

DISEÑO E IMPLEMENTACIÓN DE UN CLASIFICADOR SVM PARA LA
DETECCIÓN DE ARRITMIAS A PARTIR DEL ANÁLISIS DE SEÑALES
ELECTROCARDIOGRÁFICAS (ECG)

IVONNE NATALY MARTÍNEZ TORO

UNIVERSIDAD NACIONAL ABIERTA Y A DISTANCIA - UNAD
ESCUELA DE CIENCIAS BÁSICAS, TECNOLOGÍA E INGENIERÍA
INGENIERÍA DE TELECOMUNICACIONES
BOGOTÁ D.C.
2023

DISEÑO E IMPLEMENTACIÓN DE UN CLASIFICADOR SVM PARA LA
DETECCIÓN DE ARRITMIAS A PARTIR DEL ANÁLISIS DE SEÑALES
ELECTROCARDIOGRÁFICAS (ECG)

IVONNE NATALY MARTINEZ TORO

Trabajo de grado para optar por el título de Ingeniera de Telecomunicaciones

PAOLA ANDREA MATEUS ABAUNZA
DIRECTORA DE PROYECTO DE GRADO

UNIVERSIDAD NACIONAL ABIERTA Y A DISTANCIA - UNAD
ESCUELA DE CIENCIAS BÁSICAS, TECNOLOGÍA E INGENIERÍA
INGENIERÍA DE TELECOMUNICACIONES
BOGOTÁ D.C.
2023

NOTA DE ACEPTACIÓN

Presidente del Jurado

Jurado

Jurado

Bogotá D.C., 7 de junio de 2023

Dedicatoria

A mi abuelo, quien hace años no está conmigo, pero me llenó de valentía su historia para escoger la temática y el propósito del proyecto. Era una deuda conmigo misma y lo logré.

AGRADECIMIENTOS

Nombro inicialmente a la ingeniera Paola Mateus, quien me incentivó desde un inicio a incursionarme en esta disciplina, utilizando nuevas herramientas y algoritmos para el bien común.

A mis padres, mis más grandes agradecimientos hoy y siempre. El apoyo nunca se me fue negado y he logrado ser lo que he querido ser por ellos.

CONTENIDO

	Pág.
1. INTRODUCCIÓN.....	14
2. OBJETIVOS	15
2.1 OBJETIVO GENERAL	15
2.2 OBJETIVOS ESPECÍFICOS	15
3. PLANTEAMIENTO DEL PROBLEMA.....	16
3.1 DEFINICIÓN DEL PROBLEMA	16
3.2 JUSTIFICACIÓN	17
4. MARCO CONCEPTUAL Y TEÓRICO	19
4.1 ANTECEDENTES.....	19
4.2 MARCO CONCEPTUAL	21
5. MATERIALES Y MÉTODOS	34
5.1 MATERIALES	34
5.2 METODOLOGÍA	34
6. DESARROLLO DEL PROYECTO	46
6.1 CRONOGRAMA.....	46
6.2 RESULTADOS.....	47
6.3 ANÁLISIS DE RESULTADOS	52
7. CONCLUSIONES	57
8. RECOMENDACIONES.....	58

LISTA DE TABLAS

	Pág
Tabla 1. Fórmulas de kernels.....	30
Tabla 2. Recursos y materiales utilizados para el trabajo de grado.	34
Tabla 4. Cronograma de actividades del proyecto de grado.	46
Tabla 5. Dispersión de datos en cada validación cruzada.....	47
Tabla 6. Métricas de evaluación de las tres SVM, en cada validación cruzada.....	49
Tabla 7. Matrices de confusión para cada uno de los kernels evaluados en cada validación cruzada.....	52
Tabla 8. Métricas de proximidad de medición con cada kernel, en cada respectiva validación cruzada (k).....	53

LISTA DE FIGURAS

	Pág
Figura 1. Anatomía del corazón.....	21
Figura 2. Lectura de las ondas en el electrocardiograma y la visualización de la propagación de los impulsos dentro del corazón.	23
Figura 3. Lectura básica de intervalos rítmicos normales cardíacos en un electrocardiograma.....	24
Figura 4. Caso linealmente separable entre dos clases.	28
Figura 5. Apreciación del uso del kernel y la forma de visualización y nuevo tratamiento de la data dentro de un plano.	29
Figura 6. Tipos de kernels y discernimiento de datos según la necesidad de separación de cada punto.	31
Figura 7. Matriz de confusión y uso de clases.....	32
Figura 8. Clasificación y medida numérica de modelos con la curva ROC.....	33
Figura 9. Visualización de dos señales (sana y con arritmia), sin procesar.....	36
Figura 10. Escalograma del ventrículo 3, sin procesar, de una señal sana y otra con arritmia.	38
Figura 11. Escalograma del ventrículo 3 de dos señales procesadas, una es una señal sana y la otra presenta arritmia.....	40
Figura 12. Paso a paso del procesamiento con el algoritmo de Pan-Tompkins, en una señal ECG.....	41
Figura 13. Algoritmo de Pan-Tompkins en señales con y sin arritmia.....	43
Figura 14. Detección de ondas en señales ECG.	44

LISTA DE GRÁFICOS

	Pág
Gráfico 1. Dispersión de datos de entrenamiento en la validación cruzada 1.	48
Gráfico 2. Dispersión de datos de prueba en la validación cruzada 1.....	48
Gráfico 3. Dispersión de datos de entrenamiento en la validación cruzada 2.	48
Gráfico 4. Dispersión de datos de prueba en la validación cruzada 2.....	48
Gráfico 5. Dispersión de datos de entrenamiento en la validación cruzada 3.	48
Gráfico 6. Dispersión de datos de prueba en la validación cruzada 3.....	48
Gráfico 7. Dispersión de datos de entrenamiento en la validación cruzada 4.	49
Gráfico 8. Dispersión de datos de prueba en la validación cruzada 4.....	49
Gráfico 9. Dispersión de datos de entrenamiento en la validación cruzada 5.	49
Gráfico 10. Dispersión de datos de prueba en la validación cruzada 5.....	49
Gráfico 11. Gráfico de caja de la distribución de los resultados de Exactitud en cada Kernel, de cada validación cruzada.	54
Gráfico 12. Gráfico de caja de la distribución de los resultados de Precisión en cada Kernel, de cada validación cruzada.	55

GLOSARIO

ALGORITMO DE PAN TOMPKINS: Es un método de procesamiento de señales utilizado para la detección de los complejos QRS en las señales electrocardiográficas. Pan y Tompkins¹ indican que este algoritmo es más eficiente y preciso, gracias al ajuste automático de sus parámetros para tal fin.

APRENDIZAJE NO SUPERVISADO: Un tipo de aprendizaje de máquina en el que se proporcionan solo entradas y se espera que una SVM descubra patrones o estructuras en los datos.

APRENDIZAJE SUPERVISADO: Un tipo de aprendizaje de máquina en el que se proporcionan ejemplos etiquetados de entrada y salida a una Máquina de Soporte Vectorial.

CLASIFICADOR SVM: Algoritmo de aprendizaje automático que utiliza un enfoque de aprendizaje supervisado para realizar la clasificación de un conjunto de datos en diferentes categorías.

DESVIACIÓN ESTÁNDAR: Es una medida estadística utilizada para medir la variabilidad de un conjunto de datos. En el aprendizaje automático, la desviación estándar se utiliza para medir la variabilidad en la precisión del modelo en cada repetición de la validación cruzada.

ESCALA PARA EL KERNEL: Es un parámetro utilizado en los algoritmos de aprendizaje automático basados en SVM (Support Vector Machine) y otros modelos de aprendizaje automático que utilizan kernel. La escala del kernel determina la distancia de influencia de los puntos de datos en el modelo. Un valor más alto de la escala del kernel significa que los puntos de datos tienen una influencia más amplia en el modelo.

ESPECIFICIDAD: Capacidad de la SVM para identificar correctamente a los pacientes que no tienen IAM.

¹ PAN, Jiapu y TOMPKINS, Willis J. A Real-Time QRS Detection Algorithm. En: IEEE Transactions on Biomedical Engineering [en línea]. Marzo, 1985. BME-32, no. 3 [consultado el 11, marzo, 2023], p. 230-236. Disponible en Internet: <<https://doi.org/10.1109/tbme.1985.325532>>. ISSN 0018-9294.

EXACTITUD: Medida de la precisión de un clasificador que se refiere a la cantidad de predicciones correctas divididas por el total de datos evaluados. En cuanto a la necesidad de encontrar esta métrica, Nicholson² indica que la precisión puede decirnos inmediatamente si un modelo se está entrenando correctamente y cómo puede funcionar en general.

F1-SCORE: Es una medida de la precisión de un modelo de clasificación que tiene en cuenta tanto la precisión como la sensibilidad del modelo.

INFARTO AGUDO DE MIOCARDIO (IAM): Daño o muerte celular del tejido del corazón debido a una obstrucción del flujo sanguíneo en las arterias coronarias.

MATRIZ DE CONFUSIÓN: Tabla que se utiliza para evaluar el rendimiento de un clasificador, mostrando la cantidad de verdaderos positivos, verdaderos negativos, falsos positivos y falsos negativos.

NÚMERO DE FOLIOS K: Es una técnica utilizada en el aprendizaje automático para la validación del modelo. Se utiliza para dividir el conjunto de datos en K partes iguales, de las cuales K-1 partes se utilizan para entrenar el modelo y la parte restante se utiliza para probar el modelo. Este proceso se repite K veces, de modo que cada parte se utiliza una vez como conjunto de prueba.

ÓRDENES DEL KERNEL POLINOMIAL: En SVM, el kernel polinomial se utiliza para transformar los datos originales a una dimensión superior, para poder separar los datos en diferentes clases. El orden del kernel polinomial se refiere al grado de la función polinómica utilizada para realizar la transformación de los datos.

PRECISIÓN: Medida de la precisión de un clasificador que se refiere a la cantidad de predicciones correctas divididas por el total de predicciones realizadas.

RF: Radial Basis Function (Función de Base Radial) es una función matemática utilizada en SVM y otras técnicas de aprendizaje automático. Esta función se utiliza

² NICHOLSON, Chris V. Evaluation metrics for machine learning - accuracy, precision, recall, and F1 defined. Pathmind [página web]. [Consultado el 11, marzo, 2023]. Disponible en Internet: <<https://wiki.pathmind.com/accuracy-precision-recall-f1>>.

para transformar los datos originales a una dimensión superior, para poder separar los datos en diferentes clases.

SENSIBILIDAD: Capacidad de la SVM para detectar correctamente a los pacientes con infarto de miocardio, ya que, entre más alto el valor, maximiza la cantidad de verdaderos positivos. Por otro lado, Keldenich³ indica que no da información sobre la calidad de predicción de negativos, por lo que, debe validarse otro tipo de métricas para saber qué tan exacto en general puede ser el modelo entrenado.

TASA DE FALSOS POSITIVOS: Porcentaje de casos en los que se clasifica incorrectamente una señal como positiva (en este caso, indicando la presencia de un infarto de miocardio) cuando en realidad es negativa (sin arritmia).

TASA DE VERDADEROS POSITIVOS: Porcentaje de casos en los que se clasifica correctamente una señal como positiva (indicando la presencia de un infarto de miocardio) cuando en realidad es positiva.

VALIDACIÓN CRUZADA: Es una técnica estadística utilizada en el aprendizaje automático para evaluar el rendimiento del modelo. Data Science Team⁴ indica que se utiliza para dividir el conjunto de datos en varias partes, de las cuales una se toma para probar el modelo y las demás se utilizan para entrenar el modelo. Este proceso se repite varias veces para que cada parte se utilice una vez como conjunto de prueba.

VFP Y VPN: Valores predictivos positivo y negativo, respectivamente, que se utilizan para evaluar el rendimiento de un clasificador en términos de la precisión de las predicciones en cada categoría.

³ KELDENICH, Tom. Recall, precision, F1 score - simple metric explanation machine learning. Inside Machine Learning [página web]. (2, septiembre, 2001). [Consultado el 11, marzo, 2023]. Disponible en Internet: <<https://inside-machinelearning.com/en/recall-precision-f1-score-simple-metric-explanation-machine-learning/>>.

⁴ DATA SCIENCE TEAM. Validación cruzada K-Fold. Data science [página web]. (2020). [Consultado el 26, marzo, 2023]. Disponible en Internet: <<https://datascience.eu/es/aprendizaje-automatico/validacion-cruzada-de-k-fold/>>.

RESUMEN

El procesamiento digital de señales electrocardiográficas (ECG) es una técnica útil para detectar enfermedades cardíacas y prevenir problemas más graves como infartos de miocardio. En Colombia, las enfermedades cardiovasculares son una preocupación importante según cifras publicadas por el Ministerio de Salud. Diferentes estudios han propuesto diversas técnicas de procesamiento digital para la detección de anomalías cardíacas mediante electrocardiogramas, pero se necesitan algoritmos de aprendizaje innovadores adaptados a la realidad actual para mejorar la detección temprana de arritmias derivadas en infartos de miocardio en Colombia. Este proyecto busca diseñar un clasificador de Máquina de Soporte Vectorial (SVM) para detectar arritmias en señales ECG y analizar los resultados obtenidos, con el objetivo de detectar infartos de miocardio.

Todo este procesamiento se hace con el objetivo de analizar una base de datos que contenga señales ECG con arritmia y sin arritmia, a través del repositorio PhysioNet; acondicionar estas señales electrocardiográficas de prueba para estandarizar las diferentes señales; procesar las diferentes señales para la extracción de características de una arritmia producida por un infarto miocárdico (MI); entrenar a la máquina virtual con diferentes parámetros evaluados tanto en amplitud como en tiempo, de forma que la clasificación logre discernir entre un paciente con infarto de miocardio y una señal que demuestre que no hay arritmia; y realizar pruebas y análisis de los resultados obtenidos por el clasificador SVM.

El proyecto tiene una gran relevancia en el campo de la salud cardiovascular, ya que la detección temprana de arritmias es crucial para prevenir infartos de miocardio. A través de la implementación de un clasificador SVM, se espera mejorar la precisión en la detección de arritmias en señales ECG, lo que permitiría una detección temprana y, por lo tanto, la prevención de infartos de miocardio.

