590   2.470	2.780   1.090
Athaya <i>et al.</i>	MIMIC-III	CNN U-NET con pulsos de ECG y PPG	3.680   1.970	4.420   2.920
ATT-Net	MIMIC-III	CNN con pulsos de PPG	3.854   2.347	5.226   3.609

Tabla 27: Comparación entre los resultados obtenidos en la Evaluación de Calidad Médica en CNNs y los estudios analizados que utilizan métodos con calibración (ver Estado del Arte, Sección 7).

## 10.5. Limitaciones y Próximos Pasos

Es de importancia tener en consideración las limitaciones existentes en el proyecto y posibles futuros pasos que podrían llevarse a cabo eventualmente. Dichas potenciales modificaciones o adhesiones no fueron realizadas debido al alcance acotado del proyecto que, si bien buscó ser lo más abarcativo posible, no pudo cubrir la totalidad de los aspectos a investigar. Adicionalmente, se debe tener en cuenta que al ser un campo de estudio emergente, surgen actualizaciones de forma constante, modificando el paradigma.

Una limitación relevante a tener en cuenta en este proyecto de investigación es el sesgo de selección, ya que se debe considerar que la población de estudio se ve limitada a los pacientes que se encontraron en la Unidad de Terapia Intensiva del *Beth Israel Deaconess Medical Center* en el intervalo temporal de recolección de datos para MIMIC-III. Adicionalmente, si se tiene en cuenta que los individuos se encontraban en la UTI, no se debe pasar por alto que se trata de pacientes en condiciones graves, probablemente

con mayor prevalencia de hipertensión o incluso de otras patologías graves y que, en consecuencia, los resultados pueden verse alterados. Este sesgo particular es difícil de mitigar dada la falta de bases de datos que posean una gran cantidad de datos y señales de interés para el presente proyecto.

Un punto importante a ser tenido en cuenta es que al momento de la recolección de datos para MIMIC-III se utilizaron dos sistemas que se encuentran estandarizados y, por lo tanto, se los trató como equivalentes. De todos modos, esto podría no ser cierto, resultando en que se encuentren diferencias en la calidad y el formato de las señales adquiridas. Asimismo, se aceptó como supuesto que las mediciones de BP y PPG fueron realizadas de forma correcta, es decir que se asumió que la metodología de obtención de los datos y colocación de los sensores fue adecuada. No obstante, se tomaron ciertas medidas para mitigar algunos de los errores de medición comunes como, por ejemplo, la eliminación del corrimiento de la línea de base en la etapa de filtrado para señales de PPG (Sección 8.6.2) que, por lo general, suele surgir cuando el paciente realiza movimientos.

Relacionado a la posible fuente de error detectada previamente, se encuentra la posibilidad de que las condiciones dadas no sean las adecuadas para obtener una medición correcta utilizando un oxímetro de pulso, independientemente de la correcta colocación y manipulación del sensor. Como se mencionó en la Sección 5.2.1, las mediciones obtenidas dependen de si el individuo presenta una  $SaO_2$  entre un 70 y 100 %, si no se presentan interferencias de color (por su efecto en el sensor) y de los niveles de bilirrubina del individuo, entre otros. En el proyecto actual se asumió que en todo momento las mediciones recabadas fueron correctas, ya que resulta altamente dificultoso, para la investigación planteada, identificar las situaciones mencionadas.

En cuanto al procedimiento aplicado para la Selección de NaNs (Sección 8.6.1) se debe tener en cuenta que la metodología elegida en el actual proyecto es debatible. Si bien la decisión fue de reemplazar todos los valores NaN por 0 a modo de evitar introducir datos no reales en las señales, a futuro se podría estudiar el efecto de aplicar otros métodos como, por ejemplo, los predictores de tendencia. Sin embargo, esto implicaría la introducción de información que no se encuentra originalmente en la base de datos, lo que podría resultar una desventaja.

Considerando que el enfoque aplicado fue el de *zero padding* para que los pulsos de PPG tengan la misma duración, se debe tener en cuenta que una limitación resultante es que se agregan distintas cantidades de ceros a los distintos pulsos, dado que cada pulso difiere de 300 muestras en una cantidad distinta. Dicho esto, un posible método alternativo a futuro podría ser el de interpolar los pulsos para que todos posean una misma longitud deseada. Sin embargo, debe mencionarse que esto implicaría introducir valores muestrales que no se encontraban originalmente en la base de datos, algo que en la práctica no siempre se ve al trabajar con señales fisiológicas y que tampoco se vio aplicado en la bibliografía consultada en la Sección 7 [94].

En lo referido a la clasificación de calidad de señal (Sección 8.6.4), es pertinente mencionar que la premisa fue utilizar un algoritmo simple que sirva para clasificar segmentos lo más rápido posible y de forma robusta. El objetivo principal de esta etapa fue poder prevenir el procesamiento de señales de calidad pobre en futuras etapas del proyecto y, así, ahorrar tiempo de cómputo considerando la vasta extensión de la base de datos. Si bien eso se pudo lograr sobre los datos evaluados y con las métricas aplicadas para el entrenamiento del clasificador, se debe tener en cuenta que su funcionamiento podría verse limitado a la base de datos en la que se aplique. En otras palabras, el modelo podría fallar en clasificar correctamente las señales si se aplica sobre otros conjuntos de datos.

