

UNIVERSIDAD MIGUEL HERNÁNDEZ DE ELCHE

ESCUELA POLITÉCNICA SUPERIOR DE ELCHE

GRADO EN INGENIERÍA DE TECNOLOGÍAS DE
TELECOMUNICACIÓN



"APLICACIÓN DE HERRAMIENTAS DE
APRENDIZAJE AUTOMÁTICO PARA LA
DETECCIÓN TEMPRANA DE LA
MIOCARDIOPATÍA HIPERTRÓFICA"

TRABAJO FIN DE GRADO

Junio - 2025

AUTOR: Javier Prieto Llano

DIRECTOR/ES: Francisco Javier Gimeno Blanes
Raúl García Jiménez



ÍNDICE

1	RESUMEN	10
1	ABSTRACT	11
2	INTRODUCCIÓN	12
2.1	MOTIVACIÓN	12
2.2	EL PROBLEMA CLÍNICO	13
2.3	TÉCNICAS APLICADAS	14
3	ESTADO DEL ARTE	16
3.1	APARTADO CLÍNICO	16
3.1.1	GENERALIDADES EN LA FISIOLÓGÍA Y EL FUNCIONAMIENTO DEL CORAZÓN	16
3.1.2	EL CICLO CARDÍACO	22
3.1.3	ACTIVACIÓN CELULAR	24
3.1.4	ELECTROFISIOLÓGÍA CARDÍACA	25
3.1.5	REGULACIÓN CARDÍACA	27
3.1.6	DESCRIPCIÓN DE LA MIOCARDIOPATÍA HIPERTRÓFICA	27
3.1.7	TÉCNICAS DE DIAGNÓSTICO DE LA PATOLOGÍA	29
3.1.8	TRATAMIENTOS DE LA MIOCARDIOPATÍA HIPERTRÓFICA	65
3.2	MACHINE LEARNING COMO HERRAMIENTA DE DETECCIÓN Y DIAGNÓSTICO TEMPRANO	68
4	MATERIALES Y MÉTODOS	70
4.1	PROGRAMACIÓN ASISTIDA	70
4.1.1	CHAT GPT	71
4.1.2	GEMINI	72
4.1.3	COPILOT	72
4.1.4	ANÁLISIS DE IA	73
4.2	BASE DE DATOS	77
4.3	MÉTODOS EMPLEADOS	78
4.3.1	TEORÍA DE LAS REDES NEURONALES	78
4.3.2	TEORÍA DE LAS MÁQUINAS DE VECTORES DE SOPORTE	82
4.3.3	TEORÍA DE LOS ÁRBOLES DE DECISIÓN	88
4.3.4	TEORÍA DEL ANÁLISIS DE COMPONENTES PRINCIPALES	94
5	EXPERIMENTOS Y RESULTADOS	98
5.1	ENFOQUE EXPERIMENTAL DE LAS REDES NEURONALES (NN)	98
5.2	ENFOQUE EXPERIMENTAL DE LAS MÁQUINAS DE VECTORES DE SOPORTE (SVM)	105
5.3	ENFOQUE EXPERIMENTAL DE LOS ÁRBOLES DE DECISIÓN	109

5.4	ENFOQUE EXPERIMENTAL DEL ANÁLISIS DE COMPONENTES PRINCIPALES (PCA)	113
6	CONCLUSIONES Y PRÓXIMOS PASOS.....	118
7	BIBLIOGRAFÍA.....	121



LISTADO DE FIGURAS

Figura 1: Anatomía del corazón. [3].....	17
Figura 2: Sistema específico de conducción de impulsos eléctricos (SEC). [4]	22
Figura 3: Fases del potencial de acción transmembrana. [4].....	25
Figura 4: Capacidad de automatismo según zonas del miocardio. [4]	26
Figura 5: Diferencia entre un corazón sano y uno afectado por MCH. [7]	29
Figura 6: Árbol Genealógico.[8]	30
Figura 7: de izquierda a derecha; línea isoelectrica, deflexión positiva, negativa e isodifásica. [4]	41
Figura 8: Papel milimetrado del ECG. [4].....	42
Figura 9: Colocación de los electrodos de las extremidades. [23]	42
Figura 10: Ubicación de los electrodos precordiales. [9]	43
Figura 11: Posición anatómica de las derivaciones bipolares de miembros. [24]	45
Figura 12: Sistema Hexaxial de Bailey. [25].....	46
Figura 13: Ondas, intervalos y segmentos del ECG. [3]	47
Figura 14: ECG normal con ritmo regular.[4].....	50
Figura 15: Eje normal y sus desviaciones. [26].....	52
Figura 16: Amplitud del QRS en I y aVF. [4]	54
Figura 17: Cálculo exacto del eje eléctrico. [4].....	54
Figura 18: Triángulo formado en la Figura 17.(imagen de elaboración propia)	55
Figura 19: Características de las ondas.[28].....	57
Figura 20: Fenómenos físicos de las ondas tras un cambio de medio. [29]	59
Figura 21: Análisis con diferentes planos de estudio de un caso con MCH. [27].....	60
Figura 22: Ecocardiograma en modo M en el plano valvular mitral. [27]	61
Figura 23: Ecocardiograma en modo M en el plano de las sigmoideas aórticas. [27] ..	62
Figura 24: Alternativas terapéuticas en la MCH	66
Figura 25: Modelo de McCullogh-Pitts [41]	79
Figura 26:Red Neuronal Artificial. [41]	79
Figura 27:Principio matemático de aprendizaje para Redes Neuronales Artificiales. [41]	80
Figura 28: Clasificador con hiperplano óptimo. [44]	83
Figura 29: Clasificador que no permite separación lineal de los datos. [44].....	85
Figura 30: Aumento de dimensionalidad en los vectores de entrada. [44].....	86
Figura 31: Estructura de un árbol de decisión. [46]	89
Figura 32: Media de VPN en función del porcentaje de pruebas.....	101
Figura 33: Media de VPN en función de la combinación de neuronas.	102
Figura 34: Media de VPP en función del porcentaje de pruebas.	105
Figura 35: Media de VPP en función de la combinación de neuronas.....	105

Figura 36: Representación gráfica del modelo de SVM visto desde el plano ZY.....	107
Figura 37: Representación gráfica del modelo de SVM visto desde el plano XY	107
Figura 38: Representación gráfica del modelo de SVM visto desde el plano ZX.....	108
Figura 39: Mapa de calor de la media de VPN para pacientes del grupo 0.....	111
Figura 40: Mapa de calor de la media de VPP para pacientes del grupo 3.	111
Figura 41: Mapa de calor de la media de VPN para pacientes del grupo 0.....	112
Figura 42: Mapa de calor de la media de VPP para pacientes del grupo 3.	112
Figura 43: Predicción de la red neuronal con previo PCA para pacientes del grupo 3.	114
Figura 44: Resultado experimental al aplicar PCA sobre una máquina de vectores de soporte.	115
Figura 45: Predicción del árbol de criterio Gini con previo PCA para pacientes del grupo 1.....	116
Figura 46: Predicción del árbol de criterio de Entropía con previo PCA para pacientes del grupo 3.....	117



LISTADO DE TABLAS

Tabla 1: Estructuras cardíacas en visión anterior e inferior.	18
Tabla 2: Análisis de las ondas, intervalos y segmentos.....	51
Tabla 3: Velocidad de Propagación del sonido en diferentes tejidos y materiales.....	57
Tabla 4: Análisis de Copilot	74
Tabla 5: Análisis de Gemini	75
Tabla 6: Análisis de Chat GPT	76
Tabla 7: Conjunto pequeño de datos para la comprensión del criterio Gini.	90
Tabla 8: Conjunto pequeño de datos para la comprensión del criterio de Entropía.....	92
Tabla 9: Configuración Red Neuronal óptima para pacientes del Grupo 0.....	100
Tabla 10: Resultados Red Neuronal óptima para pacientes del Grupo 0	100
Tabla 11: Configuración Red Neuronal óptima para pacientes del Grupo 1.....	102
Tabla 12: Resultados Red Neuronal óptima para pacientes del Grupo 1	102
Tabla 13: Configuración Red Neuronal óptima para pacientes del Grupo 2.....	103
Tabla 14: Resultados Red Neuronal óptima para pacientes del Grupo 2	103
Tabla 15: Configuración Red Neuronal óptima para pacientes del Grupo 3.....	104
Tabla 16: Resultados Red Neuronal óptima para pacientes del Grupo 3	104
Tabla 17: Mejores configuraciones y resultados de los árboles de decisión.....	110
Tabla 18: Resultados experimentales al aplicar PCA sobre las redes neuronales.....	114
Tabla 19: Resultados experimentales al aplicar PCA sobre árboles de decisión con criterio GINI	115
Tabla 20: Resultados experimentales al aplicar PCA sobre árboles de decisión con criterio de Entropía	116

LISTADO DE ACRÓNIMOS

HCM / MCH	Miocardiopatía Hipertrófica
ECG	Electrocardiograma
ML	Machine Learning
DL	Deep Learning
NN	Neural Networks
SVM	Support Vectors Machine
AV	Auriculoventricular
SEC	Sistema Específico de Conducción
LPM	Latidos Por Minuto
SNA	Sistema Nervioso Autónomo
VI	Ventrículo Izquierdo
OTSVI	Obstrucción del Tracto de Salida del Ventrículo Izquierdo
ESC	Sociedad Europea de Cardiología
RM	Resonancia Magnética
MS	Muerte Súbita
IC	Insuficiencia Cardíaca
HTA	Hipertensión Arterial
DdP	Diferencia de Potencial
FC	Frecuencia Cardíaca
SCA	Síndrome Coronario Agudo
FA	Fibrilación Auricular
MLP	Multi-Layer Perceptron
PCA	Principal Component Analysis
SVD	Descomposición de Valores Singulares



1 RESUMEN

La miocardiopatía hipertrófica (MCH) es una enfermedad cardíaca genética que puede permanecer asintomática durante años, aumentando el riesgo de eventos adversos graves si no se detecta de forma temprana. Este trabajo de fin de grado propone un enfoque basado en técnicas de aprendizaje automático para apoyar la predetección de la MCH a partir de datos clínicos con múltiples variables. Para ello, he estudiado y comparado diversas herramientas de programación asistida, empleando la que he considerado la mejor, tras un amplio análisis de todas.

A partir de una base de datos compuesta por más de mil pacientes, se ha aplicado el Análisis de Componentes Principales (PCA) como técnica de reducción de dimensionalidad con el objetivo de conservar la máxima varianza explicada y facilitar la interpretación del modelo. Anteriormente, se han desarrollado y evaluado diversos modelos predictivos, hemos comparado todos los modelos entre sí y hemos comparado sus múltiples configuraciones para encontrar la óptima, los modelos empleados son: redes neuronales, máquinas de vectores de soporte y árboles de decisión. Los modelos han sido evaluados a través de métricas clínicas clave como la sensibilidad, especificidad, valor predictivo positivo (VPP) y valor predictivo negativo (VPN). Los resultados sugieren que la combinación de PCA con modelos de machine learning puede ser una herramienta más eficiente para apoyar el diagnóstico temprano de la MCH.

1 ABSTRACT

Hypertrophic cardiomyopathy (HCM) is a genetic heart disease that can remain asymptomatic for years, significantly increasing the risk of serious adverse events if not detected early. This degree thesis presents an approach based on machine learning techniques to support the early detection of HCM using clinical data with multiple variables. To this end, I have studied and compared various assisted programming tools, using the one I considered the best after a thorough analysis of all of them.

Using a database comprising over a thousand patients, Principal Component Analysis (PCA) has been applied as a dimensionality reduction technique to retain the maximum explained variance and enhance the interpretability of the models. Several predictive models have been developed and evaluated, we compared all the models with each other, as well as their multiple configurations, to identify the optimal setup. The models used were neural networks, support vector machines, and decision trees. These models were assessed using key clinical performance metrics such as sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV). The results suggest that combining PCA with machine learning models can be a more efficient tool to support the early diagnosis of HCM.

2 INTRODUCCIÓN

En este apartado vamos a tratar los motivos que me llevan a hacer este trabajo de fin de grado, la patología de la miocardiopatía hipertrófica a grandes rasgos, ¿en qué consiste? ¿Cuáles son sus principales síntomas? ¿Cómo se puede dar la enfermedad? ¿Cuáles son las pruebas de diagnóstico? Finalmente, cerraremos el apartado con las técnicas de machine learning que vamos a emplear.

2.1 MOTIVACIÓN

Los motivos para llevar a cabo este trabajo de fin de grado son muchos:

Relevancia Clínica y la Necesidad Urgente de un Diagnóstico Temprano: la miocardiopatía hipertrófica es una afección cardíaca de naturaleza genética que se desarrolla de forma silenciosa, esto puede llegar a provocar la muerte súbita cardíaca de un porcentaje de quienes la padecen, entre ellos personas jóvenes y atletas. Su silencioso desarrollo y difícil diagnóstico hacen fundamental su temprana detección, para así poder mejorar la calidad de vida del paciente. La investigación, junto con el desarrollo de las tecnologías de diagnóstico, nos pueden ayudar a detectar la enfermedad antes de que aparezcan síntomas graves.

Oportunidad para Aplicar Tecnologías Emergentes: a lo largo del desarrollo de este trabajo de fin de grado usaremos un gran conjunto de datos médicos, aplicaremos sobre este campo de datos técnicas de Machine Learning con el objetivo de mejorar el proceso de diagnóstico de la HCM. Los modelos predictivos basados en datos genéticos, electrocardiogramas, ecografías, análisis de sangre, etc, pueden facilitar la identificación temprana de la patología e incluso, la predetección de la enfermedad en personas asintomáticas.

Avances en la Cardiología Computacional: las diferentes herramientas empleadas de aprendizaje automático pueden ayudar a mejorar la comprensión acerca de la fisiología del corazón y sus diferentes afecciones. Además, estos modelos pueden facilitar el entendimiento de cómo las variaciones genéticas afectan a la estructura del corazón, también pueden realizar predicciones acerca del avance de la patología y evaluar posibles riesgos de cara al futuro.

Impacto en la Salud Pública y Reducción de Costos Sanitarios: la temprana detección de la miocardiopatía hipertrófica supone reducir la incidencia de episodios graves provocados por esta patología, disminuyendo la necesidad de aplicar cirugías invasivas, ablaciones septales con alcohol o implantación de marcapasos, debido a arritmias o insuficiencias cardíacas. Esto no solo mejora la calidad de vida del paciente, sino que supone una reducción de los costes sanitarios asociados a la atención a largo plazo.

Contribución y Ayuda al Personal Sanitario: este trabajo está motivado, en parte, por la posibilidad de ayudar al personal sanitario, desarrollando una herramienta que facilite el proceso de diagnóstico y con la menor posibilidad de fallo, de esta manera, podrán diagnosticar y recetar desde la máxima fiabilidad, lo que es una gran ventaja dado que la MCH es una fisiopatología de complicado diagnóstico.

2.2 EL PROBLEMA CLÍNICO

La miocardiopatía hipertrófica (HCM) es una enfermedad hereditaria que consiste en un engrosamiento del músculo cardíaco, provocando fatiga, cambios en el sistema cardio eléctrico, arritmias, e incluso, muerte súbita. Estudios epidemiológicos de los últimos años estiman la incidencia patológica en una de cada 500 personas en la población general. Esta patología está relacionada con mutaciones en genes que codifican proteínas del sarcómero cardíaco, el disco Z y proteínas controladoras del calcio, algunos de estos genes son bien conocidos, como los que afectan la unión de la proteína c a miosina (el gen MYBPC3). Sin embargo, muchas otras variantes también pueden ser la causa de esta enfermedad.

Para identificar algunas posibles causas, se necesitan conjuntos de datos muy grandes, generalmente se usan los que contienen información genética, esto hace que sea muy difícil encontrar patrones en los datos y relaciones entre las variables [1].

Para llevar a cabo el diagnóstico de la HCM es conveniente la realización de varias pruebas como: un ecocardiograma, un electrocardiograma (ECG) o pruebas de esfuerzo entre otras cosas, todo esto, debe ir de la mano de un análisis genético y del conocimiento de los antecedentes familiares que hayan padecido de HCM, o que hayan fallecido de forma repentina por una muerte súbita.

2.3 TÉCNICAS APLICADAS

Machine Learning (ML): hoy en día, los sistemas inteligentes que ofrecen capacidades de inteligencia artificial suelen basarse en el aprendizaje automático. El aprendizaje automático describe la capacidad de los sistemas para aprender a partir de datos de entrenamiento específicos del problema, con el fin de automatizar la construcción de modelos analíticos y resolver tareas asociadas. Algunas técnicas de machine learning son: regresión lineal, regresión logística, árboles de decisión, bosques aleatorios, máquinas de vectores de soporte, k-vecinos cercanos, redes neuronales, análisis UMAP, etc. Entre ellas he optado por utilizar las siguientes:

- Redes neuronales (neural networks): modelo matemático inspirado en el comportamiento biológico de las neuronas y en la estructura del cerebro, y que es utilizada para resolver un amplio rango de problemas.
- Máquina de vectores de soporte (support vector machine): es un algoritmo que clasifica los datos mediante la búsqueda de una línea o un hiperplano óptimos que maximice la distancia entre cada clase en un espacio N-dimensional.
- Árboles de Decisión: es un algoritmo de aprendizaje supervisado no paramétrico, que se utiliza tanto para tareas de clasificación como de regresión. Tiene una estructura jerárquica de árbol, que consta de un nodo raíz, ramas, nodos internos y nodos hoja. El aprendizaje de árboles de decisión emplea una estrategia de divide y vencerás realizando una búsqueda codiciosa para identificar los puntos de división óptimos dentro de un árbol. Este proceso de división se repite de forma descendente y recursiva hasta que todos o la mayoría de los registros se hayan clasificado con etiquetas de clase específicas.

Deep Learning (DL): el aprendizaje profundo es un concepto dentro del aprendizaje automático basado en redes neuronales artificiales. En muchas aplicaciones, los modelos de aprendizaje profundo superan a los modelos de aprendizaje automático tradicionales y a los enfoques clásicos de análisis de datos.

Los avances en el aprendizaje automático han permitido el surgimiento de sistemas inteligentes con capacidades cognitivas similares a las humanas, los cuales han

transformado sectores como la toma de decisiones empresariales, lo que mejorará la productividad. La capacidad del sistema inteligente para la resolución de problemas y mejora de la productividad se basa en modelos analíticos, que generan predicciones en base a recomendaciones o resultados similares [2].



3 ESTADO DEL ARTE

En este apartado trataremos de explicar la funcionalidad general del corazón de la forma más amplia posible, para así comprender el funcionamiento de este y viendo todo lo que es capaz de hacer, también hablaremos sobre la patología que nos compete, además de las técnicas de diagnóstico de esta enfermedad y los posibles tratamientos en caso de padecerla. También veremos cómo los procesos de aprendizaje automático son de vital importancia en el temprano diagnóstico o predetección de las enfermedades cardiovasculares, y como pueden estas investigaciones mejorar no solo la calidad de vida del paciente, sino también la administración de recursos por parte del personal sanitario.

3.1 APARTADO CLÍNICO

En este apartado introducimos los conceptos clínicos de la patología, acompañando esta información de las técnicas que se aplican para su diagnóstico y sus características, también aclararemos conceptos en relación al corazón, fisiológica y funcionalmente, toda esta información desde un punto de vista biológico nos aportará conocimientos para entender la enfermedad y los métodos de detección de la misma, lo que será fundamental para el estudio acerca de la anticipación de la MCH.

3.1.1 GENERALIDADES EN LA FISIOLOGÍA Y EL FUNCIONAMIENTO DEL CORAZÓN

El corazón es un órgano muscular hueco que actúa en el organismo como una doble bomba (Figura 1). Por un lado, impulsa la sangre hacia los pulmones para su oxigenación, dando lugar a la circulación pulmonar y, por otro, bombea la sangre oxigenada hacia el resto de zonas del organismo; originando la circulación sistemática. Cada una de estas bombas está formada por dos cavidades: una aurícula y un ventrículo [3].

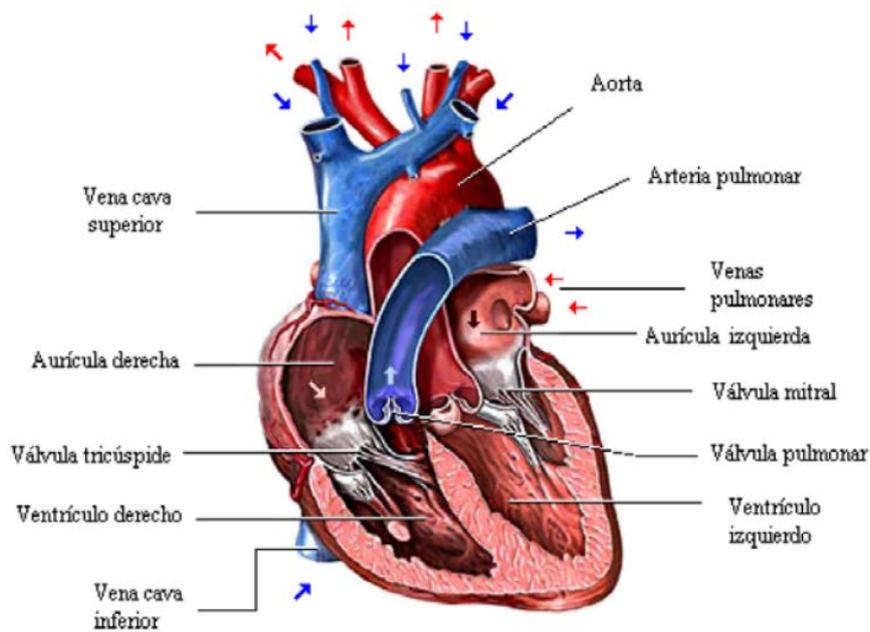
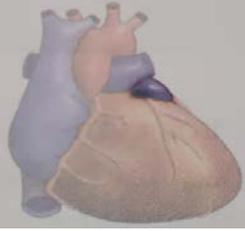
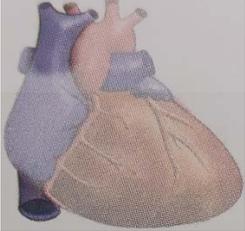
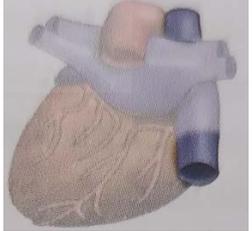
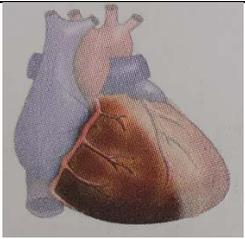
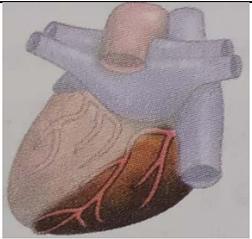
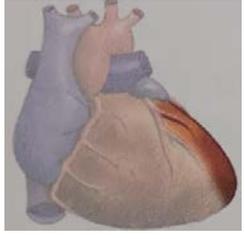
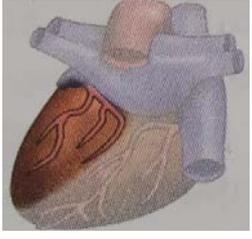
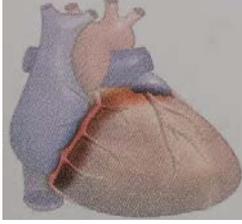
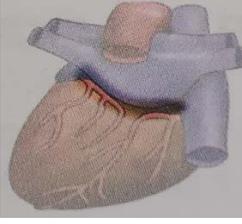
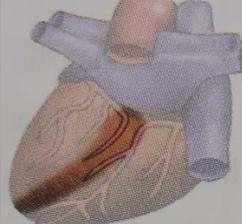
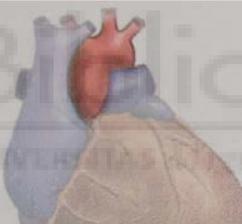
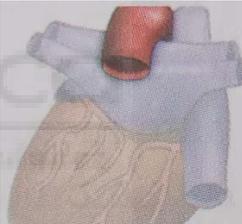


Figura 1: Anatomía del corazón. [3]

Atendiendo a su posición en el cuerpo se pueden distinguir dos caras; la cara anterior o esternal; y la posterior o diafragmática. La cara anterior está compuesta mayoritariamente por el ventrículo derecho y en menor medida por el ventrículo izquierdo. Ambos están separados por el surco interventricular, que corresponde internamente con el tabique o septum. En la cara inferior se observan ambos ventrículos en similares proporciones, así como la continuación del surco interventricular. En ambas caras se observa el surco auriculoventricular, interrumpido en la cara anterior por la salida de la arteria pulmonar y posterior a esta, la aorta. El ápex o vértice inferior corresponde a la base cardíaca, que está orientada hacia la zona posterior y derecha del cuerpo y está constituida por las aurículas. A su derecha se encuentra la entrada de las venas cavas y el resto de la base la compone la aurícula izquierda y las salidas de las cuatro venas pulmonares. En esta zona también se distingue la depresión que corresponde con el tabique interauricular. En la Tabla 1 se resumen las estructuras citadas [4].

Tabla 1: Estructuras cardíacas en visión anterior e inferior.

	Visión Anterior	Visión Inferior
Aurícula izquierda		
Aurícula derecha		
Venas pulmonares		
Venas cavas superior e inferior		
Ventrículo derecho		
Ventrículo izquierdo		

Surco auriculoventricular		
Surco interventricular		
Arterias Pulmonares		
Arteria Aorta		

Histológicamente, el corazón está compuesto por tres tipos de células; fibrosas, contráctiles y de conducción. Las primeras forman el esqueleto cardíaco, un armazón fibroso de tejido mayormente conectivo cuya función es estructural. Las células contráctiles forman el tejido muscular y se ocupan de la parte mecánica. Y, por último, las células de conducción están especializadas en el inicio y propagación de los estímulos eléctricos [4].

El músculo cardíaco está compuesto por tres capas; el endocardio, el miocardio y el epicardio. El espesor viene determinado por el miocardio, mientras que las otras dos capas tienen función de revestimiento interno y externo.

Externamente, el corazón se encuentra dentro de una bolsa denominada pericardio, que dispone de dos hojas; una interna y otra externa.

Morfología interna

El corazón está compuesto por cuatro cavidades; dos en el lado derecho y dos en el izquierdo. En la parte superior del corazón se sitúan las aurículas y en la inferior se disponen los ventrículos. Las aurículas están separadas por el tabique interauricular, y los ventrículos por el tabique interventricular o septum. La aurícula y ventrículo derechos se comunican entre sí a través de la válvula tricúspide. Así mismo, las cavidades izquierdas lo hacen mediante la válvula mitral. Las válvulas de salida de los ventrículos son la válvula pulmonar (ventrículo derecho) y la válvula aórtica (ventrículo izquierdo) [4].

Circulación sistémica y pulmonar

La aurícula derecha es la encargada de recoger la sangre corporal (no oxigenada) que proviene de las venas cavas y del seno venoso coronario, que recoge la sangre venosa del corazón. Esta sangre pasa al ventrículo derecho a través de la válvula tricúspide y desde esta cavidad se dirige a la arteria pulmonar, después de pasar por la válvula con el mismo nombre. Toda esta sangre llega a los pulmones donde se oxigena para luego volver por las venas pulmonares al corazón izquierdo. Desde la aurícula izquierda pasa al ventrículo a través de la válvula mitral para finalmente eyectarse a la circulación sistémica a través de la válvula aórtica [4].

Circulación coronaria

El sistema arterial del miocardio está formado por la arteria coronaria derecha, izquierda y sus ramas; ambas se originan en los senos de Valsalva correspondientes. El tronco de la arteria coronaria izquierda se divide a su vez en la arteria descendente anterior; que irriga la parte anterolateral, apical y septal del ventrículo izquierdo; y la arteria circunfleja que irriga la cara lateral y posterior de dicho ventrículo, así como la aurícula izquierda. La arteria coronaria derecha es la encargada junto con sus ramas de irrigar a la aurícula y ventrículo derechos, a la parte inferior del ventrículo izquierdo y a las zonas del sistema de conducción (nodo sinusal en el 50 % de las ocasiones y nodo auriculoventricular), como se analiza en [4].

Sistema de conducción

El sistema de conducción cardíaco es responsable del inicio y la conducción de los estímulos eléctricos a fin de obtener contracciones rítmicas y ordenadas del corazón (observar Figura 2).

El nodo sinusal se encuentra en la parte superior de la aurícula derecha y anterior a la vena cava superior a nivel epicárdico. Su principal característica es el automatismo de sus células, que generan entre 60-100 impulsos por minuto. Estos impulsos se distribuyen para activar la aurícula derecha a través de los llamados tractos internodulares, y se dirigen a la aurícula izquierda por el tracto interauricular (fascículo de Bachmann).

El nodo auriculoventricular se encuentra en la parte baja de la aurícula derecha, cercano al tabique interauricular. Su principal función es transmitir los impulsos desde las aurículas a los ventrículos y es la única vía de conexión entre ambas estructuras. Se encarga de realizar el retraso fisiológico de los impulsos (para sincronizar los movimientos de contracción, sístole y diástole), así como limitar la cantidad de estímulos que llegan a los ventrículos (control de taquiarritmias). Existe una desconexión funcional entre aurículas y ventrículos (esqueleto cardíaco) cuyo fin es que la única conexión entre aurículas y ventrículos sea el nodo AV, lo que permite una perfecta sincronía entre los estímulos eléctricos y los movimientos mecánicos del corazón [4].