PALABRAS CLAVE: Arritmias cardíacas, Electrocardiograma, Máquina de vectores de soporte, Tratamiento digital de imágenes, Clasificación de señales, Extracción de características

1. INTRODUCCIÓN

El procesamiento digital de señales electrocardiográficas (ECG) permite la detección de numerosas enfermedades cardíacas que podrían derivar a problemas más serios como los infartos de miocardio. En Colombia, según cifras publicadas por el Ministerio de Salud y Protección Social, Dirección de Epidemiología y Demografía⁵, la tasa de mortalidad 2020-2021 refiere a las enfermedades cardiovasculares como una situación con alto índice de afectados. Al mismo tiempo, se ha notado que diferentes algoritmos de aprendizaje intentan identificar diferentes tipos de arritmias cardíacas, pero son algoritmos muy generales que pueden ser mejorados para aplicarlos a la detección temprana de esta problemática en específico, al momento del diagnóstico.

La metodología utilizada en este proyecto involucra la programación en Matlab y la implementación de diferentes procedimientos y algoritmos para el procesamiento de señales ECG. Se realiza un preprocesamiento de las señales y se estandariza su frecuencia de muestreo a 1000 Hz utilizando la transformada Wavelet para la visualización de las frecuencias contenidas en las señales. Luego, la señal es filtrada con un filtro pasa-banda Butterworth para eliminar información poco útil y corregir la línea basal. Luego se implementa el algoritmo de Pan-Tompkins para la detección del complejo QRS en las señales ECG. Finalmente, se toman las ondas S, J y T, las más estudiadas para el diagnóstico de infartos de miocardio, para seleccionar las características de la SVM, basadas en los cambios en amplitud entre segmentos ST.

Se procesa en total 318 señales, dentro de las cuales, 253 tienen antecedentes con infartos de miocardio y 80 cuentan con control saludable. Asimismo, se utiliza el 80% de esta base de datos para entrenamiento y el 20% para la prueba. Se limpian las señales con filtros pasa-banda y se detectaron las ondas S y T para entrenar al clasificador SVM con diferentes parámetros de amplitud, lo que permite el discernir entre señales con y sin arritmia. Esta detección rápida es importante para el tratamiento oportuno y, aunque se requiere una validación por parte del personal especializado en el campo de la salud, este proyecto tiene una gran relevancia en el área de la salud cardiovascular al proporcionar una herramienta que permite una detección más eficaz por la validación con Máquinas de Vectores de Soporte.

⁵ MINISTERIO DE SALUD Y PROTECCIÓN SOCIAL DIRECCIÓN DE EPIDEMIOLOGÍA Y DEMOGRAFÍA. Mortalidad en Colombia. Periodo 2020 a 2021: Medición de la Mortalidad por todas las causas y Covid-19 [en línea]. Bogotá: [s.n.], 2022. 45 p. Informe. Disponible en Internet: <<https://www.minsalud.gov.co/sites/rid/Lists/BibliotecaDigital/RIDE/VS/ED/GCFI/mortalidad-colombia-periodo-2020-2021.pdf>>.

2. OBJETIVOS

2.1 OBJETIVO GENERAL

Diseño e implementación de un clasificador SVM para la detección de arritmias a partir del análisis de señales electrocardiográficas (ECG).

2.2 OBJETIVOS ESPECÍFICOS

Obtener una base de datos que contenga señales ECG con arritmia y sin arritmia, a través del repositorio PhysioNet.

Acondicionar las señales electrocardiográficas de prueba para estandarizar las diferentes señales.

Procesar las señales para la extracción de características de una arritmia producida por un infarto miocárdico (MI).

Entrenar a la máquina virtual con diferentes parámetros evaluados, de forma que la clasificación logre discernir entre un infarto de miocardio y una señal sin arritmia.

Realizar pruebas y análisis de los resultados obtenidos por el clasificador SVM, con las dos características indicadas en el algoritmo.

3. PLANTEAMIENTO DEL PROBLEMA

3.1 DEFINICIÓN DEL PROBLEMA

De acuerdo con la Organización Mundial de la Salud [OMS]⁶, las enfermedades cardiovasculares son la principal causa de defunción en el mundo, cobrando un promedio de 17.9 millones de vidas cada año, en donde un tercio de esas defunciones las padecen personas menores de 70 años. Al tiempo, las arritmias, las cuales son alteraciones del ritmo cardíaco, identifican cuando los impulsos eléctricos no se generan adecuadamente o se originan en un sitio erróneo, adicionando la causa de los caminos de conducción eléctrica que se pueden ver alterados. Si esto no se trata a tiempo con procedimientos de control de ritmo irregular, tal como se indica en la entrada Paro Cardíaco Súbito⁷, el corazón se puede deteriorar y/o se puede causar un daño en otro órgano importante, derivando de igual manera, en problemáticas mayores, dentro de las cuales está el paro cardíaco súbito [PCS].

Asimismo, una de las causas comunes ligadas a la muerte súbita cardíaca es la cardiopatía isquémica conocida como infarto de miocardio [IM], donde el corazón tiende a volverse grueso y se hace difícil bombear sangre. En Colombia, la tasa de mortalidad 2020-2021 tanto en hombres como mujeres refiere a las enfermedades cardiovasculares como situación con alto índice de afectados según la medición y cifras publicadas por el Ministerio de Salud⁸ dentro del informe titulado “Medición de la Mortalidad por todas las causas y Covid-19”. Entre las cifras 2020, hubo un total de 41.012 afectados por infarto agudo de miocardio, afectando más a hombres que a mujeres, presentándose el mismo caso en el 2021, en donde bajó la cantidad de afectados a 28.400, pero, del mismo modo, afectando más a hombres que a mujeres.

⁶ ENFERMEDADES CARDIOVASCULARES [Anónimo]. World Health Organization (WHO) [página web]. [Consultado el 10, mayo, 2023]. Disponible en Internet: <https://www.who.int/es/health-topics/cardiovascular-diseases#tab=tab_1>.

⁷ PARO CARDÍACO súbito [Anónimo]. The Texas Heart Institute [página web]. [Consultado el 19, mayo, 2022]. Disponible en Internet: <<https://www.texasheart.org/heart-health/heart-information-center/topics/paro-cardiaco-subito/>>.

⁸ MINISTERIO DE SALUD Y PROTECCIÓN SOCIAL DIRECCIÓN DE EPIDEMIOLOGÍA Y DEMOGRAFÍA. Mortalidad en Colombia. Periodo 2020 a 2021: Medición de la Mortalidad por todas las causas y Covid-19 [en línea]. Bogotá: [s.n.], 2022. 45 p. Informe. Disponible en Internet: <<https://www.minsalud.gov.co/sites/rid/Lists/BibliotecaDigital/RIDE/VS/ED/GCFI/mortalidad-colombia-periodo-2020-2021.pdf>>.

Respecto a los paros cardiacos súbitos y su mecanismo de MI, la mayoría de estos casos son causados por taquicardias ventriculares [TV], que se presentan cuando el ritmo cardíaco es acelerado, una segunda causa relacionada al ritmo cardíaco anormal, conocido como la fibrilación ventricular [FV] o un tercer caso por un ritmo cardíaco muy lento, el cual se conoce como bradicardia. Estas causas necesitan un diagnóstico médico urgente por parte de los especialistas dedicados a la lectura de estudios y/o gráficos que registren los movimientos del corazón, siendo el electrocardiograma [ECG] una de las herramientas más utilizadas y útiles para tales fines médicos. Asimismo, las arritmias no se presentan de la misma forma dentro de las imágenes diagnósticas y la interpretación, por diferentes causas, puede no siempre ser la correcta o puede presentar resultados tardíos.

Los tiempos de respuesta para las diferentes detecciones de arritmias se dificultan al no considerar herramientas prácticas de apoyo para el análisis de señales de electrocardiogramas y, por consiguiente, se deriva a esta problemática de infarto cardíaco súbito. Además, se debe considerar que los ECG son estudios que se realizan en un lapso corto y ambulatorio, por lo que, la monitorización debe relacionar el síntoma más exacto posible y mantener los diferentes resultados en los registros pertinentes.

Por todo lo anterior se plantea la siguiente pregunta de investigación:

¿Es posible la implementación de algoritmos supervisados para la detección de arritmias que deriven posteriormente a un paro cardíaco súbito, a partir del análisis de señales ECG con presencia de la cardiopatía isquémica, conocida como Infarto de Miocardio?

3.2 JUSTIFICACIÓN

Gracias a los diferentes avances en el campo del procesamiento digital de señales ECG, se ha podido implementar una cantidad diversa de herramientas en el campo médico que facilitan la detección temprana de enfermedades o anomalía médicas. Dentro de este campo, se encuentran los estudios relacionados con las actividades centrales del corazón. La OMS, de igual manera aconseja el uso de tecnologías básicas de salud que permitan la detección de estos tipos de riesgos por enfermedad cardiovascular, de forma que se eviten percances mayores, además de bajar índices de defunciones prematuras si se identifican a tiempo y de forma correcta.

El electrocardiograma es una fuente analógica de información patológica cardiaca que no es invasiva y es muy utilizada en el campo de la medicina hasta la actualidad. El procesamiento digital de señales ECG y su respectivo tratamiento pueden lograr acelerar los procesos de detección temprana de problemas cardíacos como los paros cardíacos súbitos y otros problemas de gran gravedad médica, demostrando, además, que los avances tecnológicos pueden mejorar la calidad de vida de las personas con diagnósticos confiables, precisos y rápidos.

Aunque, como se reitera, hay numerosos equipos capaces de identificar señales electrocardiográficas relacionadas con problemas tempranos de arritmias para su posterior clasificación según las señales ECG, también es cierto que estos esquemas no son del todo implementados en los equipos de adquisición y procesamiento de información médica y cardiaca, debido a la exactitud, complejidad y capacidad de las máquinas para discriminar la información, la cual debe ser lo más limpia y eficaz posible, evitando que las problemáticas en salud avancen y se generen valor predictivos con resultados incorrectos al diagnosticar patologías cardiacas.

Una de las herramientas más usadas en la actualidad para la clasificación de señales electrocardiográficas, son las Máquinas de vectores de Soporte [SVM], ya que esta selección puede permitir que las características de análisis sean limitadas y los métodos pueden presentar una evaluación experimental que derive a resultados más precisos. De igual manera, la clasificación de las señales durante el entrenamiento de la máquina puede variar y no verse afectado negativamente si se maneja un conjunto de datos clasificados correctamente.

4. MARCO CONCEPTUAL Y TEÓRICO

4.1 ANTECEDENTES

Validando el análisis inicial de señales ECG, Lorenzo Romero⁹ hace un análisis de las señales electrocardiográficas por medio de técnicas de procesamiento digital. Su profundización indica el uso de filtros pasa bajo (LPF) *Butterworth* para proceder a un buen muestreo y cuantificación de señales porque se acerca al concepto de filtro ideal, además de realizar una aproximación a una señal más limpia y fiel con transformadas Wavelet. Del mismo modo, las líneas base de los electrocardiogramas pueden ser validadas por medio de estas transformadas. El trabajo no se toma como un referente completo para el presente estudio, ya que no se tienen en cuenta las diferentes cardiopatías y se hace un trabajo general de análisis que sí puede ser de apoyo para utilizar los elementos de pre-procesamiento de señales ECG adecuado.

Qian Zheng et al.¹⁰, comentan en su artículo ‘A novel multi-resolution SVM (MR-SVM) algorithm to detect ECG signal anomaly in WE-CARE project’ que el uso de las Máquinas de Vectores de Soporte por Multi-Resolución puede aportar un sistema eficiente de evaluación de anomalías dentro de los electrocardiogramas y se apoya en el estudio de las ondas T y los segmentos ST. El soporte de este documento aporta una base de estudio de estas ondas y segmentos en general, ya que aportan información específica cuando se trata de enfermedades cardíacas, además, se hace una limpieza general de las señales a partir de las funciones Wavelet, del tipo *Daubechies* (trabajo ortogonal). Sin embargo, el estudio solamente se centra en encontrar un promedio de pulsaciones por minuto y su fase, la detección de la onda T y la anomalía general en el segmento T, sin especificación de una anomalía en específico.

⁹ LORENZO ROMERO, José. Análisis de señales electrocardiográficas usando técnicas de procesamiento digital [en línea]. Trabajo de grado. [s.l.]: Universitat Oberta de Catalunya, 2015 [consultado el 3, marzo, 2023]. 84 p. Disponible en Repositori Institucional (O2): <<http://hdl.handle.net/10609/40186>>.

¹⁰ QIAN ZHENG, et al. A novel multi-resolution SVM (MR-SVM) algorithm to detect ECG signal anomaly in WE-CARE project. Disponible en: 2013 ISSNIP Biosignals and Biorobotics Conference: Biosignals and Robotics for Better and Safer Living (BRC) [en línea] (18-20, febrero, 2013: Rio de Janeiro). [s.l.]: IEEE, 2013 [consultado el 3, marzo, 2023]. Disponible en Internet: <<https://doi.org/10.1109/brc.2013.6487453>>. ISBN 9781467330251.

Por otro lado, Gajare y Dey¹¹ realizan una detección de picos R en los electrocardiogramas para la clasificación de señales por medio de algoritmos de Deep Learning, de forma que se pudiese hacer una categorización de desórdenes como contracciones ventriculares prematuras. Esta detección de picos R determina la ocurrencia cardíaca y ayuda a identificar los desórdenes generales de arritmias. Por cuestiones de pre-procesamiento de las señales ECG, se usó la transformada Wavelet y se hizo eliminación de ruido a 50 Hz, por cuestiones eléctricas. Además, se hizo una eliminación de ruido en altas frecuencias al introducir ruido blanco gaussiano, aportando una señal-ruido más acorde del 12.69%, logrando un 98.32% de exactitud con su programa en Matlab. Asimismo, Hernández Casillas¹² indica en su trabajo 'Estudio de la detección y clasificación de la transmuralidad del infarto de miocardio mediante análisis de imágenes de resonancia magnética cardíaca y algoritmos de aprendizaje máquina' que la extracción de características de las arritmias se liga al hecho de cuantificar los segmentos miocárdicos. Asimismo, el trabajo, al ser un análisis de imágenes electromagnéticas, no profundiza en el hecho de analizar los electrocardiogramas, pero sí en que puede ser necesaria la información de estos para diagnosticar de forma correcta la arritmia, además de notar a detalle el volumen del órgano cardíaco del paciente.