Una limitación que no debe dejar de tenerse en consideración en la metodología aplicada para las redes CNN (Sección 8.7.3) es el hecho de que se utilizó un subconjunto de pacientes para alcanzar los objetivos de esta investigación debido a la extensión de la base de datos y la imposibilidad de procesarlos en conjunto. Si bien el subconjunto fue elegido de forma aleatoria, una posibilidad es que no sea representativo de la población y que, por ende, los resultados obtenidos no sean extrapolables al resto de los pacientes. De todas formas, se debe tener en cuenta que esta problemática también se intentó minimizar al elegir 150 nuevos pacientes aleatorios en el análisis realizado en la Evaluación de Calidad Médica en CNNs sobre la ATT-Net en la Sección 8.7.3.2.

Por otro lado, en el caso particular de las MLP, la cantidad de *features* podría haber sido aumentada haciendo un uso más extenso de la base de datos MIMIC-III: *Clinical Database*. De esta manera, se generaría un análisis de la influencia de datos clínicos adicionales en la estimación de la presión arterial, como por ejemplo la altura y el peso de los pacientes. En el caso de las CNN, no se analizó el efecto del uso de datos. Esto podría incluirse en modificaciones futuras, ya que resulta de interés el análisis del cambio en el desempeño en el caso en que se incorporen datos clínicos, tanto para el caso de las CNN como para las MLP. No obstante, no debe dejar de mencionarse que no se incluyeron más *features* clínicos debido a que se encontró que la base de datos clínica no se encuentra curada, lo que dificulta su procesamiento.

Relacionado el punto anterior, se encuentra un posible análisis para describir a la población de estudio en cuanto a sus comorbilidades, antecedentes de hipertensión y otros factores. Sin embargo, debe tenerse presente lo mencionado en la Sección 8.1.1, ya que resulta altamente dificultoso el procesamiento de esta información cuando en muchos pacientes no se encuentra, difiere la nomenclatura utilizada para los mismos términos, o se encuentran errores de escritura. Además, considerando que se trata de manipular un archivo con un peso mayor a 40 GB, el procesamiento de esta información no es trivial. Por esto, resulta un potencial paso interesante a analizar en un futuro, en donde se realice un análisis extensivo de esta base de datos y se puedan tener en cuenta todos los posibles casos presentados en el erróneo registro de la información.

Un punto importante a tener en cuenta a la hora de trabajar con *features* es el uso de distintas herramientas para la selección de *features*. Si bien se decidió utilizar el

*F1-Score*, a futuro sería de interés estudiar el uso de otras herramientas como el Análisis de Componentes Principales (PCA). Este último es un método de *feature extraction* que permite disminuir la dimensionalidad del problema. A grandes rasgos, el método genera nuevas variables independientes y las combina con las variables pre-existentes. Se analiza la relación que tienen estas variables y se obtienen los componentes principales. Por último, se determina la cantidad de *features* a descartar [95].

Otra limitación presente en la etapa de las MLPs es que sólo se entrenaron de forma generalizada, sin calibración. La decisión de no entrenar las MLPs con un enfoque calibrado se tomó en base a varias razones. En primer lugar, los primeros resultados obtenidos con las redes MLP están lejos de cumplir con el Criterio Simplificado. En segundo lugar, la bibliografía consultada indica que las CNNs son más prometedoras para cumplir con el objetivo propuesto. Por último, el preprocesamiento de datos para las MLPs es complejo y conlleva un costo computacional adicional en comparación con las CNNs ya que implica la extracción de *features*. Teniendo en cuenta lo expuesto anteriormente, se decidió omitir el estudio de redes MLPs calibradas y trabajar directamente con CNNs. Como trabajo futuro se propone el estudio de redes MLPs calibradas y la comparación del desempeño de éstas con las redes ya analizadas, sean MLPs o CNNs.

Un punto importante a mencionar es que la determinación de las arquitecturas de las ANNs a utilizar constó de un análisis de una cantidad de arquitecturas y de parámetros finita. Esto fue realizado de esta forma debido al límite de tiempo de cómputo por lo que no puede descartarse que exista alguna solución que presente mejores resultados que los obtenidos en el presente proyecto. A futuro resultaría enriquecedor realizar una búsqueda exhaustiva de hiperparámetros y poder comparar distintas arquitecturas y sus respectivos rendimientos. De todas formas, resulta destacable hacer mención que de acuerdo a Vladimir Vapnik la resolución del problema de la búsqueda de la arquitectura es un proceso heurístico, así como también la minimización del error en el entrenamiento de la ANN [96].

Además, se encuentra la limitación de la falta de detalle en la bibliografía presente al momento de realizar este proyecto de investigación. Es de interés analizar los modelos propuestos en los trabajos anteriores a este sobre la base de datos utilizada a modo de verificación de las investigaciones y sus resultados presentados. Sin embargo, esto no fue posible debido a la imposibilidad de reproducibilidad mencionada en Estado del Arte (Sección 7).