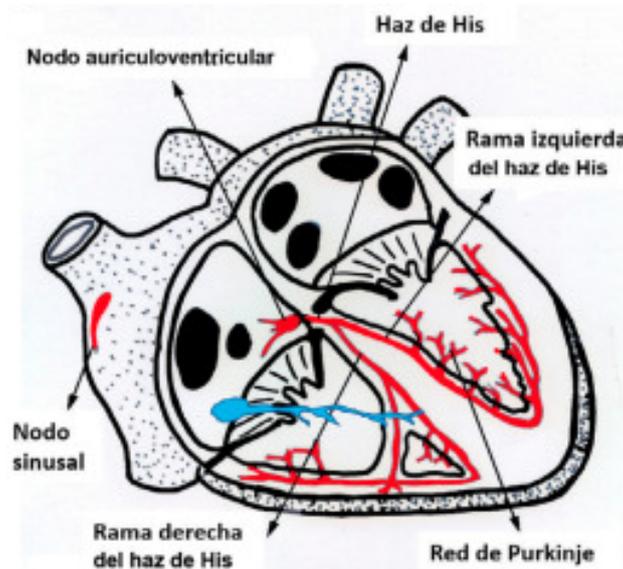


Figura 2: Sistema específico de conducción de impulsos eléctricos (SEC). [4]

El Haz de His continúa al nodo AV, tiene un trayecto común y posteriormente se bifurca en sus ramas derecha e izquierda que recorren el septo interventricular. La rama izquierda se divide en dos fascículos, anterior y posterior.

Las células de Purkinje son la parte distal del sistema específico de conducción tras las ramas del Haz de His. Son el tejido de transición que permite el paso de impulsos eléctricos de las células encargadas de conducirlos al miocardio contráctil. Adquiere una morfología de red que ocupa los dos tercios cercanos al ápex de los ventrículos. Esta configuración permite que el impulso eléctrico alcance una gran cantidad de miocardio en un espacio corto de tiempo, produciendo una contracción sincrónica y efectiva de los ventrículos.

3.1.2 EL CICLO CARDÍACO

Como se ha mencionado anteriormente, el corazón es una bomba de distribución de sangre por el organismo. Esta actividad cardíaca se realiza de forma cíclica; las aurículas se encargan de recoger la sangre para proveer a los ventrículos y que estos puedan distribuir la sangre tanto a la circulación sistémica como a la pulmonar, tal y como se explica en [4]. A este proceso de llenado y vaciado de las cavidades se le conoce como latido cardíaco y, al conjunto de eventos eléctricos, mecánicos y hemodinámicos que ocurren durante cada latido se le denomina ciclo cardíaco.

El ciclo cardíaco consta de dos fases; sístole y diástole. La diástole es el proceso de llenado que coincide con la relajación muscular. La sístole consiste en el vaciado por contracción muscular y posterior eyección.

Ciclo auricular

Las aurículas se encargan de recibir la sangre para, posteriormente, proceder con el llenado ventricular. Existe un periodo prolongado de relajación auricular (diástole) y uno corto de contracción o sístole auricular. Alrededor del 80-90% del volumen sanguíneo llega a los ventrículos de forma pasiva durante la diástole auricular (de ahí su mayor duración) y, el volumen restante, se desplaza gracias a la contracción o sístole auricular.

Ciclo ventricular: Diástole ventricular

Corresponde al periodo de relajación y llenado, y está dividida en dos fases:

- Durante la primera, llamada relajación isovolumétrica, la presión de los ventrículos desciende por debajo de la presión aórtica, provocando un flujo retrógrado que favorece el cierre de la válvula aórtica (originando el segundo ruido cardíaco). Por décimas de segundo las válvulas de salida y de entrada están cerradas y los ventrículos relajados, por lo que el volumen se mantiene constante. Cuando la presión de los ventrículos desciende lo suficiente se abren las válvulas auriculoventriculares iniciando la segunda fase.
- La segunda fase, llenado ventricular, se divide a su vez en tres; llenado rápido, llenado lento y llenado activo. Tras la contracción, las paredes del ventrículo se relajan y se expanden haciendo un efecto de succión (por descenso de presiones) sobre las aurículas, moviendo la mayor parte del volumen. A ello le sigue el llenado lento, en el que ambas cavidades se comportan como una y el paso de sangre es mínimo. Finalmente, el llenado activo coincide con la contracción auricular, que mueve el resto del volumen hacia los ventrículos.

Sístole ventricular

Es el periodo de contracción y vaciado, consta de dos fases; la fase de contracción isovolumétrica, en la que los ventrículos empiezan a contraerse cerrando las válvulas auriculoventriculares (produciendo el primer ruido cardíaco), pero sin tener suficiente presión para abrir las válvulas de salida; y el periodo de eyección, en el que la presión del ventrículo supera la de la arteria. El segundo periodo se divide a su vez en tres fases; expulsión rápida, expulsión lenta y protodiástole, tal y como se recoge en [4].

3.1.3 ACTIVACIÓN CELULAR

La actividad eléctrica de las células miocárdicas se origina a partir de desplazamientos iónicos. En condiciones de reposo, las concentraciones de sodio y calcio son elevadas en el medio extracelular, mientras que el potasio predomina en el interior celular. Esta distribución iónica genera una diferencia de cargas eléctricas, con el exterior de la célula más positivo respecto al interior, que se encuentra cargado negativamente. Este fenómeno se denomina diferencial de potencial de membrana, que en reposo presenta un valor aproximado de -90 milivoltios [4].

Las células contráctiles permanecen en estado de reposo hasta que reciben un estímulo eléctrico que supere su potencial umbral. En caso de que la señal eléctrica no alcance dicho límite, la célula no se activará y, en consecuencia, no se genera la contracción muscular.

Cuando una célula miocárdica es activada, se abren sus canales de sodio, permitiendo la entrada pasiva de este ion al interior celular, que en condiciones normales presenta una carga eléctrica negativa. Este ingreso de sodio facilita el intercambio de otros iones y da inicio al proceso de contracción celular. Este fenómeno se denomina despolarización.

Tras la sístole, la célula cardíaca activa el intercambio de iones mediante la bomba de sodio-potasio, que restituye el potencial negativo del interior celular respecto al medio extracelular. Este proceso coincide con la relajación de las células, lo que permite la diástole de la cavidad cardíaca. Esta fase se conoce como repolarización.

La secuencia de despolarización seguida de repolarización conforma el denominado potencial de acción transmembrana, que representa el intervalo durante el cual la célula

se encuentra activada. En las fases iniciales de este potencial, las células cardíacas son incapaces de responder a estímulos adicionales, hallándose en lo que se conoce como periodo refractario absoluto. No obstante, durante la repolarización, a medida que se inicia la salida de potasio y se restablece la negatividad intracelular, un estímulo de gran intensidad puede inducir una nueva despolarización celular, incluso si la célula aún no ha alcanzado su estado de reposo. Esta etapa se denomina periodo refractario relativo.

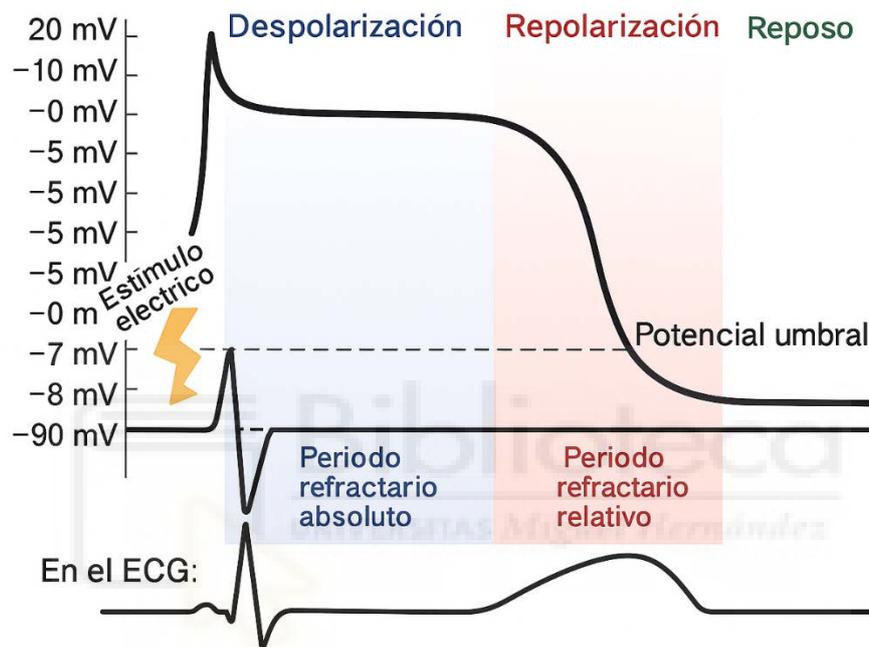


Figura 3: Fases del potencial de acción transmembrana. [4]

3.1.4 ELECTROFISIOLOGÍA CARDÍACA

Para comprender el funcionamiento cardíaco es importante conocer las características electrofisiológicas del mismo.

Como se ha visto anteriormente, gran parte del tejido cardíaco es capaz de contraerse, existiendo además un sistema específico encargado del inicio y conducción de los estímulos eléctricos.

Este sistema de conducción tiene unas propiedades específicas respecto al resto de células no cardíacas.

Automatismo o cronotropismo

Se denomina automatismo a la capacidad de ciertas células cardíacas para autoestimularse o generar despolarizaciones rítmicas que determinan la frecuencia cardíaca. Esta propiedad es característica de las estructuras que conforman el sistema específico de conducción, las cuales presentan un potencial de acción transmembrana distinto al del músculo cardíaco.

Como se cita en [4], en condiciones normales el nodo sinusal se encarga de controlar la frecuencia cardíaca, llevando a cabo entre 60-100 latidos por minuto (lpm). En condiciones anómalas otras zonas del sistema de conducción pueden tomar el mando del automatismo (Figura 4). El nodo AV realiza alrededor de 45-50 lpm, el Haz de His en torno a 40-45 lpm, las fibras de Purkinje entre 35-40 lpm y las células miocárdicas no especializadas en la conducción de los impulsos sobre 30-35 lpm.

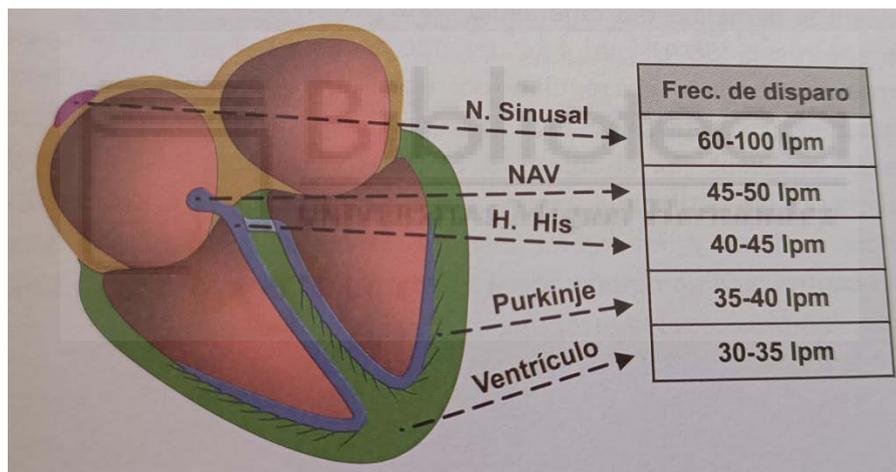


Figura 4: Capacidad de automatismo según zonas del miocardio. [4]

Excitabilidad o batmotropismo

Se denomina excitabilidad a la capacidad que tienen las células de responder a estímulos generando una respuesta.

Conductibilidad o dromotropismo

Es la capacidad que tienen las células cardíacas para conducir el potencial de acción excitando a las células contiguas.

Contractilidad o inotropismo

Se denomina contractilidad a la velocidad de acortamiento de las fibras musculares independientemente de su longitud. En resumen, es la capacidad que tienen las fibras para contraerse con más o menos intensidad, este fenómeno varía con la frecuencia cardíaca, a mayor frecuencia mayor fuerza de contracción y a la inversa.

3.1.5 REGULACIÓN CARDÍACA

El corazón posee un sistema de regulación autónomo propio, es decir, no se controla de forma voluntaria. Este sistema genera descargas y contracciones rítmicas en función de la demanda del organismo. Esta regulación depende del sistema nervioso autónomo (SNA), dicho sistema se divide en sistema nervioso simpático y parasimpático. El primero de ellos aumenta la contractilidad y la frecuencia cardíaca, mientras que, el SNA parasimpático lleva a cabo efectos contrarios al anterior, disminuye la frecuencia cardíaca y la contractilidad. En condiciones normales el control lo lleva el SN parasimpático [4].

3.1.6 DESCRIPCIÓN DE LA MIOCARDIOPATÍA HIPERTRÓFICA

La miocardiopatía hipertrófica (MCH) es la enfermedad miocárdica primaria genética más frecuente, con una prevalencia estimada en la población adulta de 1/500. Está causada por mutaciones en diversos genes que codifican para proteínas del sarcómero y caracterizada por una hipertrofia del ventrículo izquierdo (VI) simétrica o asimétrica con o sin obstrucción al tracto de salida del VI (OTSVI), en ausencia de otra enfermedad cardíaca o sistémica capaz de producirla. La fisiopatología de la MCH es compleja y multifactorial; destacando la disfunción diastólica, OTSVI, isquemia miocárdica, insuficiencia mitral y arritmias como los principales mecanismos que contribuyen a la sintomatología. La manifestación clínica más llamativa de la MCH es la muerte súbita de origen cardíaco, de la que se describe una incidencia en torno al 2-3% anual en adultos. Los síntomas clínicos más relevantes son disnea, dolor torácico, presíncope o síncope [5].

La MCH conlleva un aumento considerable del grosor de la pared del ventrículo izquierdo (VI) (observar Figura 5), que no puede explicarse por condiciones de carga anómalas. La Sociedad Europea de Cardiología (ESC) delimita el engrosamiento según sean pacientes

adultos o pediátricos. Para los primeros, establece que sea ≥ 15 mm y para los segundos en ≥ 2 desviaciones estándar con respecto a la media del grosor para su edad [6].

Una mutación presente, en un paciente con MCH, disminuye la incertidumbre diagnóstica y permite hacer un seguimiento adecuado, como se explica en [6]. El problema surge cuando presenta mutación y no hay clínica de MCH, o cuando hay clínica de MCH y no aparece mutación, ya que la ausencia de la misma no excluye la enfermedad.

En la MCH se observan numerosas anormalidades electrocardiográficas, que están determinadas por la extensión, el grado y la distribución de la hipertrofia del miocardio comprometido, la presencia de fibrosis y/o necrosis del músculo cardíaco y la aparición de trastornos de la conducción intraventricular.

Para el diagnóstico de la patología, el ecocardiograma transtorácico es el método no invasivo más utilizado que nos aporta información sobre la morfología y el tipo de MCH, la función ventricular diastólica y sistólica, la presencia y gravedad de la OTSVI (Obstrucción del Tracto de Salida del Ventrículo Izquierdo), el grado de insuficiencia mitral, algunos aspectos fisiopatológicos sobre la respuesta aguda y crónica a medidas terapéuticas, el pronóstico y el screening en familias de portadores de la enfermedad con el fin de identificar formas subclínicas. Sin embargo, la naturaleza tridimensional de la resonancia magnética cardíaca (cardio-RM) permite una definición más precisa del lugar y la extensión de la hipertrofia, en especial en la pared anterolateral, basal y en el ápex; zonas difíciles de evaluar por ecocardiografía, lo que con frecuencia hace que se pase por alto el diagnóstico [5].

Miocardiopatía hipertrófica

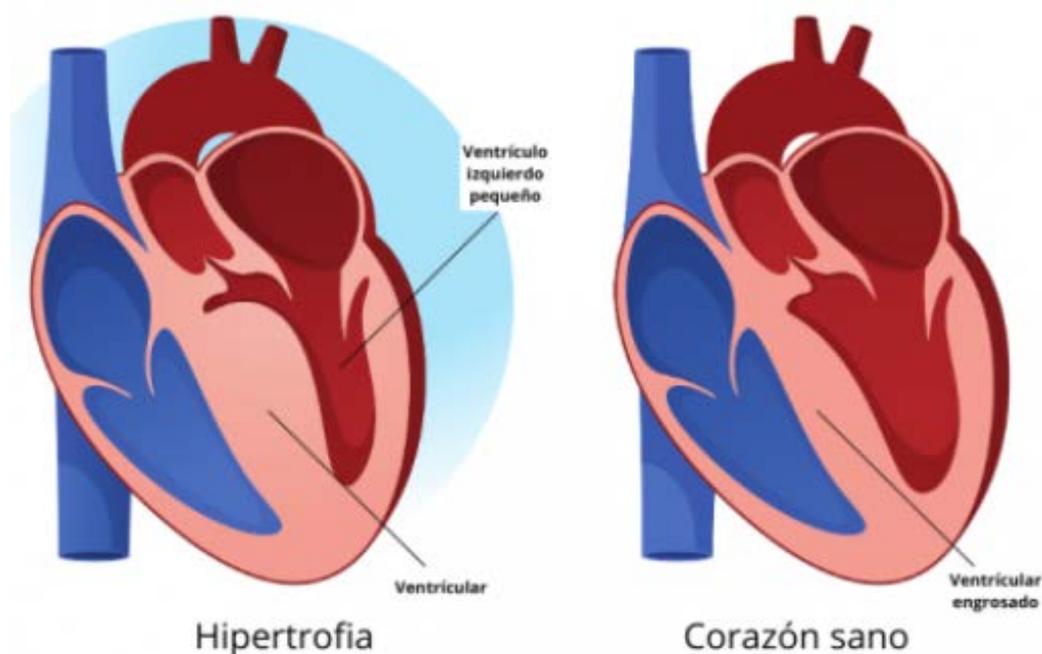


Figura 5: Diferencia entre un corazón sano y uno afectado por MCH. [7]

3.1.7 TÉCNICAS DE DIAGNÓSTICO DE LA PATOLOGÍA

El paso inicial en la valoración de una miocardiopatía con un posible origen genético será la elaboración de una historia familiar detallada. Se recomienda la elaboración de un árbol familiar que incluya al menos tres generaciones. Será necesario identificar enfermedades cardíacas conocidas o síntomas que pudieran ser de origen cardíaco: muerte súbita (MS), insuficiencia cardíaca (IC), trasplante, implante de marcapasos o desfibrilador implantable e ictus en individuos jóvenes [8]. Realizar el árbol genealógico también puede permitir identificar un patrón de herencia (Figura 6). La ausencia de datos sobre familiares afectados no descarta por sí misma la posibilidad de enfermedad genética, ya que pueden existir mutaciones.

En el caso de la MCH, se encuentra un origen genético en un 60% de los casos, las variantes patogénicas identificadas con mayor frecuencia se encuentran en los genes MYH7 y MYBPC3 y suponen el 80% de los casos en los que se alcanza un diagnóstico genético. Existen hasta 57 genes que se han asociado a la MCH tanto de origen en el sarcómero como relacionada con síndromes que presentan hipertrofia ventricular

izquierda. De estos, 8 genes han sido clasificados como definitivos para el diagnóstico genético de la MCH. Los paneles diagnósticos genéticos deben incluir aquellos genes sobre los que existe evidencia robusta de asociación con la enfermedad, mientras que se debe ser más cauteloso a la hora de incluir genes con evidencia moderada o menor para evitar el diagnóstico de variantes de significado incierto. También es necesario incluir genes que se asocian a enfermedades con fenotipo de hipertrofia ventricular (fenocopias) como el gen GLA de la enfermedad de Fabry, como bien explican en [8].

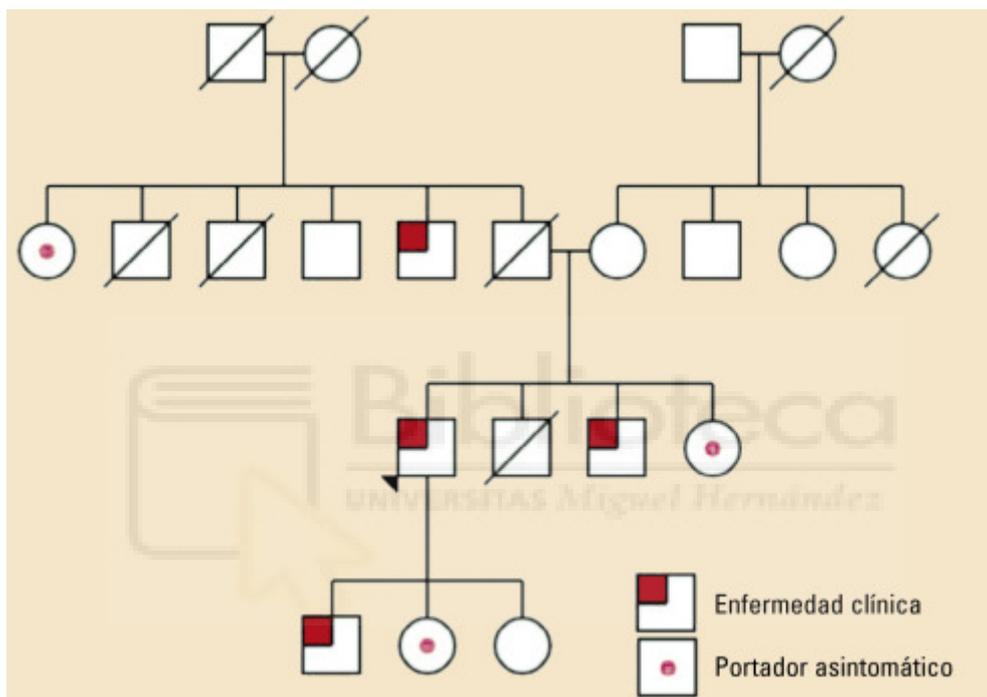


Figura 6: Árbol Genealógico.[8]

En la Figura 6 podemos observar un ejemplo de árbol genealógico en el que se recogen datos sobre la enfermedad en 4 generaciones de la misma familia. Los varones se representan por un cuadrado, las mujeres por un círculo. Los sujetos que tienen la enfermedad clínica se presentan con un cuadrado rojo. Los sujetos portadores asintomáticos se representan por un punto rosa.

Las pruebas complementarias como el ecocardiograma, el electrocardiograma y la resonancia cardíaca permitirán identificar el fenotipo morfológico de la miocardiopatía, además de las ya mencionadas, las analíticas de sangre pueden proporcionar información importante sobre el estado de salud del paciente y ayudar al médico al diagnóstico de la enfermedad y evaluación del tratamiento.

3.1.7.1 ANÁLISIS DE SANGRE

El análisis de sangre es una herramienta de la medicina moderna que proporciona información acerca del estado de salud del paciente, esto puede ayudar al médico a identificar una enfermedad, establecer su gravedad, elegir un tratamiento adecuado y evaluar la posible respuesta del paciente ante dicho tratamiento, además de poder observar estado de salud general del paciente [9].

El proceso del análisis de sangre consiste en la extracción de una pequeña muestra de la misma mediante una aguja que se introduce en una vena del brazo, la cantidad de sangre a extraer dependerá del tipo de análisis a realizar. Una vez obtenida la sangre se guarda en unos tubos especiales que contienen anticoagulantes y otros aditivos para evitar la coagulación de la sangre y preservar la muestra.

Mediante un análisis de sangre se pueden diagnosticar una amplia gama de enfermedades, algunas de estas son: anemia, infecciones, leucemia, lupus, diabetes, etc.

A continuación, se muestra como a partir de una analítica de sangre se pueden determinar algunas enfermedades:

- Anemia: se detecta a través de un recuento bajo de glóbulos rojos o hemoglobina, puede deberse a falta de hierro o de vitamina B12. Esta enfermedad impide que el cuerpo transporte suficiente oxígeno a los tejidos.
- Diabetes: los niveles de glucosa en sangre pueden indicar diabetes tipo I o tipo II.
- Enfermedades hepáticas: las pruebas de función hepática pueden revelar daño hepático, afecciones como las hepatitis e incluso, ayudar a detectar una cirrosis hepática. Las pruebas de función hepática consisten en la detección de los niveles de enzimas, como la ALT, la AST y la bilirrubina.
- Enfermedades renales: las pruebas de función renal pueden evaluar daños en los riñones, estas pruebas miden la concentración en sangre de creatinina y de

nitrógeno ureico para determinar la capacidad de los riñones para poder filtrar la sangre y eliminar residuos.

- Enfermedades cardiovasculares: los análisis del colesterol, tanto HDL como LDL, y de los triglicéridos ayudan a evaluar el riesgo de enfermedades cardíacas. Aunque un análisis de sangre por sí solo no puede detectar una miocardiopatía, sirve como herramienta complementaria, en los análisis se puede medir niveles de ciertas sustancias en sangre que pueden indicar daño cardíaco, como la troponina, o inflamación, como los marcadores inflamatorios, además, unos niveles altos de triglicéridos y/o de LDL pueden favorecer la aparición de enfermedades cardiovasculares.
- Infecciones: el recuento de glóbulos blancos (leucocitos) pueden indicar la presencia de infecciones.
- Problemas de coagulación: las pruebas de coagulación pueden revelar trastornos que afectan a la coagulación de la sangre, como la hemofilia.
- Cáncer de sangre: los análisis de sangre marcan indicios de que una persona puede padecer distintos tipos de cáncer de sangre, como por ejemplo la leucemia.
- Enfermedades autoinmunes: las pruebas de anticuerpos pueden ayudar a diagnosticar enfermedades autoinmunes como el lupus y la artritis reumatoide. Estas pruebas consisten en detectar la presencia de anticuerpos específicos en el cuerpo. Los anticuerpos son proteínas que el sistema inmunológico produce para combatir agentes extraños como virus, bacterias y otras sustancias.
- Enfermedades tiroideas: se lleva a cabo la medición de los niveles de hormonas tiroideas, la TSH, la T3 y la T4, una vez se evalúa la función tiroidea, en caso de presentar niveles anormales, se puede padecer de enfermedades como el hipertiroidismo o el hipotiroidismo.

Variables de la Base de Datos presentes en un análisis de sangre

Ahora vamos a hablar de las variables clínicas que aparecen en un análisis de sangre y que, además, están presentes en la base de datos que tenemos. Todas las variables a las que se va a hacer mención se han empleado en este TFG para los procedimientos de machine learning y deep learning, para así poder llevar a cabo los modelos de predicción de la MCH.

Las variables anteriormente mencionadas son las que siguen:

- **Glucosa:** es el sustrato energético más importante del organismo. En la mitocondria celular y a través del ciclo de Krebs se transforma en energía, en forma de moléculas de ATP (adenosín trifosfato), para ser utilizada en las reacciones químicas que requieran tal aporte de energía [10].

La concentración de glucosa en sangre se mueve entre límites estrechos, sin ocasionar sintomatología. Los niveles de glucemia están regulados por la acción hipoglucemiante de la insulina y la acción hiperglucemiante del glucagón, cortisol, catecolaminas y hormona del crecimiento. La determinación de la glucemia es útil para el diagnóstico y monitorización de las enfermedades metabólicas de los carbohidratos como la diabetes mellitus. El valor de referencia de la glucosa en sangre debe estar comprendido entre 74-106 mg/dL.

- **Creatinina:** es el producto resultante del catabolismo muscular, formándose a partir del fosfato de creatina que contiene el músculo. Tras pasar a la sangre se elimina por el riñón. La creatinina filtra libremente en el glomérulo y no es reabsorbida por el túbulo.

La determinación de creatinina es el mejor indicador de la función renal. Su concentración depende de la masa muscular y en mucha menor medida de la ingesta de proteínas [10]. El valor de referencia de la creatinina en sangre debe estar comprendido entre 0'7-1'2 mg/dL.

- **Ácido Úrico:** es el producto final del metabolismo de las purinas y uno de los marcadores biológicos de reconocido papel patogénico desde la antigüedad. La relación entre niveles séricos de ácido úrico y enfermedad cardiovascular es conocida desde hace tiempo, en 1879 el Dr. Frederik Akbar Mohamed señalaba en *Lancet* que muchos de sus pacientes hipertensos venían de familias gotosas y planteaba la hipótesis de que el ácido úrico pudiera ser un factor determinante del desarrollo de la hipertensión arterial (HTA). En la actualidad existe evidencia que vincula niveles elevados de ácido úrico con el desarrollo de HTA. La observación de que la hiperuricemia precede a la aparición de la HTA indica que puede desempeñar un potencial papel patogénico y que no es simplemente uno de los efectos de la hipertensión [11].

El valor de referencia del ácido úrico está comprendido entre 3'4-7 mg/dL.

- **Proteínas Totales:** la proteína sérica o proteínas totales es la suma de las concentraciones de todas las proteínas presentes en el suero. Se conocen más de 125 proteínas, cada una de las cuales tiene funciones muy diversas: transportadoras, proenzimas, inhibidores enzimáticos, factores de la coagulación o anticuerpos. Su valor de referencia se encuentra entre 64-84 g/L[10].
- **Sodio:** es el principal ion del líquido extracelular e interviene decisivamente en mantener una relación adecuada entre la cantidad de agua y sales minerales del organismo, siendo el factor determinante del que depende la distribución de ambos elementos entre el espacio vascular y el intersticial.

El paso del sodio a través de la pared celular se produce mediante la bomba de sodio, un sistema activo con consumo de energía. El sodio es excretado mediante el filtrado glomerular y reabsorbido en su mayoría a nivel del asa de Henle, a través de un mecanismo regulado por la aldosterona [10].

El valor de referencia del sodio está comprendido entre 131-155 mmol/L

- **Potasio:** es un ion (catión) eminentemente intracelular, cuyo nivel plasmático está regulado fundamentalmente por su excreción renal. El riñón pierde aproximadamente 50 mmol de potasio al día. Por las heces la pérdida es habitualmente muy pequeña, pero puede verse aumentada en procesos diarreicos [10].

El potasio tiene un papel importante en la excitabilidad celular y en la contracción muscular, por lo que se requiere mantener su concentración plasmática dentro de unos límites precisos. Dichos límites de referencia son de 3'3-5'1 mmol/L.

- **Calcio:** es el catión más cuantioso del organismo (representa aproximadamente entre 1 y 2 kg en el adulto) y en su regulación intervienen dos hormonas, la calcitonina y la parathormona, así como la vitamina D.

El calcio plasmático supone tan solo un 2% del total del organismo. El restante 98% se encuentra en el esqueleto óseo.