Finalmente, se toma como referente el trabajo de Kohli & Verma¹³, donde la máquina de soporte trabaja con clasificación multi-clase y se compone de 6 clases: normal, cambios isquémicos y antiguo infarto de miocardio inferior, bradicardia sinusal, bloqueo de rama derecha y otras (demás clases de arritmias); también, analiza el porcentaje de precisión de detección de cada una, encontrando una mejor clasificación entre los tipos de arritmia normal, con cambios isquémicos y de bloqueo de rama derecha, aunque su porcentaje general de precisión es más bajo que en los anteriores estudios. La información valiosa del documento se presenta en la onda que evalúa (onda P), ya que puede no tomarse muy en cuenta para el estudio final, limitando la cantidad de información que validará la SVM. Lo anterior, teniendo en cuenta que la clase 3, de infarto de miocardio, tuvo menos evaluación

¹¹ GAJARE, Amogh y DEY, Hrishikesh. MATLAB-based ECG R-peak Detection and Signal Classification using Deep Learning Approach. *En*: 2021 IEEE Bombay Section Signature Conference (IBSSC) [en línea] (18-20, noviembre, 2021: Gwalior, India). [s.l.]: IEEE, 2021 [consultado el 3, marzo, 2023]. Disponible en Internet: <<https://doi.org/10.1109/ibssc53889.2021.9673301>>.

¹² HERNÁNDEZ CASILLAS, Andrea. Estudio de la detección y clasificación de la transmuralidad del infarto de miocardio mediante análisis de imágenes de resonancia magnética cardíaca y algoritmos de aprendizaje máquina [en línea]. Tesis de Máster. Valencia: Universitat Politècnica de València, 2020 [consultado el 10, mayo, 2022]. 118 p. Disponible en Repositorio institucional UPV: <<http://hdl.handle.net/10251/174516>>.

¹³ KOHLI, N. y VERMA, NK. Arrhythmia classification using SVM with selected features. *En*: International Journal of Engineering, Science and Technology [en línea]. 17, agosto, 2012. vol. 3, no. 8 [consultado el 3, marzo, 2023]. Disponible en Internet: <<https://doi.org/10.4314/ijest.v3i8.10>>. ISSN 2141-2839.

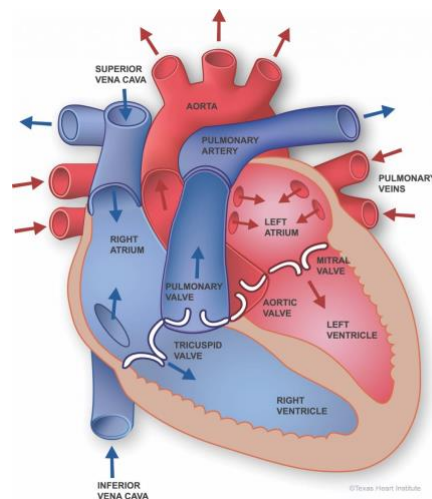
en el estudio y los diferentes métodos de clasificación multi-clase en SVM mostraron resultados inferiores que, en parte, no son incluidos en la conclusión de precisión.

4.2 MARCO CONCEPTUAL

4.2.1 Anatomía del corazón. Primeramente, hay dos movimientos identificados que hará el corazón: sístole y diástole. Si se bombea la sangre, el primer movimiento será conocido como sístole y pasará a diástole cuando se empieza a inflar por la recepción de sangre.

El corazón humano está dividido en cuatro cámaras: las dos aurículas y los dos ventrículos. Las aurículas se encuentran en la parte superior del corazón y están separadas por el tabique interauricular. Los ventrículos se encuentran en la parte inferior del corazón y están separados por el tabique interventricular. Cada aurícula se comunica con su ventrículo correspondiente a través de una válvula atrioventricular: la válvula tricúspide en el lado derecho y la válvula mitral en el lado izquierdo. Estas válvulas permiten que la sangre fluya en una sola dirección, desde las aurículas hacia los ventrículos, tal como se visualiza en la Figura 1.

Figura 1. Anatomía del corazón.



Fuente: THE TEXAS HEART INSTITUTE. Anatomía del corazón. The Texas Heart Institute [página web]. [Consultado el 19, marzo, 2023]. Disponible en Internet: <https://www.texasheart.org/heart-health/heart-information-center/topics/anatomia-del-corazon/>.

Además de las aurículas y los ventrículos, el corazón cuenta con otros componentes importantes. La aorta es la arteria principal que sale del corazón y lleva la sangre oxigenada hacia el resto del cuerpo. La aorta se divide en varias ramas más pequeñas para suministrar sangre a diferentes partes del cuerpo.

4.2.2 Conducción eléctrica en el corazón. Para realizar el proceso anterior, el corazón se contrae de forma cíclica y luego se expande. El procedimiento variará dependiendo de la conducción eléctrica que pase a través de las fibras musculares del corazón. Por esta razón, el sistema de conducción es auto excitable.

4.2.3 Electrocardiogramas (ECG). Para evaluar la conducción eléctrica del corazón, se registra la información en un electrocardiograma y así se logra hacer una detección de anomalías o enfermedades cardíacas. Este es un instrumento mecánico que realiza trazos del ciclo cardíaco. Cada cuadro del electrocardiograma corresponde a 0.5mV y 0.2 segundos. A continuación, se nombran los componentes evaluados en un ECG:

4.2.3.1 Complejo QRS. Mide el tiempo total de despolarización ventricular y se mide desde el inicio de la onda Q hasta el final de la onda S.

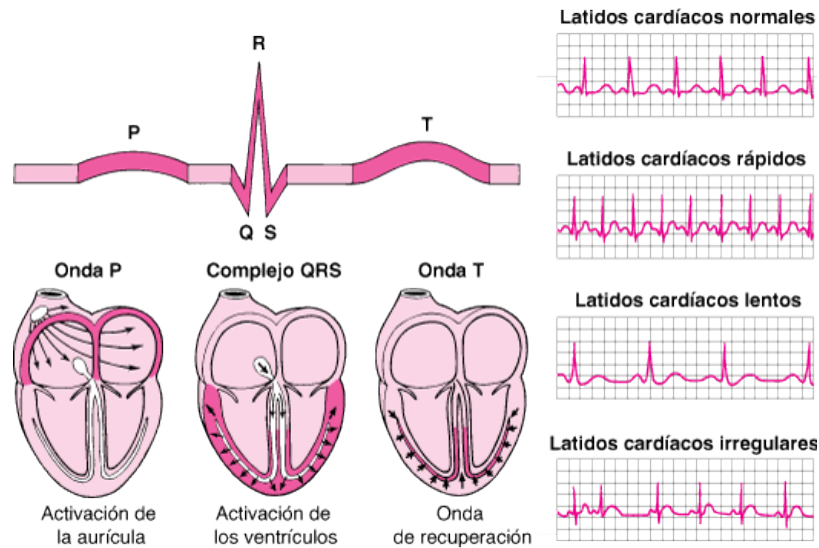
4.2.3.2 Onda J. Villamil y Alvarado indican en su texto “Onda J: predictor de muerte súbita”, que esta es una “onda con deflexión positiva que une la parte final del complejo QRS con el inicio del segmento S-T”.

4.2.3.3 Segmento ST. Representa el inicio de la repolarización ventricular lenta.

4.2.3.4 Onda T. Muestran la repolarización ventricular en el corazón.

En la figura 2 se pueden visualizar las diferentes ondas y el complejo QRS, para mayor entendimiento de su representación en un electrocardiograma:

Figura 2. Lectura de las ondas en el electrocardiograma y la visualización de la propagación de los impulsos dentro del corazón.



Fuente: ECG: lectura de las ondas - Manuales Merck versión para el público general [Anónimo]. Manual Merck versión para el público general [página web]. [Consultado el 4, marzo, 2023]. Disponible en Internet: <<https://www.merckmanuals.com/es-us/hogar/multimedia/figure/ecg-lectura-de-las-ondas>>.

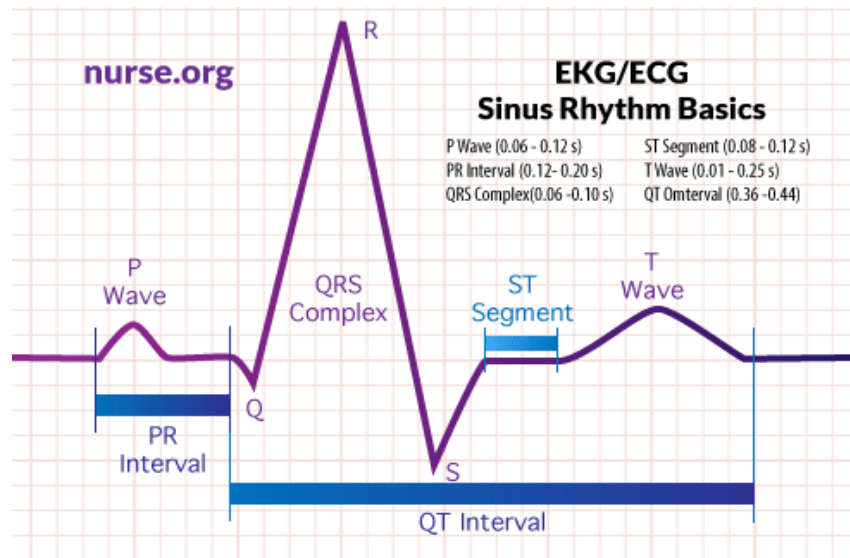
Kathleen¹⁴ y Prieto-Solís¹⁵ indican diferentes valores normales de grabación cuando el paciente no presenta anomalías. Estos se anexan a continuación y varios de los mismos podrán ser visualizados en la Figura 2:

- Intervalo QRS: 0.06-0.10 segundos.
- Segmento ST: 0.34-0.44 segundos (0.46 segundos en mujeres adultas). Valor tomado a 60 ms de la onda J.
- Onda T: 0.01 a 0.25 segundos.

¹⁴ KATHLEEN, Gaines. How To Read An Electrocardiogram (EKG/ECG). Nurse.org [página web]. (9, abril, 2022). [Consultado el 4, marzo, 2023]. Disponible en Internet: <<https://nurse.org/articles/how-to-read-an-ECG-or-EKG-electrocardiogram/>>.

¹⁵ PRIETO-SOLÍS, José A.; BENITO, Natividad y MARTÍN-DURÁN, Rafael. Diagnóstico electrocardiográfico de la obstrucción del tronco coronario izquierdo mediante el análisis vectorial del segmento ST y el complejo QRS. *En: Revista Española de Cardiología [en línea].* Febrero, 2008. vol. 61, no. 2 [consultado el 4, marzo, 2023], p. 137-145. Disponible en Internet: <<https://doi.org/10.1157/13116201>>. ISSN 0300-8932.

Figura 3. Lectura básica de intervalos rítmicos normales cardíacos en un electrocardiograma.



Fuente: KATHLEEN, Gaines. How To Read An Electrocardiogram (EKG/ECG). Nurse.org [página web]. (9, abril, 2022). [Consultado el 4, marzo, 2023]. Disponible en Internet: <https://nurse.org/articles/how-to-read-an-ECG-or-EKG-electrocardiogram/>.

La onda J, aunque muchas veces no se grafica, es un predictor de muerte súbita y es útil para encontrar el segmento ST y validar si hay anomalías cardíacas al polarizar y despolarizar.

4.2.4 Infarto de miocardio. es una afección cardíaca en la que una porción del músculo cardíaco muere debido a una falta de flujo sanguíneo. Esto puede ocurrir cuando una arteria coronaria se obstruye por una acumulación de placa o un coágulo de sangre.

Existen diferentes tipos de infarto de miocardio, que se clasifican según la ubicación del área afectada del corazón. A continuación, se nombran los tipos más comunes:

- Anterior
- Anterior e inferior
- Anterior y posterior
- anterolateral
- Antero septal

- Antero-septo-lateral
- Inferior
- Infero-lateral
- Infero-posterior
- Infero-posterior e inferior
- Infero-posterolateral
- Lateral
- Posterior
- Posterolateral

4.2.5 Paros cardiacos súbitos.

Según lo anotado por The Texas Heart Institute¹⁶, refiere a cuando el corazón deja de latir y, por tanto, la sangre no llega al cerebro y a otros órganos del cuerpo. Se diferencia del ataque cardíaco en que este último detiene el flujo de sangre en una parte del corazón (Paro cardíaco súbito, s/f).

4.2.6 Arritmias cardiacas.

Son aquellas alteraciones que se presentan al evaluar el ritmo cardiaco. Generalmente, se puede encontrar un ritmo más lento o rápido del corazón, según el tipo de afección.

4.2.7 Pre-procesamiento de señales.

Refiere al conjunto de técnicas de recolección de información, de forma que se pueda adquirir una señal de entrada estable. Venkatesan et al¹⁷ indican que un preprocesamiento de señales ECG se basa en la eliminación primaria del ruido blanco, además de que es importante adicionar transformadas Wavelet para poder extraer características.

¹⁶ PARO CARDÍACO súbito [Anónimo]. The Texas Heart Institute [página web]. [Consultado el 19, mayo, 2022]. Disponible en Internet: <<https://www.texasheart.org/heart-health/heart-information-center/topics/paro-cardiaco-subito/>>.

¹⁷ VENKATESAN, C., *et al.* ECG signal preprocessing and SVM classifier-based abnormality detection in remote healthcare applications. *En: IEEE Access* [en línea]. 2018. vol. 6 [consultado el 11, marzo, 2023], p. 9767-9773. Disponible en Internet: <<https://doi.org/10.1109/access.2018.2794346>>. ISSN 2169-3536.

4.2.7.1 Eliminación de ruido. Las imágenes tenderán a verse afectadas por la instrumentación mecánica y en sí el procesamiento aplicado. La imagen se altera al no encontrarse un color y luz estable en los píxeles.

4.2.7.2 Transformada wavelet. Es un tipo de filtro básico que logrará una reducción lineal o no lineal, dependiendo del análisis y transformada de la señal. La función wavelet logrará una composición o recomposición de la señal a partir de una entrada de muestra con características específicas. En el modo de actuación, Venkatesan¹⁸ afirma que la transformada de wavelet se aplica de manera sucesiva en los coeficientes de aproximación para recuperar la señal original sacrificando una pequeña cantidad de información.

4.2.7.3 Filtros. Al considerar el ruido por movimiento de los objetos evaluados, la calidad y el tipo de procesador que se esté usando para dibujar una determinada señal, se encontrarán ruidos que deterioran una señal y por lo tanto, la evaluación de la misma se hace más tediosa.