Una posibilidad a explorar en futuros trabajos es la de redondear los valores de BP utilizados. En el presente estudio, se trabajó con todos los decimales. Si bien esta decisión se basó en la idea de obtener resultados lo más precisos posibles, se debe considerar que los decimales no suelen ser de importancia a la hora de brindar un diagnóstico clínico a un paciente y, por lo tanto, una opción a explorar a futuro es la de trabajar sin decimales en los valores de BP o, incluso, trabajar con una menor cantidad. En línea con esta idea, también se podría considerar un enfoque en donde lo relevante no sea obtener un valor



de presión arterial preciso sino un rango en el cual el valor se encuentra. Esto también presenta utilidad a la hora del diagnóstico clínico, si bien no se realizó en el presente proyecto porque el alcance del mismo es estimar la presión arterial con calidad médica. De todas formas, a futuro esa opción podría ser interesante de analizar a fines de su utilidad práctica.

Si bien ya se mencionó el posible espacio de mejora existente en explorar más cantidad de arquitecturas de ANNs, también se debe tener en cuenta, a futuro, la implementación de otras configuraciones u otros tipos de redes. Una posibilidad, por ejemplo, es la de mantener una red para la SBP y otra para la DBP y analizar si esta opción presenta un mejor desempeño. Por otro lado, también podría armarse una ANN que consista de dos ramas: una a la que ingrese la señal de de PPG y otra a la que ingresen datos clínicos del paciente. Esta podría ser una forma interesante de introducir los datos clínicos de los pacientes, incluso pudiéndose utilizar en la rama en la que ingresa la señal de PPG la ATT-Net, que presentó el mejor rendimiento en este proyecto.

Teniendo en vista un cambio del tipo de red, podría explorarse el uso de las redes GRU o, incluso, una combinación de CNN con GRU, considerando que en Estado del Arte (Sección 7) se pudo ver que están ganando terreno en el área de aplicación. Dicho esto, se debe tener en cuenta que los resultados de los estudios que utilizan las redes tipo GRU presentan resultados con que cumplen con el Criterio Simplificado.

Un siguiente paso luego de lograr la calidad médica calibrando para cada individuo podría ser el de aplicar *transfer learning*. En el marco de este proyecto, esto podría implementarse, por ejemplo, logrando entrenar una red con una base de datos y luego realizando el *transfer learning* con pacientes diferentes a aquellos comprendidos en el conjunto de entrenamiento. Esto implicaría un paso intermedio entre la necesidad de calibración de las redes para individuos en particular y una ANN que logre estimar la BP sin necesidad de calibración previa. Dicho esto, el próximo paso lógico e ideal es la construcción de una ANN entrenada con una cantidad determinada de pacientes que pueda estimar, con calidad médica, los valores de SBP y DBP para un individuo cualquiera, no incluido en el conjunto de entrenamiento. De esta forma, se obtendría una herramienta valiosa a la hora del diagnóstico clínico. Se debe tener en cuenta que para poder lograr esto, se deberían analizar las alternativas y limitaciones planteadas anteriormente a modo de lograr obtener el mejor modelo posible, aplicando la mejor metodología posible.

Una de las principales ventajas del método que se buscó desarrollar en el proyecto es que, a futuro, presenta el potencial para integrarse en una interfaz implementada con la finalidad de obtener la señal de PPG y, así, devolver un valor de BP pulso a pulso. De esta forma, se podría lograr obtener un método alternativo de medición de presión arterial que sea no invasivo y, además, que presente resultados con calidad médica. Para esto, sería necesario realizar un análisis similar al encontrado en la Sección 8.7.3.2 referida a la evaluación de calidad médica. Adicionalmente, esto implicaría determinar las acciones a llevar a cabo por el algoritmo del preprocesamiento, como por ejemplo qué acción tomar

frente a un segmento de señal que se clasifica de mala calidad en un contexto *online*. Estas decisiones deberían ser evaluadas junto con un médico especialista en el campo de estudio. Además, en este proyecto las CNNs se encuentran entrenadas y evaluadas con pulsos temporalmente consecutivos. Podría ser interesante evaluar el rendimiento de estas sobre pulsos temporalmente lejanos a con los que se ha entrenado y analizar la distribución del error a medida que pasa el tiempo, lo cual se asemeja a lo que pasaría en una aplicación en tiempo real. De obtener un nivel de calidad adecuado según los estándares pertinentes, el método podría ser integrado con el *software* presente en los oxímetros para así utilizarse tanto en un ámbito hospitalario como en uno doméstico. Considerando que existen aplicaciones en dispositivos móviles que levantan la curva de PPG para medir la frecuencia cardíaca, a futuro se podría incluso pensar en una interfaz en un *smartphone* que permita la medición de la BP. Debido a la alta disponibilidad de este tipo de tecnologías, se garantizaría el registro continuo de la BP a bajo costo, convirtiendo a esta técnica en una de posible acceso universal y consumo masivo. Se debe tener en cuenta que se encuentran equipos que utilizan métodos alternativos sin manguito (*cuffless*), con necesidad de calibración, actualmente aprobados por la FDA. Sin embargo, pueden involucrar el uso de curvas de PPG en conjunto con información adicional (como por ejemplo curvas de ECG), así como también la utilización de sensores de presión, lo que resulta de mayor complejidad que lo propuesto en esta investigación [97].