En el plasma el calcio se halla bajo tres formas distintas: libre o iónico, unido a las proteínas (principalmente la albúmina) y formando complejos solubles con el bicarbonato, el fosfato, el citrato o el sulfato.

Las funciones del calcio son múltiples. Además de ser un constituyente de la estructura ósea, estabiliza la membrana celular, interviene en la coagulación sanguínea, en la excitabilidad neuromuscular y en la transmisión nerviosa [10].

El valor de referencia del calcio está comprendido entre 8'5-10'5 mg/dL.

- **Hierro:** es un metal imprescindible en la hemoglobinogénesis, ya que forma parte de la estructura de la hemoglobina. En el plasma se encuentra en muy pequeña cantidad y circula vehiculado por la transferrina [10].

El valor de referencia del hierro es de 7-35 μ mol/L.

- **Ferritina:** es una proteína de depósito de hierro. Hay dos formas de almacenamiento de hierro: una forma movilizable como ferritina y la fracción insoluble como hemosiderina. La ferritina está compuesta de una cubierta proteínica formada por 24 subunidades que rodea un núcleo de fosfato de óxido férrico, se ha demostrado que cuando la molécula de ferritina está completamente saturada puede contener más del 20% de su masa en hierro [12][13].

El valor de referencia de la ferritina está comprendido entre 30-400 µg/L.

- **Transferrina:** El hierro es transportado por la transferrina, que es una glicoproteína de aproximadamente 80 kDa de peso molecular, sintetizada en el hígado, que posee 2 dominios homólogos de unión para el hierro férrico (Fe^{3+}). Esta proteína toma el hierro liberado por los macrófagos producto de la destrucción de los glóbulos rojos o el procedente de la mucosa intestinal, se ocupa de transportarlo y hacerlo disponible a todos los tejidos que lo requieren [14].
- **Troponina:** es una proteína globular de gran tamaño reguladora de la contracción del músculo cardíaco. Contiene tres subunidades polipeptídicas: troponina C (fijadora de calcio), troponina I (inhibidora de la interacción actina-miosina) y troponina T (fijadora de tropomiosina), las cuales son liberadas hacia el torrente sanguíneo durante un infarto al miocardio, por pérdida de la integridad de la membrana celular.

Su medición está fuertemente recomendada en la literatura internacional. Actualmente existen inmunoensayos para la detección tanto de troponina I como T, mostrando ambas moléculas características diagnósticas similares, considerándose los biomarcadores más sensibles y específicos de daño miocárdico[15].

- **Triglicéridos:** están formados por la esterificación de la glicerina con tres ácidos grasos. En realidad, no son un constituyente único sino una mezcla de constituyentes bioquímicos según los ácidos grasos que los forman. Se depositan en el tejido adiposo en donde constituyen una importante reserva de energía [10].

El valor de referencia de los triglicéridos es deseable que sea <150 mg/dL.

- **Colesterol:** circula en la sangre unido a las lipoproteínas, de manera que las de baja densidad (LDL) transportan aproximadamente el 70 %, las de alta densidad (HDL) transportan entre un 15-20 %, y el resto circula unido a las de muy baja densidad (VLDL) y a los quilomicrones. El aumento de la concentración sérica del colesterol se ha asociado a un riesgo especialmente alto de sufrir enfermedades cardiovasculares, por lo que su control y regulación en la población general tiene alto interés epidemiológico [10].

El valor de referencia para el colesterol total es <200 mg/dL, para el HDL es deseable que sea >40 mg/dL mientras que, para el LDL debe estar por debajo de los 120 mg/dL.

- **NTproBNP:** el péptido natriurético cerebral N-terminal (NTproBNP) es útil en el diagnóstico de insuficiencia cardíaca. Cuando comparamos a los sujetos hipertensos con los normotensos, encontramos valores mayores de NTproBNP en los hipertensos [16]. El BNP aumenta de forma proporcional a la disfunción ventricular izquierda y la gravedad de la insuficiencia cardíaca [17].
- **GPT:** el glutamato pirúvico transaminasa (GPT) es una enzima intracelular, exclusivamente citoplasmática. Se encuentra en concentraciones elevadas en los tejidos con un metabolismo de los aminoácidos activo como es el hígado, el músculo esquelético y cardíaco o el riñón. Las concentraciones catalíticas de esta enzima en el suero corresponden básicamente al catabolismo diario de las células de estos tejidos y cuando se produce una lesión celular aumenta su concentración [10].

El valor de referencia para el glutamato pirúvico transaminasa es de 0-41 U/L, siendo U la unidad enzimática.

- **GOT:** el glutamato oxalacetato transaminasa (GOT) es una enzima que se concentra en el corazón, riñones y cerebro. Si los valores son anormales, puede indicarnos diferentes enfermedades o lesiones hepáticas. La elevación no es una enfermedad en sí, más bien son los síntomas y la evolución de ciertas enfermedades del hígado; debido a una alimentación pobre en minerales y proteínas o, simplemente por el consumo excesivo de alcohol y de ciertos medicamentos.

Un valor adecuado para esta transaminasa puede ser de entre 6 y 40 unidades enzimáticas por litro en los hombres y de entre 6 y 34 unidades enzimáticas por litro en las mujeres [18].

- **TSH:** la medición de TSH (hormona estimulante de la tiroides) en muestras de sangre es esencial para evaluar la función tiroidea y el equilibrio hormonal. La TSH, producida por la glándula pituitaria, regula la producción de hormonas tiroideas, que a su vez afectan el metabolismo y la energía del cuerpo. Los niveles elevados de TSH en la sangre pueden indicar hipotiroidismo, que es la incapacidad de la glándula tiroides para producir suficiente cantidad de esta hormona. Por otro lado, un nivel bajo de TSH puede indicar hipertiroidismo, que es una producción excesiva de hormona tiroidea. Las mediciones de TSH brindan información importante sobre el equilibrio hormonal [9].

El valor de referencia de la hormona TSH está comprendido entre 0'35-4'94 μ UI/mL.

- **Ancho de Distribución Eritrocitaria:** describe la variación porcentual (siendo estadísticamente un coeficiente de variación) del tamaño de los glóbulos rojos, donde su fórmula es: $[\text{desviación estándar}/\text{volumen corpuscular medio}] \times 100$. Depende de cada laboratorio clínico, pero generalmente, su valor normal va desde 11% a 15%. Tanto una excesiva destrucción como también el déficit de producción de hematíes, conllevan un aumento de su valor. Este parámetro se introdujo como ayuda al diagnóstico diferencial de las anemias hipocrómicas; sin embargo, su uso real ha sido limitado a la diferenciación entre betatalasemia y

anemia ferropénica, y es casi un parámetro desconocido al que muy pocos clínicos prestan atención al examinar el hemograma. Un valor de ancho de distribución eritrocitaria menor al límite inferior, no se ha correlacionado con procesos fisiopatológicos ni enfermedades subyacentes y ha sido considerado sin relevancia clínica. Por el contrario, el aumento sobre el límite superior tiene varias implicancias clínicas denominándose anisocitosis, reflejando una gran diferencia de tamaños entre los glóbulos rojos circulantes [19].

- **Plaquetas:** son elementos sanguíneos anucleados derivados de la fragmentación de sus células precursoras: los megacariocitos. Tienen forma de disco biconvexo de 2-3 μm de diámetro, su concentración normal en sangre periférica es de 150 a 400 $\times 10^3/\mu\text{L}$ y su vida media es de entre 7 y 10 días.

La participación de las plaquetas en los diferentes procesos de la hemostasia depende de 3 eventos fundamentales: la adhesión plaquetaria, el cambio de forma y la agregación plaquetaria. Se ha planteado que las plaquetas proporcionan, en cuestión de segundos, la primera respuesta celular del organismo después que se produce la ruptura de la integridad vascular, lesión de algún tejido o de ambos elementos. De esta forma, ellas pueden recubrir y sellar cualquier tipo de estas lesiones e iniciar por un mecanismo extra hemostático el proceso de curación de los sitios dañados [20].

Tanto los niveles altos como los bajos de plaquetas pueden indicar un problema médico, como un trastorno hemorrágico, una infección o un trastorno de la médula ósea. Esta prueba es importante para diagnosticar trastornos sanguíneos y evaluar la capacidad del cuerpo para controlar el sangrado.

- **Leucocitos:** los leucocitos son células nucleadas en la sangre, carente de hemoglobina. Los leucocitos se dividen en: granulocitos, monocitos y linfocitos [21]. Miden 8 o 9 μm de diámetro y son por tanto mayores que los hematíes. Un milímetro cúbico de sangre suele contener entre 5000 y 10000 glóbulos blancos. Entre las funciones más importantes figuran la fagocitosis de bacterias, hongos, virus y cuerpos extraños, destoxicación de las proteínas tóxicas producidas en las

reacciones alérgicas y las lesiones celulares y el desarrollo de la inmunidad. Un recuento bajo o alto de los glóbulos blancos nos puede indicar un trastorno en el sistema inmunológico, como puede ser la leucemia [22].

Los valores de referencia de los leucocitos están comprendidos entre $4'5 \cdot 10^3$ – $11 \cdot 10^3$ glóbulos blancos/ μ L.

- **Fibrinógeno:** es una proteína plasmática esencial para la coagulación de la sangre, que es convertida en fibrina por la trombina en presencia de iones calcio, la fibrina es una proteína filamentosa insoluble que proporciona su carácter semisólido al coágulo sanguíneo [22].

Tanto los niveles bajos como los altos de fibrinógeno pueden indicar un problema médico, como un trastorno hemorrágico, inflamación crónica o un mayor riesgo de coágulos sanguíneos [9].

Los valores de referencia de cada una de las variables clínicas los he sacado de los laboratorios clínicos del hospital general universitario Dr. Balmis, que pertenece a la Conselleria de Sanidad Universal y Salud Pública de la Generalitat Valenciana. El servicio de análisis clínico de este laboratorio está certificado por Bureau Veritas según la norma internacional ISO 9001:2015.

3.1.7.2 ELECTROCARDIOGRAMA

El electrocardiograma (ECG) es la representación gráfica de las señales eléctricas de todas las células miocárdicas. Se trata de un registro en el que se observan las variaciones de voltaje en relación con el tiempo. Estas variaciones de voltaje se corresponden con el paso de los impulsos eléctricos a través de las distintas estructuras cardíacas. Dicha información eléctrica queda reflejada en el papel milimetrado del ECG mediante un trazado con distintas deflexiones.

Se observará una línea basal plana cuando las células están en estado de reposo eléctrico, y unas ondas positivas, negativas o complejas con fases positivas y negativas cuando la célula se activa eléctricamente por los fenómenos de despolarización y repolarización (Figura 7).

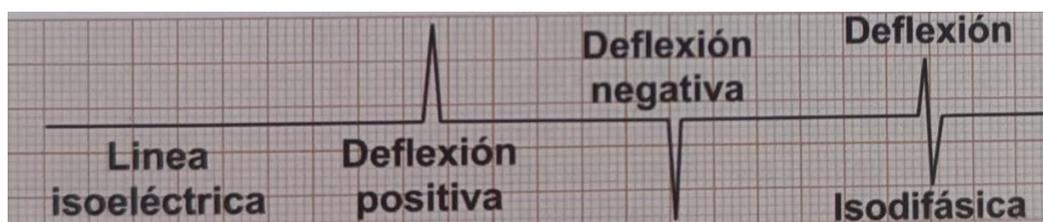


Figura 7: de izquierda a derecha; línea isoelectrica, deflexión positiva, negativa e isodifásica. [4]

En la Figura 7 podemos observar claramente un papel cuadriculado, éste se llama papel de registro. El papel de registro del electrocardiograma es un papel milimetrado que nos facilita la medición de la amplitud de las ondas y su duración temporal. Cada cuadrado pequeño mide 1 mm y cada 5 milímetros tenemos una línea más gruesa que define otro cuadrado mayor.

El eje vertical mide la amplitud eléctrica que genera el corazón. Generalmente el desplazamiento de 1 mm en vertical corresponde a 0'1 mV de voltaje, por lo que 1 mV serán 10 mm en la cuadrícula.

El eje horizontal mide el tiempo. El desplazamiento de 1mm en dicho eje se corresponde con 0'04 segundos, lo que quiere decir que 25 mm de desplazamiento en el eje horizontal equivalen a 1 segundo. La duración de un ECG normal suele ser de 10 segundos (250 mm).

Ambos parámetros, voltaje y tiempo, pueden ser modificados en función de la información que se desee analizar con mayor precisión. Por defecto, la mayoría de los electrocardiógrafos están calibrados para registrar a 10 mm/mV y 25 mm/segundo.

En la Figura 8 se puede ver a la perfección una representación gráfica de lo explicado anteriormente, donde aparecen los desplazamientos que hay que realizar para obtener determinadas medidas de tensión y de tiempo.

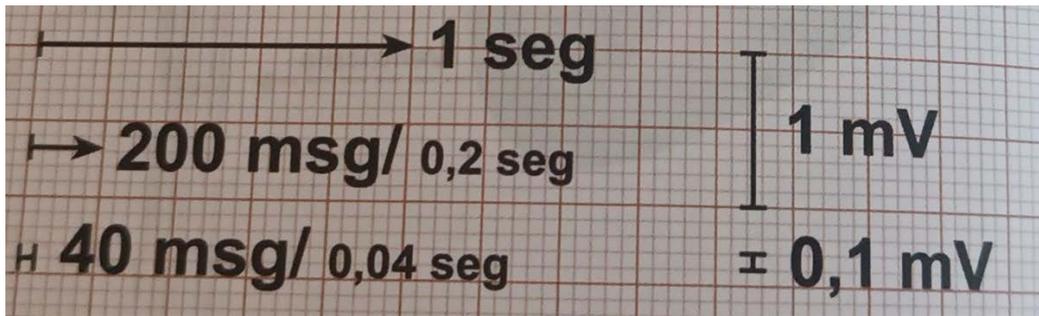


Figura 8: Papel milimetrado del ECG. [4]

Se debe tener muy presente que la técnica a la hora de realizar un electrocardiograma debe ser estricta y rigurosa. Es fundamental la correcta colocación de los electrodos exploradores, por lo que es crucial ser metódicos y sistemáticos para evitar errores que dificulten y tergiversen la interpretación.

Colocación de los electrodos

Como se analiza en [4], los electrodos del ECG ponen en contacto las señales eléctricas del corazón y el electrocardiógrafo. Para la realización del ECG se colocan diez electrodos; cuatro en las extremidades y seis precordiales. De ellos se obtienen 12 derivaciones que se estudiarán posteriormente. La ubicación de estos electrodos es universal y se debe ser riguroso en su colocación para evitar errores diagnósticos.

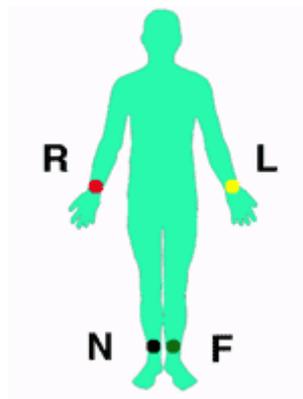


Figura 9: Colocación de los electrodos de las extremidades. [23]

Para empezar, se colocarán los cuatro electrodos de las extremidades, siempre evitando prominencias óseas.

- En el brazo derecho se ubicará el electrodo rojo (R).
- En el brazo izquierdo el amarillo (L).
- En el tobillo derecho el negro (N).
- En el tobillo izquierdo el verde (F).

En pacientes que, por la razón que sea, no sea posible colocar los electrodos en los tobillos (por amputaciones, por ejemplo), se ubicarán en el muñón, o en la región más cercana, teniendo en cuenta que habrá que dejar el electrodo a la misma altura que el electrodo contralateral.

Los seis electrodos precordiales se sitúan de la siguiente forma (observar Figura 10):

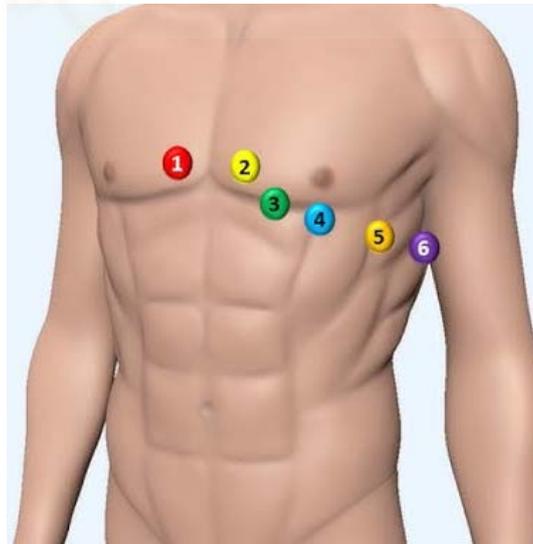


Figura 10: Ubicación de los electrodos precordiales. [9]

V1: 4º espacio intercostal en la línea paraesternal derecha.

V2: 4º espacio intercostal en la línea paraesternal izquierda.

V3: entre V2 y V4.

V4: 5º espacio intercostal en la línea medio clavicular.

V5: 5º espacio intercostal en la línea axilar anterior.

V6: 5º espacio intercostal en la línea axilar media.

Derivaciones del electrocardiograma

Las derivaciones cardíacas son el registro de la diferencia eléctrica entre dos puntos; que puede ser entre dos electrodos (derivaciones bipolares), o entre un electrodo y un punto neutro (derivaciones monopolares). Es importante saber que las 12 derivaciones del ECG nos muestran el mismo estímulo eléctrico desde distintos puntos de vista.

Las corrientes eléctricas generadas por el corazón tienen una representación vectorial. La suma de estas fuerzas genera un vector con una magnitud, dirección y sentido. Si imaginamos que este vector tiene forma de flecha, su tamaño será la magnitud y su orientación será la dirección y el sentido.

Todo electrodo que observe como se acerca hacia él la cabeza del vector, inscribirá una deflexión positiva, y aquel que observe como se aleja la cola del vector inscribirá una deflexión negativa. En el caso de que el electrodo esté en un punto medio, la deflexión será isodifásica, es decir, tendrá un componente positivo y otro negativo [4].

Los electrodos de las extremidades registran las derivaciones de los miembros, que exploran el plano frontal del corazón:

- Derivaciones bipolares: I, II y III.
- Derivaciones monopolares ampliadas: aVR, aVL y aVF.

Derivaciones bipolares

Las derivaciones bipolares muestran la diferencia de potencial (DdP) entre dos puntos:

- I: DdP entre brazo derecho e izquierdo. Vector dirigido a 0°.
- II: DdP entre brazo derecho y pierna izquierda. Vector dirigido a 60°.
- III: DdP entre brazo izquierdo y pierna izquierda. Vector dirigido a 120°.

Los ejes de las tres derivaciones bipolares de los miembros (es decir, las líneas imaginarias que unen los polos positivo y negativo de cada derivación) forman un triángulo hipotético, que se denomina triángulo de Einthoven (atender a la figura 11). Según Einthoven estas derivaciones tienen una proporción matemática ($II = I + III$) que nos sirve para conocer la correcta colocación de los electrodos.

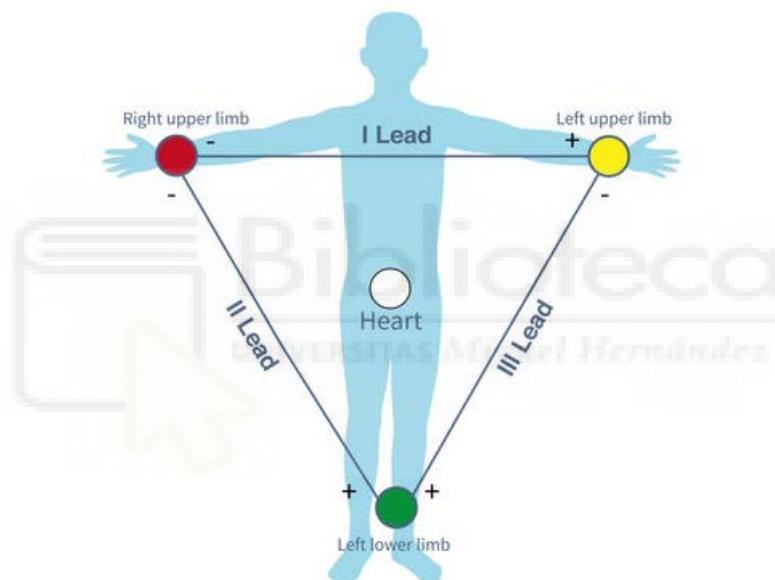


Figura 11: Posición anatómica de las derivaciones bipolares de miembros. [24]

Derivaciones monopolares aumentadas

Estas derivaciones muestran la DdP entre un punto neutro con valor de cero (situado en el centro del triángulo de Einthoven) y el electrodo de cada extremidad

- aVR. Potencial del brazo derecho. Vector dirigido a -150°.
- aVL. Potencial del brazo izquierdo. Vector dirigido a -30°.
- aVF. Potencial de la pierna izquierda. Vector dirigido a 90°.

En los vértices del triángulo de Einthoven se sitúan las tres derivaciones aumentadas de los miembros. Estas seis derivaciones conforman el sistema hexaxial de Bailey, que está dividido en ángulos de 30° , donde cada uno de ellos corresponde a una derivación del triángulo de Einthoven. Este sistema es primordial a la hora de calcular el eje cardíaco.

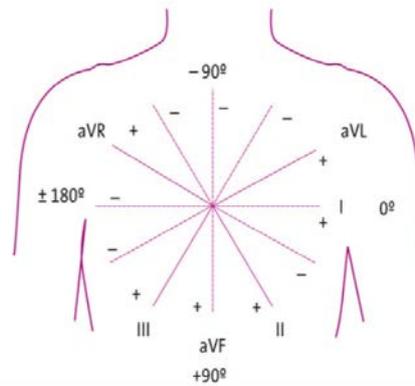


Figura 12: Sistema Hexaxial de Bailey. [25]

Ondas, intervalos y segmentos

La activación de las distintas estructuras cardíacas conlleva fenómenos de despolarización y repolarización con los consecuentes eventos mecánicos. El registro de estos sucesos por medio de electrodos da lugar al electrocardiograma. Dicho registro, según [4], implica distintas deflexiones en forma de ondas, intervalos y segmentos que se verán en la siguiente figura:

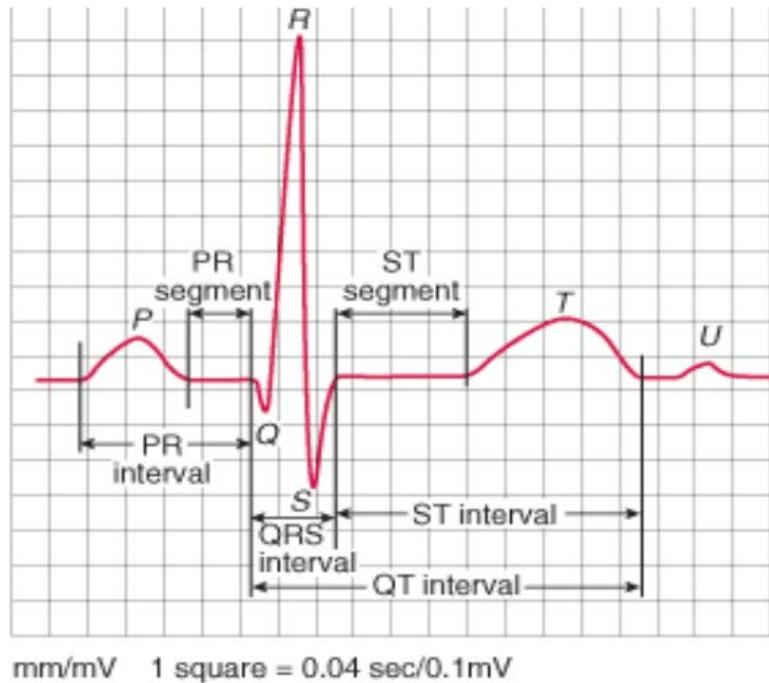


Figura 13: Ondas, intervalos y segmentos del ECG. [3]

Las ondas son las distintas deflexiones positivas y negativas que se inscriben en el ECG producto de los potenciales de acción durante la estimulación miocárdica.

La onda P es la primera onda del ciclo y representa la despolarización auricular. Su parte inicial corresponde a la despolarización de la aurícula derecha y la parte final a la de la aurícula izquierda. Sus características son las siguientes:

- Duración menor de 0'1 segundos.
- Voltaje máximo de 0'25 mV.
- Positiva en todas las derivaciones salvo en aVR (negativa) y V1, donde es difícil establecer un criterio.

La onda Q es la primera deflexión negativa del complejo QRS. La presencia de una onda Q no indica un infarto, se considera patológica cuando dura más de 0'04 segundos o su amplitud es mayor del 25% del QRS.

El complejo QRS representa la despolarización ventricular. Como se puede observar en la Figura 13, la primera deflexión negativa es la onda Q, la primera onda positiva es la R y la onda negativa que se encuentra detrás de la R se denomina onda S. Cuando existe una segunda onda R o S se les llama R' o S'.

- El complejo QRS tiene una duración temporal de entre 0'6 y 0'1 segundos.
- Su morfología depende de la derivación observada.

La onda T refleja la repolarización ventricular. Como norma general tendrá la misma polaridad que el QRS y será positiva en todas las derivaciones salvo en aVR. Su amplitud máxima es de unos 5 mm (0'5 mV) en el plano frontal y unos 15 mm (1'5 mV) en el horizontal.

La onda U es la deflexión que sigue a la onda T y aunque su origen es desconocido, podría representar la repolarización de los músculos papilares.

Intervalos y segmentos

Los segmentos de un ECG son las líneas que unen ondas entre sí sin incluirlas. Por otro lado, los intervalos son segmentos que incluyen una o más ondas.

El intervalo RR mide el tiempo entre dos ondas R. En condiciones normales debe ser regular (rítmico) y su longitud se modifica con la frecuencia cardíaca (FC).

El intervalo PR representa la suma de la despolarización auricular y el retraso fisiológico del estímulo a su paso por el nodo AV.

- Dicho intervalo mide desde el inicio de la onda P hasta el inicio de la onda Q o de la onda R.
- Duración entre 0'12 y 0'2 segundos (3-5mm).

El intervalo QT mide la despolarización y repolarización ventricular. Esta última varía en función de la frecuencia cardíaca, por lo que debe ser corregido (QTc) mediante la fórmula de Bazett ($QTc = QT / \sqrt{RR}$).

- Comprende desde el inicio de la onda Q hasta el final de la onda T.
- QTc es normal entre 340-450 milisegundos y como norma general no debería ser mayor que la mitad de la distancia entre dos QRS.

Cuando se habla del segmento ST es imprescindible conocer lo que es el punto J. El punto J se sitúa donde finaliza la onda S y comienza el segmento ST.

- El segmento ST comprende desde el punto J hasta el inicio de la onda T.
- Un supra o infra desnivel mayor de 1 mm en el plano frontal o 2 mm en el plano horizontal puede ser patológico.

Electrocardiograma normal

El ritmo normal del corazón se denomina ritmo sinusal. Los criterios del ritmo sinusal son los siguiente:

- Ondas P seguidas del complejo QRS.
- Ritmo regular, de 60 a 100 lpm.
- Onda P positiva en I, II, III y aVF. Onda P negativa en aVR.

Cálculo de la frecuencia cardíaca

En el ECG normal un segundo equivale a 5 cuadrados grandes del papel milimetrado, por lo que un minuto estará compuesto por 300 cuadrados grandes ($60 \text{ segundos} \cdot 5 \text{ cuadrados/segundo} = 300 \text{ cuadrados}$). Sabiendo lo anterior, podemos calcular la FC dividiendo 300 entre el número de cuadrados entre dos complejos RR.

Para facilitar el cálculo sería lógico escoger un complejo QRS que coincida con una de las líneas gruesas del papel milimetrado, una vez hecho eso contamos los cuadrados grandes que hay hasta el siguiente complejo QRS, debemos tener en cuenta que cada cuadrado pequeño es equivalente a 0'2 cuadrados grandes. Para facilitar la comprensión del concepto observar la Figura 14.



Figura 14: ECG normal con ritmo regular.[4]

En la Figura 14 podemos observar un ECG normal donde su ritmo es periódico, es decir, sus complejos QRS están equidistantes. Procedemos con el cálculo de la FC, seleccionamos el 3er complejo QRS de la Figura 14 (el de en medio, dado que coincide con una línea gruesa), y contamos los cuadrados grandes que hay hasta el siguiente complejo, en esta ocasión hay 4 cuadrados grandes y solo uno pequeño, como hemos mencionado anteriormente un cuadrado pequeño equivale a 0'2 cuadrados grandes, luego la distancia entre ambas señales es de 4'2 cuadrados grandes. Finalmente, llevamos a cabo la división: $300/4'2 = 71$ o 72 lpm.

Ahora supongamos que el paciente no tiene ritmos regulares, la herramienta anterior dejaría de ser fiable. En ese caso, lo ideal sería contar todos los complejos QRS del ECG y multiplicarlo por 6, ya que la duración del ECG es de 10 segundos.

Finalmente, pongámonos en el peor de los casos, el paciente no tiene ritmos regulares, y además, no conocemos la duración del ECG. En ese caso, como sabemos que un cuadrado grande equivale a 200 milisegundos, podemos deducir que 30 cuadrados grandes serán 6 segundos ($30 \text{ cuadrados} \cdot 200 \text{ ms/cuadrado} = 6000 \text{ milisegundos} = 6 \text{ segundos}$). Sabiendo que 30 cuadrados grandes son 6 segundos, solo nos haría falta contar cuantos complejos hay en 30 cuadrados y multiplicarlo por 10 para obtener la FC en lpm.

Valoración de las ondas, intervalos y segmentos

En primer lugar, se debe identificar la presencia o no de ondas P, además de si estas son sinusales o no. También se analizará su morfología, que puede indicar hipertrofias o latidos ectópicos.

Del complejo QRS se calculará su duración, para saber si es ancho o estrecho y poder identificar ritmos ventriculares, ectópicos o bloqueos de rama. También se medirá la amplitud de los complejos (hipertrofias) y se tendrá en cuenta la presencia de ondas Q patológicas.