4.2.8 Machine learning y reconocimiento de patrones.

4.2.8.1 Validación cruzada. es una técnica utilizada en el aprendizaje automático para evaluar la capacidad predictiva de un modelo en datos no vistos. Consiste en dividir el conjunto de datos en varios conjuntos de entrenamiento y prueba, ajustar el modelo a cada uno de ellos y evaluar su rendimiento en los conjuntos de prueba.

La validación cruzada implica dividir el conjunto de datos en varios subconjuntos más pequeños, y luego entrenar y evaluar el modelo en cada uno de estos subconjuntos de forma iterativa. Esto permite una evaluación más precisa de la precisión del modelo y ayuda a evitar el sobreajuste (cuando el modelo se ajusta demasiado a los datos de entrenamiento y no se generaliza bien a los datos nuevos).

¹⁸ VENKATESAN, C., *et al.* ECG signal preprocessing and SVM classifier-based abnormality detection in remote healthcare applications. *En: IEEE Access* [en línea]. 2018. vol. 6 [consultado el 11, marzo, 2023], p. 9767-9773. Disponible en Internet: <<https://doi.org/10.1109/access.2018.2794346>>. ISSN 2169-3536.

Una vez que la SVM ha sido entrenada, se utilizan los datos resultantes para predecir las etiquetas de un conjunto de datos de evaluación. La evaluación se realiza midiendo la precisión de las predicciones en el conjunto de datos de evaluación y comparándolas con las etiquetas verdaderas conocidas de dicho conjunto de datos.

4.2.8.1.1. Validación cruzada K-Fold. Este método es más efectivo que los estudios anteriores para identificar las muestras mal etiquetadas y mejorar la precisión de clasificación de la señal de ECG. Además, se describe el método de validación cruzada K-fold como una técnica eficaz para evitar el sobreaprendizaje y el subaprendizaje y obtener resultados precisos. Además, el método k-fold de validación cruzada es especialmente útil cuando se dispone de un conjunto de datos pequeño, ya que permite aprovechar al máximo la información disponible y obtener una mejor estimación del desempeño del modelo.

Li y Cui¹⁹ indican que, en comparación con los estudios anteriores, el método con este tipo de validación cruzada puede lograr una mayor tasa de identificación de muestras mal etiquetadas y suele ser un valor numérico mayor a 3, el cual debe ser indicado durante la programación.

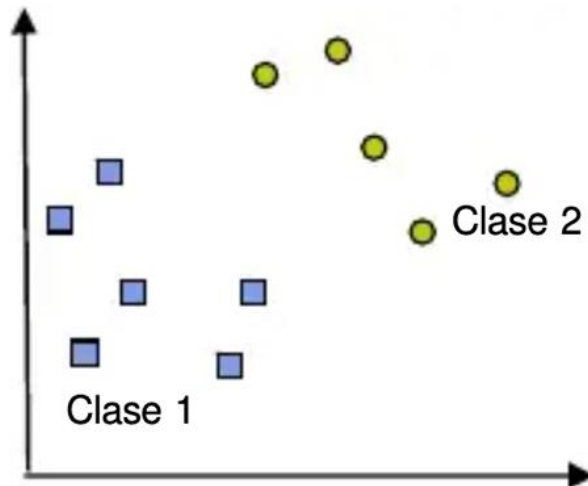
4.2.8.2 Máquinas de Vectores de Soporte. Son un algoritmo de aprendizaje supervisado que se utiliza para la clasificación y regresión de datos. El objetivo de una SVM es encontrar el hiperplano que mejor separe los datos de diferentes clases en un espacio de características. El hiperplano es una frontera de decisión que se utiliza para clasificar nuevos datos. El algoritmo busca el hiperplano que tenga la mayor distancia posible con los puntos de datos más cercanos a cada clase. Esto se conoce como máxima margen de separación.

¹⁹ LI, Yaoguang y CUI, Wei. Identifying the mislabeled training samples of ECG signals using machine learning. *En: Biomedical Signal Processing and Control* [en línea]. Enero, 2019. vol. 47 [consultado el 4, marzo, 2023], p. 168-176. Disponible en Internet: <<https://doi.org/10.1016/j.bspc.2018.08.026>>. ISSN 1746-8094.

4.2.8.3 Extracción de características. Consiste en la extracción de secuencias en imágenes, de modo que la información se asocia a características específicas en una figura o señal bidimensional analizada.

4.2.8.4 Clasificación por SVM. Según Betancourt²⁰, una Máquina de Soporte Vectorial [SVM] mapea los puntos de información de entrada y los dispone en un espacio, el cual tendrá características y dimensiones especiales, tal como se visualiza en la figura 4. La clasificación se hace entre clases (también conocidas como 'etiquetas'), las cuales tendrán diferencias entre sí, aunque esta diferencia definirá su posición en el plano, encontrando el caso de clasificación separable y no separable. Teniendo la complicación de no poder separar por completo las clases, se utilizan los kernels, los cuales calcularán una función $k()$ y así lograr la separación de características.

Figura 4. Caso linealmente separable entre dos clases.

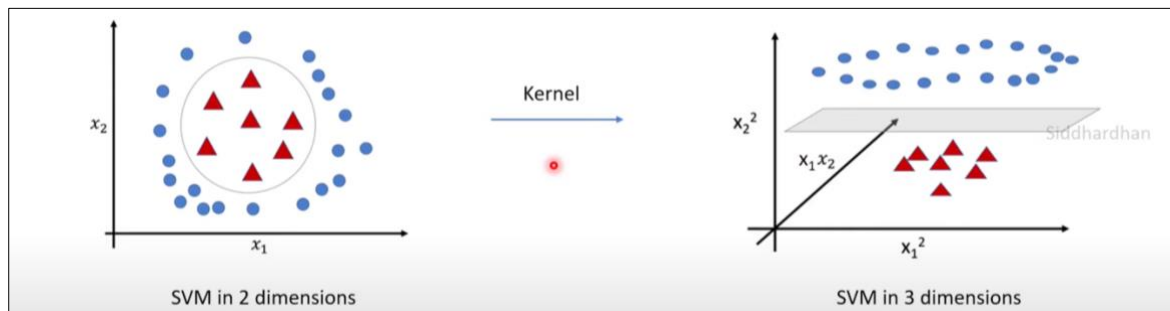


Fuente: BETANCOURT, GUSTAVO A. LAS MÁQUINAS DE SOPORTE VECTORIAL (SVMs). En: Scientia Et Technica [en línea]. 2005. vol. 1, no. 27 [consultado el 10, mayo, 2022], p. 6. Disponible en Internet: <<https://doi.org/10.22517/23447214.6895>>. ISSN 0122-1701.

²⁰ BETANCOURT, GUSTAVO A. LAS MÁQUINAS DE SOPORTE VECTORIAL (SVMs). En: Scientia Et Technica [en línea]. 2005. vol. 1, no. 27 [consultado el 10, mayo, 2022], p. 6. Disponible en Internet: <<https://doi.org/10.22517/23447214.6895>>. ISSN 0122-1701.

4.2.8.5 Kernel. es una función matemática utilizada para medir la similitud entre dos puntos de datos en un espacio dimensional elevado. Los kernels se utilizan comúnmente en SVM para convertir el problema de clasificación lineal en uno no lineal. Los kernels comúnmente utilizados son el kernel RBF, el kernel polinomial y el kernel gaussiano, aunque hay variantes como la función de Sigmoid para el Kernel, la cual no será tratada en este proyecto. Para la visualización de la información, la composición se realiza tanto en un plano bidimensional como tridimensional, encontrando diferencias notables para el tratamiento de la información, tal como se puede apreciar en la figura 5, a continuación:

Figura 5. Apreciación del uso del kernel y la forma de visualización y nuevo tratamiento de la data dentro de un plano.



Fuente: SIDDHARDHAN. 7.3.3. support vector machine - kernels [video]. YouTube. (29, octubre, 2021). [Consultado el 18, marzo, 2023]. 19:51 min. Disponible en Internet: <<https://www.youtube.com/watch?v=WLYFhROVM18>>.

Asimismo, teniendo en cuenta las variables y la vista del plano tridimensional gracias a los kernels, Siddhardhan²¹ también incluye las tres fórmulas que son la base para cálculos y la respectiva elección del Kernel a utilizar en todo el proyecto. Por otro lado, también se utiliza el Kernel gaussiano, el cual transforma los vectores de características en un espacio no lineal y Esterhuizen²² anexa la fórmula con la cual actúa este Kernel sobre las clases. Estas cuatro fórmulas se compilan en la siguiente tabla 1:

²¹ SIDDHARDHAN. 7.3.3. support vector machine - kernels [video]. YouTube. (29, octubre, 2021). [Consultado el 18, marzo, 2023]. 19:51 min. Disponible en Internet: <<https://www.youtube.com/watch?v=WLYFhROVM18>>.

²² ESTERHUIZEN, Dewald. C# how to: calculating gaussian kernels. Software by Default [página web]. (8, junio, 2013). [Consultado el 11, abril, 2023]. Disponible en Internet: <<https://softwarebydefault.com/2013/06/08/calculating-gaussian-kernels/>>.

Tabla 1. Fórmulas de kernels.

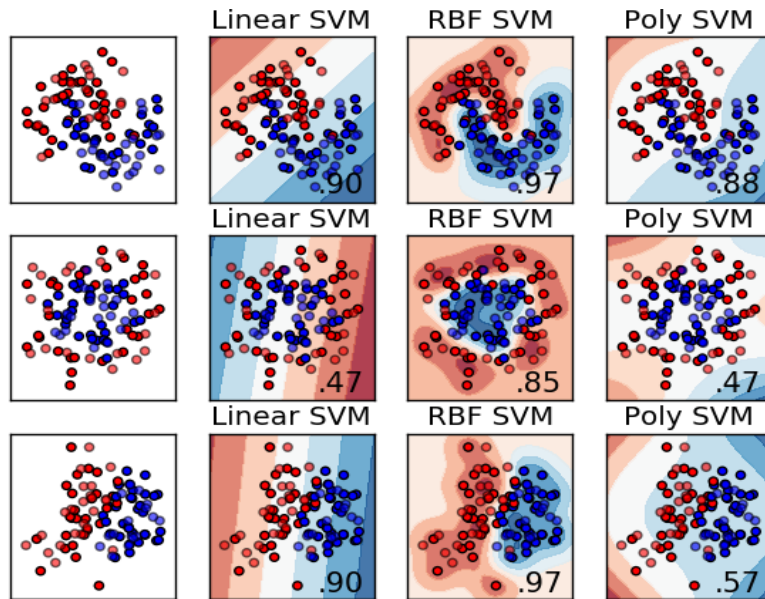
Tipo de Kernel	Fórmula
Lineal	$K(x_1, x_2) = x_1^T x_2$
Polinomial	$K(x_1, x_2) = (x_1^T x_2 + r)^d$
Función de Base Radial (RBF)	$K(x_1, x_2) = \exp(-\gamma \cdot \ x_1 - x_2\ ^2)$
Gaussiano	$K(x_1, x_2) = x_1^T x_2$

Como se aprecia, el kernel lineal calcula la similitud entre dos vectores como el producto punto de estos y el Kernel polinomial es similar al lineal, pero con la posibilidad de mapear los datos a un espacio de características de mayor dimensión mediante la adición de una constante y un exponente; No obstante, el kernel gaussiano también es similar al lineal, pero con la diferencia de que este utiliza una función de distribución gaussiana para medir la similitud entre dos vectores. Por otro lado, el kernel RBF (de Función de Base Radial) utiliza una función de distribución de probabilidad gaussiana para medir la similitud entre dos vectores y no es simétrico.

Según Sánchez, la explicación pertinente de cada uno de los kernels que se dispusieron en la anterior tabla sería la siguiente: “Kernel lineal, Kernel radial y Kernel polinomial. El Kernel lineal es equivalente a un *support vector classifier*, mientras que el Kernel radial establece los límites de forma radial y el Kernel polinomial tiene límites más flexibles”²³, los cuales, podrán ser visualizados en la Figura 6. A esto se le suma el Kernel gaussiano, el cual fue explicado en el anterior párrafo.

²³ ESTERHUIZEN, Dewald. C# how to: calculating gaussian kernels. Software by Default [página web]. (8, mayo, 2013). [Consultado el 11, abril, 2023]. Disponible en Internet: <<https://softwarebydefault.com/2013/06/08/calculating-gaussian-kernels/>>.

Figura 6. Tipos de kernels y discernimiento de datos según la necesidad de separación de cada punto.



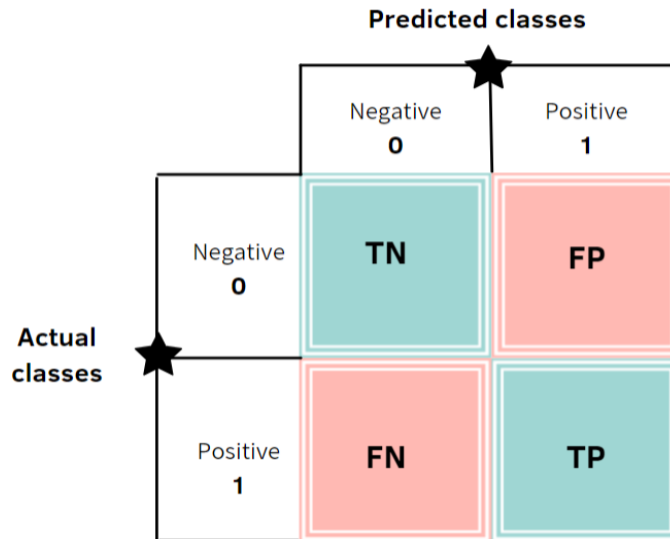
Fuente: RESIDENTMARIO. Kernels and support vector machine regularization. Kaggle: Your Machine Learning and Data Science Community [página web]. (14, abril, 2018). [Consultado el 19, marzo, 2023]. Disponible en Internet: <<https://www.kaggle.com/code/residentmario/kernels-and-support-vector-machine-regularization>>.

4.2.8.6 Segmentación y normalización de señal. La segmentación de imágenes es posible validando sus características y diferenciales entre sí. En SVM, el método solo es estable cuando los vectores de características son de baja dimensión y no hay estabilidad en características.