## 11. Conclusiones

La búsqueda de técnicas para la medición de BP de forma continua y no invasiva ha cobrado importancia en los últimos años debido a las cifras reportadas por la OMS de las complicaciones asociadas a la hipertensión y la tasa de mortalidad ligada a la misma. La demanda de dispositivos que permitan el registro continuo expone la necesidad de generar métodos con alta portabilidad y poca dependencia de la colocación del dispositivo para la medición. Es por esto que se propone utilizar la señal de PPG obtenida a partir del oxímetro de pulso, el cual es fácil de usar, y, mediante el uso de redes neuronales artificiales, estimar la BP.

A principios del desarrollo del presente proyecto se hizo una consulta bibliográfica y, luego, se decidió comenzar analizando el desempeño de redes MLP para cumplir este objetivo. Cabe destacar que la información presente al momento de este proyecto, tal como se comentó en la Sección 7, es incompleta. Es por esto que se partió de las arquitecturas encontradas en la literatura para obtener una nueva. Teniendo en cuenta los resultados presentados en la Sección 9.4 y el análisis hecho en la Sección 10.3, se puede concluir que la red MLP propuesta no es capaz de estimar de forma precisa la BP a partir de la PPG. Esto puede observarse en la Tabla 20, donde los errores obtenidos no logran cumplir con el Criterio Simplificado explicado en la Sección 6.1. Adicionalmente, a partir del análisis de la selección de *features* se obtuvo que, al entrenar la misma red utilizando los 6 más relevantes, los resultados no fueron significativamente diferentes, reforzando que dichos *features* son los que aportan la mayor cantidad de información, en comparación, a la hora de la estimación de la BP.

A partir de que los resultados obtenidos con las MLPs no fueron exitosos y del aumento de publicaciones que utilizan CNNs para la estimación de BP, se resolvió estudiar la implementación de este tipo de ANNs. En base a la bibliografía consultada, se pudo observar que los mejores resultados de las distintas redes propuestas se obtienen cuando éstas son calibradas. Por esto, en una primera instancia, se entrenaron y validaron las redes con esta modalidad. En esta etapa del proyecto, se implementaron dos CNNs denominadas Red V1 y Red V2.

De acuerdo a los resultados obtenidos para la Red V1 presentados en la Tabla 21, se concluye que la Red V1 con mejor rendimiento es la que utiliza un N de 400 pulsos normalizados, la cual logra cumplir el Criterio Simplificado. Los resultados alcanzados para la Red V2, expuestos en la Tabla 22, permiten concluir que todas las configuraciones analizadas cumplen con el Criterio Simplificado. Si bien no se advierten diferencias entre las distintas combinaciones estudiadas al hacer foco únicamente en el desempeño, al incorporar la CCS, CCD y CCT se encuentra que el mejor rendimiento es alcanzado al utilizar un N de 400 con normalización.

Posterior a estos análisis, se procedió a comparar las mejores versiones obtenidas

para la Red V1 y para la Red V2. A partir de este proceso, se concluye que la CNN con desempeño superior es la Red V2 con un N de 400 y normalización de pulsos de entrada, renombrada como la ATT-Net por ser la abreviación del término *Artificial Tailored Tensiometer*. Por esto, se tomó la ATT-Net para avanzar con el procedimiento explicado en la Sección 8.8.2 con el objetivo de verificar el cumplimiento de los requisitos propuestos por el estándar AAMI-ISO en cuanto a la calidad médica (ver Sección 6.2). El método demostró ser capaz de obtener mediciones con un error tal que sea posible la validación de la técnica mediante el estándar AAMI-ISO 81060-2:2018.

En conclusión, el método propuesto en la Sección 8.7.3.2 es capaz de tomar mediciones con calidad médica según lo establecido por el estándar AAMI-ISO 81060-2:2018 y ser incorporado a un sistema para la medición continua y no invasiva de la BP. La técnica también demuestra tener alta concordancia con la medición de referencia y adecuada precisión para la categorización de la misma. Esto es de gran importancia clínica ya que el diagnóstico y tratamiento de los pacientes no se verá afectado, en la mayoría de los casos, por errores en la estimación. Si bien el caso presentado implica la necesidad de calibración previa, se abre el panorama a futuras investigaciones referidas a la cuestión. Consecuentemente, se confirma la hipótesis planteada en el presente trabajo referida a la posibilidad de estimar la BP con calidad médica a partir de las curvas de PPG mediante el uso de ANNs.

En el proyecto actual se desarrolló un método para la estimación de la BP con calidad médica (de acuerdo a la AAMI-ISO) con necesidad de calibración previa. Esta calibración necesitaría de una recolección de datos simultáneos de señal de PPG y de valores de BP del paciente. Sin embargo, una vez que esto se haya realizado y que la red se encuentre entrenada, no sería necesario repetir el proceso cada vez que el individuo necesite una medición de BP, es decir que, a futuro, es una inversión de tiempo y de practicidad del método. Ya sea logrando que la red estime la BP para pacientes sin tener información previa de ellos, utilizando *transfer learning* o necesitando una calibración previa para cada individuo, esta herramienta podría representar un paso previo en la integración de una nueva funcionalidad del oxímetro de pulso: la medición de la presión arterial.