De la onda T se valorará sobre todo su amplitud, polaridad y morfología. Se observará si es picuda o si por el contrario está aplanada, además de ver si es positiva o negativa. Cambios en la onda T pueden indicar alteraciones electrolíticas o isquémicas.

Del intervalo PR se medirá su duración para saber si está alargado o acortado, esto puede indicar anomalías en el nodo AV o existencia de vías accesorias.

El intervalo QT mide como norma general la mitad de la distancia entre dos QRS. Se deberá calcular el QTc para detectar si está alargado o acortado, lo que puede indicar alteraciones electrolíticas y síndromes potencialmente mortales.

En cuanto al segmento ST se medirán infra o supra desniveles del punto J. Su elevación o descenso es clave en el diagnóstico de la cardiopatía isquémica, entre otras patologías [4].

En la Tabla 2 se resumen los aspectos a valorar:

Tabla 2: Análisis de las ondas, intervalos y segmentos.

Ondas	¿Qué analizar?	Posible patología
Onda P	¿Está presente? ¿Es sinusal?	Fibrilación Auricular (FA), Flutter. Hipertrofias
QRS	¿Es el ancho > 0'12s (3mm)? ¿La onda Q es patológica?	Bloqueos Rama Síndrome Coronario Agudo (SCA)
Onda T	¿Es picuda o aplanada? ¿Es negativa?	Hiper/hipopotasemia Pericarditis, SCA
Intervalo PR	¿Es > 0'2s (5mm) o < 0'12 (3mm)?	Bloqueos AV, Preexcitación
Intervalo QTc	¿Es > 0'45s?	Síndrome de QT largo

Segmento ST	¿Elevado o descendido? ¿1mm en el plano frontal o 2mm en el horizontal?	SCA

Cálculo del eje cardíaco

El cálculo del eje eléctrico es uno de los pasos más complejos a la hora de la interpretación de un ECG. Aunque los electrocardiógrafos modernos lo calculan y lo reflejan en la leyenda, a continuación, veremos algunas herramientas para localizarlo.

El eje eléctrico es la dirección del vector resultante de la despolarización de los ventrículos, o lo que es lo mismo, hacia donde se dirige el estímulo eléctrico de los ventrículos [4].

Según los valores que tome el eje cardíaco se clasifican del siguiente modo:

- Eje normal: de -30° a $+90^\circ$.
- Eje desviado a la izquierda: de -30° a -90° .
- Eje desviado a la derecha: de $+90^\circ$ a $+180^\circ$.
- Eje inverso derecho: de -90° a -180° .

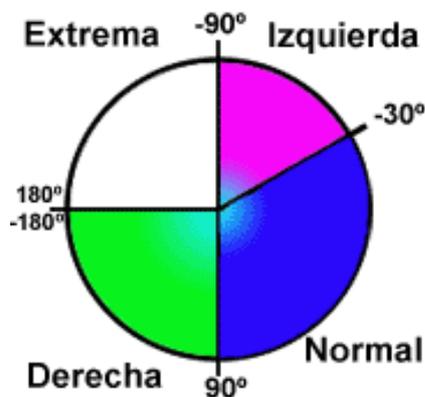


Figura 15: Eje normal y sus desviaciones. [26]

La Figura 15 es una representación muy clara de la clasificación realizada previamente.

Las derivaciones que se utilizan para calcular el eje eléctrico son las seis derivaciones de los miembros (I, II, III, aVR, aVL y aVF) que forman el triángulo de Einthoven y el sistema hexaxial (observar Figura 12).

Para calcular el eje de una forma sencilla bastaría con analizar la polaridad del QRS de las derivaciones I y aVF. Al verificar si estas dos derivaciones son positivas o negativas y combinando el resultado se podrá saber en qué cuadrante del sistema se halla el eje, debido a que son perpendiculares en el sistema hexaxial.

- Si I y aVF son positivas el eje es normal, está en el cuadrante inferior derecho (ver la Figura 15).
- Si I es positiva y aVF negativa el eje es desviado a la izquierda, cuadrante superior derecho (ver Figura 15).
- Si I es negativa y aVF positiva el eje es desviado a la derecha, cuadrante inferior izquierdo (ver Figura 15).
- Si I y aVF son negativas el eje es extremo derecho (o inverso derecho), cuadrante superior izquierdo (ver Figura 15).

Por último, en [4] se explica que para calcular los grados hacia donde se dirige el eje de forma exacta hay que medir la amplitud (en milímetros) del complejo QRS en I y aVF. Hay que tener en cuenta que los complejos pueden ser positivos, negativos o isodifásicos. En este último caso se hallará la diferencia de milímetros de voltaje entre el componente positivo y el negativo. Una vez calculados los milímetros de amplitud trasladaremos los valores a los ejes I y aVF de un sistema hexaxial en un papel milimetrado (teniendo en cuenta si el valor es positivo o negativo). Finalmente, se trazarán las perpendiculares a esos valores. El eje se dirigirá al punto en el que coincidan ambas perpendiculares. (ver Figura 16 y Figura 17).

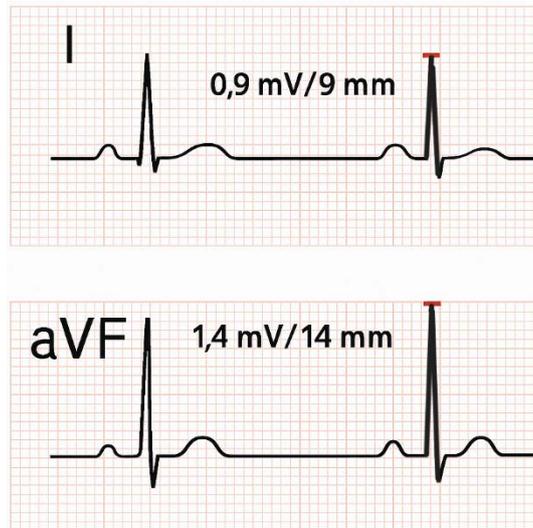


Figura 16: Amplitud del QRS en I y aVF. [4]

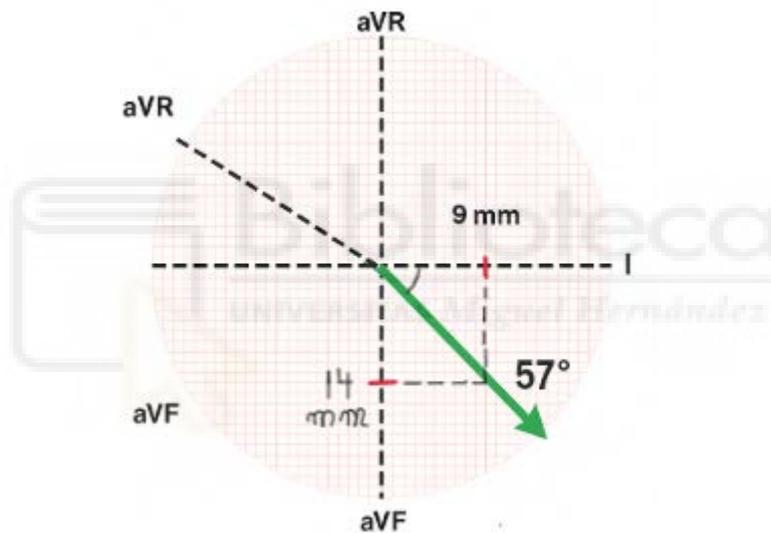


Figura 17: Cálculo exacto del eje eléctrico. [4]

En la Figura 16 trazamos las amplitudes del QRS en I y aVF, en I son 9mm (0'9mV) y en aVF son 14 mm (1'4 mV). En la Figura 17 marcamos las longitudes que hemos obtenido, en sus ejes correspondientes (son las líneas rojas), trazamos las perpendiculares a esas marcas rojas y obtenemos un punto donde se cortan (el punto que atraviesa la flecha verde).

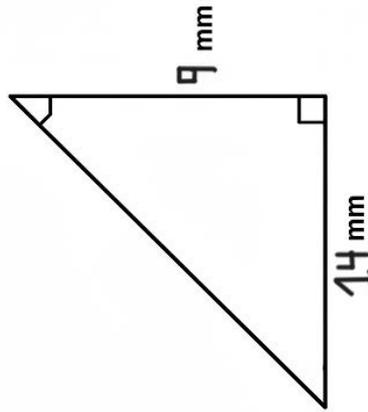


Figura 18: Triángulo formado en la Figura 17.(imagen de elaboración propia)

Una vez tenemos el punto de corte de las rectas perpendiculares trazadas, tiramos una flecha que va desde el origen de coordenadas hasta el punto de corte, obtenemos un triángulo rectángulo como el de la Figura 18. Para sacar el ángulo marcado en la Figura 18 aplicamos trigonometría básica pues conocemos las longitudes de ambos catetos:

$$\tan \alpha = \text{cateto opuesto} / \text{cateto contiguo} = 14/9 = 1'5555\dots$$

$$\alpha = \text{arco Tan} (1'555\dots) = 57'26^\circ$$

3.1.7.3 ECOCARDIOGRAMA

La ecografía o ECO, también conocida como ultrasonografía, es una técnica que utiliza ondas de ultrasonidos para crear imágenes en tiempo real de las estructuras internas del cuerpo [27].

Esta tecnología no invasiva es revolucionaria, puesto que proporciona una visión detallada de órganos, tejidos y flujos sanguíneos, en tiempo real, a pie de cama y sin la necesidad de radiación ionizante.

La base fundamental de la ecografía radica en la emisión de ondas de ultrasonido desde un transductor, que luego interactúan con los tejidos internos. Estas ondas son reflejadas de nuevo al transductor después de encontrarse con diferentes densidades de tejido, y la información recopilada se traduce en imágenes visuales que pueden ser interpretadas por profesionales de la salud [27].

Fundamentos físicos

El sonido es una onda mecánica, es decir, requiere de un medio material para propagarse (a diferencia de las ondas electromagnéticas no se pueden propagar en el vacío). El oído humano tiene la capacidad de detectar sonidos con una frecuencia comprendida entre 20 Hz y 20 KHz. Los sonidos con una frecuencia superior se denominan ultrasonidos y no son detectados por el ser humano.

El transductor empleado en las ecografías es el ecógrafo, este dispositivo transmite pulsos cortos de ultrasonido al organismo y mide el tiempo y de qué manera retornan. Todo ello es posible gracias a las ondas sonoras emitidas por el material piezoeléctrico (cerámica sintética de Zirconato titanato de plomo o PZT) que se encuentra dentro de las sondas o transductor del ecógrafo. Cuando se le aplica una corriente eléctrica esta cerámica sintética (cristal piezoeléctrico) comienza a expandirse y contraerse con rapidez, generando vibraciones que producen ondas de ultrasonido que chocan con las estructuras corporales y retornan a la sonda ecográfica de nuevo. La señal recibida es analizada por un procesador y en función de su amplitud se muestra una imagen en escala de grises.

Las ondas, tanto electromagnéticas como de sonido, tienen una serie de características:

- **Periodo (T):** es el tiempo (en segundos) transcurrido entre dos puntos equivalentes de la onda.
- **Amplitud:** es la altura o tamaño máximo que alcanza una onda. Hace referencia a la intensidad del sonido y se mide en decibelios (dB).
- **Longitud de onda (λ):** es la distancia que recorre la onda en un periodo.
- **Frecuencia (f):** es el número de periodos que tenemos por unidad de tiempo. Se calcula haciendo la inversa del periodo y se mide en hertzios. Las ondas de alta frecuencia proporcionan una imagen más detallada, pero penetran menos en los tejidos, mientras que, las ondas de baja frecuencia tienen una mayor capacidad de penetración, pero ofrecen menos detalle

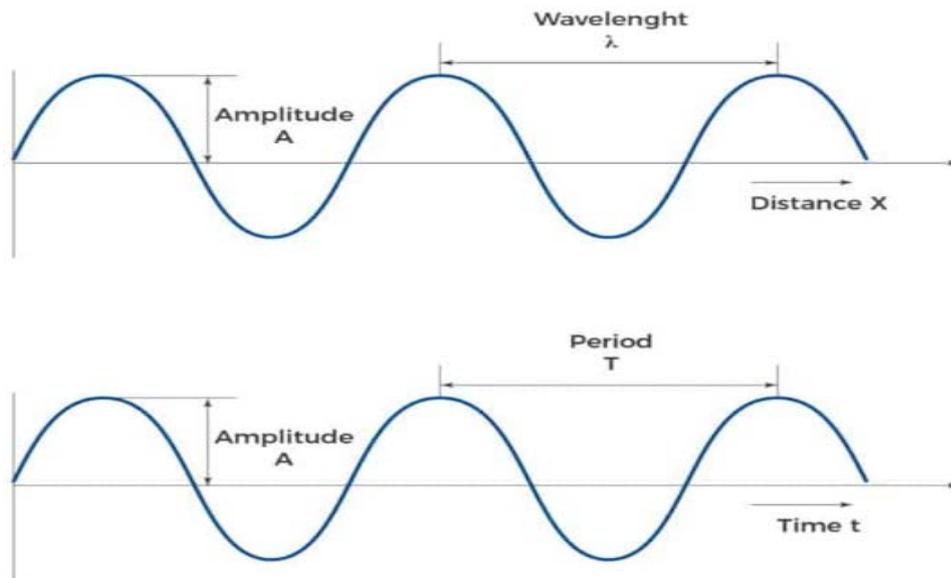


Figura 19: Características de las ondas.[28]

En los tejidos biológicos las ondas de ultrasonido se desplazan a una velocidad constante, que suele ser aproximadamente de 1540 metros por segundo [27]. La comprensión de esta velocidad es esencial para calcular la distancia entre el transductor y las estructuras anatómicas, lo que permite una representación precisa en la imagen ecográfica. La velocidad de propagación exacta de la onda de sonido viene dada por las características del medio como la rigidez o la densidad.

Tabla 3: Velocidad de Propagación del sonido en diferentes tejidos y materiales.

Tejido o Material	Velocidad de Propagación (m/s)
Aire	330
Grasa	1450
Agua	1495
Riñones	1561
Sangre	1570
Músculo	1585
Metal	4000
Hueso	4080

Las ondas se transmiten en los diferentes tejidos o materiales dependiendo de los siguientes parámetros:

- **Potencia:** es la energía aplicada sobre un tejido en un momento concreto. Al ser energía por unidad de tiempo, su unidad en el sistema internacional es el watio [W].
- **Intensidad:** es la cantidad de potencia por unidad de superficie [W/cm²].

Es importante conocer todas estas propiedades físicas de los ecógrafos para poder obtener el resultado de la forma más precisa posible, siempre buscando la seguridad del paciente. También es fundamental el conocimiento acerca de las propiedades de los tejidos, éstas son:

- **Reflectividad:** se refiere a la capacidad de un tejido para reflejar las ondas ultrasónicas. Los tejidos con mayor densidad tienden a reflejar más, apareciendo más brillantes en la imagen, mientras que, los tejidos menos densos permiten una mayor profundidad de penetración, apareciendo más oscuros en la imagen.
- **Atenuación:** es la pérdida de energía que sufren las ondas de ultrasonido al atravesar los tejidos, esto hace que la onda pierda amplitud (dB). Las ondas de alta frecuencia se absorben más fácilmente, y eso hace que penetren menos en los tejidos en comparación con las ondas de baja frecuencia.
- **Impedancia:** es la resistencia que ofrece un tejido a la propagación del sonido por el mismo.
- **Refracción:** proceso por el cual las ondas cambian de dirección al atravesar de un medio material a otro.
- **Difracción:** fenómeno por el cual las ondas se curvan, cambian su dirección o amplitud al encontrarse con obstáculos en su trayectoria.
- **Dispersión:** es la reflexión de la onda en múltiples direcciones.
- **Ecogenicidad:** es la capacidad de un tejido para producir eco.

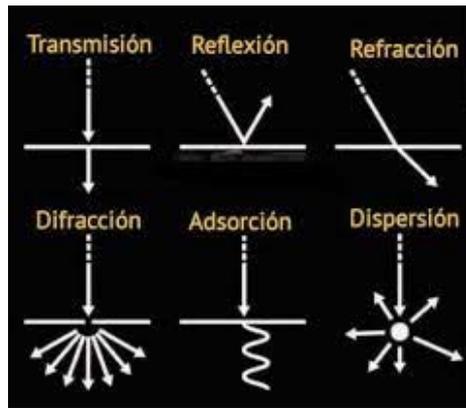


Figura 20: Fenómenos físicos de las ondas tras un cambio de medio. [29]

Información de la ecocardiografía en la MCH

Los hallazgos ecocardiográficos son muy numerosos y establecen claramente el diagnóstico y la fisiopatología de la enfermedad.

La hipertrofia del miocardio izquierdo es sin duda el dato más importante que define la enfermedad. Existe una hipertrofia que puede ser muy variable y que ha permitido hacer diferentes tipos de clasificaciones en base a la distribución segmentaria de la hipertrofia. Hay casos en que sólo está afectado el septo (Figura 21), en otros casos se afecta tanto el septo como la cara lateral o prácticamente la hipertrofia es concéntrica. En algunos casos existe un patrón muy inusual que corresponde a la miocardiopatía hipertrófica apical. Cuando situamos el traductor en el plano de cuatro cámaras podemos registrar un miocardio apical muy engrosado con colapso sistólico de la cavidad, como se explica en [30].

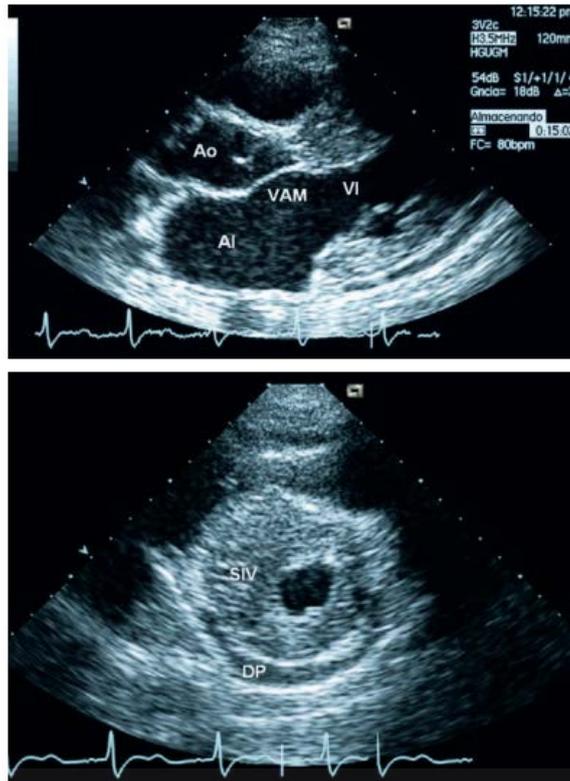


Figura 21: Análisis con diferentes planos de estudio de un caso con MCH. [27]

En la Figura 21 contamos con 2 imágenes:

- Imagen superior: plano paraesternal longitudinal del ventrículo izquierdo en diástole. Puede observarse el importante aumento de espesor del septo interventricular.
- Imagen inferior: plano paraesternal transversal. Permite localizar la extensión de la hipertrofia que afecta tanto al septo anterior como al posterior.

Para acabar con la Figura 21, nombraremos las acotaciones que aparecen en la misma: AI: aurícula izquierda; VI: ventrículo izquierdo; SIV: septo interventricular; Ao: aorta; DP: derrame pericárdico.

El patrón típico característico de la miocardiopatía hipertrófica con modo M (Figura 22), es la presencia de un movimiento anterior sistólico de la válvula mitral que puede llegar a contactar con el septo interventricular. Este movimiento anterior sistólico mitral va a ser el responsable de la obstrucción dinámica al tracto de salida del ventrículo izquierdo

y marca una de las imágenes más típicas y características de la miocardiopatía hipertrófica.

El movimiento sistólico va a producir indirectamente otro de los hallazgos típicos de la enfermedad como es la insuficiencia mitral. En este caso, el ecocardiograma Doppler color nos va a permitir, no sólo el diagnóstico, sino la cuantificación de la misma, igualmente con técnica con modo M podemos analizar como la válvula aórtica se abre al inicio de la sístole y en mesosístole hay un cierre abrupto de la misma. Este dato está relacionado con la rápida disminución del flujo a través de la válvula aórtica, que produce la obstrucción del tracto de salida del ventrículo izquierdo debido al movimiento anterior sistólico mitral (Figura 23) [30].

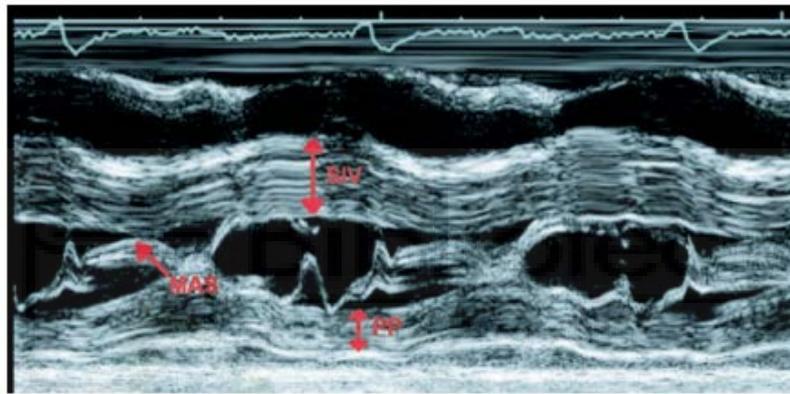


Figura 22: Ecocardiograma en modo M en el plano valvular mitral. [27]

En la Figura 22 se observa a un claro aumento del grosor septal con menos afectación de pared posterior. También puede verse el movimiento anterior sistólico mitral con contacto de la valva anterior con el septo. La terminología aparecida en la imagen es la siguiente: S: septo interventricular; PP: pared posterior del ventrículo izquierdo; MAS: movimiento anterior sistólico de la valva anterior mitral.

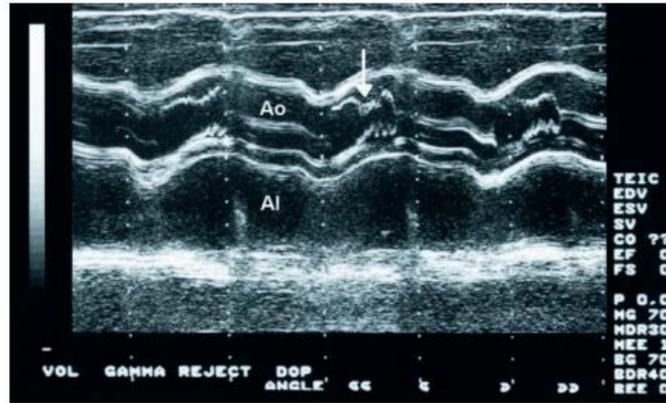


Figura 23: Ecocardiograma en modo M en el plano de las sigmoideas aórticas. [27]

En la Figura 23 se observa colapso de las sigmoideas aórticas durante mesosístole (flecha) con reapertura posterior. Este patrón es típico de obstrucción dinámica del tracto de salida del ventrículo izquierdo y suele darse en los pacientes con obstrucción dinámica importante.

La ecocardiografía es el procedimiento básico de elección para el diagnóstico y cuantificación de la miocardiopatía hipertrófica, no existe ningún otro método ni sangriento ni sin derramamiento de sangre que nos ofrezca tanta información fisiopatológica.

De hecho, el conocimiento de la fisiopatología en las distintas variantes de la enfermedad ha sido una de las máximas aportaciones de la ecocardiografía en el estudio de las miocardiopatías, según [30]. Puesto que la enfermedad es hereditaria, la ecocardiografía sin duda va a ser la técnica para el screening de pacientes que tengan un familiar con afectación de miocardiopatía hipertrófica. En el momento actual se recomienda realizar un estudio ecocardiográfico a todos aquellos familiares en primer grado de un paciente con miocardiopatía hipertrófica. Igualmente, la ecocardiografía puede ser muy útil para evaluación a la terapia farmacológica o a la más compleja terapia de implantación de marcapaso o de cirugía cardíaca, en todos los casos el ecocardiograma nos va a demostrar la presencia de la obstrucción y el resultado de las maniobras terapéuticas realizadas.

Variables de la Base de Datos presentes en un Ecocardiograma

Ahora vamos a hablar de las variables clínicas que aparecen en un ecocardiograma y que, además, están presentes en la base de datos que tenemos. Todas las variables a las que se va a hacer mención se han empleado en este TFG para los procedimientos de machine learning y deep learning, para así poder llevar a cabo los modelos de predicción de la MCH. Las variables son las que siguen:

- **Volumen Telediastólico Del Ventrículo Izquierdo (VTDVI):** representa el volumen de sangre presente en el ventrículo izquierdo, es importante conocer el valor (en mililitros) para poder evaluar su capacidad para recibir sangre y retenerla antes de que sea bombeada al resto del cuerpo.
- **Septo Telediastólico del Ventrículo Izquierdo:** es el tejido muscular que separa el ventrículo izquierdo del ventrículo derecho en la fase final de la diástole (llenado). Más específicamente, se trata del diámetro o grosor del septo interventricular en el momento en que el ventrículo izquierdo ha alcanzado su volumen de máximo llenado.

El septo telediastólico se puede evaluar mediante ecocardiograma y otros estudios para diagnosticar y monitorizar enfermedades cardíacas, como la hipertrofia ventricular izquierda. Un septo que se mueve de forma anormal durante la sístole puede indicar problemas en la contracción del ventrículo, mientras que, un septo grueso o rígido puede dificultar el llenado del ventrículo, lo que podría ser indicativo de disfunción diastólica.

- **Pared Telediastólica Posterior del Ventrículo Izquierdo:** se refiere a la pared del ventrículo izquierdo que se encuentra en la parte posterior del corazón en el momento de la diástole, cuando el ventrículo está relajado y se llena de sangre.

La pared posterior del ventrículo izquierdo puede ser utilizada para medir el grosor de la pared y la velocidad de relajación, lo que puede ayudar a identificar problemas en la capacidad del ventrículo izquierdo para relajarse y llenarse de sangre.

- **Masa Ventrículo Izquierdo:** está directamente relacionada con la función ventricular izquierda, indica el peso total del VI del corazón, que es el principal responsable de bombear la sangre al resto del organismo. La masa ventricular izquierda depende del tamaño de la persona, edad o su sexo, cuando dicha masa es inapropiada debido a un aumento de la misma puede aumentar el riesgo de padecer enfermedades cardiovasculares [31].
- **Fracción de Eyección del Ventrículo Izquierdo (FEVI):** es un índice de función ventricular, básico a la hora de analizar el funcionamiento cardíaco, se obtiene de los volúmenes ventriculares.

Es un parámetro muy útil para valorar el estado del corazón como bomba, los valores normales, están alrededor de 0,70 con el límite inferior de la normalidad de 0,52. La ventaja del ecocardiograma bidimensional es que puede ser utilizado para el cálculo de la fracción de eyección aún en presencia de zonas de contracción segmentaria alterada [30].

- **Velocidad Máxima de la Válvula Mitral:** es un parámetro medido mediante Doppler pulsado, este método mide la velocidad de flujo de sangre a través de la válvula mitral. La válvula mitral está situada entre la aurícula y el ventrículo izquierdos, y su función es permitir que la sangre fluya desde la aurícula al ventrículo cuando el corazón se relaja.

Medir la velocidad máxima de la válvula mitral es esencial para evaluar la función de la válvula mitral y detectar problemas como la insuficiencia mitral severa o la estenosis mitral.

- **Relación E/A:** se refiere a la relación entre la velocidad de la onda E y la onda A en el flujo de entrada mitral. La onda E representa el llenado rápido ventricular durante la fase diastólica, mientras que la onda A representa la contracción auricular. La relación E/A es un parámetro importante para evaluar la función diastólica del ventrículo izquierdo. Esta relación de ondas se puede utilizar para:

estimar la presión de llenado del VI, identificar patrones de llenado diastólico y monitorizar la función diastólica del VI.

- **Relación E/E lateral:** La razón entre la velocidad diastólica E del flujo mitral y la velocidad diastólica e' del anillo lateral de la válvula mitral (relación E/e'), es el mejor índice no invasivo para detectar elevación aguda de la presión de llenado del VI. La relación E/E lateral es un predictor útil de remodelación del VI, indicando pacientes con un mayor riesgo cardiovascular [32].

3.1.8 TRATAMIENTOS DE LA MIOCARDIOPATÍA HIPERTRÓFICA

La incertidumbre terapéutica en la MCH deriva de su fisiopatología diversa y espectro clínico heterogéneo. En la actualidad se acepta el tratamiento medicamentoso con betabloqueantes, verapamilo, disopiramida y amiodarona, la cirugía para el manejo de síntomas refractarios en las formas obstructivas y el implante de cardio desfibrilador, la amiodarona sería para pacientes con alto riesgo de muerte súbita. Como alternativas en las formas obstructivas se pueden plantear tanto la estimulación bicameral como la ablación septal con alcohol. Gracias al análisis realizado en [33] podemos observar la Figura 24 para ver las alternativas terapéuticas en la MCH.