4.2.8.7 Métricas de la SVM: La matriz de confusión tiene cuatro elementos, que se dividen en dos categorías: verdaderos positivos (TP), verdaderos negativos (TN), falsos positivos (FP) y falsos negativos (FN).

En la matriz de confusión, cada fila representa las instancias en una clase real, mientras que cada columna representa las instancias en una clase predicha. Por lo tanto, los elementos en la diagonal principal representan los aciertos de la SVM, mientras que los elementos fuera de la diagonal principal representan las clasificaciones incorrectas. Su distribución se puede visualizar en la siguiente figura:

Figura 7. Matriz de confusión y uso de clases.



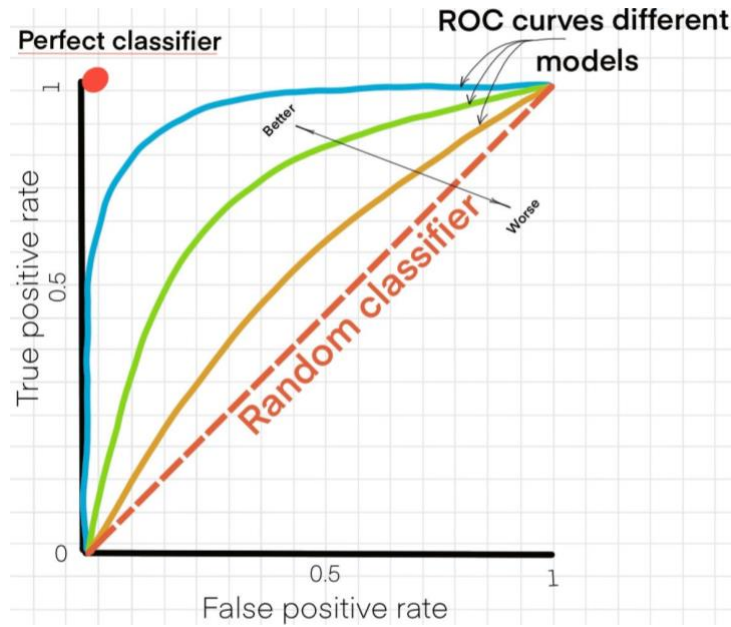
Fuente: EDITORIAL TEAM. How to evaluate you model using the Confusion Matrix. Towards AI [página web]. (23, febrero, 2021). [Consultado el 18, marzo, 2023]. Disponible en Internet: <<https://towardsai.net/p/data-science/how-to-evaluate-you-model-using-the-confusion-matrix>>.

4.2.8.8 Curva ROC (Receiver Operating Characteristic, por sus siglas en inglés). Es una herramienta utilizada en la evaluación de modelos de clasificación. Es una representación gráfica que muestra la relación entre la tasa de verdaderos positivos (sensibilidad) y la tasa de falsos positivos (1 - especificidad) para diferentes valores de umbral de clasificación.

En esta herramienta de evaluación, cada punto representa un valor de umbral de clasificación diferente y se calcula trazando una línea entre el punto (0,0) y el punto (1,1) en un gráfico de coordenadas cartesianas. Un modelo ideal tendría una curva ROC que se acercara lo más posible al punto (0,1), que representaría una sensibilidad del 100% y una especificidad del 100%.

Esta también proporciona una medida numérica de la calidad del modelo de clasificación como se aprecia en la Figura 8: el área bajo la curva (AUC). Un AUC cercano a 1 indica un modelo con una alta capacidad de discriminación, mientras que un AUC cercano a 0,5 indica un modelo que no es mejor que un clasificador aleatorio.

Figura 8. Clasificación y medida numérica de modelos con la curva ROC.



Fuente: EDITORIAL TEAM. How to evaluate you model using the Confusion Matrix. Towards AI [página web]. (23, febrero, 2021). [Consultado el 18, marzo, 2023]. Disponible en Internet: <<https://towardsai.net/p/data-science/how-to-evaluate-you-model-using-the-confusion-matrix>>.

5. MATERIALES Y MÉTODOS

5.1 MATERIALES

Tabla 2. Recursos y materiales utilizados para el trabajo de grado.

RECURSOS NECESARIOS		
RECURSO	DESCRIPCION	PRESUPUESTO (\$)
Equipo Humano	Director y estudiante.	\$0
Equipos y Software	Equipo de cómputo y software Matlab con licencia otorgada por la universidad.	\$0
Viajes y Salidas de Campo	N/A	\$0
Materiales y suministros	Electrocardiogramas con cardiopatía isquémica de infarto de miocardio extraídas de la base de datos PhysioNet.	\$0
Bibliografía	IEEE, Repositorios académicos.	\$0
TOTAL		\$0

5.2 METODOLOGÍA

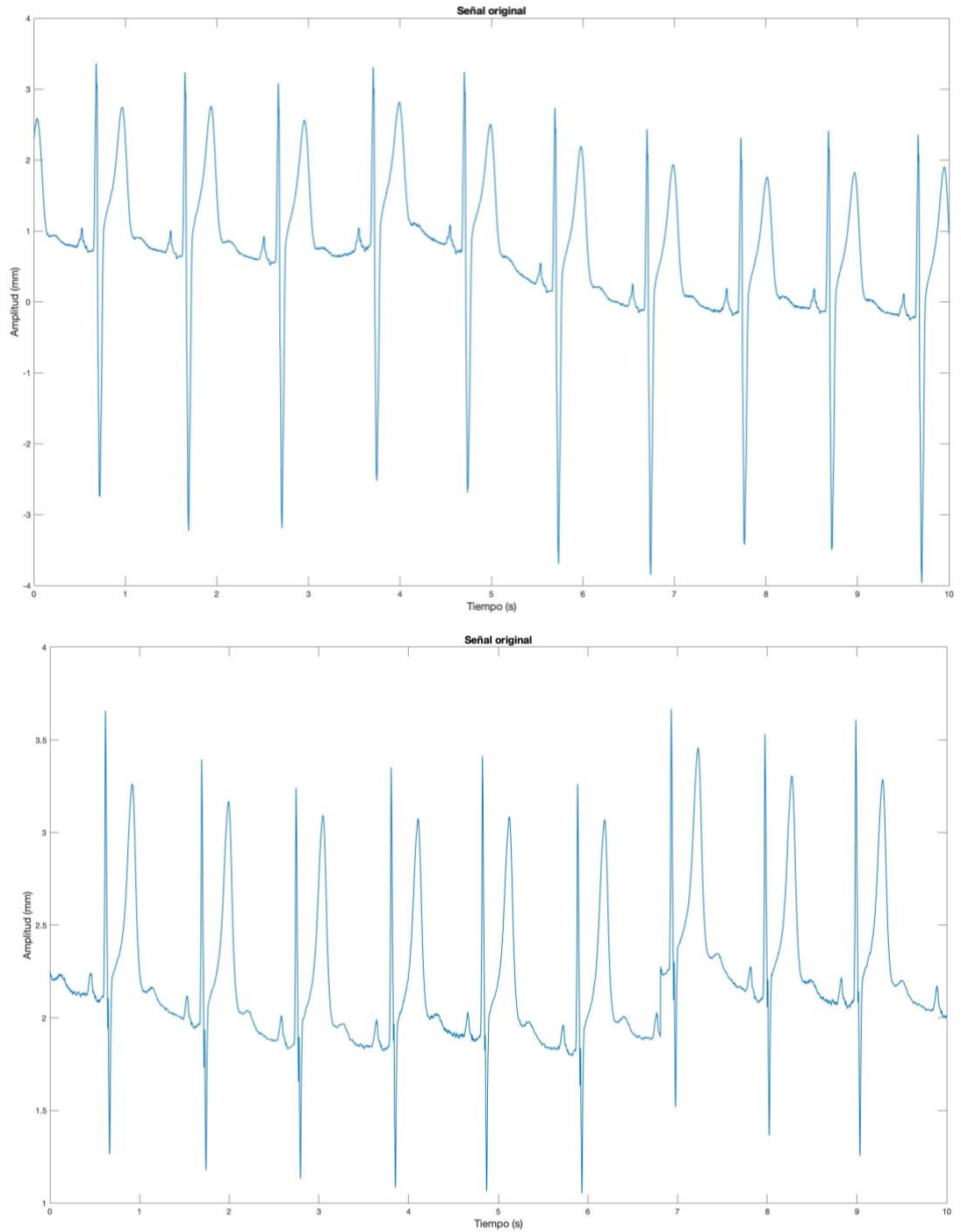
5.2.1 Tipo de investigación. Es un proyecto de investigación que se centra en resolver un problema práctico específico en el campo de la salud por medio de la detección rápida de arritmias conocidas como cardiopatías isquémicas (infartos de miocardio) mediante el uso de un clasificador SVM.

5.2.2 Población y muestra. La población son todas las personas que podrían sufrir de arritmias y tener un diagnóstico de infarto de miocardio. Las muestras utilizadas se toman del repositorio “PTB Diagnostic ECG Database”, en el sitio web PhysioNet, las cuales se tomaron de pacientes reales ingresados y, por tanto, da una mayor validez a los resultados obtenidos.

5.2.3 Análisis y preprocesamiento de los datos. El algoritmo se implementa en su totalidad utilizando el entorno de Matlab. Se debe hacer un preprocesamiento de las señales ECG para lograr extraer la señal del ventrículo 3 [V3] y luego estandarizar su frecuencia de muestreo a 1000 Hz.

En la figura 9, se pueden observar dos señales originales sin procesamiento, la de la izquierda presenta control sano y la señal de la derecha presenta diagnóstico de arritmia.

Figura 9. Visualización de dos señales (sana y con arritmia), sin procesar.

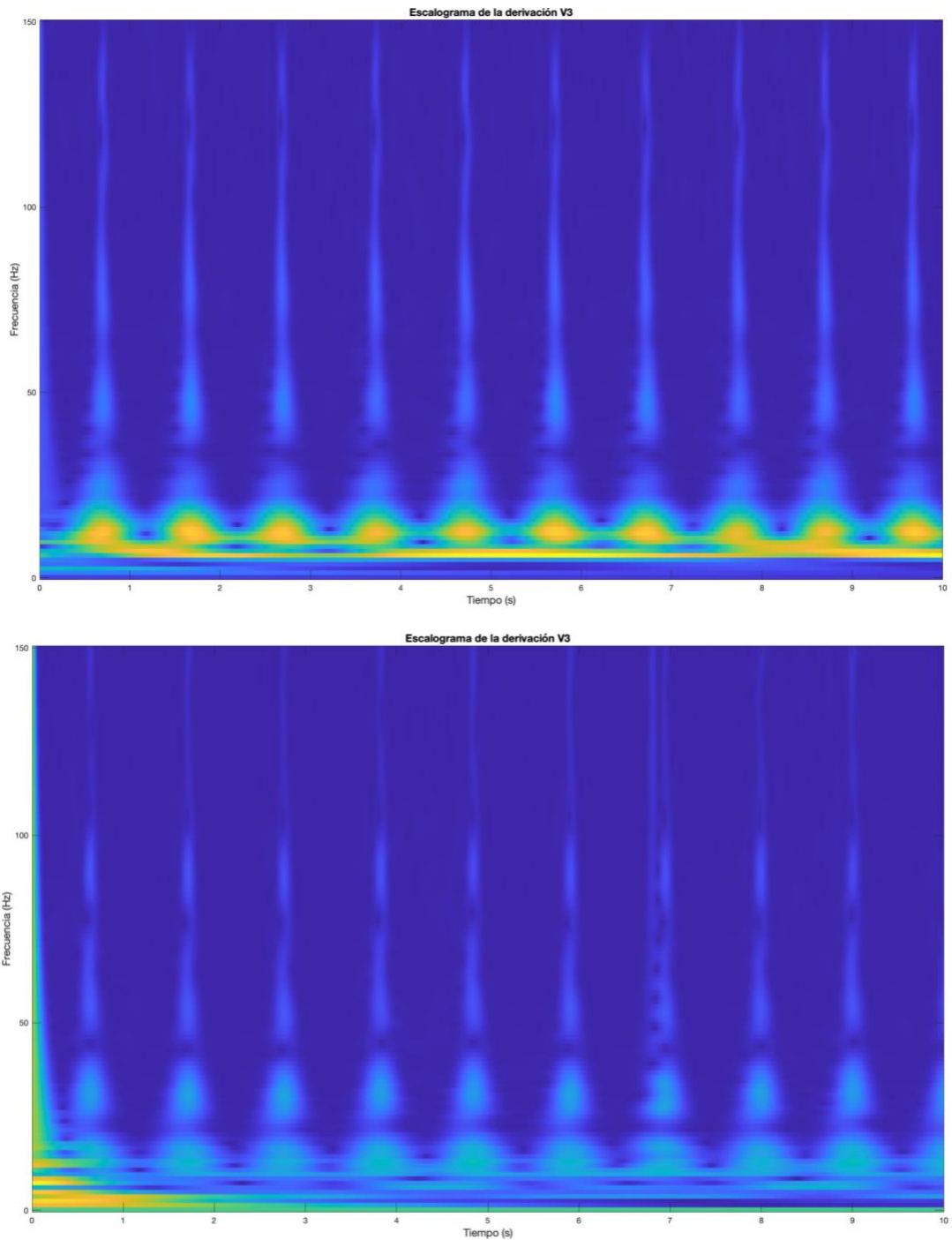


Nota. Señal sin arritmia (superior) y con arritmia (inferior). Estas señales sin procesar pueden presentar variaciones en amplitud y ruido, por la falta de corrección de la línea basal o isoelectrica.

Teixeira, C. A., et al.²⁴ utilizan la transformada Wavelet de Gabor para visualizar la información en el espectro frecuencial y este mismo método es utilizado en este proyecto para lograr encontrar el rango de frecuencias que contiene la mayor energía (Ver figura 10). Esto mismo es útil para eliminar el ruido, eliminar la línea basal y lograr visualizar la señal sin el ruido adquirido por los componentes físicos usados al momento de realizar las mediciones.

²⁴ TEIXEIRA, C. A., *et al.* EPILAB: A software package for studies on the prediction of epileptic seizures. *En: Journal of Neuroscience Methods* [en línea]. Septiembre, 2011. vol. 200, no. 2 [consultado el 4, marzo, 2023], p. 257-271. Disponible en Internet: <<https://doi.org/10.1016/j.jneumeth.2011.07.002>>. ISSN 0165-0270.

Figura 10. Escalograma del ventrículo 3, sin procesar, de una señal sana y otra con arritmia.