## 12. Referencias

- [1] World Health Organization. (2013). Información general sobre la hipertensión en el mundo. Una enfermedad que mata en silencio, una crisis de salud pública mundial. Día Mundial de la Salud.
- [2] O'Brien, E., Atkins, N., Stergiou, G., Karpettas, N., Parati, G., Asmar, R., ... & Working Group on Blood Pressure Monitoring of the European Society of Hypertension. (2010). European Society of Hypertension International Protocol revision 2010 for the validation of blood pressure measuring devices in adults. *Blood pressure monitoring*, 15(1), 23-38.
- [3] Mynurse.ai Editorial Team. (2021). Blood pressure cuff hurting your arm when inflated? Fit is key. Consultado el 25 de marzo de 2021 de <https://mynurse.ai/blog/blood-pressure-cuff-fit-arm-pain/>
- [4] Athaya, T., & Choi, S. (2021). An estimation method of continuous non-invasive arterial blood pressure waveform using photoplethysmography: A U-Net architecture-based approach. *Sensors*, 21(5), 1867.
- [5] Pickering, D., & Stevens, S. (2013). How to measure and record blood pressure. *Community eye health*, 26(84), 76.
- [6] McGhee, B. H., & Bridges, E. J. (2002). Monitoring arterial blood pressure: what you may not know. *Critical Care Nurse*, 22(2), 60-79.
- [7] Jagadeesh, A. M., Singh, N. G., & Mahankali, S. (2012). A comparison of a continuous noninvasive arterial pressure (CNAP™) monitor with an invasive arterial blood pressure monitor in the cardiac surgical ICU. *Annals of cardiac anaesthesia*, 15(3), 180
- [8] Freeman, B., & Berger, J. (2016). *Anesthesiology core review: Part two advanced exam*. McGraw Hill Professional.
- [9] Brar, K. S., & Ramesh, S. (2003). Technique of Blood Pressure Measurement. *Medical journal, Armed Forces India*, 59(1), 51.
- [10] Nickson, C. (2020). Arterial line and Pressure Transducer. *Life in the Fast Lane* • LITFL. Consultado el 13 de abril de 2020 de <https://litfl.com/arterial-line-and-pressure-transducer/>
- [11] Medición de la presión arterial. medlineplus. Consultado el 9 de febrero de 2021 de <https://medlineplus.gov/spanish/ency/article/007490.htm>

- 
- [12] Merck Sharp & Dohme Corp. Figure: Medición de la presión arterial. Manual MSD versión para público general. Consultado el 9 de octubre de 2021 de [https://www.msdmanuals.com/es-mx/hogar/multimedia/figure/cvs\\_measuring\\_blood\\_pressure\\_es](https://www.msdmanuals.com/es-mx/hogar/multimedia/figure/cvs_measuring_blood_pressure_es)
- [13] OMS. (2013). La OMS insta a eliminar gradualmente los termómetros y los dispositivos de medición de la presión arterial que contienen mercurio para 2020. <https://www.who.int/es/news/item/11-10-2013-who-calls-for-the-phase-out-of-mercury-fever-thermometers-and-blood-pressure-measuring-devices-by-2020>
- [14] Stergiou, G. S., Palatini, P., Asmar, R., Bilo, G., De La Sierra, A., Head, G., ... & O'Brien, E. (2018). Blood pressure monitoring: theory and practice. European Society of Hypertension working group on blood pressure monitoring and cardiovascular variability teaching course proceedings. *Blood pressure monitoring*, 23(1), 1-8.
- [15] Low Blood Pressure: A Side Effect of Treatment. (2020). Breastcancer.Org. Consultado el 18 de noviembre de 2020, de [https://www.breastcancer.org/treatment/side\\_effects/low\\_blood\\_press](https://www.breastcancer.org/treatment/side_effects/low_blood_press)
- [16] Understanding Blood Pressure Readings. (2016). Consultado el 4 de noviembre de 2020 de <https://www.heart.org/en/health-topics/high-blood-pressure/understanding-blood-pressure-readings>
- [17] Health Threats From High Blood Pressure. (2016). Consultado el 4 de noviembre de 2020 de <https://www.heart.org/en/health-topics/high-blood-pressure/health-threats-from-high-blood-pressure>
- [18] D'Agostino, R. B., Grundy, S., Sullivan, L. M., Wilson, P., & CHD Risk Prediction Group. (2001). Validation of the Framingham coronary heart disease prediction scores: results of a multiple ethnic groups investigation. *Jama*, 286(2), 180-187.
- [19] Mansia, G., De Backer, G., Dominiczak, A., Cifkova, R., Fagard, R., Germano, G. European Society of Cardiology. (2007). 2007 ESH-ESC Guidelines for the management of arterial hypertension: the task force for the management of arterial hypertension of the European Society of Hypertension (ESH) and of the European Society of Cardiology (ESC). *Blood pressure*, 16(3), 135-232
- [20] UK, N. G. C. (2019). Hypertension in Adults: Diagnosis and Management.
- [21] Coca, A., Bertomeu, V., Dalfó, A., Esmatjes, E., Guillén, F., Guerrero, L., ... & Suárez, C. (2007). Automedida de la presión arterial. Documento de Consenso Español 2007. *Revista Española de Geriatria y Gerontología*, 42(2), 115-128.