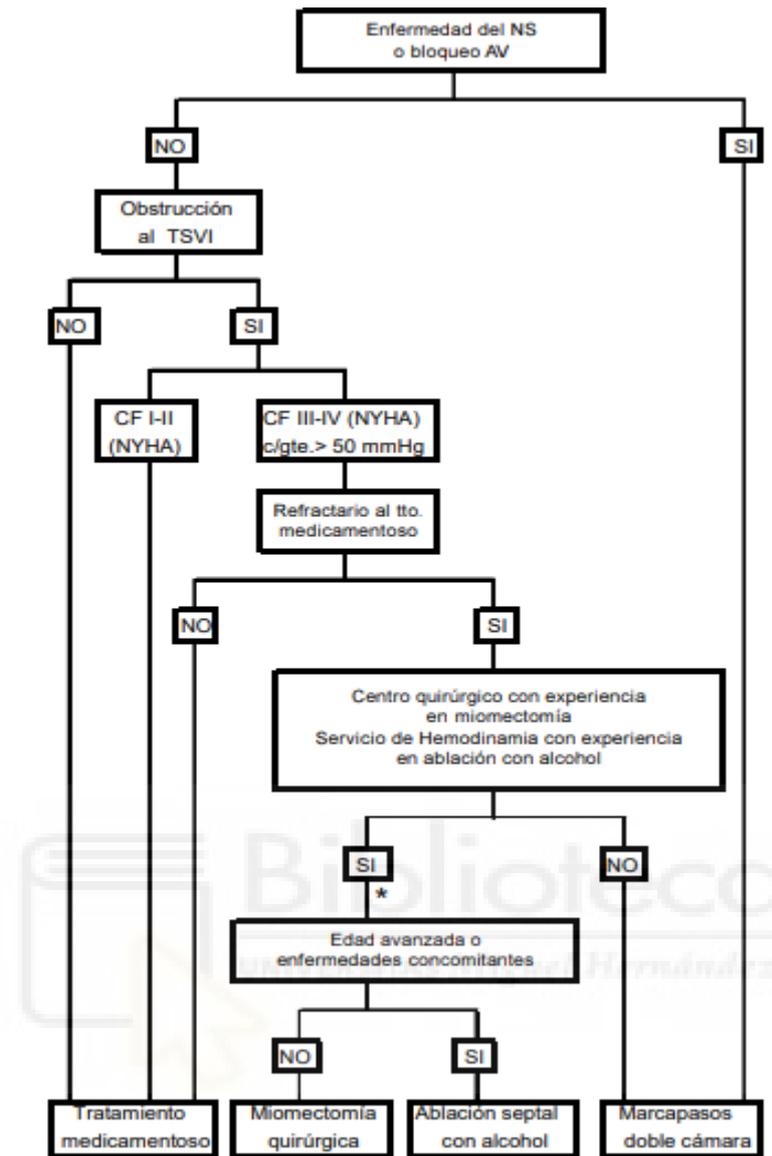


Figura 24: Alternativas terapéuticas en la MCH

Cirugía

La miotomía-miomectomía septal ventricular (operación de Morrow) representa el mejor tratamiento para pacientes con gradiente en condiciones basales (≥ 50 mm Hg) y síntomas severos de insuficiencia cardíaca refractaria al tratamiento médico. La cirugía mejora los síntomas en el 70% de los pacientes y el gradiente basal se reduce en el 90%, con mejoría en la calidad de vida y excelentes resultados a largo plazo. El mecanismo por el cual la cirugía mejora la función ventricular y los síntomas probablemente se deba a la reducción o normalización de la presión sistólica y de la presión de fin de diástole con una influencia favorable en el llenado ventricular izquierdo y en el déficit de oxígeno que condiciona isquemia miocárdica [33].

Implante de marcapasos

Recientemente algunas publicaciones mencionan el beneficio de la estimulación bicameral con un intervalo AV óptimo en pacientes con MCH obstructiva que proporcione un máximo ensanchamiento del complejo QRS, este hecho provocaría una despolarización precoz de la punta del ventrículo izquierdo, que produce un movimiento paradójico del septum y aumento del diámetro del tracto de salida. Como consecuencia, disminuyen la velocidad del flujo y el gradiente en el tracto de salida. Por otra parte, la sístole auricular aumenta la presión de llenado, con incremento del volumen de fin de diástole y, por ley de Starling, el volumen sistólico. El Comité de estimulación cardíaca del ACC/AHA recomienda el implante de marcapasos DDD en MCH obstructiva con disfunción del nódulo sinusal o bloqueo AV y en pacientes sintomáticos, refractarios al tratamiento médico, con obstrucción severa del tracto de salida del VI [33].

Ablación septal con alcohol

Consiste en la inyección de alcohol dentro de la primera rama perforante septal mayor de la arteria descendente anterior, para producir un infarto de miocardio en el septum ventricular subaórtico. Las células mueren con cicatrización y reemplazo por fibrosis, adelgazamiento septal, agrandamiento del tracto de salida, disminución de la aceleración de la eyección, que lleva a la reducción de la obstrucción al flujo de salida. La adición de un infarto septal puede actuar como sustrato para producir arritmias ventriculares malignas sostenidas, por lo que es prudente evitar este procedimiento en pacientes jóvenes [33].

Tratamiento farmacológico

Como explica Jesumond Phanius en [34], el tratamiento mediante el uso de fármacos es la primera opción ante el diagnóstico de la MCH, éstos suelen ser betabloqueantes no vasodilatadores, bloqueadores de los canales de calcio no dihidropiridínicos y en segunda línea la disopiramida. Algunos de estos medicamentos son:

- El verapamilo: se usa para tratar la presión arterial alta y controlar la angina.

- La disopiramida: actúa a nivel del corazón como agente antiarrítmico clase I, usado principalmente en el tratamiento de taquicardias ventriculares
- La amiodarona: se usa para tratar y prevenir ciertos tipos graves y posiblemente mortales de arritmia ventricular.

3.2 MACHINE LEARNING COMO HERRAMIENTA DE DETECCIÓN Y DIAGNÓSTICO TEMPRANO

Encontrar nuevos enfoques para la detección y el diagnóstico temprano puede considerarse urgente, ya que las enfermedades cardiovasculares siguen siendo una de las principales causas de mortalidad a nivel internacional. Recientemente se ha establecido una nueva revelación que aplica la importancia del aprendizaje automático (ML) como una forma innovadora de tecnología que puede tener un impacto significativo en el sector sanitario, en particular en la predicción y el diagnóstico de enfermedades cardiovasculares. Como resultado, en comparación con los métodos descendentes y tradicionales, el aprendizaje automático puede evaluar las características complejas de los historiales clínicos de los pacientes basándose en grandes conjuntos de datos y algoritmos, esto se traduce en una mayor precisión y eficiencia diagnósticas. Varios algoritmos de aprendizaje automático son útiles para predecir enfermedades cardíacas, con altos índices de precisión.

Algunos ejemplos de clasificadores que han demostrado resultados prometedores son las máquinas de vectores de soporte (SVM), los bosques aleatorios (RF), que es una técnica que consiste en múltiples árboles de decisión, y las redes neuronales. En algunos casos, la precisión de estos clasificadores ha alcanzado el 98,7 %. El perfil de riesgo cardiovascular de una persona puede evaluarse completamente, gracias a la capacidad del aprendizaje automático para incluir numerosos parámetros clínicos, como la edad, la presión arterial, los niveles de colesterol, las lecturas del electrocardiograma, etc.

Este enfoque multifactorial proporciona un conocimiento más completo del estado de salud del paciente, lo que también permite diagnosticar posibles problemas cardíacos de forma más temprana y precisa. Además, el aprendizaje automático facilita el proceso de diagnóstico y mejora la terapia individualizada al identificar los factores de riesgo específicos de cada paciente. Con una intervención rápida, es posible minimizar

considerablemente la morbilidad y la mortalidad asociadas a las enfermedades cardíacas, según explican en [35].

Esta habilidad es muy beneficiosa en el sector sanitario, permite a los profesionales sanitarios ajustar las terapias a las necesidades particulares de cada paciente, mejorando los resultados y optimizando la asignación de recursos.

En el campo del aprendizaje automático (ML), la optimización de hiperparámetros es un proceso crucial que implica encontrar las configuraciones de parámetros más efectivas para el algoritmo de aprendizaje. Para cualquier modelo de ML que empleemos en este TFG será necesario encontrar la configuración óptima del mismo, para así poder obtener los mejores valores predictivos posibles [36].

Para mejorar la robustez de los modelos, es recomendable entrenar el conjunto de datos haciendo distintas divisiones de entrenamiento y prueba de los mismos.



4 MATERIALES Y MÉTODOS

En este apartado vamos a ver los diferentes tipos de IA que podemos emplear para ayudarnos en caso de necesitarlo, también vamos a entrar a valorarlas para ver con qué modelo nos quedamos como asistente en la programación. Todo el código de este trabajo de fin de grado se lleva a cabo en un software de Google llamado “Google colaboratory”, dicho software ya nos incluye una versión de Gemini, al tener cuenta con la universidad, dicha versión de Gemini es más avanzada que la que tiene “Google colaboratory” por defecto. Para finalizar el apartado, entraré a evaluar la base de datos que se nos proporciona para poder realizar este trabajo de investigación.

4.1 PROGRAMACIÓN ASISTIDA

Según nos explica Tomás De Camino-Beck, profesor de la universidad Cenfotec de Costa Rica en [37], el desarrollo de lenguajes de programación responde a la necesidad humana de expresar instrucciones de manera precisa. Aunque los lenguajes de alto nivel han facilitado este proceso, siguen enfocándose en el “cómo” en lugar del “qué”, sin capturar directamente la intención del programador. Con los Modelos de Lenguaje Grandes (LLMs), la programación ha evolucionado, las instrucciones pueden formularse en lenguaje natural. Esto reduce la necesidad de especificaciones procedimentales detalladas y prioriza la expresión clara de objetivos. Para aprovechar la programación asistida por inteligencia artificial (IA), es fundamental considerar:

- 1° La formulación clara del texto dentro del prompt.
- 2° La descomposición modular del problema
- 3° La implementación de pruebas rigurosas.

La IA generativa mejora la programación asistida apoyada sobre la inferencia de LLMs, permitiendo que el código refleje mejor los objetivos del programador y facilitando su adopción en entornos profesionales. Es por esto que vamos a emplear diferentes tipos de IAs para la programación necesaria a lo largo de este trabajo.

4.1.1 CHAT GPT

Chat GPT es un modelo de lenguaje de inteligencia artificial desarrollado por OpenAI. Funciona mediante el uso de técnicas de procesamiento del lenguaje natural y se entrena en una gran cantidad de texto para generar respuestas a preguntas o sugerencias dadas por los usuarios. Chat GPT está construido sobre la familia GPT-3.5, considerado uno de los modelos de lenguaje más avanzados hasta la fecha. Gracias al aprendizaje automático, los modelos de lenguaje de gran tamaño, como Chat GPT, pueden llevar a cabo diversas tareas, como resumir, ampliar información, traducir textos, hacer predicciones, crear guiones, escribir código y generar discursos con gran precisión.

Uno de los mayores desafíos de Chat GPT es que, aunque puede generar respuestas coherentes y a veces sorprendentes, estas respuestas pueden no ser precisas ni confiables. La falta de contexto y la naturaleza limitada del entrenamiento pueden llevar a respuestas erróneas o inapropiadas. El sesgo de los datos de entrenamiento también puede generar problemas de sesgo en las respuestas proporcionadas, lo que puede conducir a respuestas insensibles u ofensivas. Además, es importante tener en cuenta las limitaciones lingüísticas de Chat GPT, ya que se ha entrenado principalmente en inglés y su rendimiento en determinados idiomas podría ser limitado.

Según [38], a la hora de interactuar con Chat GPT es esencial proporcionar un contexto claro y preciso. Un mensaje claro y conciso ayudará a garantizar que Chat GPT entienda el tema o la tarea que se le plantea y sea capaz de generar una respuesta adecuada. Es importante ser concretos y específicos al momento de formular preguntas, evitando preguntas vagas o confusas que puedan generar respuestas poco precisas. La inteligencia artificial es tan buena como la información que se le brinda, por lo que es fundamental proporcionarle un marco adecuado para que pueda dar una respuesta precisa y relevante. En resumen, los "Prompts" deben incluir: (a) tema, (b) estilo, (c) tono, (d) contexto y (e) información de contexto.

Por lo tanto, es crucial prestar atención a la calidad de los prompts y asegurarse de que estén bien contruidos para lograr una conversación exitosa en ChatGPT.

4.1.2 GEMINI

Gemini se basa en un modelo de aprendizaje automático (transformador) que solo decodifican, pero con algunos cambios para que funcionen mejor en las unidades de procesamiento tensorial (TPU en inglés). Pueden manejar hasta 32.768 tokens de contexto, usando una técnica llamada atención multi-consulta.

Hay dos versiones pequeñas de Gemini: Nano-1 (1.8 mil millones de parámetros) y Nano-2 (3.25 mil millones de parámetros), que se obtienen a partir de modelos Gemini más grandes, pensados para dispositivos de borde como los teléfonos inteligentes.

Gemini es multimodal, lo que quiere decir que puede procesar diferentes tipos de entrada. No importa el orden ni la combinación de los modos, lo que permite tener una conversación multimodal. Las imágenes pueden tener distintas resoluciones, y los vídeos se tratan como una serie de imágenes. El audio se muestrea a 16 kHz y luego se transforma en una secuencia de tokens por el Modelo Universal de Habla. Gemini se entrena con un conjunto de datos que también es multimodal y multilingüe, formado por documentos web, libros y código, e incluye datos de imagen, audio y vídeo.

La segunda generación de Gemini ("Gemini 1.5") tiene dos modelos. Gemini 1.5 Pro es un modelo multimodal de mezcla dispersa de expertos, con una longitud de contexto en los millones, mientras que Gemini 1.5 Flash está destilado de Gemini 1.5 Pro, con una longitud de contexto superior a 2 millones [39].

4.1.3 COPILOT

Microsoft Copilot es una potente herramienta de inteligencia artificial integrada en Microsoft 365, diseñada específicamente para potenciar la productividad en el uso de estas aplicaciones. Esta tecnología revolucionaria está disponible en toda la suite Office de Microsoft, incluyendo Word, Excel, OneNote, PowerPoint y Teams, entre otras.

Con Copilot puedes generar automáticamente código. Además, Copilot también tiene la capacidad de identificar errores en documentos y proyectos, ofreciendo soluciones o correcciones sugeridas, lo que evita errores costosos y mejora la precisión del trabajo. Copilot funciona interactuando mediante "prompts" textuales, es decir, mediante comandos que se realizan con un lenguaje natural.

La principal diferencia entre ChatGPT y Microsoft Copilot es que ChatGPT es un modelo de inteligencia artificial conversacional desarrollado por OpenAI, mientras que Microsoft Copilot es una herramienta diseñada para mejorar la productividad dentro del ecosistema de Microsoft 365, utilizando los datos de la empresa bajo las directivas de seguridad establecidas. Además, Copilot funciona utilizando el modelo de lenguaje de inteligencia artificial GPT, el mismo que utiliza ChatGPT, de hecho, su principal característica es que utiliza GPT-4, la misma versión que usa ChatGPT Plus.

Mientras ChatGPT se enfoca en la generación de texto general y respuestas conversacionales, Microsoft Copilot está diseñado para ayudar a los usuarios en tareas empresariales, utilizando datos internos de manera segura y eficiente [40].

4.1.4 ANÁLISIS DE IA

Pese a no ser competencia de este TFG realizar un análisis de IA, hemos entrado a valorar las diferentes características entre Copilot, Gemini y Chat GPT, algunas de estas características han sido: el tiempo que emplean, la cantidad de errores que cometen, los comentarios que añaden o la cantidad de líneas de código utilizadas. Una vez realizada esta tarea llego a la conclusión de que Chat GPT es la mejor herramienta de las 3 para la programación asistida, pensé que al utilizar Copilot una versión de GPT-4 sería la mejor de las 3, pero nada más lejos, pese a utilizar “prompts” extensos y claros de lo que quiero tiende a cometer muchos errores. Gemini comete más errores que Chat GPT a la hora de producir código desde cero, sin embargo, encuentra y corrige mejor los errores que Chat GPT. Como conclusión de todo esto, a la hora de necesitar ayuda con el código recurriré a Chat GPT como asistente de programación, en caso de que haya que corregir errores puntuales o no ser capaz de ver dónde está el error, recurriré a la opción de Gemini que aparece en Google Colaboratory con la cuenta de la universidad.

Este análisis de IA se ha puesto en práctica traduciendo un código de Matlab a Python, en el cual llevamos a cabo un análisis exploratorio. Hemos valorado determinadas características de cada uno de los modelos de IA (Copilot, Gemini y Chat GPT) y los hemos metido en una tabla de Excel, el resultado ha sido el siguiente (las “X” indican que no podemos saber el resultado porque el modelo no ha hecho bien el código):

Tabla 4: Análisis de Copilot

	Tiempo empleado (minutos)	Nº Líneas de código	Nº Líneas de Comentarios
Carga de datos y selección de variables	5	72	37
Representación de Histogramas	15	22	4
Representación Normal – Gaussian Equivalente	80	39	5
Representación Normal con todas las variables	No lo hace bien	X	X
Estudio edad de diagnóstico	No lo hace bien	X	X
Separabilidad	No lo hace bien	X	X
Variables con una separabilidad mínima	65	15	4
Severidad de Fenotipo	No lo hace bien	X	X
Mortalidad	7	16	4

Tabla 5: Análisis de Gemini

	Tiempo empleado (minutos)	Nº Líneas de código	Nº Líneas de Comentarios
Carga de datos y selección de variables	1	81	39
Representación de Histogramas	23	32	9
Representación Normal – Gaussian Equivalente	50	24	16
Representación Normal con todas las variables	No lo hace bien	X	X
Estudio edad de diagnóstico	5	41	25
Separabilidad	60	46	29
Variables con una separabilidad mínima	10	19	5
Severidad de Fenotipo	35	35	9
Mortalidad	3	13	12

Tabla 6: Análisis de Chat GPT

	Tiempo empleado (minutos)	Nº Líneas de código	Nº Líneas de Comentarios
Carga de datos y selección de variables	1	79	36
Representación de Histogramas	5	33	11
Representación Normal – Gaussian Equivalente	45	24	8
Representación Normal con todas las variables	120	34	14
Estudio edad de diagnóstico	13	34	14
Separabilidad	38	41	21
Variables con una separabilidad mínima	8	15	6
Severidad de Fenotipo	18	33	19
Mortalidad	3	10	8

Si observamos la Tabla 4, la Tabla 5 y la Tabla 6 aparecen los valores de tiempo empleado, líneas de código utilizadas y cantidad de comentarios para los 3 modelos y para cada bloque que empleé en el análisis exploratorio. Es por estos resultados que he utilizado Chat GPT como ayudante en la programación en algún experimento.

En el caso de Copilot (Tabla 4), es el modelo que más errores comete, hasta el punto de no ser capaz de generar el código que haga lo que requiere el apartado. Por el lado positivo, los códigos que ha generado bien los ha comprimido en un menor número de líneas que Gemini.

Para Gemini (Tabla 5), tenemos que es un modelo más capaz que Copilot en la generación del código, pero menos que Chat GPT, emplea más tiempo que Chat GPT y también utiliza un mayor número de líneas de código, por otra parte, Gemini es, de los 3, el modelo más explicativo en lo que hace, siendo el que más comentarios adjunta de los 3.

Finalmente, Chat GPT (Tabla 6), es el modelo más completo de los 3. Ha sido capaz de generar un código correcto en cada apartado, emplea menos tiempo que los otros modelos, dado que comete menos errores, genera un menor número de líneas de código, sin embargo, es menos explicativo que Gemini, pues genera menos comentarios.

4.2 BASE DE DATOS

El conjunto de datos utilizados para este trabajo de investigación pertenece al Hospital Clínico Universitario Virgen de la Arrixaca de la Región de Murcia. Gracias a ellos contamos con hasta 159 datos clínicos y genéticos de 1053 pacientes. Es una base de datos extensa, aunque no emplearemos todos los datos que nos proporciona.

Tenemos datos de un amplio rango de personas, comprendidas en edades de entre 11 y 88 años, la edad promedio de los 1053 pacientes es de 58 años, predominan claramente pacientes de entre 50 y 70 años. La edad promedio de diagnóstico de la MCH en pacientes afectados por la enfermedad es de 58'59 años.

En este trabajo estamos realizando distinciones entre los pacientes y organizándolos en 4 posibles grupos:

- **Grupo 0.** Paciente no afectado. Nuestra base de datos cuenta con 36 pacientes de este tipo, un 3'42% del total, la edad media de este grupo es de 56'61 años.
- **Grupo 1.** Paciente portador genético sin cardiopatía, es decir, genéticamente predispuesto a poder sufrir la patología, pero sin sintomatología de la afección. Nuestra base de datos cuenta con 41 pacientes de este tipo, un 3'89% del total, la edad media de este grupo es de 37'32 años, siendo claramente la muestra de pacientes de menor edad.
- **Grupo 2.** Paciente portador con posibilidad de ser afectado, pruebas dudosas y genética positiva. Nuestra base de datos cuenta con 66 pacientes de este tipo, un

6'27% del total, la edad media de este grupo es de 63'2 años, siendo este grupo el que posee una media de edad mayor.

- **Grupo 3.** Paciente afectado, pruebas afectadas y portador genético. Nuestra base de datos cuenta con 902 pacientes de este tipo, eso representa un 85'66% del total de nuestra base de datos y es, con diferencia, nuestro grupo de pacientes mayoritario.

A parte de la relación entre edad y afectación y la distribución de grupos clínicos nuestra base de datos cuenta con muchísima más información: tiempos de seguimientos, datos del electrocardiograma (ECG), datos acerca del complejo QRS, datos genéticos, datos de analíticas (calcio, glucosa, creatinina, sodio, potasio, hierro, ferritina, transferrina, triglicéridos, colesterol, TSH, hemoglobina, etc), también contamos con un código del paciente en el sistema y un identificador en la base de datos de la Comunidad Autónoma de la Región de Murcia con el fin de anonimizar a los pacientes. Además de la edad a la que se diagnosticó la afección, tenemos también la edad a la que se detectó el fenotipo de la enfermedad. Podemos observar también parámetros bioquímicos como la concentración de dióxido de carbono en sangre. Todos los parámetros anteriormente mencionados son algunos de los 159 datos clínicos y genéticos con los que contamos para nuestro trabajo de investigación.

4.3 MÉTODOS EMPLEADOS

Ahora vamos a ver en que consiste cada modelo que hemos utilizado, los fundamentos matemáticos que hay detrás de cada uno de estos modelos predictivos al ser programados, esto es importante para saber lo que hace cada uno de ellos, dado que en el código no se puede ver, simplemente se implementa una función determinada.

4.3.1 TEORÍA DE LAS REDES NEURONALES

Las neuronas funcionan procesando información. Reciben y proporcionan información en forma de picos. Las redes neuronales tienen una inclinación a adquirir el conocimiento a través de la experiencia, el cual es almacenado, al igual que en un cerebro humano.

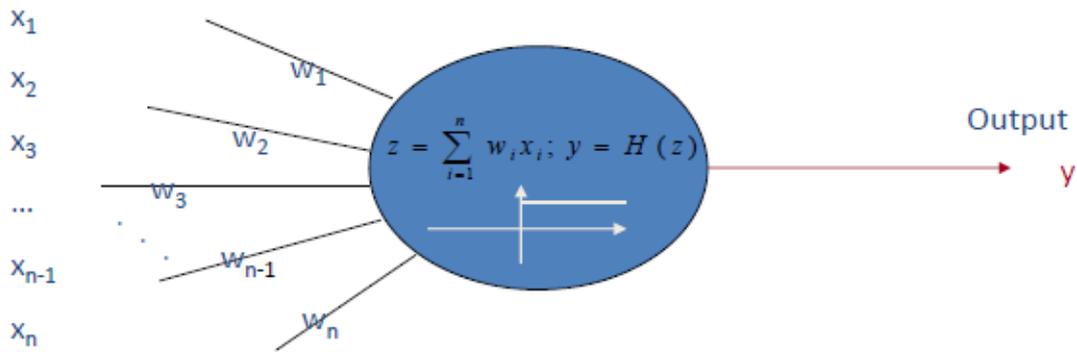


Figura 25: Modelo de McCulloch-Pitts [41]

En la Figura 25 podemos ver un vector de entrada (x_1, x_2, \dots, x_n), un vector de peso sináptico (w_1, w_2, \dots, w_n) y una salida (y) que será una función que dependa de x y w ($y = f(x, w)$).

En el modelo de McCulloch-Pitts de la Figura 25:

- La fuerza sináptica se traduce como pesos sinápticos.
- Al producto positivo entre la tasa de picos entrantes y el peso sináptico correspondiente se le conoce como excitación.
- Al producto negativo entre la tasa de picos entrantes y el peso sináptico correspondiente se le conoce como inhibición.

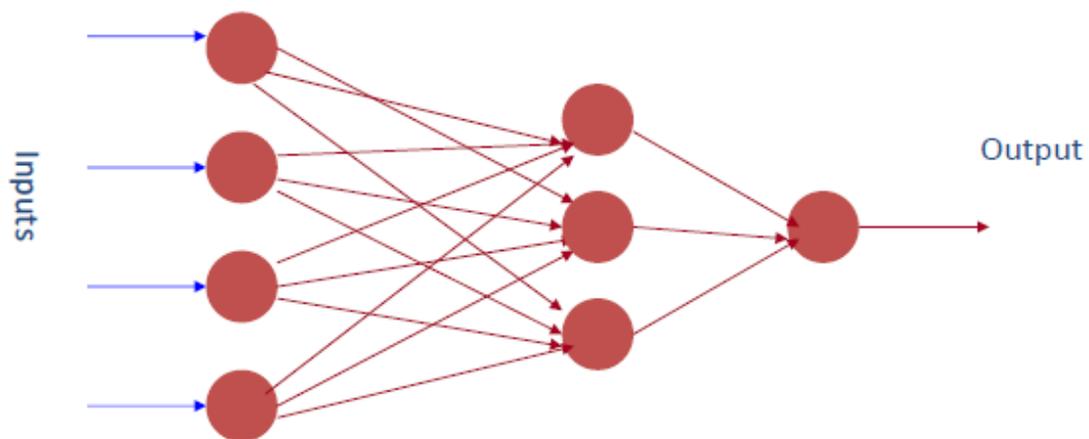


Figura 26: Red Neuronal Artificial. [41]

Una red neuronal artificial se compone de muchas neuronas artificiales interconectadas según una arquitectura de red específica, como se puede observar en la Figura 26, estas redes neuronales nos servirán para realizar predicciones.

La energía de una red neuronal artificial es una medida del error de rendimiento de una determinada tarea. Para optimizar el rendimiento de nuestra red aplicamos un modelo matemático.

Fundamentos Matemáticos

Supongamos que tenemos una red neuronal como la de la Figura 26, cada unidad básica (cada neurona artificial) se denomina perceptrón, en esta ocasión, nuestra red es del tipo MLP (perceptrón multicapa), pues tiene 2 capas ocultas antes de obtener la salida. En la primera capa contamos con 4 neuronas o perceptrones, cada una de ellas tendrá un valor de entrada asignado y otro valor correspondiente a la matriz de peso, las distintas salidas de cada una de las neuronas servirán de entradas para las neuronas de la siguiente capa, estas neuronas de la segunda capa tendrán asignado un valor de la matriz de peso, que junto con los valores de entrada nos proporcionarán una salida concreta (observar Figura 27 para una mejor comprensión del proceso).

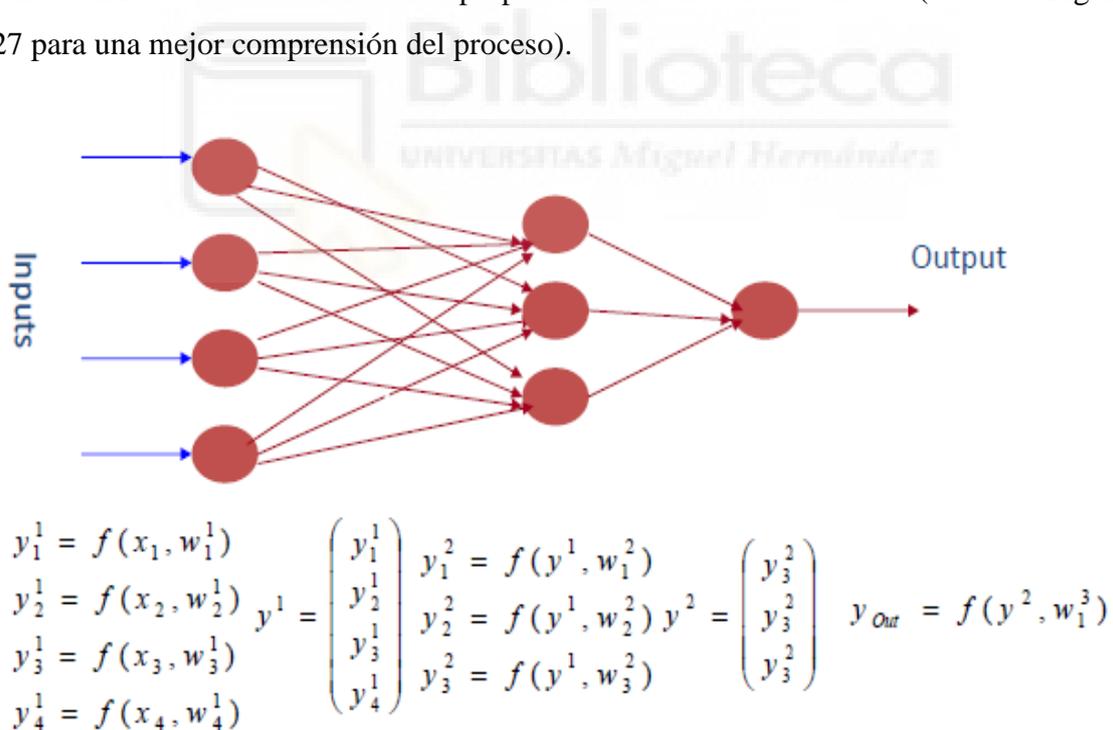


Figura 27: Principio matemático de aprendizaje para Redes Neuronales Artificiales. [41]

La salida de la primera capa oculta será la tangente hiperbólica del producto de la matriz de datos de entrada por la matriz de pesos más una matriz de “sesgo” o “matriz de bias”, evidentemente para que se pueda llevar a cabo este producto matricial el número de

columnas de la matriz de entrada tendrá que ser igual que el de filas de la matriz de pesos, teniendo el resultado las dimensiones del número de filas de la matriz de entrada por el número de columnas de la matriz de pesos, una vez hemos obtenido el producto de matrices sumamos los sesgos de la capa oculta, al resultado final se le aplicará la tangente hiperbólica. La matriz resultante durante todo este proceso será la matriz de entrada de la siguiente capa oculta, repitiéndose este mismo proceso para cada una de las capas presentes en la red. Finalmente, se obtiene una salida para cada posible entrada.