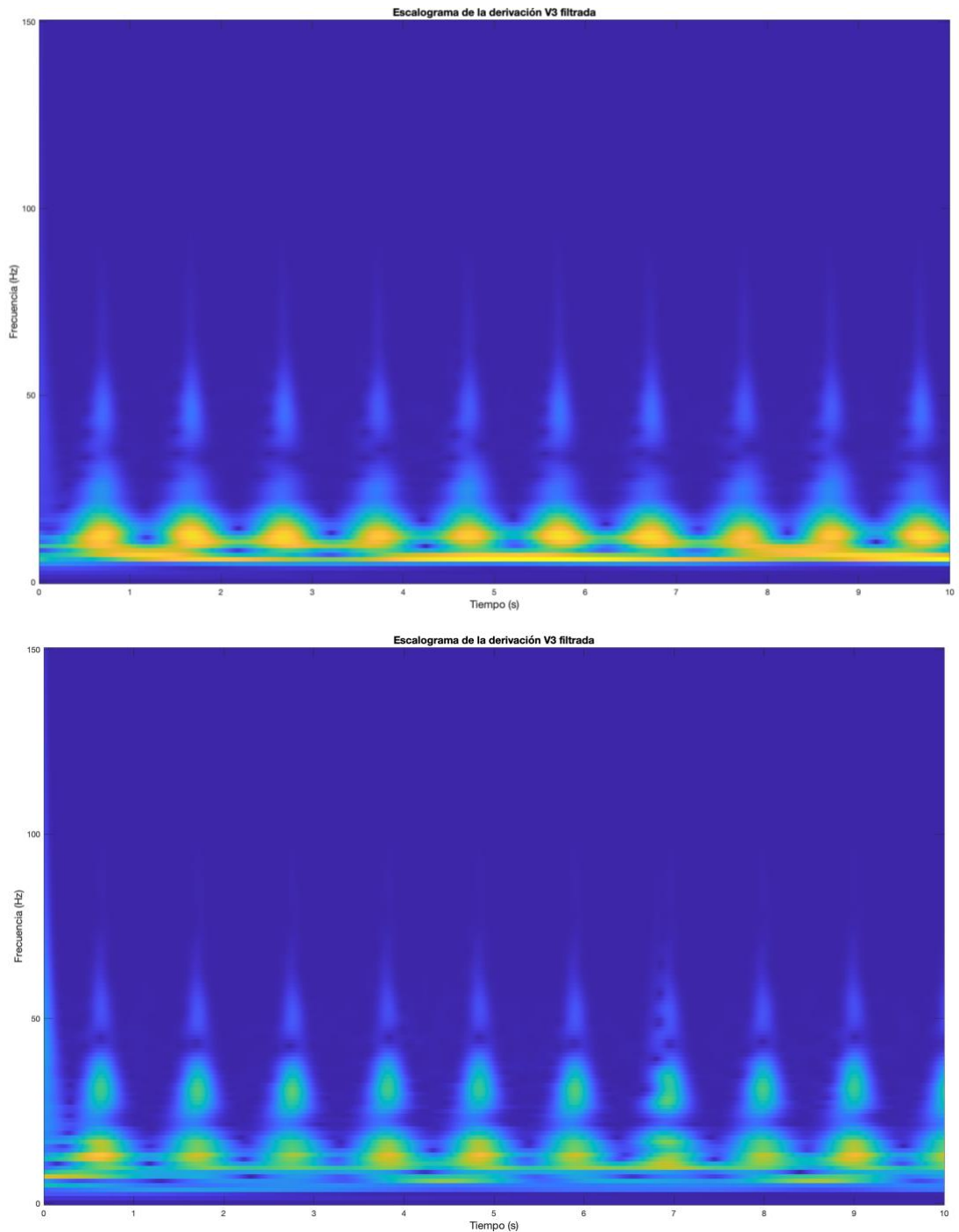


Nota. Señales sin arritmia (superior) y con arritmia (inferior), antes de la corrección de la línea basal.

Seguido de esto, se diseña un filtro pasa-banda Butterworth de orden 6 con frecuencia de 4 hasta 55 Hz, con el fin de realizar la corrección de la línea basal (también llamada línea de base). Esta corrección se puede apreciar en la figura 11, la cual representa de nuevo las mismas señales (con y sin arritmia) pero filtradas.

Visualmente, el ruido es reducido, pero no de forma abrupta, haciendo que la señal no sufra modificaciones en cuanto a contenido, que más adelante perjudiquen el procesamiento de las señales.

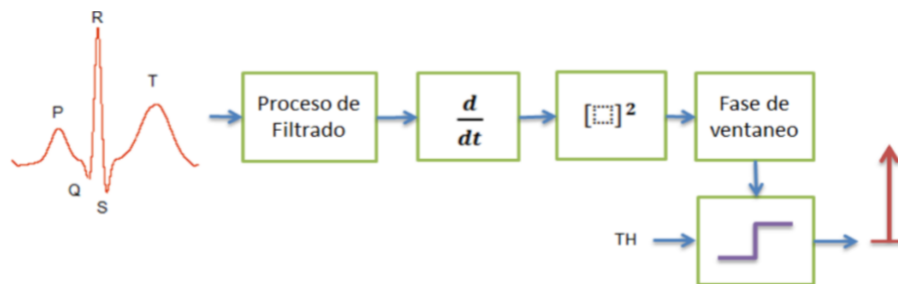
Figura 11. Escalograma del ventrículo 3 de dos señales procesadas, una es una señal sana y la otra presenta arritmia.



Nota. Señales sin arritmia (superior) y con arritmia (inferior) filtradas para la corrección de la línea basal.

Luego se utiliza el algoritmo de Pan-Tompkins, que es una técnica ampliamente utilizada para la detección del complejo QRS en señales ECG. Este algoritmo es útil para lograr detectar la onda o pico R, que es la que marca el inicio de la ventana para encontrar las siguientes ondas. Seguido, se puede encontrar la onda S, la cual es la onda que representa los valles en la señal y es de vital importancia para posteriormente encontrar la onda T, la cual, como se visualiza en el marco teórico, es aquella onda que se presenta como el siguiente pico más pronunciado después de esa onda S. Asimismo, entre la onda S y la onda T, si hay arritmias, se encuentra una onda pronunciada, pero de menor amplitud que la onda T, que se conoce como la onda J. El paso a paso de este algoritmo se visualiza en la Figura 12.

Figura 12. Paso a paso del procesamiento con el algoritmo de Pan-Tompkins, en una señal ECG.

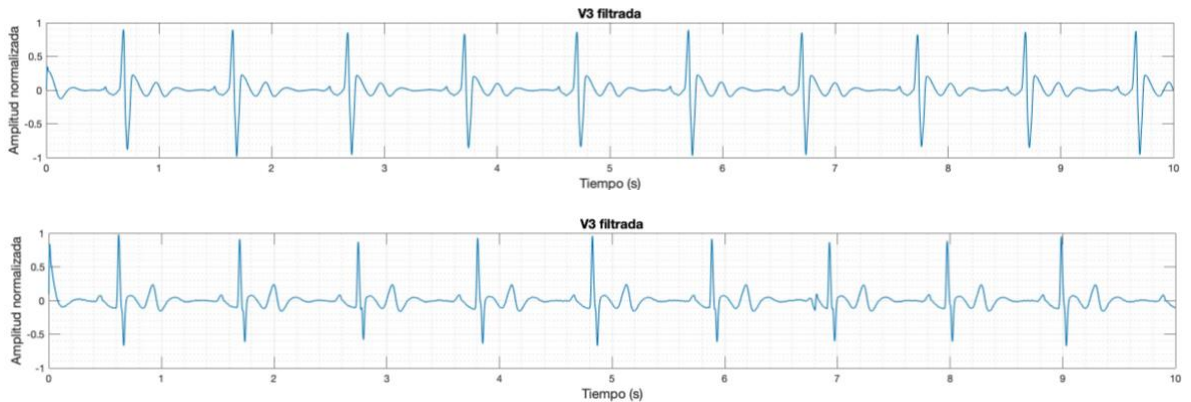


Fuente: ALGORITMO DE Pan Tompkins (QRS) [Anónimo]. Google Docs [página web]. [Consultado el 1, abril, 2023]. Disponible en Internet: <https://docs.google.com/document/d/1Wxa7u0gtGcF_A-ayon7RUWa0ml0yvvcFW05APOE9vD8/preview?hgd=1>.

Cada paso del algoritmo se explica a continuación:

- Filtro: primero realiza un filtrado de la señal normalizada, el cual se realiza y se visualiza en las anteriores figuras y tiene como objetivo el eliminar ruido y corregir la línea basal o isoeletrica que se causa por ser una señal tomada desde dispositivos electrónicos análogos. Este filtrado se visualiza en la Figura 13.

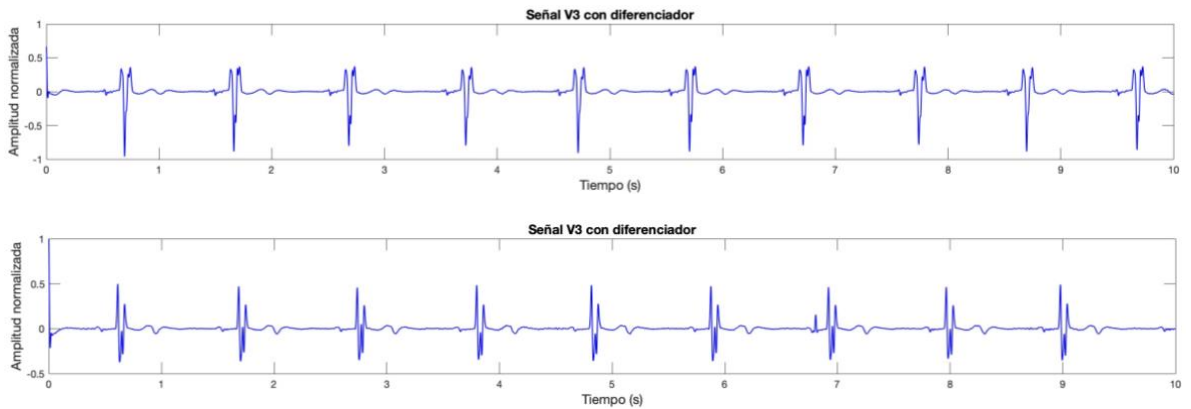
Figura 13. Filtrado en señales sin y con arritmia.



Nota. Señales sin arritmia (superior) y con arritmia (inferior) filtradas y sin ruido, para poder proseguir con el algoritmo de Pan-Tompkins.

- Derivación de la señal: es necesario para pronunciar los picos y dar paso a poder encontrar las ondas R en toda la señal. Este procedimiento se visualiza en la Figura 14.

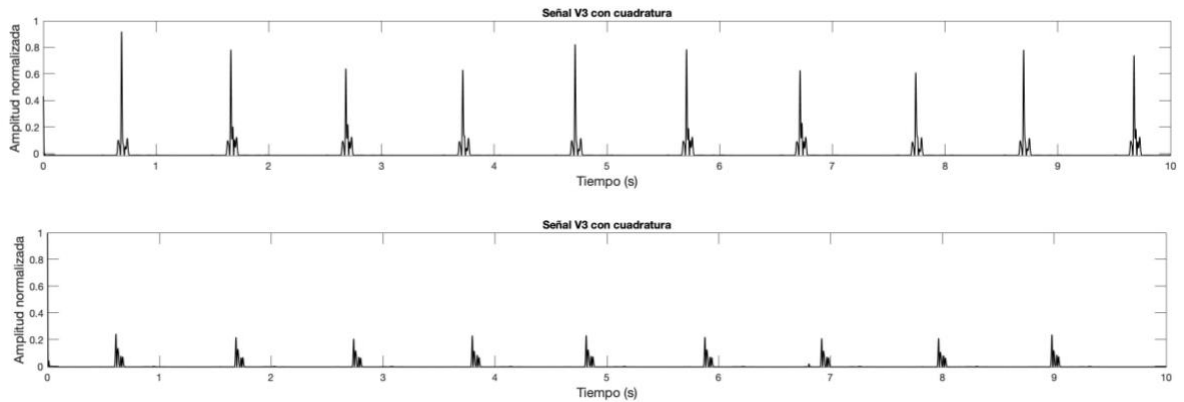
Figura 14. Derivación de señales con y sin arritmia.



Nota. Señales sin arritmia (superior) y con arritmia (inferior) procesadas y con el filtro de derivación de Pan-Tompkins.

- Elevación al cuadrado: se eleva el resultado anterior al cuadrado para que los valores pasen a ser netamente positivos y los resultados se aprecian en la Figura 15, tanto para la señal con control sano como para la señal con arritmia.

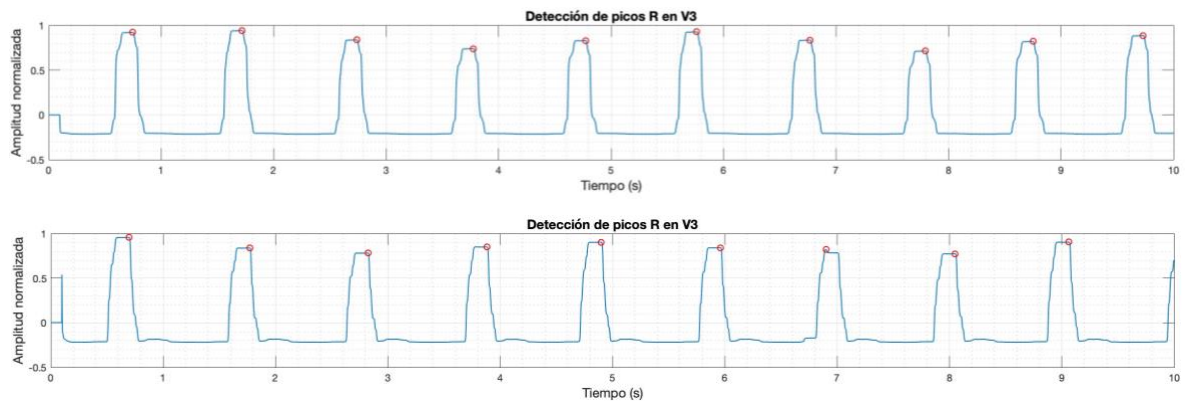
Figura 15. Cuadratura en señales con y sin arritmia.