- 
- [22] Álvarez Batard, G., Ruso, R., Pérez Caballero, M. D., & Fernández Arias, M. A. (2006). Manual clínico del monitoreo ambulatorio de la presión arterial. *Revista Cubana de Medicina*, 45(4), 0-0.
- [23] O'BRIEN, E. (1988). Dippers and non-dippers. *Lancet*, 2, 397.
- [24] Orphanidou, C. (2017). Signal quality assessment in physiological monitoring: state of the art and practical considerations. Springer
- [25] Allen, J. (2007). Photoplethysmography and its application in clinical physiological measurement. *Physiological measurement*, 28(3), R1.
- [26] Sukor, J. A., Redmond, S. J., & Lovell, N. H. (2011). Signal quality measures for pulse oximetry through waveform morphology analysis. *Physiological measurement*, 32(3), 369.
- [27] Narayan, R. (2018). *Encyclopedia of biomedical engineering*. Elsevier.
- [28] Goel, V. K., & Ferrara, L. (2007). *Wiley Encyclopedia of Biomedical Engineering*, vols 1-6. JAMA, 297(14), 1601-1604. ISO 690
- [29] US Food and Drug Administration. (2021). Pulse oximeter accuracy and limitations: FDA safety communication. FDA. February, 19.
- [30] Martínez, G., Howard, N., Abbott, D., Lim, K., Ward, R., & Elgendi, M. (2018). Can photoplethysmography replace arterial blood pressure in the assessment of blood pressure?. *Journal of clinical medicine*, 7(10), 316.
- [31] Choi, W., & Cho, J. H. (2017). Correlation of peak time shift in blood pressure waveform and PPG based on compliance change analysis in RLC Windkessel model. *Current Optics and Photonics*, 1(5), 529-537.
- [32] McEniery, C. M., Cockcroft, J. R., Roman, M. J., Franklin, S. S., & Wilkinson, I. B. (2014). Central blood pressure: current evidence and clinical importance. *European heart journal*, 35(26), 1719-1725.
- [33] Queensland Brain Institute (2019). What is a Neuron? The University of Queensland. Consultado el 13 de junio 2021 de <https://qbi.uq.edu.au/brain/brain-anatomy/what-neuron>
- [34] Urban Child Institute. Consultado el 9 de octubre de 2020 de <http://www.urbanchildinstitute.org>
- [35] Stanford University (2019). Convolutional Neural Networks for Visual Recognition.

- 
- [36] Aggarwal, C. C. (2018). Neural networks and deep learning. Springer, 10, 978-3.
- [37] Salas, R. (2004). Redes neuronales artificiales. Universidad de Valparaíso. Departamento de Computación, 1.
- [38] Wang, Z., & Bovik, A. C. (2009). Mean squared error: Love it or leave it? A new look at signal fidelity measures. *IEEE signal processing magazine*, 26(1), 98-117.
- [39] Qi, J., Du, J., Siniscalchi, S. M., Ma, X., & Lee, C.-H. (2020). On Mean Absolute Error for Deep Neural Network Based Vector-to-Vector Regression. *IEEE Signal Processing Letters*, 1-1. doi:10.1109/lsp.2020.3016837
- [40] Mean squared logarithmic error (MSLE). Peltarion Platform. Peltarion. Consultado el 9 de marzo de 2021 de [https://peltarion.com/knowledge-center/documentation/modeling-view/build-an-ai-model/loss-functions/mean-squared-logarithmic-error-\(msle\)](https://peltarion.com/knowledge-center/documentation/modeling-view/build-an-ai-model/loss-functions/mean-squared-logarithmic-error-(msle))
- [41] Sharma, S. (2017). Activation functions in neural networks. *Towards Data Science*, 6.
- [42] Feng, J., & Lu, S. (2019). Performance analysis of various activation functions in artificial neural networks. In *Journal of physics: conference series* (Vol. 1237, No. 2, p. 022030). IOP Publishing.
- [43] Serengil, S. (2020). ELU as a Neural Networks Activation Function. Consultado el 3 de marzo de 2021 de <https://sefiks.com/2018/01/02/elu-as-a-neural-networks-activation-function/>
- [44] Ruder, S. (2016). An overview of gradient descent optimization algorithms. *arXiv preprint arXiv:1609.04747*.
- [45] Wilamowski, B. M. (2009). Neural network architectures and learning algorithms. *IEEE Industrial Electronics Magazine*, 3(4), 56-63.
- [46] Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., ... & Fei-Fei, L. (2015). Imagenet large scale visual recognition challenge. *International journal of computer vision*, 115(3), 211-252.
- [47] Panarello, B. (2017). Estudio de arquitecturas convolucionales con propiedades de invarianza respecto a la rotación aplicado a la clasificación morfológica de galaxias (Doctoral dissertation, Universidad de Buenos Aires).
- [48] Albawi, S., Mohammed, T. A., & Al-Zawi, S. (2017). Understanding of a convolutional neural network. In *2017 International Conference on Engineering and Technology (ICET)* (pp. 1-6). Ieee.