Cada capa oculta presenta una matriz de entrada, una matriz de peso y una matriz de “bias” diferente.

Para este experimento, vamos a utilizar el algoritmo L-BFGS (Limited memory – Broyden - Fletcher - Goldfarb - Shanno), este algoritmo es un método de optimización quasi-Newton de funciones con un gran número de parámetros o de una gran complejidad. Se trata de un método que hace un uso limitado de memoria y que permite obtener el mínimo de una función diferenciable. Para cada iteración el algoritmo busca una aproximación de la matriz Hessiana, concretamente de su inversa. Si la función tiene N variables, la matriz Hessiana tiene N^2 elementos. Si N es grande, el tiempo necesario para calcular toda la matriz de forma exacta puede ser prohibitivo, es por esto que se busca una aproximación [42]. L-BFGS es un algoritmo ideal para problemas en los que hay que tratar con muchas variables, como va a ser nuestro caso, además es un algoritmo eficiente en cuanto a memoria se refiere. Alguno de los posibles inconvenientes es que puede dar problemas para funciones no diferenciables y no puede resolver problemas con restricciones (por ejemplo, $x < 10$ sería una restricción que no podríamos añadir).

La función de activación de nuestra red neuronal multicapa va a ser la tangente hiperbólica [$\tanh = (e^x - e^{-x}) / (e^x + e^{-x})$] esta función tiene una salida centrada en 0, lo que ayudará a que el gradiente no esté sesgado en una dirección, los resultados de la tangente hiperbólica están comprendidos entre -1 y 1, lo que nos va a favorecer, porque en la última salida de la red, y solo en la última, vamos a usar la función `np.floor()` de la librería Numpy de Python, que va a redondear hacia abajo los números, es decir, todos los valores que sean negativo van a ser -1, y todos los valores positivos van a ser 0, si al vector de resultados, una vez se le ha aplicado la tangente hiperbólica, le

sumamos una unidad a cada elemento obtenemos un array de ceros y unos, que para nuestro caso, nos indicará si el paciente pertenece a un determinado grupo (un 1), o si no pertenece a ese grupo (un 0).

Todo este proceso matemático se puede ver de una forma más resumida como:

$$\mathbf{z} = \sum_{i=1}^n \mathbf{w}_i \mathbf{x}_i \quad \text{donde } \mathbf{w} \text{ es la matriz de pesos y } \mathbf{x} \text{ es la matriz de datos de entrada}$$

“z” será una matriz, a la que habrá que sumarle la matriz de bias o de sesgo

$$\mathbf{Z} = \mathbf{z} + \mathbf{B} \quad \text{donde } \mathbf{Z} \text{ es la matriz resultante de la suma y } \mathbf{B} \text{ es nuestra matriz de sesgo.}$$

Finalmente, la salida será: $\mathbf{y} = \mathbf{H}(\mathbf{Z})$ donde H va a ser nuestra función de activación, “y” será la salida y Z es la matriz resultante de la suma de arriba.

4.3.2 TEORÍA DE LAS MÁQUINAS DE VECTORES DE SOPORTE

Las máquinas de vectores de soporte son un conjunto de algoritmos de aprendizaje supervisado desarrollados por Vladimir Vapnik, son utilizados específicamente en clasificación y regresión.

Las SVM han demostrado tener un gran desempeño en muchas aplicaciones del mundo real. Sin embargo, a pesar de sus buenos fundamentos teóricos y buen desempeño al generalizar, no son adecuadas para clasificación con grandes conjuntos de datos, ya que la matriz del kernel crece de forma cuadrática con el tamaño del conjunto de datos, provocando que el entrenamiento de las SVM sobre conjuntos de datos grandes sea un proceso muy lento. En la literatura, la mayoría de los métodos de agilización de SVM usan el conjunto de datos entero con el objetivo de obtener buenas precisiones de clasificación. Sin embargo, la principal desventaja de las SVM es debido a su excesivo costo computacional en conjuntos de datos grandes. Algunos algoritmos han sido propuestos en la literatura, pero el tiempo de entrenamiento sigue siendo muy grande y prohibitivo en conjuntos de datos grandes según explica en su tesis doctoral el Dr. Jair Cervantes Canales en [43].

hiperplano de la Figura 28 es el que maximiza el margen de distancia entre ambos conjuntos de datos.

El vector W que aparece en la Figura 28 siempre tiene que ser normal al hiperplano.

El hiperplano separa las muestras positivas de las negativas, por lo tanto, los puntos que se encuentran en el hiperplano deben cumplir la siguiente ecuación:

$$\mathbf{w}x + \mathbf{b} = 0$$

La distancia o margen entre hiperplanos positivos y negativos viene dado por:

$$m = \frac{2}{\|\mathbf{w}\|}$$

Nos interesa maximizar ese margen, o lo que es lo mismo, minimizar la norma euclídea del vector W ($\min\{\|\mathbf{w}\|\}$), para así obtener el mejor hiperplano.

W puede calcularse empleando multiplicadores de Lagrange, como bien explica el Dr. Jair en [43]. El Lagrangiano es:

$$L(\omega, \mathbf{b}, \alpha) = \frac{1}{2} \|\mathbf{w}\|^2 - \sum_{i=1}^n \alpha_i (y_i (\omega x_i + \mathbf{b}) - 1)$$

Donde w es el vector de pesos, el que es normal al hiperplano, b es el sesgo, X_i e Y_i son el conjunto de datos, con $y_i \in \{-1, 1\}$ y $x_i \in \mathbb{R}$.

Para minimizar el Lagrangiano con respecto a w derivamos e igualamos a 0, observar la derivada a continuación:

$$\frac{\partial L}{\partial \omega} = \omega - \sum_{i=1}^n \alpha_i y_i x_i = 0$$

Recordamos que la derivada de un producto es la derivada del primer término multiplicando al segundo sin derivar más el primer término sin derivar por la derivada del segundo término. Si atendemos a la ecuación anterior llegamos a la siguiente expresión:

$$W = \sum_{i=1}^n a_i y_i x_i$$

Donde alfa es el multiplicador de Lagrange y va asociado a la restricción del dato en concreto.

Toda esta matemática nos sirve siempre y cuando los datos se puedan separar linealmente.

¿Qué pasa cuando los datos no se pueden separar linealmente? (Observar la Figura 29)

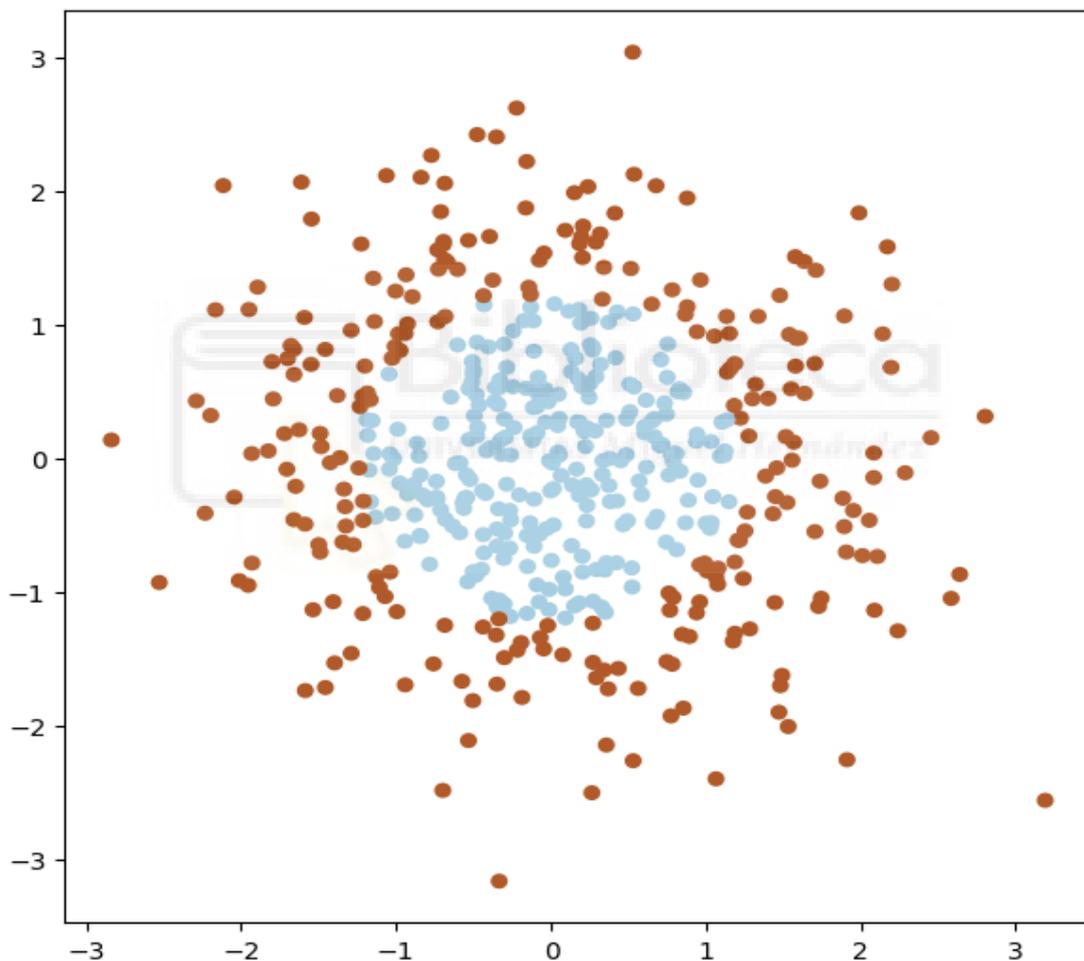


Figura 29: Clasificador que no permite separación lineal de los datos. [44]

En el caso de que ocurra algo parecido a lo que tenemos en la Figura 29, para permitir cierta flexibilidad, las SVM manejan un parámetro C que controla la compensación entre los errores de entrenamiento y los márgenes rígidos, creando así un margen flexible que permite algunos errores en la clasificación.

$$wx_i + b \geq 1 - \varepsilon_i \quad \forall y_i = 1$$

$$wx_i + b \leq -1 + \varepsilon_i \quad \forall y_i = -1$$

$$\varepsilon_i \geq 0 \quad \forall i$$

Cabe destacar que si $\varepsilon_i = 0$, no hay error, por lo que buscaríamos minimizar la siguiente expresión:

$$\frac{1}{2} \|w\|^2 + c \sum_{i=1}^n \varepsilon_i$$

La constante C determina la holgura del margen.

- Si C tiene un valor alto los datos de entrenamiento se clasificarán correctamente
- Si C tiene un valor bajo es posible que haya demasiados datos de entrenamiento mal clasificados.

Cuando los datos no pueden separarse linealmente, como es el caso de la Figura 29, se realiza un cambio de espacio mediante una función de transformación, que aumenta la dimensionalidad de los vectores de entrada a un espacio que puede separarse linealmente mediante un hiperplano. Esta función se denomina kernel.

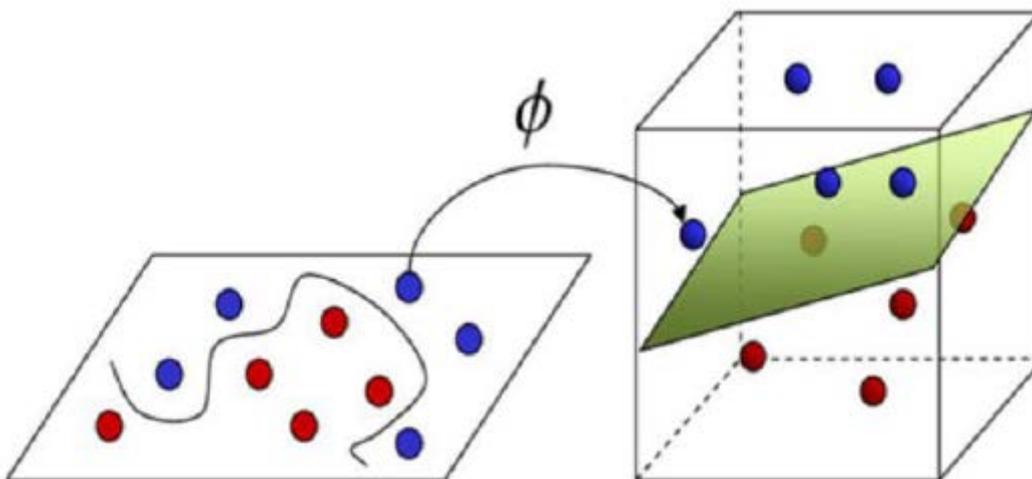


Figura 30: Aumento de dimensionalidad en los vectores de entrada. [44]

El espacio de características original siempre puede mapearse a un espacio de características de mayor dimensión donde el conjunto de entrenamiento sea separable, tal y como ocurre en la Figura 30.

Hay diferentes tipos de funciones del kernel:

- Kernel polinomial de grado d (utilizado en reconocimiento de imágenes).

$$k(\mathbf{x}, \mathbf{y}) = (\mathbf{x}^T \mathbf{y} + 1)^d$$

- Kernel Base Radial (su espacio de entrada es de dimensión infinita).

$$k(\mathbf{x}, \mathbf{y}) = e^{-\frac{\|\mathbf{x}-\mathbf{y}\|^2}{2\sigma^2}}$$

- Kernel sinusoidal con parámetros K y q .

$$k(\mathbf{x}, \mathbf{y}) = \text{tgh}(K\mathbf{x}^T \mathbf{y} + \theta)$$

Siendo **tgh** la tangente hiperbólica.

Cuando tenemos dos conjuntos de datos, tenemos que clasificar cada nuevo dato en una región u otra, para clasificar un nuevo punto se tiene en cuenta el signo de la siguiente función:

$$f(\mathbf{x}) = \sum_{i=1}^{N_s} \alpha_i y_i \phi(s_i) \phi(\mathbf{x}) + \mathbf{b} = \sum_{i=1}^{N_s} \alpha_i y_i k(s_i, \mathbf{x}) + \mathbf{b}$$

Donde, s_i son los vectores de soporte, y la función de transformación ϕ está definida por el kernel K :

$$\phi(\mathbf{x}) = k(s_i, \mathbf{x})$$

Si se obtiene un número negativo como resultado de la función $f(\mathbf{x})$, el punto se clasifica como -1 o su clase equivalente, y si se obtiene un número positivo, el punto se clasifica como +1 o su clase equivalente.

¿Qué pasaría si hubiese más de dos clases de datos? SVM está enfocado a clasificar dos clases de datos, podríamos enfocarlo de alguna manera para clasificar 3 clases de datos, pero en caso de que tuviésemos muchas variables para clasificar los datos se nos haría imposible llevar a cabo una representación para encontrar el hiperplano óptimo, dado que como máximo podemos representar en 3 dimensiones.

Si por ejemplo tuviésemos 3 clases de datos, podríamos enfrentar una clase contra el resto, es decir, hacer un grupo con dos de las clases y la tercera clase de forma individual.

SVM 1 vs 2&3; SVM 2 vs 1&3; SVM 3 vs 1&2

Otra manera sería enfrentar una clase contra otra y ver las 3 posibles representaciones.

SVM 1 vs 2; SVM 2 vs 3; SVM 1 vs 3

4.3.3 TEORÍA DE LOS ÁRBOLES DE DECISIÓN

Un árbol de decisión es un modelo de predicción cuyo objetivo principal es el aprendizaje inductivo a partir de observaciones y construcciones lógicas. Son muy similares a los sistemas de predicción basados en reglas, que sirven para representar y categorizar una serie de condiciones que suceden de forma sucesiva para la solución de un problema. Constituyen probablemente el modelo de clasificación más utilizado y popular. El conocimiento obtenido durante el proceso de aprendizaje inductivo se representa mediante un árbol. Un árbol gráficamente se representa por un conjunto de nodos, hojas y ramas. El nodo principal o raíz es el atributo a partir del cual se inicia el proceso de clasificación; los nodos internos corresponden a cada una de las preguntas acerca del atributo en particular del problema. Cada posible respuesta a los cuestionamientos se representa mediante un nodo hijo. Las ramas que salen de cada uno de estos nodos se encuentran etiquetadas con los posibles valores del atributo. Los nodos finales o nodos hoja corresponden a una decisión, la cual coincide con una de las variables clase del problema a resolver (observar la Figura 31).

Este modelo se construye a partir de la descripción narrativa de un problema, ya que provee una visión gráfica de la toma de decisión, especificando las variables que son evaluadas, las acciones que deben ser tomadas y el orden en el que la toma de decisión será efectuada. Cada vez que se ejecuta este tipo de modelo, sólo un camino será seguido

dependiendo del valor actual de la variable evaluada. Los valores que pueden tomar las variables para este tipo de modelos pueden ser discretos o continuos [45].

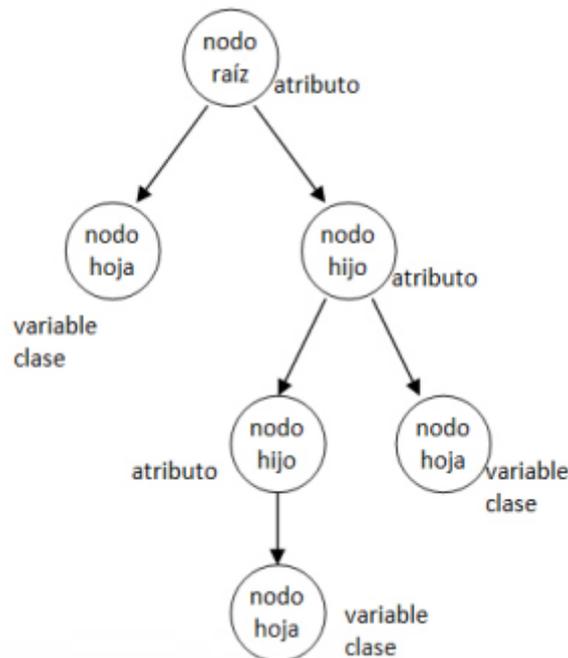


Figura 31: Estructura de un árbol de decisión. [46]

Un algoritmo de generación de árboles de decisión consta de 2 etapas: la primera corresponde a la inducción del árbol y la segunda a la clasificación. En la primera etapa se construye el árbol de decisión a partir del conjunto de entrenamiento; comúnmente cada nodo interno del árbol se compone de un atributo de prueba y la porción del conjunto de entrenamiento presente en el nodo es dividida de acuerdo con los valores que pueda tomar ese atributo. La construcción del árbol inicia generando su nodo raíz, eligiendo un atributo de prueba y dividiendo el conjunto de entrenamiento en dos o más subconjuntos; para cada partición se genera un nuevo nodo y así sucesivamente. Cuando en un nodo se tienen objetos de más de una clase se genera un nodo interno; cuando contiene objetos de una clase solamente, se forma una hoja a la que se le asigna la etiqueta de la clase. En la segunda etapa del algoritmo cada objeto nuevo es clasificado por el árbol construido; después se recorre el árbol desde el nodo raíz hasta una hoja, a partir de la que se determina la membresía del objeto a alguna clase. El camino a seguir en el árbol lo determinan las decisiones tomadas en cada nodo interno, de acuerdo con el atributo de prueba presente en él.

Nuestro objetivo va a ser evaluar el desempeño de clasificación de nuestros datos clínicos a partir de los resultados obtenidos en la aplicación de algoritmos basados en árboles de decisión, para poder determinar si esta técnica de clasificación puede ser una herramienta de soporte y ayuda eficaz en el tratamiento y diagnóstico médico.

Fundamentos Matemáticos

En primer lugar, en nuestro experimento vamos a implementar dos tipos de árboles de decisión, según su medida de impureza, uno de ellos tendrá medida de impureza Gini, y el otro la tendrá de entropía.

Comenzaremos viendo en que consiste un árbol de decisión con el criterio Gini a nivel matemático.

$$GINI = 1 - \sum_{i=1}^c p_i^2$$

Donde P_i es la proporción de elementos de la clase i en el nodo y C es el número de clases.

Para entender lo que hace este modelo a nivel matemático pondremos un ejemplo simple. Supongamos un conjunto de datos pequeños como el siguiente:

Tabla 7: Conjunto pequeño de datos para la comprensión del criterio Gini.

Edad del estudiante	¿Ha estudiado el examen?	¿Aprueba el examen? (Clase)
22	Si	Si
25	No	No
28	Si	Si
35	No	No
40	Si	Si

Calculamos la impureza Gini del nodo raíz. Tenemos 5 muestras totales, 3 aprueban y 2 no lo hacen.

$$GINI = 1 - \left(\left(\frac{3}{5} \right)^2 + \left(\frac{2}{5} \right)^2 \right) = 1 - (0,36 + 0,16) = 0,48$$

Evaluamos posibles divisiones de datos, aquí lo haremos según la edad y según si estudia o no.

- Opción A: dividir en función de si estudia

Rama izquierda: no estudia = 0 ; tenemos 2 casos y ninguno aprueba.

$$GINI_{izq} = 1 - \left(\frac{2}{2} \right)^2 = 0$$

Rama derecha: si estudia = 1 ; tenemos 3 casos y aprueban todos.

$$GINI_{der} = 1 - \left(\frac{3}{3} \right)^2 = 0$$

Calculamos el total: $GINI_T = \frac{2}{5} \cdot (0) + \frac{3}{5} \cdot (0) = 0$

División perfecta, pura al 100%.

- Opción B: dividimos por edad < 30 años

Rama izquierda: (22, 25, 28) ; clase aprobados (1, 0, 1) aprueban el primero y el tercero, el segundo suspende.

$$GINI_{izq} = 1 - \left(\left(\frac{2}{3} \right)^2 + \left(\frac{1}{3} \right)^2 \right) = 0,444$$

Rama derecha: (35, 40) ; clase aprobados (0, 1) el primero suspende, el segundo aprueba.

$$GINI_{der} = 1 - \left(\left(\frac{1}{2} \right)^2 + \left(\frac{1}{2} \right)^2 \right) = 0,5$$

Calculamos el total: $GINI_T = \frac{3}{5} \cdot 0,444 + \frac{2}{5} \cdot 0,5 = 0,466$

Nos interesa tener un Gini total bajo ($0 < 0'466$), luego el algoritmo elige si estudia o no como la mejor variable para el primer nodo (reduce más la impureza). Este método se repite recursivamente en cada nodo hijo.

Ahora veremos a nivel matemático como funcionaría un árbol de decisión con criterio de entropía.

$$E = - \sum_{i=1}^n p_i \log_2(p_i)$$

Cuanto más balanceadas las clases, mayor entropía. Si la entropía es 0, el nodo es puro (solo una clase), pero si la entropía es 1, las clases están distribuidas al 50%.

Veamos un ejemplo sencillo de árbol de decisión con criterio de entropía, supongamos un conjunto de datos como el siguiente:

Tabla 8: Conjunto pequeño de datos para la comprensión del criterio de Entropía.

Edad	Ingresos (alto/bajo)	¿Se va de compras?
Joven	Bajo	No
Joven	Bajo	No
Mediana	Alto	Si
Mayor	Alto	Si
Mayor	Bajo	Si
Mayor	Bajo	No
Mediana	Bajo	Si
Joven	Alto	Si
Joven	Alto	Si
Mayor	Alto	No

La entropía del nodo raíz es:

$$E = -0,6 \log_2(0,6) - 0,4 \log_2(0,4) \approx 0,971$$

La P_i del sumatorio indica el número de personas que se van de compras y el número de personas que no se van de compras.

- Evaluamos la variable Edad:

Tenemos 4 jóvenes, los 2 primeros no se van de compras, los otros dos sí, luego la clase

es (0, 0, 1, 1).
$$E_J = \frac{-2}{4} \log_2(0,5) - \frac{2}{4} \log_2(0,5) = 1$$

Tenemos 2 personas de mediana edad, los dos se van de compras, luego la clase es (1, 1).

$$E_{MED} = \frac{-2}{2} \log_2(1) = 0$$

Tenemos 4 personas mayores, los 2 primeros se van de compras, los otros 2 no, luego la

clase es (1, 1, 0, 0).
$$E_M = -\frac{2}{4} \log_2(0,5) - \frac{2}{4} \log_2(0,5) = 1$$

Calculamos el total:
$$E_T = \frac{4}{10} \cdot 1 + \frac{2}{10} \cdot 0 + \frac{4}{10} \cdot 1 = 0,8$$

- Ahora evaluamos la variable ingreso:

Tenemos 5 personas con ingresos altos, de los cuales 4 se van de compras, luego la clase

es (1, 1, 1, 1, 0).
$$E_A = \frac{-4}{5} \log_2(0,8) - \frac{1}{5} \log_2(0,2) \approx 0,722$$

Tenemos 5 personas con ingresos bajos, de los cuales 2 se van de compras, luego la clase

es (0, 0, 1, 0, 1).
$$E_B = \frac{-3}{5} \log_2(0,6) - \frac{2}{5} \log_2(0,4) = 0,971$$

Calculamos el total:
$$E_T = \frac{5}{10} \cdot 0,722 + \frac{5}{10} \cdot 0,971 \approx 0,847$$

Finalmente, evaluamos la ganancia de información para ambas variables.

- Para la variable edad: $0,971 - 0,8 = 0,171$
- Para la variable ingreso: $0,971 - 0,847 = 0,124$

Dado que $0,171 > 0,124$ el árbol de decisión seleccionaría la edad como la primera variable de división, porque reduce más la incertidumbre (entropía) en la variable objetivo.

Espero que estos 2 ejemplos prácticos simples, uno para cada criterio, sirva para entender el funcionamiento del árbol de decisión a nivel matemático.

4.3.4 TEORÍA DEL ANÁLISIS DE COMPONENTES PRINCIPALES

El Análisis de Componentes Principales (PCA) es una técnica estadística de síntesis de la información, o reducción de la dimensión (número de variables). Es decir, ante un banco de datos con muchas variables, el objetivo será reducirlas a un menor número perdiendo la menor cantidad de información posible.

Los nuevos componentes principales o factores serán una combinación lineal de las variables originales, y además serán independientes entre sí.

Un aspecto clave en PCA es la interpretación de los factores, ya que ésta no viene dada a priori, sino que será deducida tras observar la relación de los factores con las variables iniciales (habrá, pues, que estudiar tanto el signo como la magnitud de las correlaciones), tal y como explica Manuel Terrádez, autor de [47].

Al ser PCA una técnica de reducción de la dimensionalidad para sintetizar la información (no es un modelo para que generemos predicciones), tendremos que juntar esta técnica con las redes neuronales, máquinas de vectores soporte o árboles de decisión para obtener una predicción, es decir, reducimos la dimensionalidad de nuestra base de datos con PCA, y esos nuevos datos serán ahora la entrada de cualquiera de los métodos predictivos anteriormente mencionados. Veremos si los resultados experimentales aplicando PCA a nuestros modelos predictivos son mejores o peores que sin aplicarlo.

Fundamentos Matemáticos

El algoritmo de PCA se compone de 5 pasos basados en las ideas matemáticas de varianza y covarianza además de vectores y valores propios. Los pasos son los siguientes:

1º Normalización de los datos:

$$Z = (\text{valor} - \text{media}) / \text{desviación estándar}$$

2º Cálculo de la matriz de covarianza: es una matriz de dimensiones $p \times p$ simétrica cuyas entradas son las covarianzas asociadas a todos los posibles pares de variables originales. Para un conjunto de datos tridimensionales con 3 variables x, y, z , la matriz de covarianza es una matriz de datos de 3×3 como la siguiente.

$$C = \begin{pmatrix} \mathit{cov}(x, x) & \mathit{cov}(x, y) & \mathit{cov}(x, z) \\ \mathit{cov}(x, y) & \mathit{cov}(y, y) & \mathit{cov}(y, z) \\ \mathit{cov}(x, z) & \mathit{cov}(z, y) & \mathit{cov}(z, z) \end{pmatrix}$$

Las covarianzas son conmutativas, es decir, $\mathit{cov}(a, b) = \mathit{cov}(b, a)$ luego se queda una matriz simétrica. Hay que saber también que $\mathit{cov}(a, a) = \mathit{var}(a)$, luego la matriz se queda de la siguiente manera:

$$C = \begin{pmatrix} \mathit{VAR}(x) & \mathit{cov}(x, y) & \mathit{cov}(x, z) \\ \mathit{cov}(x, y) & \mathit{VAR}(y) & \mathit{cov}(y, z) \\ \mathit{cov}(x, z) & \mathit{cov}(y, z) & \mathit{VAR}(z) \end{pmatrix}$$

3º Cálculo de autovectores y autovalores: tenemos que resolver la siguiente ecuación matricial.

$$|A - \lambda \cdot I|$$

Donde A es nuestra matriz de covarianza, λ son nuestros autovalores e I es la matriz identidad. Una vez hemos calculado el determinante de $A - \lambda \cdot I$ igualamos a 0, se nos debe quedar una ecuación de grado 3 (dado que tenemos un conjunto de datos tridimensionales con 3 variables), resolvemos la ecuación bien aplicando un factor común que nos reduzca la ecuación a una de segundo grado, o bien con el método de Ruffini, siempre y cuando la ecuación de grado 3 tenga término independiente.

Debemos tener tres valores de λ como soluciones, aunque puede ser que alguna de estas soluciones esté repetida. Realizamos la ecuación matricial $A - \lambda \cdot I$ para los valores de λ que nos han salido como soluciones, para cada valor de λ obtendremos una matriz 3×3 , que multiplicaremos por un vector columna (x, y, z) , para así conformar un sistema de 3 ecuaciones con 3 incógnitas. Resolvemos cada uno de los sistemas y obtenemos para cada uno de ellos una solución de “ x ”, otra de “ y ” y otra de “ z ”, las

soluciones de cada sistema formarán un vector concreto, estos vectores se usarán para formar la matriz de proyección.