Nota. Señales sin arritmia (superior) y con arritmia (inferior) procesadas con el algoritmo de Pan-Tompkins hasta la elevación al cuadrado.

- Integración de la señal: finalmente detecta los picos por medio de un umbral y una ventana integradora móvil. El umbral encuentra picos máximos a partir del 58% de la amplitud de la señal y con una ventana de 200 muestras. Esta integración se visualiza en la Figura 16.

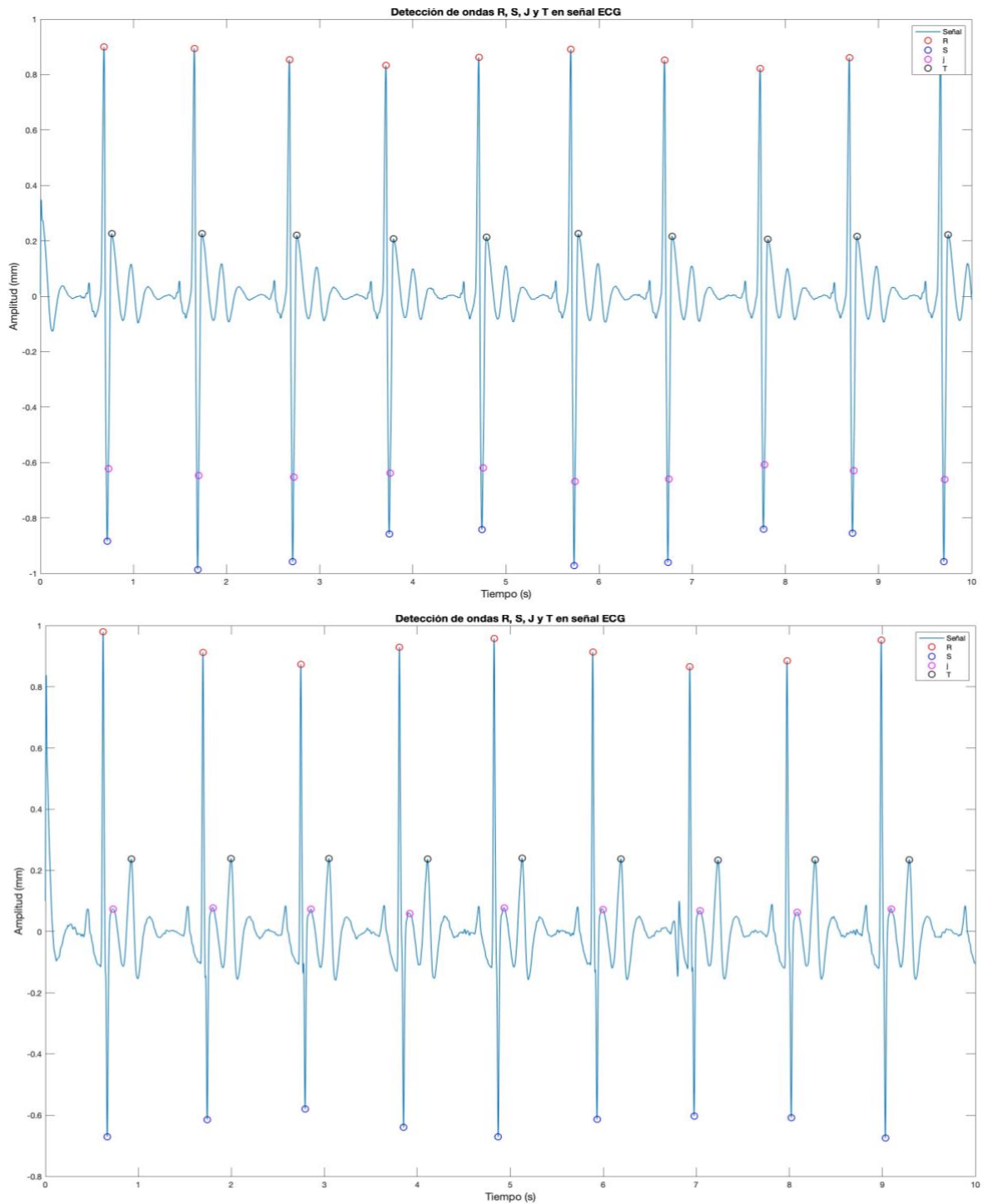
Figura 16. Algoritmo de Pan-Tompkins en señales con y sin arritmia.



Nota. Señales sin arritmia (superior) y con arritmia (inferior) procesadas con el algoritmo de Pan-Tompkins hasta el proceso de integración y detección de picos R.

Finalmente, la detección de las ondas R, S, J y T se logra apreciar en la Figura 14, en donde se visualiza la onda filtrada, sin ruido y detectando cada onda tanto en señales sanas como con arritmia detectada.

Figura 17. Detección de ondas en señales ECG.



Nota. Detección de ondas R, S, J y T en dos señales, sin arritmia (superior) y con arritmia (inferior).

Cuando la señal ya presenta un acondicionamiento por procesamiento de señal, se procede a encontrar características básicas de la arritmia producida por infarto agudo de miocardio. El entrenamiento de la máquina puede tomar estas características y así logra procesar de forma adecuada las señales finales de prueba.

El segmento ST, se reitera, conecta el complejo QRS y la onda T. Si hay deflexión del punto u onda J, se puede predecir una enfermedad cardíaca estructural. Por esto mismo, se escogen tres características relacionadas con el segmento ST y sus variaciones en amplitud, con base en la teoría de valoración del Segmento ST²⁵ y el '*Fourth Universal Definition of Myocardial Infarction Consensus Document*'²⁶, el cual incorpora la clasificación actual de los Infartos Agudos de Miocardio y los criterios de evaluación:

1. Elevación del segmento ST > de 1 mm en dos derivaciones contiguas.
2. Elevación del segmento ST > de 2 mm en derivaciones no necesariamente contiguas (V2 causa errores).
3. Descenso del segmento ST < de 1 mm en dos derivaciones contiguas.

Posteriormente al entrenamiento de la máquina SVM, la clasificación se evalúa, de forma que, teniendo los parámetros básicos de lectura ECG con pacientes sin anomalías, discierne de estas e identifique aquellos pacientes que tengan lecturas con presencias de tendencia a infartos agudos de miocardio.

²⁵ VALORACIÓN DEL Segmento ST [Anónimo]. My EKG, La web del Electrocardiograma [página web]. [Consultado el 1, mayo, 2023]. Disponible en Internet: <<https://www.my-ekg.com/como-leer-ekg/segmento-st.html>>.

²⁶ THYGESEN, Kristian, *et al.* Fourth universal definition of myocardial infarction (2018). *En: Circulation* [en línea]. 13, noviembre, 2018. vol. 138, no. 20 [consultado el 5, junio, 2023]. Disponible en Internet: <<https://doi.org/10.1161/cir.0000000000000617>>. ISSN 1524-4539.

6. DESARROLLO DEL PROYECTO

6.1 CRONOGRAMA

Tabla 3. Cronograma de actividades del proyecto de grado.

ACTIVIDAD	ME S 1	ME S 2	ME S 3	ME S 4	ME S 5	ME S 6	ME S 7	ME S 8	ME S 9	ME S 10	ME S 11	ME S 12
Obtener una base de datos que contenga señales ECG con arritmia y sin arritmia, a través del repositorio Physionet.	X											
Acondicionar las señales electrocardiográficas de prueba para estandarizar las diferentes señales.		X	X	X								
Procesar las señales para la extracción de características de una arritmia producida por un infarto miocárdico (MI).				X	X	X	X					
Entrenar a la máquina virtual con diferentes parámetros							X	X	X	X		

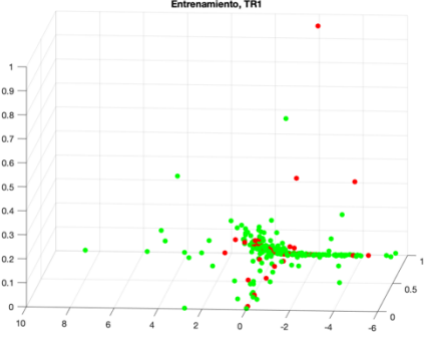
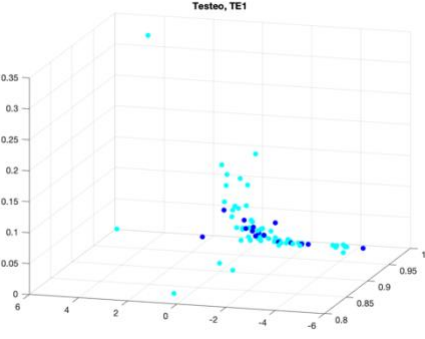
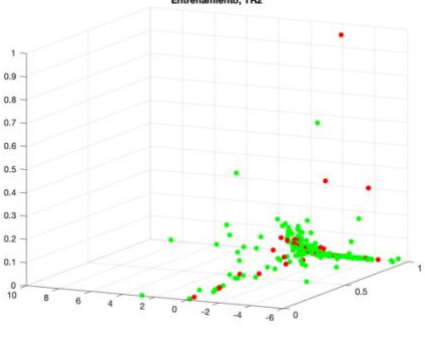
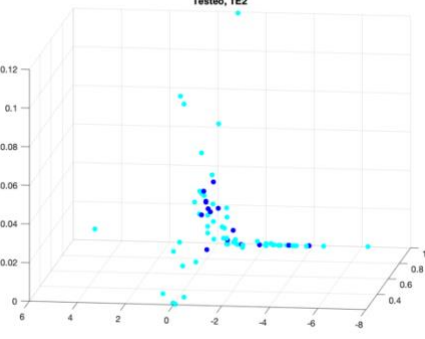
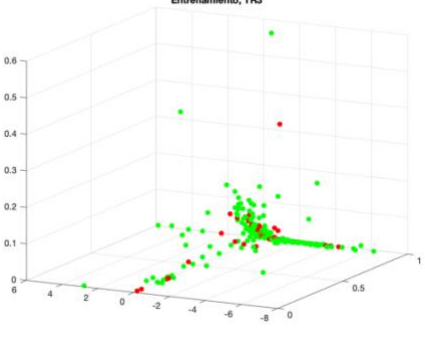
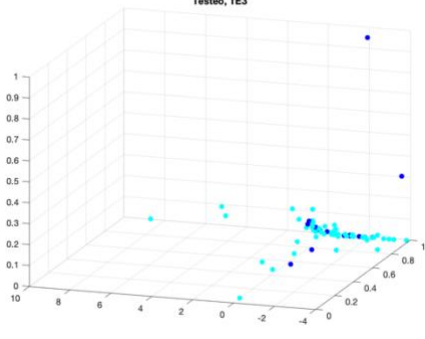
evaluados, de forma que la clasificación logre discernir entre un infarto de miocardio y una señal sin arritmia.												
Realizar pruebas y análisis de los resultados obtenidos por el clasificador SVM.										X	X	X

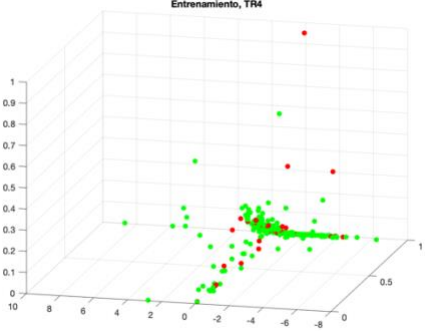
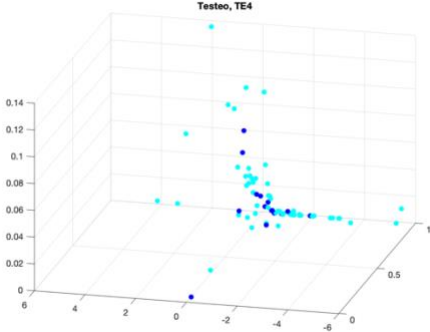
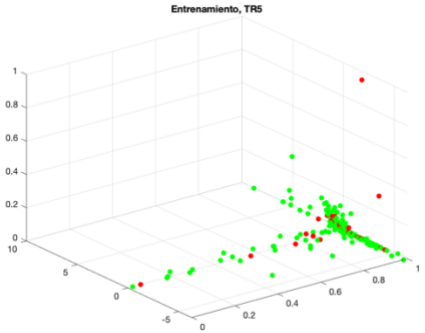
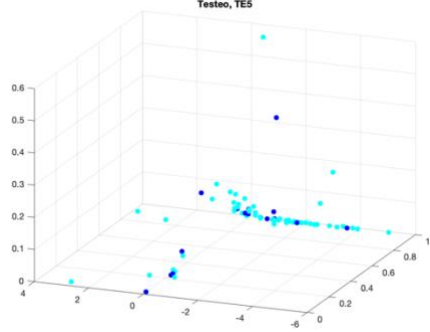
6.2 RESULTADOS

Con el fin de evitar que el algoritmo este sobre entrenado, se realiza una validación cruzada con cinco iteraciones diferentes, usando el 80% de las señales para el entrenamiento y el 20% para prueba. En todas las validaciones cruzadas se utilizan 64 señales sanas y 255 enfermas para el entrenamiento, mientras que, para la prueba se utilizan 16 señales sanas y 63 enfermas.

Tabla 4. Dispersión de datos en cada validación cruzada.

K	GRÁFICA DE DISPERSIÓN DE LOS DATOS DE ENTRENAMIENTO	GRÁFICA DE DISPERSIÓN DE LOS DATOS DE PRUEBA
---	---	--

1	<p>Gráfico 1. Dispersión de datos de entrenamiento en la validación cruzada 1.</p> 	<p>Gráfico 2. Dispersión de datos de prueba en la validación cruzada 1.</p> 
2	<p>Gráfico 3. Dispersión de datos de entrenamiento en la validación cruzada 2.</p> 	<p>Gráfico 4. Dispersión de datos de prueba en la validación cruzada 2.</p> 
3	<p>Gráfico 5. Dispersión de datos de entrenamiento en la validación cruzada 3.</p> 	<p>Gráfico 6. Dispersión de datos de prueba en la validación cruzada 3.</p> 

4	<p>Gráfico 7. Dispersión de datos de entrenamiento en la validación cruzada 4.</p> 	<p>Gráfico 8. Dispersión de datos de prueba en la validación cruzada 4.</p> 
5	<p>Gráfico 9. Dispersión de datos de entrenamiento en la validación cruzada 5.</p> 	<p>Gráfico 10. Dispersión de datos de prueba en la validación cruzada 5.</p> 

El análisis se realiza con tres kernels: polinomial, RBF y gaussiano. Los resultados de las métricas de rendimiento, tal como lo son la especificidad, exactitud, precisión y sensibilidad; además del F1-Score, el Valor Falso Positivo y el Valor Falso Negativo (Ver tabla 5). La matriz de confusión de cada validación cruzada y para los tres kernels evaluados se aprecia en la Tabla 6.