- [49] Gurucharan, M. K. (2021). Basic CNN Architecture: Explaining 5 Layers of Convolutional Neural Network. UpGrad Blog. Consultado el 15 de enero de 2021, de <https://www.upgrad.com/blog/basic-cnn-architecture/>
- [50] Kiranyaz, S., Avci, O., Abdeljaber, O., Ince, T., Gabbouj, M., & Inman, D. J. (2021). 1D convolutional neural networks and applications: A survey. *Mechanical systems and signal processing*, 151, 107398.
- [51] Ramele, R. (2018). Histogram of gradient orientations of EEG signal plots for brain computer interfaces (Doctoral dissertation, INSTITUTO TECNOLÓGICO DE BUENOS AIRES BUENOS AIRES, ARGENTINA).
- [52] Stergiou, G. S., Alpert, B., Mieke, S., Asmar, R., Atkins, N., Eckert, S., ... & O'Brien, E. (2018). A universal standard for the validation of blood pressure measuring devices: Association for the Advancement of Medical Instrumentation/European Society of Hypertension/International Organization for Standardization (AAMI/ESH/ISO) Collaboration Statement. *Hypertension*, 71(3), 368-374.
- [53] Association for the Advancement of Medical Instrumentation. Consultado el 2 de febrero de 2021 de <https://www.aami.org/>
- [54] Part A - Introduction to AAMI. AAMI. Consultado el 12 de julio de 2021 de <https://www.aami.org/standards.old/participate-in-standards-development.old/effective-participation-in-standards/resources-for-new-committee-members/part-A—introduction-to-aami>
- [55] White, W. B., Berson, A. S., Robbins, C., Jamieson, M. J., Prisant, L. M., Roccella, E., & Sheps, S. G. (1993). National standard for measurement of resting and ambulatory blood pressures with automated sphygmomanometers. *Hypertension*, 21(4), 504-509.
- [56] Stergiou, G. S., Alpert, B., Mieke, S., Asmar, R., Atkins, N., Eckert, S., ... & O'Brien, E. (2018). A universal standard for the validation of blood pressure measuring devices: Association for the Advancement of Medical Instrumentation/European Society of Hypertension/International Organization for Standardization (AAMI/ESH/ISO) Collaboration Statement. *Hypertension*, 71(3), 368-374.
- [57] O'Brien, E., Pickering, T., Asmar, R., Myers, M., Parati, G., Staessen, J., ... & Gerin, W. (2002). Working Group on Blood Pressure Monitoring of the European Society of Hypertension International Protocol for validation of blood pressure measuring devices in adults. *Blood pressure monitoring*, 7(1), 3-17.
- [58] Association for the Advancement of Medical Instrumentation. (2013). Non-invasive sphygmomanometers—Part 2: clinical investigation of automated measurement type

- ANSI/AAMI. ISO 81060-2/ANSI-AAMI, 2nd ed. Arlington, VA: AAMI.
- [59] Xing, X., & Sun, M. (2016). Optical blood pressure estimation with photoplethysmography and FFT-based neural networks. *Biomedical optics express*, 7(8), 3007-3020.
  - [60] Priyanka, K. N. G., Chao, P. C. P., Tu, T. Y., Kao, Y. H., Yeh, M. H., Pandey, R., & Eka, F. P. (2018). Estimating blood pressure via artificial neural networks based on measured photoplethysmography waveforms. In *2018 IEEE SENSORS* (pp. 1-4). IEEE.
  - [61] Slapničar, G., Mlakar, N., & Luštrek, M. (2019). Blood pressure estimation from photoplethysmogram using a spectro-temporal deep neural network. *Sensors*, 19(15), 3420.
  - [62] Lin, W. H., Chen, F., Geng, Y., Ji, N., Fang, P., & Li, G. (2021). Towards accurate estimation of cuffless and continuous blood pressure using multi-order derivative and multivariate photoplethysmogram features. *Biomedical Signal Processing and Control*, 63, 102198.
  - [63] Yousef, Q., Reaz, M. B. I., & Ali, M. A. M. (2012). The analysis of PPG morphology: investigating the effects of aging on arterial compliance. *Measurement Science Review*, 12(6), 266.
  - [64] Wang, C., Yang, F., Yuan, X., Zhang, Y., Chang, K., & Li, Z. (2020). An end-to-end neural network model for blood pressure estimation using ppg signal. In *Artificial Intelligence in China* (pp. 262-272). Springer, Singapore.
  - [65] Esmaelpoor, J., Moradi, M. H., & Kadkhodamohammadi, A. (2020). A multistage deep neural network model for blood pressure estimation using photoplethysmogram signals. *Computers in Biology and Medicine*, 120, 103719.
  - [66] Jeong, D. U., & Lim, K. M. (2021). Combined Deep CNN-LSTM Network-based Multitasking Learning Architecture for Noninvasive Continuous Blood Pressure Estimation using Difference in ECG-PPG Features
  - [67] Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention* (pp. 234-241). Springer, Cham.
  - [68] Johnson, A. E., Pollard, T. J., Shen, L., Li-Wei, H. L., Feng, M., Ghassemi, M., ... & Mark, R. G. (2016). MIMIC-III, a freely accessible critical care database. *Scientific data*, 3(1), 1-9.