Nota: los sistemas de 3 ecuaciones con 3 incógnitas no tienen por qué ser siempre sistemas compatibles determinados con soluciones concretas, puede ser que el sistema sea incompatible o que sea compatible indeterminado y tengamos infinitas soluciones, y, por tanto, infinitos posibles vectores que cumplan las condiciones de nuestro sistema. Para asegurarnos de conocer con qué tipo de sistema de ecuaciones estamos trabajando es recomendable aplicar el teorema de **Rouché-Frobenius**, este teorema nos permitirá clasificar el sistema, y en caso de que el sistema sea compatible indeterminado también nos dice que grado de dependencia tiene.

Formamos la matriz de proyección poniendo en la primera columna el vector de soluciones del 1er sistema, seguimos el mismo proceso con los otros dos sistemas para completar la segunda y tercera columna. En caso de que alguno de los sistemas sea compatible indeterminado, cogemos cualquier vector de soluciones que cumpla el sistema de entre los infinitos que hay y lo colocamos en la columna que le corresponda en la matriz de proyección. Se tendrá que cumplir la siguiente condición:

$$D = P^{-1} \cdot A \cdot P$$

Donde “P” es nuestra matriz de proyección, “P⁻¹” es su inversa, “A” es la matriz de covarianza y “D” debe salirnos una matriz como la siguiente:

$$D = \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{pmatrix}$$

Siendo λ_1 , λ_2 y λ_3 las soluciones a la ecuación de grado 3 que nos salía al igualar a cero el determinante $|A - \lambda \cdot I|$.

4º Selección de componentes principales: se seleccionan los autovectores correspondientes a los autovalores más grandes (los vectores correspondientes a los valores de lambda más grandes). Estos autovectores representan los componentes principales que explican la mayor parte de la varianza en los datos.

5º Transformación de los datos: se proyectan los datos originales sobre los autovectores seleccionados para obtener el nuevo conjunto de variables.

Matriz de datos originales \cdot P (matriz de proyección) = matriz de datos proyectados



5 EXPERIMENTOS Y RESULTADOS

En esta parte del trabajo vamos a narrar como hemos llevado a cabo los experimentos. También hablaremos del enfoque experimental que se le va a dar a cada uno de estos métodos y, finalmente, de los resultados producidos por cada modelo, resultados que se entenderán y comentarán para así poder llegar a una conclusión, de la que se hablará en el apartado 6.

5.1 ENFOQUE EXPERIMENTAL DE LAS REDES NEURONALES (NN)

Sabemos que la salida de la red neuronal va a ser un 0 o un 1, luego no podemos establecer de forma directa a que grupo pertenece cada paciente de los 4 posibles (0,1,2,3), pues solo podemos abarcar 2 posibilidades (0,1), debido a esto vamos a realizar diversas redes neuronales aplicadas a cada uno de los grupos, de manera que podamos saber si un paciente pertenece o no a un determinado grupo.

En primer lugar, voy a importar la base de datos en formato Excel y le voy a asignar una variable, ahí tendremos los datos de los 1053 pacientes. Una vez tengo los datos realizo una división de los mismos, declaro una variable objetivo, que va a ser una columna en concreto de la base datos, la de los grupos a los que pertenecen los pacientes, y otra variable para el resto de los datos. Con la división ya hecha, recorro el vector objetivo, de manera que si encuentro un número que sea igual al de un grupo en concreto me devuelve un true, mientras que si encuentro un número que sea distinto al grupo pertinente me devuelve un false, de esta manera tengo un vector de booleanos (true/false) que convertiré a números enteros, cambiando los trues por unos y los false por ceros.

A la hora de trabajar con nuestra base de datos ha surgido un problema, hay celdas que presentan NaNs (Not a Number), los NaN se pueden sustituir de varias formas, realizando la media de la columna en la que se encuentren y sustituyéndolo por este valor, sustituyéndolos por ceros o directamente eliminarlos, este último método no es viable a excepción de en la variable objetivo, esto se debe a que cada columna tiene un número determinado de NaNs, si los eliminamos cada columna tendrá una cantidad de datos distinta, haciendo imposible la formación de la matriz y su posterior producto con otra

matriz. En este experimento los sustituiremos por ceros, esto nos evitará problemas más adelante.

A continuación, se implementan varios bucles anidados, con el objetivo de probar varias configuraciones de redes neuronales, cada una de ellas con parámetros diferentes. Un bucle indicará el porcentaje de entrenamiento y test de los datos, es decir, que cantidad de datos va a utilizar el modelo para entrenar y que cantidad va a emplear en comprobar si la salida del modelo de predicción se corresponde con el modelo real. Utilizaremos dos bucles más, dado que nuestra red neuronal artificial va a tener dos capas ocultas, uno de los bucles indicará el número de neuronas de la primera capa, y el otro, indicará el número de neuronas de la otra capa. Todos estos bucles nos permitirán evaluar 1440 configuraciones distintas de nuestra red neuronal.

Dentro de los bucles configuraremos la red neuronal, para ello empleamos la función “MLPClassifier”, dentro de la función indicamos el algoritmo L-BFGS, la tangente hiperbólica como función de activación, los parámetros que hacen referencia al número de neuronas (que se irán actualizando en cada iteración de los bucles) y declaramos una semilla nula, para que así en cada configuración de la red coja un conjunto de datos aleatorios, si pusiéramos semilla siempre cogería el mismo conjunto de datos para entrenamiento y el mismo conjunto de datos para el test.

Llevaremos a cabo otro bucle, este iterará cada configuración 10 veces, pero mezclaremos los datos con el comando “shuffle” entre iteración e iteración, de manera que el resultado cambiará cada vez, nosotros calcularemos la media y la desviación estándar de las 10 salidas para cada configuración.

Después de todo lo anterior, dividiremos los datos, los estandarizaremos, realizaremos el entrenamiento de la red y su predicción.

Evaluamos los resultados con la función “classification_report”, esto nos devolverá unos resultados en forma de “precision(1/0)” y “recall (1/0)”. Transformamos estos valores a VPP (valor predictivo positivo), VPN (valor predictivo negativo), sensibilidad y especificidad, la transformación será de la siguiente manera:

- Precision en 0 = especificidad
- Precision en 1 = sensibilidad

- Recall en 0 = VPN
- Recall en 1 = VPP

Finalmente, sacamos los valores medios y las desviaciones estándar de los 4 parámetros mencionados arriba y cada una de las configuraciones para los que tenemos esos 4 valores. Formamos un “dataframe” con los resultados y los guardamos en formato “.xlsx” para así poder descargar los resultados en Excel y poder llevar a cabo un análisis estadístico en profundidad.

He insertado en el Excel de cada grupo de pacientes varias gráficas, haciendo referencia al VPP, VPN, sensibilidad y especificidad, tanto su media como su desviación estándar, para cada combinación de neuronas y para cada porcentaje de entrenamiento y test. Para no saturar la memoria de figuras y tablas solo comentaré la mejor configuración de cada grupo.

Resultados Experimentales del Grupo 0

Para el grupo 0 la mejor red neuronal artificial se produce en la iteración 58 (fila 59 del Excel), la configuración y resultados han sido los siguientes:

- Configuración:

Tabla 9: Configuración Red Neuronal óptima para pacientes del Grupo 0

iteración	% test	% entrenamiento	Neuronas capa 1	Neuronas capa 2
58	5%	95%	22	47

- Resultados:

Tabla 10: Resultados Red Neuronal óptima para pacientes del Grupo 0

VPP	VPN	Sensibilidad	Especificidad	Media total
0'625	0'978	0'51875	0'99	0'778

Para los pacientes del grupo 0, que son los no afectados por la miocardiopatía hipertrófica, tenemos que:

- De todos los casos que nuestro modelo ha predicho como pertenecientes al grupo 0, el 62'5% lo eran.
- De todos los casos que nuestro modelo ha predicho como no pertenecientes al grupo 0, el 97'8% efectivamente no lo eran.
- La sensibilidad es una métrica que hace referencia a como de bien identifica un modelo correctamente los casos positivos reales. Viene dada por: $S = \frac{vp}{vp+FN}$ donde vp son los verdaderos positivos y FN son los falsos negativos. En nuestra red óptima para el grupo 0 la sensibilidad es del 51'875%.
- La especificidad es una métrica que hace referencia a la capacidad del modelo para identificar correctamente los casos negativos. Viene dada por: $E = \frac{vN}{vN+FP}$ donde vN son los verdaderos negativos y FP son los falsos positivos. En nuestra red óptima para el grupo 0 la especificidad es del 99%.

La media de estos 4 parámetros es de un **77'8%**.

Como en el grupo 0 son los pacientes no afectados, voy a insertar las gráficas de la media de la VPN en función del porcentaje de entrenamiento y prueba y en función de la combinación de neuronas.

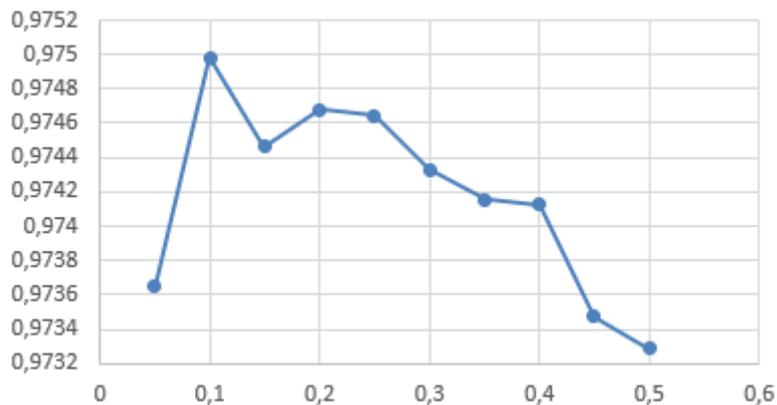


Figura 32: Media de VPN en función del porcentaje de pruebas.

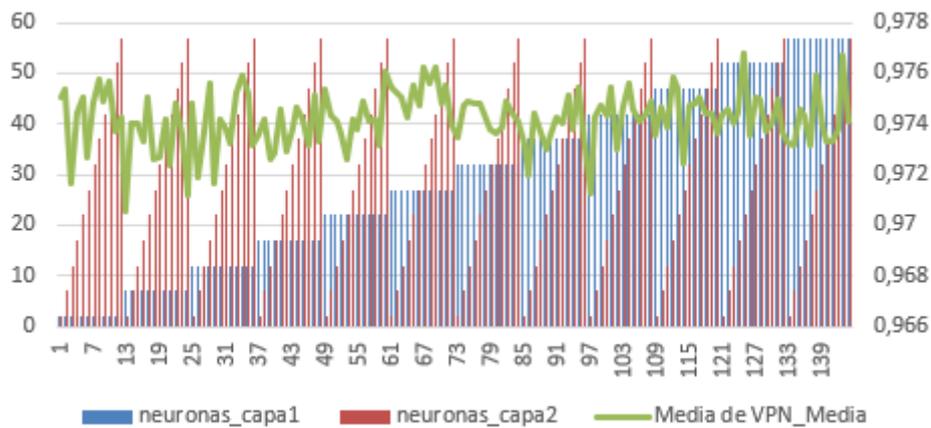


Figura 33: Media de VPN en función de la combinación de neuronas.

Resultados Experimentales del Grupo 1

Para el grupo 1 la red neuronal artificial óptima se produce en la iteración 113 (fila 114 en el Excel), la configuración y resultados han sido los siguientes:

- Configuración:

Tabla 11: Configuración Red Neuronal óptima para pacientes del Grupo 1

iteración	% test	% entrenamiento	Neuronas capa 1	Neuronas capa 2
113	5%	95%	47	22

- Resultados:

Tabla 12: Resultados Red Neuronal óptima para pacientes del Grupo 1

VPP	VPN	Sensibilidad	Especificidad	Media total
0'45	0'9823	0'6481	0'96	0'7603

Para los pacientes del grupo 1, que son pacientes portadores genéticos sin cardiopatía, es decir, genéticamente predispuestos a poder sufrir la patología, pero sin sintomatología de la afección, tenemos que:

- El modelo predictivo ha acertado el 45% de los pertenecientes al grupo 1.

- Del conjunto de pacientes no pertenecientes al grupo 1, el 98'23% ha sido acertado por el modelo predictivo.
- Nuestra sensibilidad es del 64'81% y nuestra especificidad del 96%.

La media de estos 4 parámetros es de un **76'03%**.

Resultados Experimentales del Grupo 2

Para el grupo 2 la red neuronal artificial óptima se produce en la iteración 32 (fila 33 en el Excel), la configuración y resultados han sido los siguientes:

- Configuración:

Tabla 13: Configuración Red Neuronal óptima para pacientes del Grupo 2

iteración	% test	% entrenamiento	Neuronas capa 1	Neuronas capa 2
32	5%	95%	12	37

- Resultados:

Tabla 14: Resultados Red Neuronal óptima para pacientes del Grupo 2

VPP	VPN	Sensibilidad	Especificidad	Media total
0'5983	0'9588	0'5595	0'9609	0'7694

Para los pacientes del grupo 2, que son los portadores con posibilidad de ser afectado, pruebas dudosas y genética positiva, tenemos que:

- El modelo predictivo identifica correctamente al 59'83% de los pacientes pertenecientes al grupo 2.
- El modelo ha identificado correctamente al 95'88% de los pacientes que no pertenecen al grupo 2.
- La sensibilidad en esta ocasión es del 55'95% y la especificidad del 96'09%.

La media de estos 4 parámetros es de un **76'94%**.

Resultados Experimentales del Grupo 3

Para el grupo 3 la red neuronal artificial óptima se produce en la iteración 135 (fila 136 en el Excel), la configuración y resultados han sido los siguientes:

- Configuración:

Tabla 15: Configuración Red Neuronal óptima para pacientes del Grupo 3

iteración	% test	% entrenamiento	Neuronas capa 1	Neuronas capa 2
135	5%	95%	57	12

- Resultados:

Tabla 16: Resultados Red Neuronal óptima para pacientes del Grupo 3

VPP	VPN	Sensibilidad	Especificidad	Media total
0'9262	0'7352	0'9673	0'5429	0'7929

Para los pacientes del grupo 3, que son los afectados, pruebas afectadas y portadores genéticos, tenemos que:

- El modelo predictivo identifica correctamente al 92'62% de los pacientes afectados por MCH. Los pacientes del grupo 3 son el grupo mayoritario en nuestra base de datos, así que es importante tener un valor predictivo positivo elevado, en otras palabras, tener un VPP alto significa que tenemos muy pocos falsos positivos.
- El modelo ha identificado con éxito a un 73'52% de los pacientes que no pertenecen al grupo 3, como este porcentaje es menor que el del valor predictivo positivo ($VPP > VPN$) sabemos que hay más falsos negativos que falsos positivos.
- La sensibilidad es del 96'73% y la especificidad es del 54'29%.

La media de estos 4 parámetros es de un **79'29%**.

Como en el grupo 3 son los pacientes afectados, voy a insertar las gráficas de la media de la VPP en función del porcentaje de entrenamiento y prueba y en función de la combinación de neuronas.

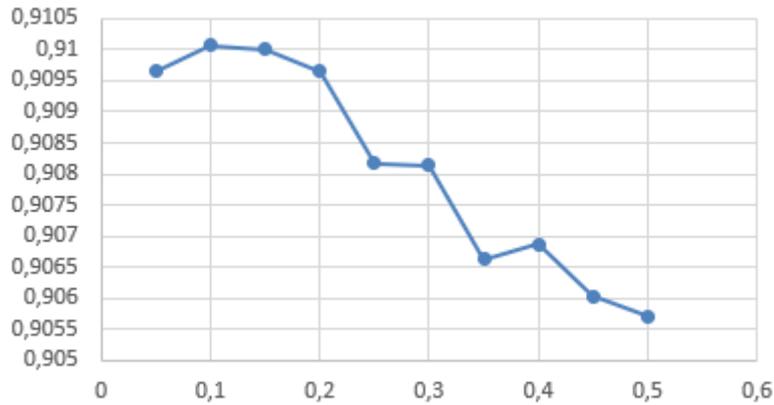


Figura 34: Media de VPP en función del porcentaje de pruebas.

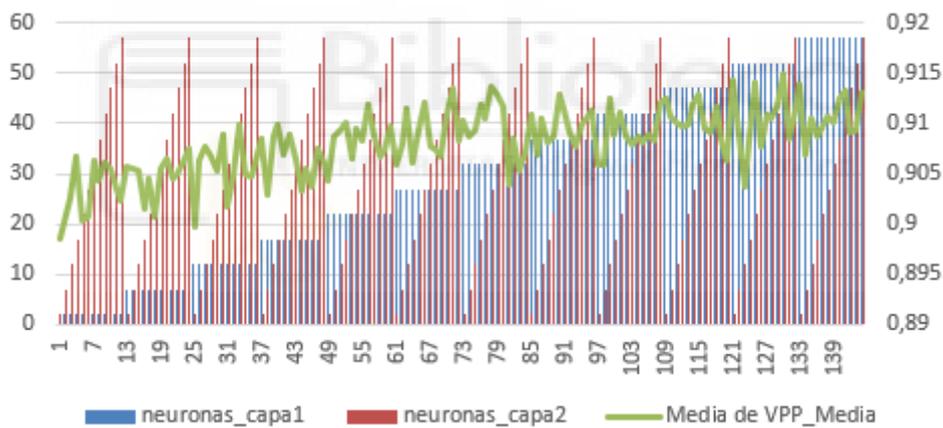


Figura 35: Media de VPP en función de la combinación de neuronas.

5.2 ENFOQUE EXPERIMENTAL DE LAS MÁQUINAS DE VECTORES DE SOPORTE (SVM)

El enfoque experimental no es muy diferente al empleado con las redes neuronales. En primer lugar, debemos saber que no vamos a poder representar en un plano o en 3 dimensiones nuestros datos de entrada, ni mucho menos trazar un hiperplano, como he explicado anteriormente, para llevar a cabo una representación gráfica de los datos de entrada podemos contar con 2 o 3 variables, en nuestra base de datos contamos con más de 70 variables.

Una vez he importado todas las librerías y he asignado la base de datos a una variable, suprimo los NaN de nuestra variable objetivo, que recuerdo que es la variable que contiene la información de a qué grupo pertenecen los pacientes.

Separamos los datos, asignamos una variable al objetivo y otra variable al resto del conjunto de datos. En la variable objetivo ya hemos suprimido los NaN, en el resto del conjunto de datos los sustituiremos por ceros.

Con la función “train_test_split” separamos nuestros datos en un conjunto de prueba y en otro de entrenamiento.

Estandarizamos los conjuntos de prueba y entrenamiento de la variable que contiene al conjunto de datos completo menos la variable objetivo (el grupo del paciente). Definimos una matriz de ceros de dimensiones 9x9 que será nuestra matriz de precisión.

Creamos 2 arrays, uno de 50 elementos distribuidos logarítmicamente entre 10^{-10} y 10^{-1} y otro de 9 elementos que va desde 0.1 hasta 10 de forma equiespaciada.

Preparamos los datos para el ajuste de parámetros, creando todas las combinaciones posibles de los valores “c” (parámetro de regularización) y “g” (influencia del kernel) y una matriz para almacenar los resultados de precisión para cada combinación.

Creamos una instancia de SVM con la función svm.SVC, esta función va a aplicar todo el proceso matemático explicado anteriormente, el kernel que utilizaremos será el de base radial, ya que el espacio de los datos de entrada es grande. Entrenamos el modelo de SVM y calculamos su predicción sobre los datos estandarizados, calculamos la precisión del modelo para cada combinación.

Finalmente, tomo los resultados de precisión obtenidos para cada combinación de parámetros y los represento, también imprimo por pantalla la matriz de precisión.

Resultados Experimentales

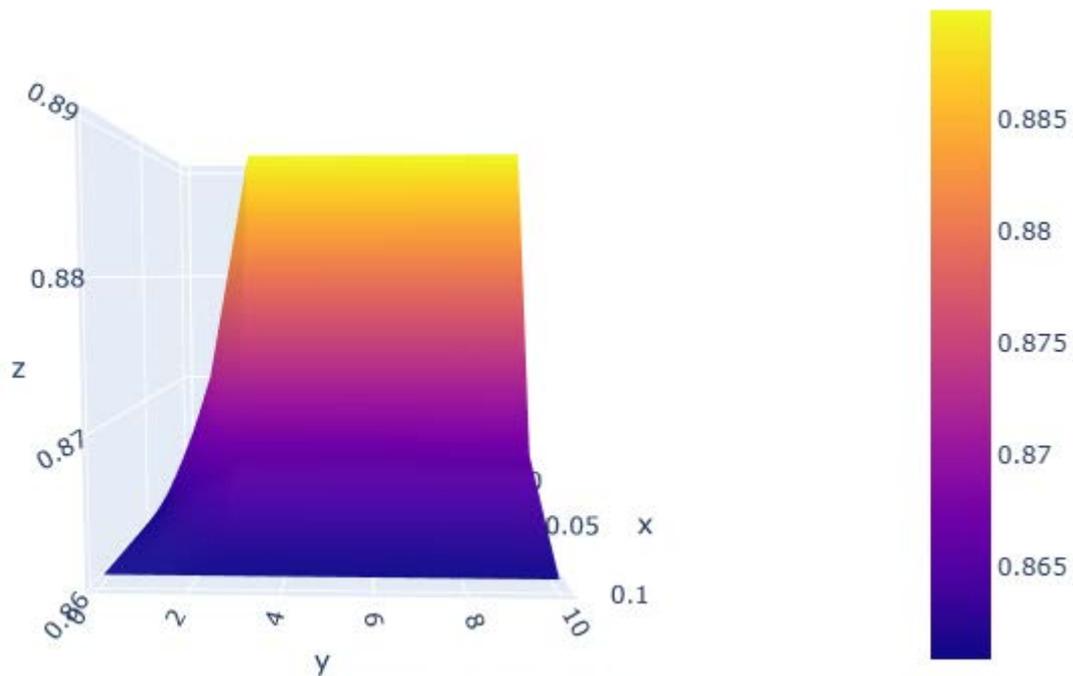


Figura 36: Representación gráfica del modelo de SVM visto desde el plano ZY

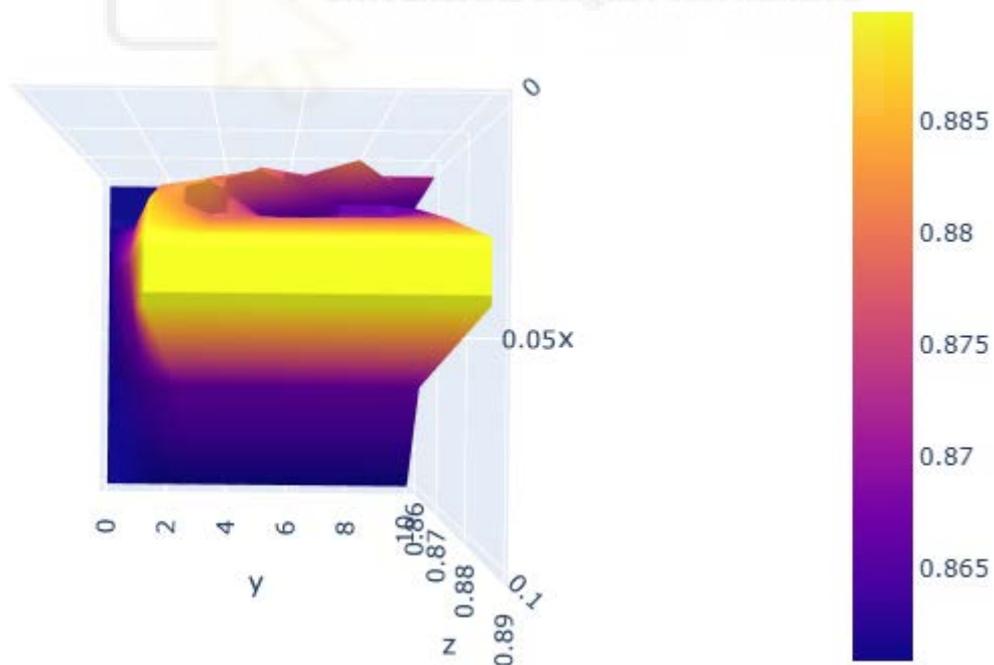


Figura 37: Representación gráfica del modelo de SVM visto desde el plano XY

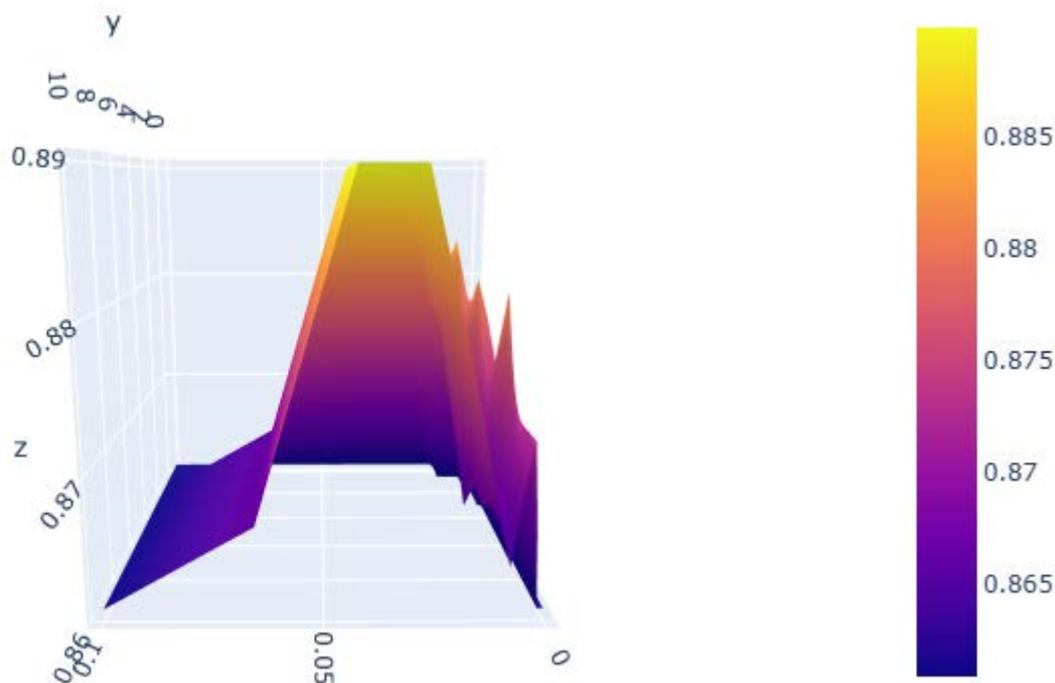


Figura 38: Representación gráfica del modelo de SVM visto desde el plano ZX

En la Figura 36, Figura 37 y Figura 38 podemos apreciar la representación gráfica de los resultados, como la representación es tridimensional he creído conveniente adjuntar las 3 figuras anteriores, donde se pueden apreciar la vista de los tres planos posibles.

Los ejes X e Y representan valores del parámetro de regularización (c) y los del parámetro de influencia del kernel (g) respectivamente. El eje Z representa la precisión de nuestras predicciones, nos interesa encontrar los valores del eje X e Y que maximizan este valor y de cuanto es el mismo.

En la Figura 34 se puede apreciar para que valores del eje X la representación alcanza su máximo en el eje Z y en la Figura 32 se puede ver los valores del eje Y que maximizan los valores del eje Z.

El eje Z alcanza su valor máximo $\forall y \in [2.575, 10]$; $\forall x \in [0.0281, 0.0429]$

Dicho valor máximo es de **0.8899**, luego con SVM hemos alcanzado una precisión del **88.99 %** aproximadamente.

5.3 ENFOQUE EXPERIMENTAL DE LOS ÁRBOLES DE DECISIÓN

El enfoque de este modelo es parecido al empleado en redes neuronales, importamos las librerías y la base de datos, dividimos el conjunto de datos en una variable objetivo y en otra variable que tendrá los datos restantes. En la variable objetivo eliminamos los NaN y en la otra variable, la del conjunto de datos, sustituiremos los NaN por ceros.

Vamos a emplear 3 bucles anidados, el primero va a servir para indicar el porcentaje de datos utilizados para las pruebas y el entrenamiento, el segundo va a ir incrementado el parámetro de profundidad del árbol, el tercero va a realizar 10 iteraciones de nuestro árbol, dichas iteraciones cogerán cada vez un conjunto de datos aleatorios, pues hemos eliminado la semilla, este tercer bucle servirá para que podamos calcular la media y la desviación estándar del valor predictivo positivo, el valor predictivo negativo, la sensibilidad y la especificidad. Entre los 3 bucles se llevarán a cabo 810 iteraciones, 81 configuraciones distintas que se realizarán 10 veces (cada vez con un conjunto de datos distinto).

Dentro de los bucles llevaremos a cabo la división de datos en entrenamiento y prueba (sin emplear semilla), luego usaremos la función “DecisionTreeClassifier”, en esta función emplearemos primero el criterio de Gini y después el de entropía. Entrenamos el modelo y realizamos su predicción.

Clasificamos los resultados en VPP, VPN, sensibilidad y especificidad, llevamos a cabo su media y su desviación estándar y guardamos estos resultados en un “dataframe” que exportaremos en formato “.xlsx”, para así tenerlo todo en Excel, tanto las distintas configuraciones del modelo como sus resultados.

Resultados Experimentales

Expresamos los resultados mediante mapas de calor, es una manera de ver los resultados muy intuitiva además de visual. Cada grupo de pacientes tendrá cuatro mapas de calor, estos representarán el valor medio de la VPP, VPN, sensibilidad y especificidad, los resultados serán dependientes del porcentaje de entrenamiento y prueba y de la

profundidad del árbol, de manera que cada mapa de calor representará los resultados con dos decimales de precisión para las 81 configuraciones posibles.

Con el fin de no saturar la memoria de este TFG con figuras, insertaré las configuraciones que dan los mejores resultados, además de los mejores resultados, esto la haré para cada grupo y para los dos criterios empleados.