- [69] Goldberger, A. L., Amaral, L. A., Glass, L., Hausdorff, J. M., Ivanov, P. C., Mark, R. G., ... & Stanley, H. E. (2000). PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. *circulation*, 101(23), e215-e220.
- [70] HIPAA. (2021). HIPAA Home. HHS.Gov. <https://www.hhs.gov/hipaa/index.html>
- [71] Kluyver, T., Ragan-Kelley, B., Pérez, F., Granger, B. E., Bussonnier, M., Frederic, J., ... & et al. (2016). Jupyter Notebooks-a publishing format for reproducible computational workflows. In *ELPUB* (pp. 87-90).
- [72] Boderó, E. M., López, M. P., Congacha, A. E., Cajamarca, E. E., & Morales, C. H. (2020). Google Colaboratory como alternativa para el procesamiento de una red neuronal convolucional. *Revista ESPACIOS*, 41(07).
- [73] Challenger-Pérez, I., Díaz-Ricardo, Y., & Becerra-García, R. A. (2014). El lenguaje de programación Python. *Ciencias Holguín*, 20(2), 1-13.
- [74] Dessau, R. B., & Pipper, C. B. (2008). "R-project for statistical computing. *Ugeskrift for laeger*, 170(5), 328-330.
- [75] Peralta, D., Herrera-Poyatos, A., & Herrera, F. Un Estudio sobre el Preprocesamiento para Redes Neuronales Profundas y Aplicación sobre Reconocimiento de Dígitos Manuscritos.
- [76] van Gent, P., Farah, H., van Nes, N., & van Arem, B. (2019). HeartPy: A novel heart rate algorithm for the analysis of noisy signals. *Transportation research part F: traffic psychology and behaviour*, 66, 368-378.
- [77] Sviridova, M., & Sakai, K. (2015). Human photoplethysmogram: New insight into chaotic characteristics. *Chaos, Solitons & Fractals*, 77, 53-63.
- [78] Kamal, A. A. R., Harness, J. B., Irving, G., & Mearns, A. J. (1989). Skin photoplethysmography: A review. *Computer Methods and Programs in Biomedicine*, 28, 257-269.
- [79] Smith, S. W. (1997). The scientist and engineer's guide to digital signal processing (Vol. 14, p. 626). San Diego: California Technical Pub.
- [80] Liang, Y., Elgendi, M., Chen, Z., & Ward, R. (2018). An optimal filter for short photoplethysmogram signals. *Scientific data*, 5, 180076.
- [81] A. (2021). ¿Cómo se evalúan los modelos de Machine Learning en el sector sanitario? Sigesa. Consultado el 4 de junio de 2021 de <https://www.siges.com/como-se-evaluan-los-modelos-de-machine-learning-en-el-sector-sanitario/>

- 
- [82] Krishna, H. S. V., Muthukaruppan, M. S., Paruthi, P., & Veezhinathan, M. (2013). Performance evaluation of PPG signal using time-frequency features. *Int J Eng Res Technol*, 2, 988-992.
- [83] Elgendi, M. (2012). On the analysis of fingertip photoplethysmogram signals. *Current cardiology reviews*, 8(1), 14-25.
- [84] Bishop, S. (2011, 6 mayo). Elevated Heart Rate Most Likely Caused by Medical Condition. Mayo Clinic. <https://newsnetwork.mayoclinic.org/discussion/elevated-heart-rate-most-likely-caused-by-medical-condition/>
- [85] Widened Pulse Pressure: Definition, Causes & Treatments. (2019). New Health Advisor. Consultado el 15 de octubre de 2020 de <https://www.newhealthadvisor.org/Widened-Pulse-Pressure.html>
- [86] van Gent, P. (2020, 4 abril). Algorithm functioning — Python Heart Rate Analysis Toolkit 1.2.5 documentation. Python Heart Rate Analysis Toolkit's Documentation. <https://python-heart-rate-analysis-toolkit.readthedocs.io/en/latest/algorithmfunctioning.html#peak-detection>
- [87] Takazawa K TN, Fujita M, Matsuoka O, Saiki T, Aikawa M, Tamura S, Ibukiyama C. Assessment of vasocative agents and vascular aging by the second derivative of photoplethysmogram waveform. *Hypertension*. 1998;32:365-70.
- [88] Imanaga I, Hara H, Koyanagi S, Tanaka K. Correlation between wave components of the second derivative of plethysmogram and arterial distensibility. *Jpn Heart J* 1998;39:775-84.
- [89] Otsuka T, Kawada T, Katsumata M, Ibuki C. Utility of second derivative of the finger photoplethysmogram for the estimation of the risk of coronary heart disease in the general population. *Circ J* 2006; 70: 304-10.
- [90] Mousavi, S. S., Firouzmand, M., Charmi, M., Hemmati, M., Moghadam, M., & Ghorbani, Y. (2019). Blood pressure estimation from appropriate and inappropriate PPG signals using A whole-based method. *Biomedical Signal Processing and Control*, 47, 196-206.
- [91] Giavarina, D. (2015). Understanding bland altman analysis. *Biochemia medica*, 25(2), 141-151.
- [92] Zhang, X., Zou, J., He, K., & Sun, J. (2015). Accelerating very deep convolutional networks for classification and detection. *IEEE transactions on pattern analysis and machine intelligence*, 38(10), 1943-1955.

- [93] Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.
- [94] Rim, B., Sung, N. J., Min, S., & Hong, M. (2020). Deep learning in physiological signal data: A survey. *Sensors*, 20(4), 969.
- [95] Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: a review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 374(2065), 20150202. doi:10.1098/rsta.2015.0202
- [96] Vapnik, V. (2013). *The nature of statistical learning theory*. Springer science & business media.
- [97] Tamura, T. (2021). Cuffless blood pressure monitors: Principles, standards and approval for medical use. *IEICE Transactions on Communications*, 104(6), 580-586.