Tabla 17: Mejores configuraciones y resultados de los árboles de decisión

Grupo	Criterio	Profundidad del árbol	% test	% entrenamiento	VPP	VPN	Sens.	Esp.
0	Gini	7	5%	95%	0'75	0'99	0'73	0'99
1	Gini	6	5%	95%	0'72	0'98	0'71	0'98
2	Gini	4	15%	85%	0'78	0'95	0'27	0'99
3	Gini	10	5%	95%	0'91	0'75	0'95	0'61
0	Entropía	8	5%	95%	0'71	0'99	0'58	1
1	Entropía	7	10%	90%	0'79	0'98	0'69	0'99
2	Entropía	10	5%	95%	0'56	0'96	0'57	0'96
3	Entropía	10	10%	90%	0'96	0'66	0'95	0'72

Atendiendo a la Tabla 17 podemos observar claramente las mejores configuraciones del árbol para cada grupo y criterio, además de sus resultados para nuestros 4 parámetros.

He almacenado los resultados experimentales para cada iteración de los bucles en mapas de calor, veamos algunos en las siguientes figuras:

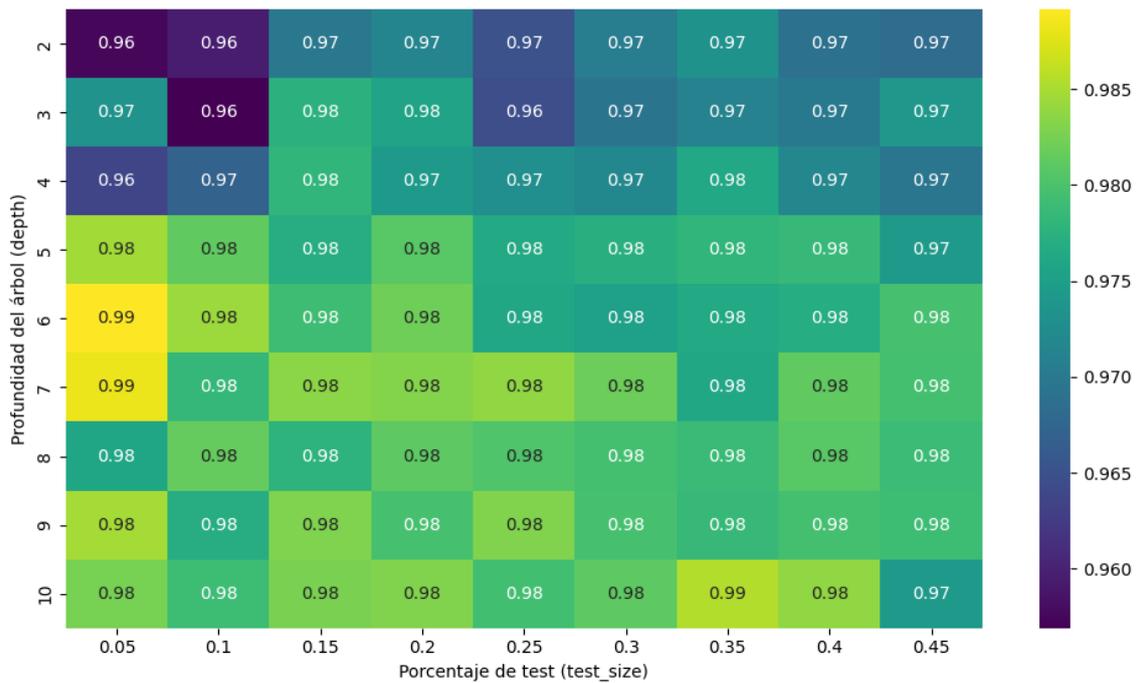


Figura 39: Mapa de calor de la media de VPN para pacientes del grupo 0.

En la Figura 39 podemos ver para pacientes del grupo 0 el valor predictivo negativo para cada porcentaje de entrenamiento y prueba y para cada profundidad del árbol con una medida de impureza **Gini**. En la Figura 40 observamos lo mismo, pero para el valor predictivo positivo en pacientes del grupo 3.



Figura 40: Mapa de calor de la media de VPP para pacientes del grupo 3.



Figura 41: Mapa de calor de la media de VPN para pacientes del grupo 0.

En la Figura 41 podemos ver para pacientes del grupo 0 el valor predictivo negativo para cada porcentaje de entrenamiento y prueba y para cada profundidad del árbol con una medida de impureza de **Entropía**. En la Figura 42 observamos lo mismo, pero para el valor predictivo positivo en pacientes del grupo 3.

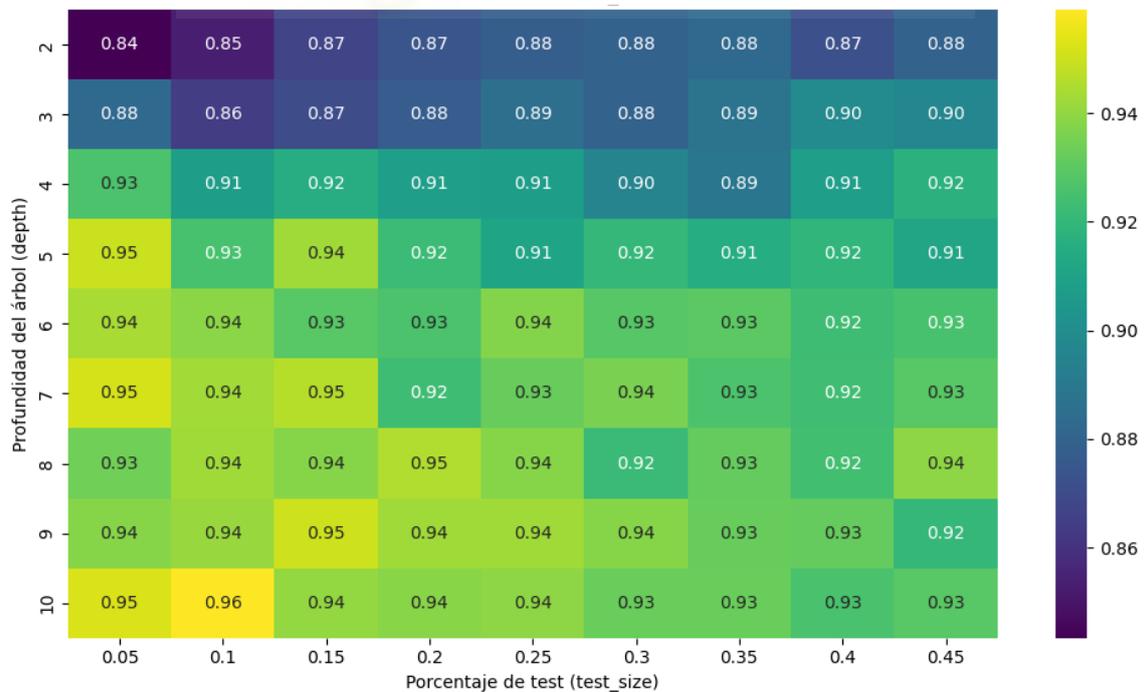


Figura 42: Mapa de calor de la media de VPP para pacientes del grupo 3.

5.4 ENFOQUE EXPERIMENTAL DEL ANÁLISIS DE COMPONENTES PRINCIPALES (PCA)

Lo primero que vamos a hacer es comprobar cuantas variables necesito para cubrir, al menos, el 90% de la varianza, para ello vamos a realizar una descomposición de valores singulares (SVD). La SVD se lleva a cabo mediante la función “`np.linalg.svd`”, esta función nos pedirá como entrada nuestra matriz de datos original, y nos devolverá 3 matrices. La primera matriz que devuelve la función es la que contiene los vectores singulares izquierdos de la matriz de entrada como columnas, se denomina matriz “U”, la segunda es un array unidimensional (vector) que contiene los valores singulares de la matriz de entrada, se denomina matriz “S”, la tercera es la matriz traspuesta de los vectores singulares derechos de la matriz de entrada, se denomina matriz “Vt”.

Los autovalores se sacarán elevando al cuadrado el vector “S”, ahora podemos sacar la proporción de varianza total, dividiendo los autovalores entre la suma global del vector, al hacer la división elemento a elemento del vector obtenemos un array donde cada componente representa el porcentaje de varianza total que tenemos.

Calculamos la varianza acumulada, sumando progresivamente los elementos de nuestro array de porcentaje de varianza, calculamos cuantas componentes (variables) necesitamos para cubrir, al menos, el 90% de la varianza de nuestra base de datos. Resulta que, con 47 variables ya cubrimos más del 90% de la información, saber esto evitará redundancia en los datos, hará que el problema sea más liviano y que el código compile antes, perdiendo menos del 10% de la información.

Calculamos las proyecciones de las variables originales en el espacio de los componentes principales y hacemos exactamente lo mismo para los pacientes.

Sabiendo que ahora empleamos las primeras 47 variables, y que para esa cantidad de datos hemos calculado las proyecciones sobre los pacientes, ahora vamos a aplicar todos los experimentos anteriores habiendo utilizado el método de PCA, y observaremos si los resultados son mejores o no cuando aplicamos este método de forma previa.

Resultados Experimentales de PCA & NN

Hemos empleado la configuración de red neuronal óptima de cada grupo de pacientes, el resultado ha sido el siguiente:

Tabla 18: Resultados experimentales al aplicar PCA sobre las redes neuronales

Grupo	VPP	VPN	Sensibilidad	Especificidad
0	1	1	1	1
1	0'75	0'9796	0'75	0'9796
2	1	0'96	0'6	1
3	1	1	1	1

En la Tabla 10, la Tabla 12, la Tabla 14 y la Tabla 16 se observan los resultados experimentales de las redes neuronales de cada grupo de pacientes, sin haber aplicado el método de PCA. Observando la Tabla 18 podemos ver como los resultados han mejorado notablemente al aplicar el análisis de componentes principales.

Veamos en la siguiente figura el resultado exacto de la predicción.

Biblioteca
UNIVERSITAS Miguel Hernández

```

Predicción de la red: [1 1 1 1 0 1 1 1 1 1 1 0 1 1 1 1 1 1 0 1 1 1 1 1 1 0 0 1 1 1 1 1 1 0
 1 1 1 1 1 1 1 1 1 1 1 0 1 1 1 1]
salida real: [1 1 1 1 0 1 1 1 1 1 1 0 1 1 1 1 1 1 1 0 1 1 1 1 1 1 0 0 1 1 1 1 1 1 0
 1 1 1 1 1 1 1 1 1 1 1 0 1 1 1 1]
VPP: 1.0
VPN: 1.0
Sensibilidad 1.0
Especificidad 1.0
    
```

Figura 43: Predicción de la red neuronal con previo PCA para pacientes del grupo 3.

Resultados Experimentales de PCA & SVM

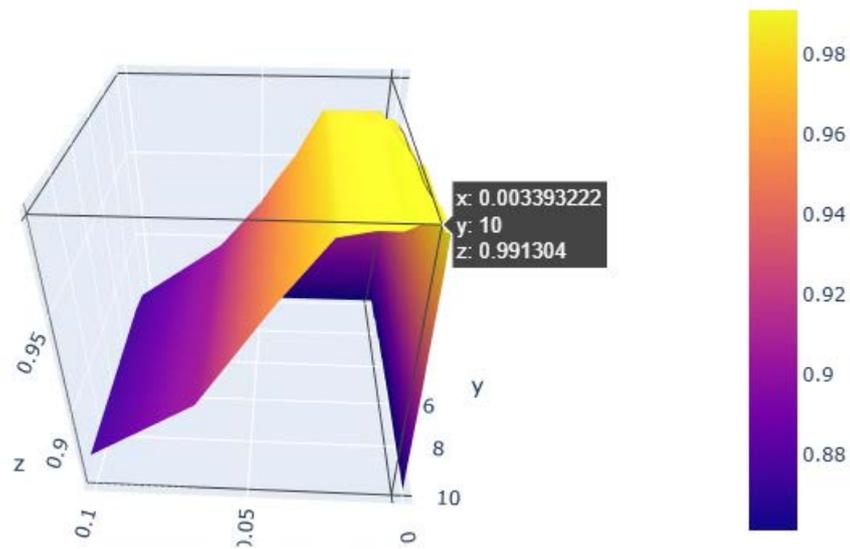


Figura 44: Resultado experimental al aplicar PCA sobre una máquina de vectores de soporte.

En la Figura 36, Figura 37 y Figura 38 se aprecian los resultados de SVM sin aplicar PCA, en esa ocasión la precisión máxima alcanzaba un valor del 88'99 %. Si ahora nos fijamos en la Figura 44 , podemos ver como la precisión máxima alcanza el 99'13 % (mirar el valor del eje Z), luego el resultado mejora ostensiblemente al aplicar PCA.

Resultados Experimentales de PCA & Árbol de Decisión con Criterio GINI

Hemos empleado la configuración de árbol de decisión óptima para cada grupo de pacientes, el resultado ha sido el siguiente:

Tabla 19: Resultados experimentales al aplicar PCA sobre árboles de decisión con criterio GINI

Grupo	VPP	VPN	Sensibilidad	Especificidad
0	0'5	1	1	0'9608

6 CONCLUSIONES Y PRÓXIMOS PASOS

Como conclusión, el estudio que se ha realizado consta de unos muy buenos resultados experimentales, esto se debe, en parte, a la realización de un análisis de IA que nos ha ayudado a elegir a un buen asistente en la programación. En el caso de las redes neuronales los resultados son relativamente buenos, una vez hemos encontrado la configuración óptima de la red para cada grupo. Estos resultados deben ser interpretados también para el grupo de pacientes con el que estamos trabajando, es decir, si trabajo con el grupo 0, que son los pacientes no afectados, me interesa tener un valor predictivo negativo alto, mientras que, si trabajo con el grupo 3, que son las personas afectadas por MCH (genética positiva y pruebas afectadas), me interesa tener un valor predictivo positivo elevado. Es por esto que, pese a no ser números elevados para todos los parámetros en este experimento, considero buenos los valores que han salido, pues los que interesan que sean altos lo son.

Para nuestro segundo experimento (SVM), la precisión toma valores superiores al 88 % para su configuración óptima, es un porcentaje alto de precisión y considero que se puede tener muy en cuenta este valor experimental.

En los experimentos 3 y 4, que son los árboles de decisión, con distintas medidas de impureza, considero que los parámetros poseen valores bastante buenos para cada grupo y para ambas medidas de impureza, además, al igual que con las redes neuronales, para pacientes no afectados tenemos un muy buen valor predictivo negativo, acompañado también de un buen valor predictivo positivo, en el caso de los pacientes afectados nuestro valor predictivo positivo es muy alto, y está acompañado de un alto valor predictivo negativo.

Para nuestro último experimento, que consiste en aplicar el método de análisis de componentes principales a nuestros experimentos previos, reducimos considerablemente el número de variables a emplear a la entrada de cada experimento, con ello evitamos la redundancia de información y el ruido, y conservamos más del 90 % de la información, debido a que tenemos aún un alto porcentaje de la varianza de los datos. Este experimento ha salido con grandes resultados, ha aumentado considerablemente todos los parámetros de las redes neuronales para cada grupo de pacientes, ha incrementado la precisión de la máquina de vectores de soporte en más de un 10 %, pasando de un 88'99 % de precisión

a un 99'13%, en el caso de los árboles de decisión no se ha producido un gran cambio en los resultados, al aplicar PCA se mejoran algunos parámetros de determinados grupos, pero empeoran otros, de manera que ha sido el experimento al que menos ha beneficiado aplicar el análisis de componentes principales.

Con lo comentado hasta hora, es fácil deducir que el método de PCA mejora a grandes rasgos los valores de nuestros experimentos, llegando a obtener valores realmente buenos en las redes neuronales y SVM, con valores de predicción de incluso el 100% de los casos. Hemos llegado a sacar muestras de más de 50 pacientes y en todos los casos ha acertado si pertenecía o no la persona a un determinado grupo, luego considero fundamental la aplicación de PCA a la hora de realizar investigaciones de este tipo si se quieren llevar a la práctica.

De cara al futuro, en caso de llevarse a la práctica, podríamos mejorar los experimentos realizando más configuraciones. En el caso de las redes neuronales aumentamos el porcentaje de prueba de 5% en 5%, podríamos ir incrementando de uno en uno en lugar de 5 en 5, el número de neuronas en cada capa podría incrementarse de uno en uno también y en lugar de llegar a 60 neuronas como máximo que fuesen muchas más, también podríamos aumentar el número de capas ocultas de la red. Las distintas combinaciones de configuraciones para cada experimento son infinitas, podemos encontrar una que, incluso, nos de valores mejores a los que tenemos ahora. Yo no he podido llevar a cabo esto que digo, debido a que si aumentamos el número de iteraciones de los bucles el proceso de compilación se vuelve más complejo y, sobre todo, más largo, a mi ordenador ya le ha costado más de una hora algunas compilaciones de redes neuronales con 1440 iteraciones, luego en caso de querer mejorar este aspecto sería conveniente poder acceder a un ordenador más potente.

Otro experimento que se podría plantear es juntar la salida de todos los experimentos posibles, supongamos que tenemos 7 experimentos (no es el caso de este TFG), si sacamos por pantalla a los 50 primeros pacientes de la base de datos, podemos observar que experimentos lo dan por perteneciente a un grupo (1) o lo rechazan de ese grupo (0), con que 4 experimentos nos den como salida un 1 o un 0 ya tenemos mayoría. El objetivo de este nuevo experimento sería evaluar las salidas de todos los experimentos para el mismo grupo de pacientes y sacar por pantalla el que sea mayoría, si fuesen 7 experimentos, como el caso propuesto, la mayoría se alcanzaría con 4, luego en cuanto se

obtengan 4 unos o más sacaríamos por pantalla un 1 y, al contrario, en cuanto se obtengan 4 ceros o más sacaríamos por pantalla un 0.

Para acabar con este apartado, podemos ver estos experimentos de mil formas diferentes y podemos emplear infinitas configuraciones para cada uno de ellos, el objetivo siempre será maximizar nuestros valores predictivos y, por supuesto, llevarlos a la práctica para finalmente poder comprobar que lo que se está haciendo funciona correctamente, y si sirve para poder realizar un diagnóstico temprano de la miocardiopatía hipertrófica, para así poder reducir la morbilidad y mortalidad de la población frente a esta afección cardíaca.



7 BIBLIOGRAFÍA

- [1] A. Kazemi-Diaz *et al.*, “An Embedding Approach for Biomarker Identification in Hypertrophic Cardiomyopathy,” *Computing in Cardiology*, vol. 50, 2023, doi: 10.22489/CinC.2023.123.
- [2] C. Janiesch, P. Zschech, and K. Heinrich, “Machine learning and deep learning,” *Electronic Markets*, vol. 31, no. 3, pp. 685–695, Sep. 2021, doi: 10.1007/s12525-021-00475-2.
- [3] F.-J. Gimeno-Blanes, “Detección Anticipada y Mecanismos del Síncope Vasovagal en la Prueba de Basculación Mediante Tratamiento Digital de Señal,” 2009.
- [4] M. Rojas Jiménez, V. Medina Carrillo, and C. Esteve Piñeiro, *Electrocardiografía clínica para Enfermería*. 2020.
- [5] C. Santos Molina Mazón, A. Carlos Santos Molina Mazón, X. Martín Cerezo, L. Álvarez Vivas, and S. Asensio Flores, “74/ Enfermería en Cardiología,” vol. 22, no. 66, p. 74, 2015.
- [6] I. Gómez Arraiz, E. Barrio Ollero, and A. Gómez Peligros, “Pruebas genéticas en la miocardiopatía hipertrófica: beneficios, limitaciones y aplicaciones en la práctica clínica,” *Medicina de Familia. SEMERGEN*, vol. 44, no. 7, pp. 485–491, Oct. 2018, doi: 10.1016/j.semerg.2018.03.002.
- [7] H. universitario 12 de Octubre, “Cirugía de la miocardiopatía hipertrófica obstructiva.” [Online]. Available: <https://www.comunidad.madrid/hospital/12octubre/profesionales/cirugia-miocardiopatia-hipertrofica-obstruccion>
- [8] A. G. Campos, M. P. Domínguez, B. D. Fernández, and J. R. González-Juanatey, “Protocolo diagnóstico de las miocardiopatías genéticas,” *Medicine - Programa de Formación Médica Continuada Acreditado*, vol. 13, no. 42, pp. 2481–2485, Oct. 2021, doi: 10.1016/j.med.2021.09.024.
- [9] M. Luis Alonso, “APLICACIÓN DE TÉCNICAS DE APRENDIZAJE ESTADÍSTICO PARA LA DETECCIÓN Y CLASIFICACIÓN DE LA MIOCARDIOPATÍA HIPERTRÓFICA DE ORIGEN HEREDITARIO,” p. 79, 2023.
- [10] P. Travé i Mercadé, “Interpretación clínica de las pruebas analíticas y su aplicación en Atención Farmacéutica,” p. 44, 2009.
- [11] J. Cebollada and J. A. Gimeno, “Ácido úrico como factor de riesgo cardiovascular,” *Hipertensión y Riesgo Vascular*, vol. 29, no. 2, pp. 36–43, Apr. 2012, doi: 10.1016/j.hipert.2012.04.001.
- [12] Erramouspe B, “Determinación de Ferritina Sérica,” *LABORATORIO HEMATOLOGIA*, vol. 16, pp. 122–123, 2012.
- [13] L. C. Hematólogo, “Ferritina,” *Med. lab*, pp. 493–496, 2017, Accessed: May 05, 2025. [Online]. Available: www.edimeco.com
- [14] M. C. Mariela, F. Barrios, D. H. Gautier, D. Gómez, D. Norma, and F. Delgado, “Metabolismo del hierro,” *Revista Cubana de Hematología, Inmunología y Hemoterapia*, vol. 16, no. 3, pp. 149–160, 2000, [Online]. Available: http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S0864-

- 02892000000300001&lng=es&nrm=iso&tlng=es
- [15] A. M. GUZMÁN D and T. QUIROGA G, “Troponina en el diagnóstico de infarto al miocardio: Consideraciones desde el laboratorio clínico,” *Revista médica de Chile*, vol. 138, no. 3, Mar. 2010, doi: 10.4067/S0034-98872010000300020.
 - [16] M. Rivera *et al.*, “Hipertensión y valores de NT-proBNP. Su importancia en el diagnóstico de insuficiencia cardíaca,” *Revista Española de Cardiología*, vol. 57, no. 5, pp. 396–402, Jan. 2004, doi: 10.1016/S0300-8932(04)77124-9.
 - [17] J. Osca *et al.*, “Péptido cerebral natriurético. Valor diagnóstico en la insuficiencia cardíaca,” *Revista Española de Cardiología*, vol. 55, no. 1, pp. 7–15, Jan. 2002, doi: 10.1016/S0300-8932(02)76547-0.
 - [18] “TRANSAMINASAS » Significado, Valores y Enfermedades Asociadas,” 2023. [Online]. Available: <https://www.todoellas.com/salud/transaminasas/>
 - [19] H. Alcaíno, J. Pozo, M. Pavez, and H. Toledo, “Ancho de distribución eritrocitaria como potencial biomarcador clínico en enfermedades cardiovasculares,” *Revista médica de Chile*, vol. 144, no. 5, pp. 634–642, May 2016, doi: 10.4067/S0034-98872016000500012.
 - [20] Ms. M. F.-B. Dra. Norma Fernández-Delgado, Dr.Cs. Prof. Porfirio Hernández-Ramírez, “Espectro funcional de las plaquetas: de la hemostasia a la medicina regenerativa,” *Instituto de Hematología e Inmunología. La Habana, Cuba.*, La Habana, Cuba, p. 17, 2012. [Online]. Available: http://scielo.sld.cu/scielo.php?pid=S0864-02892012000300002&script=sci_arttext&tlng=en
 - [21] “Nueva Enciclopedia Larousse,” *Larousse*. 1981.
 - [22] “ENCICLOPEDIA DE MEDICINA Y ENFERMERIA MOSBY,” *MOSBY'S MEDICAL & NURSING DICTIONARY*. 1988.
 - [23] L. W. del Electrocardiograma, “Electrodos del Electrocardiograma.” [Online]. Available: <https://www.my-ekg.com/generalidades-ekg/electrodos-ekg.php>
 - [24] Shutterstock, “No Title.” [Online]. Available: <https://www.shutterstock.com/es/image-illustration/einthoven-triangle-assumes-that-left-upper-2228232255>
 - [25] D. A. Uherek and D. A. Gómez, “ECG Pediátrico.” [Online]. Available: http://www.saludinfantil.org/Seminarios_Pediatria/Cardiologia/ECG_Pediatrico_AUherek.pdf
 - [26] H. P. M.D, “El Eje Cardíaco y como calcularlo.” [Online]. Available: <https://cerebromedico.com/electrocardiograma/eje-cardiaco>
 - [27] E. Bernal Pérez, M. Camós Ejarque, J. M. López-Reina Roldán, H. Moreda Aragón, M. Rojas Jiménez, and J. Palamidessi Domínguez, *Guía integral de ecografía aplicada a intervenciones enfermeras*, Fuden. 2023. [Online]. Available: <https://www.fuden.es/curso/guia-integral-de-ecografia-aplicada-a-intervenciones-enfermeras/>
 - [28] Tpempresas, “Introducción a las comunicaciones inalámbricas.” [Online]. Available: <https://tpempresas.com/introduccion-a-las-comunicaciones-inalambricas>
 - [29] “TEMA 7. ECOGRAFÍA, FUNDAMENTOS Y MANEJO BÁSICO.” [Online]. Available: https://www.salusplay.com/img-apuntes/ampliadas/5744_8332.jpg

- [30] M. Á. García Fernández, J. L. Zamorano Gómez, and J. A. García Robles, *Manual de ecocardiografía: indicaciones e interpretación en la práctica clínica*, EDIMED. Madrid, 2004.
- [31] E. M. Escudero, O. A. Pinilla, and I. L. Ennis, “Masa ventricular izquierda inapropiada en una población de adultos jóvenes,” *Revista Española de Cardiología*, vol. 65, no. 9, pp. 855–856, Sep. 2012, doi: 10.1016/j.recesp.2012.01.012.
- [32] S. H. Barberato, A. Moraes de Souza, C. Ortiz Costantini, and C. R. Frack Costantini, “Relación E/e’ en la Predicción de la Remodelación del Ventrículo Izquierdo Después de Infarto Agudo del Miocardio,” p. 7, 2013, [Online]. Available: <http://departamentos.cardiol.br/dic/publicacoes/revistadic/revista/2013/espanhol/Revista03/05-original-relato.pdf>
- [33] A. Fernández *et al.*, “Alternativas terapéuticas en la miocardiopatía hipertrófica,” *Revista Argentina de Cardiología*, vol. 71, pp. 294–301, 2003.
- [34] J. Phanius, “Mavacamten : nueva estrategia terapéutica sobre el manejo de la miocardiopatía hipertrófica obstructiva,” 2024, [Online]. Available: <https://repositorio.uai.edu.ar/handle/123456789/3011>
- [35] J. Pant, D. Singh, V. Sharma, H. K. Pant, and J. Bhatt, “A Machine-Learning Approach to Detect Heart Disease,” in *2024 8th International Conference on Electronics, Communication and Aerospace Technology (ICECA)*, IEEE, Nov. 2024, pp. 860–863. doi: 10.1109/ICECA63461.2024.10801012.
- [36] D. Purwanto, S. C. Hidayati, D. I. Ricoida, and K. A. Putri, “Optimized Machine Learning Models for Heart Disease Prediction: A Performance Analysis,” in *2024 International Seminar on Application for Technology of Information and Communication (iSemantic)*, IEEE, Sep. 2024, pp. 559–562. doi: 10.1109/iSemantic63362.2024.10762500.
- [37] T. De-Camino-Beck, “PROGRAMACIÓN ASISTIDA POR INTELIGENCIA ARTIFICIAL GENERATIVA,” 2025, [Online]. Available: <https://cpic-sistemas.or.cr/revista/index.php/technology-inside/article/view/114>
- [38] M. A. Morales-Chan, “Explorando el potencial de Chat GPT: Una clasificación de Prompts efectivos para la enseñanza,” p. 8, 2023.
- [39] “Gemini (modelo de lenguaje).” [Online]. Available: [https://es.wikipedia.org/wiki/Gemini_\(modelo_de_lenguaje\)](https://es.wikipedia.org/wiki/Gemini_(modelo_de_lenguaje))
- [40] J. Valero, “¿Qué es y cómo funciona Microsoft Copilot? El asistente de IA de Microsoft 365.” [Online]. Available: <https://blog.beservices.es/blog/que-es-como-funciona-microsoft-copilot>
- [41] U. M. Hernández, “Lesson 6: Neural Network an Introduction.”
- [42] “L-BFGS - Wikipedia, la enciclopedia libre.” Accessed: Jun. 02, 2025. [Online]. Available: <https://es.wikipedia.org/wiki/L-BFGS>
- [43] J. Cervantes Canales and D. D. Xiaou Li Zhang Wen Yu Liu México, “Tesis que presenta,” 2009.
- [44] U. M. Hernández, “Lesson 7: Support Vector Machines.”
- [45] M. Garcia, J. Munoz, R. Reveles, S. Saldivar, and C. Tapia, “Artículo Original,” *Rev Med Urug*, vol. 27, no. 4, pp. 228–235, 2011, Accessed: Jun. 07, 2025. [Online]. Available: www.uv.mx/rm

- [46] “Estructura de un árbol de decisión.” [Online]. Available: <https://www.researchgate.net/publication/285054363/figure/fig2/AS:391476732284942@1470346734856/Figura-2-Estructura-de-un-arbol-de-decision.png>
- [47] M. Terrádez Gurra, “Análisis De Componentes Principales Esquema De Contenidos,” *Proyecto e-Math Secretaría de Estado de Educación y Universidades (MECD)*, p. 11, 2005, Accessed: Jun. 07, 2025. [Online]. Available: https://www.uoc.edu/in3/emath/docs/Componentes_principales.pdf
- [49] A. Gil Izquierdo, ““Detección automática en tiempo real de la activación del nervio frénico durante el procedimiento de crioablación para el tratamiento de la fibrilación auricular mediante técnicas de procesamiento de la señal y aprendizaje automático,”” p. 96, 2024.