Tabla 5. Métricas de evaluación de las tres SVM, en cada validación cruzada.

VALIDACIÓN CRUZADA	POLINOMIAL	RBF	GAUSSIANO
--------------------	------------	-----	-----------

1	Especificidad: 0.500000 Exactitud: 0.544304 Precisión: 0.813953 Sensibilidad: 0.555556 F1 score: 0.660377 Valor Falso Positivo (VFP): 0.186047 Valor Falso Negativo (VFN): 0.444444	Especificidad: 0.187500 Exactitud: 0.835443 Precisión: 0.828947 Sensibilidad: 1.000000 F1 score: 0.906475 Valor Falso Positivo (VFP): 0.171053 Valor Falso Negativo (VFN): 0.000000	Especificidad: 0.000000 Exactitud: 0.797468 Precisión: 0.797468 Sensibilidad: 1.000000 F1 score: 0.887324 Valor Falso Positivo (VFP): 0.202532 Valor Falso Negativo (VFN): 0.000000
2	Especificidad: 0.625000 Exactitud: 0.518987 Precisión: 0.837838 Sensibilidad: 0.492063 F1 score: 0.620000 Valor Falso Positivo (VFP): 0.162162 Valor Falso Negativo (VFN): 0.507937	Especificidad: 0.000000 Exactitud: 0.784810 Precisión: 0.794872 Sensibilidad: 0.984127 F1 score: 0.879433 Valor Falso Positivo (VFP): 0.205128 Valor Falso Negativo (VFN): 0.015873	Especificidad: 0.000000 Exactitud: 0.797468 Precisión: 0.797468 Sensibilidad: 1.000000 F1 score: 0.887324 Valor Falso Positivo (VFP): 0.202532 Valor Falso Negativo (VFN): 0.000000
3	Especificidad: 0.812500 Exactitud: 0.329114 Precisión: 0.812500 Sensibilidad: 0.206349 F1 score: 0.329114	Especificidad: 0.000000 Exactitud: 0.784810 Precisión: 0.794872 Sensibilidad: 0.984127 F1 score: 0.879433	Especificidad: 0.000000 Exactitud: 0.797468 Precisión: 0.797468 Sensibilidad: 1.000000 F1 score: 0.887324

	<p>Valor Falso Positivo (VFP): 0.187500</p> <p>Valor Falso Negativo (VFN): 0.793651</p>	<p>Valor Falso Positivo (VFP): 0.205128</p> <p>Valor Falso Negativo (VFN): 0.015873</p>	<p>Valor Falso Positivo (VFP): 0.202532</p> <p>Valor Falso Negativo (VFN): 0.000000</p>
4	<p>Especificidad: 0.250000</p> <p>Exactitud: 0.569620</p> <p>Precisión: 0.773585</p> <p>Sensibilidad: 0.650794</p> <p>F1 score: 0.706897</p> <p>Valor Falso Positivo (VFP): 0.226415</p> <p>Valor Falso Negativo (VFN): 0.349206</p>	<p>Especificidad: 0.000000</p> <p>Exactitud: 0.797468</p> <p>Precisión: 0.797468</p> <p>Sensibilidad: 1.000000</p> <p>F1 score: 0.887324</p> <p>Valor Falso Positivo (VFP): 0.202532</p> <p>Valor Falso Negativo (VFN): 0.000000</p>	<p>Especificidad: 0.125000</p> <p>Exactitud: 0.822785</p> <p>Precisión: 0.818182</p> <p>Sensibilidad: 1.000000</p> <p>F1 score: 0.900000</p> <p>Valor Falso Positivo (VFP): 0.181818</p> <p>Valor Falso Negativo (VFN): 0.000000</p>
5	<p>Especificidad: 0.312500</p> <p>Exactitud: 0.632911</p> <p>Precisión: 0.803571</p> <p>Sensibilidad: 0.714286</p> <p>F1 score: 0.756303</p> <p>Valor Falso Positivo (VFP): 0.196429</p> <p>Valor Falso Negativo (VFN): 0.285714</p>	<p>Especificidad: 0.000000</p> <p>Exactitud: 0.772152</p> <p>Precisión: 0.792208</p> <p>Sensibilidad: 0.968254</p> <p>F1 score: 0.871429</p> <p>Valor Falso Positivo (VFP): 0.207792</p> <p>Valor Falso Negativo (VFN): 0.031746</p>	<p>Especificidad: 0.000000</p> <p>Exactitud: 0.797468</p> <p>Precisión: 0.797468</p> <p>Sensibilidad: 1.000000</p> <p>F1 score: 0.887324</p> <p>Valor Falso Positivo (VFP): 0.202532</p> <p>Valor Falso Negativo (VFN): 0.000000</p>

Tabla 6. Matrices de confusión para cada uno de los kernels evaluados en cada validación cruzada.

MATRICES DE CONFUSIÓN							
VALIDACIÓN CRUZADA		POLINOMIAL		RBF		GAUSSIANO	
		Negativo	Positivo	Negativo	Positivo	Negativo	Positivo
1	Negativo	8	8	3	13	0	16
	Positivo	28	35	0	63	0	63
2	Negativo	10	6	0	16	0	16
	Positivo	32	31	1	62	0	63
3	Negativo	13	3	0	16	0	16
	Positivo	50	13	1	62	0	63
4	Negativo	4	12	0	16	2	14
	Positivo	22	41	0	63	0	63
5	Negativo	5	11	0	16	0	16
	Positivo	18	45	2	61	0	63

6.3 ANÁLISIS DE RESULTADOS

Concluyendo la información anterior, se anexan los siguientes resultados de métricas de proximidad con cada Kernel, en cada una de las cinco validaciones cruzadas:

Tabla 7. Métricas de proximidad de medición con cada kernel, en cada respectiva validación cruzada (k).

MÉTRICAS DE PROXIMIDAD DE MEDICIÓN CON CADA KERNEL						
VALIDACIÓN CRUZADA	POLINOMIAL		RBF		GAUSSIANO	
	EXACTITUD (%)	PRECISIÓN (%)	EXACTITUD (%)	PRECISIÓN (%)	EXACTITUD (%)	PRECISIÓN (%)
1	54.43	81.39	83.54	82.89	79.75	79.75
2	51.89	83.78	78.48	79.49	79.75	79.75
3	33.91	81.25	78.48	79.49	79.75	79.75
4	56.96	77.36	79.75	79.75	82.28	81.82
5	63.29	80.36	77.21	79.22	79.75	79.75

De la anterior tabla, se puede indicar que los resultados no obedecen a una naturaleza lineal y el kernel RBF puede ser más flexible en cuanto a la captura de las relaciones no lineales complejas que pueden encontrarse en señales ECG.

Asimilando los datos anteriores, se crean gráficas de caja, las cuales permiten discernir mejor la exactitud de cada Kernel y su precisión.

Gráfico 11. Gráfico de caja de la distribución de los resultados de Exactitud en cada Kernel, de cada validación cruzada.

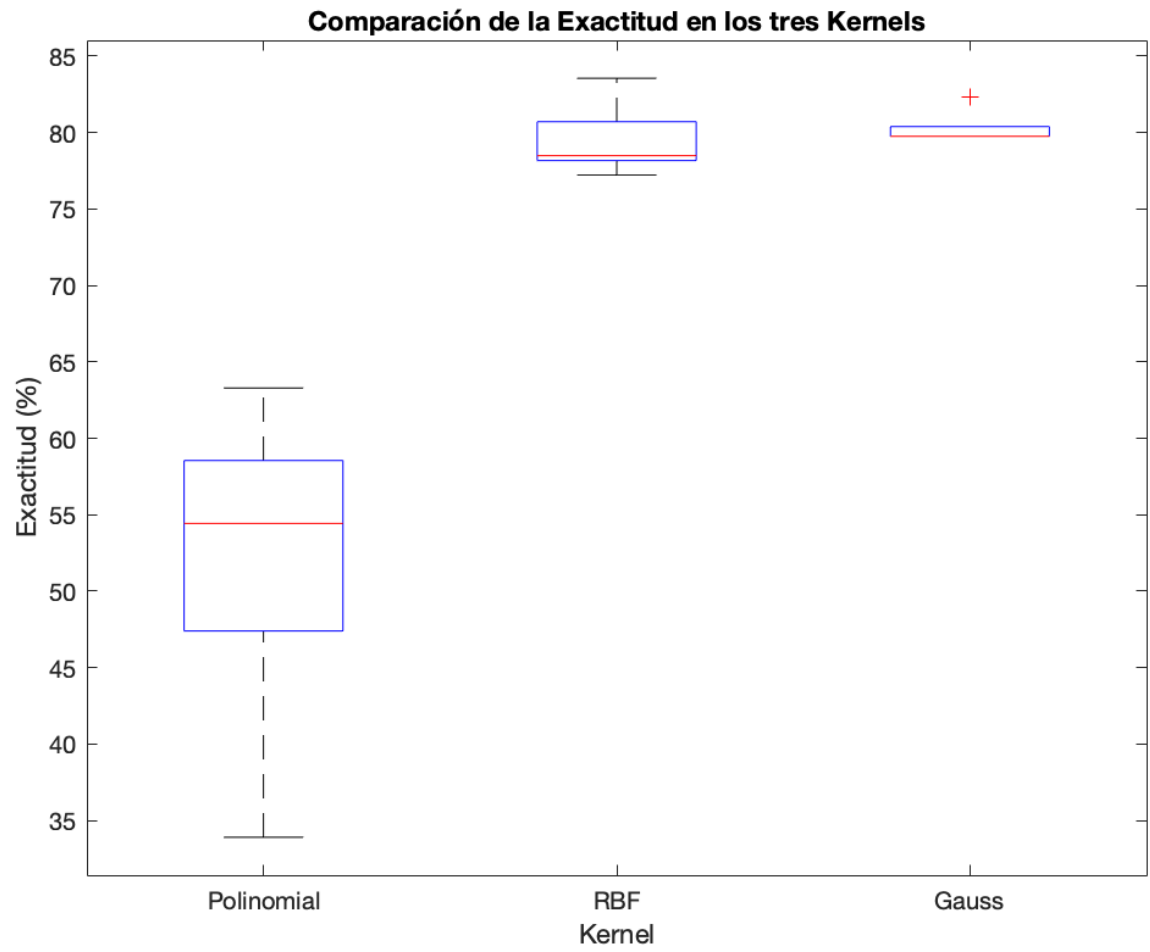
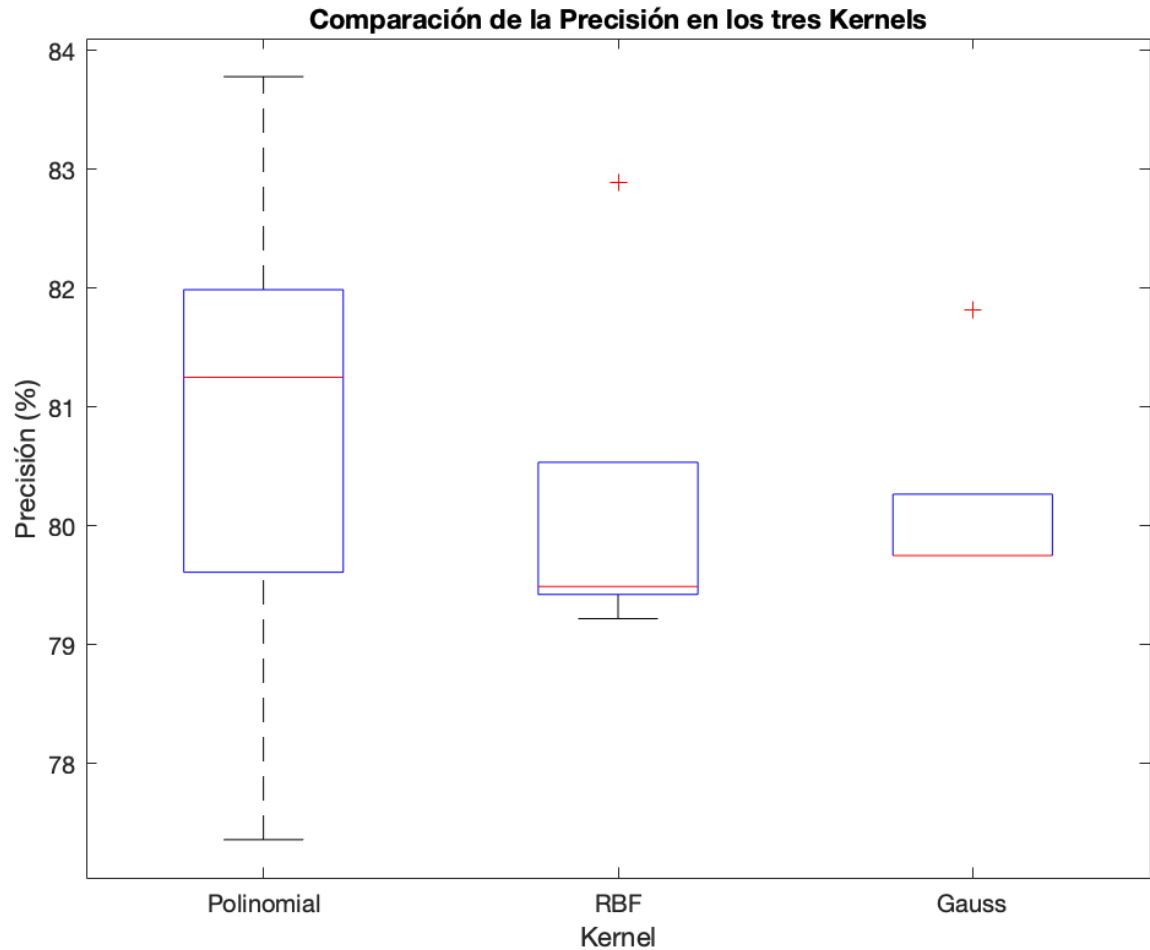


Gráfico 12. Gráfico de caja de la distribución de los resultados de Precisión en cada Kernel, de cada validación cruzada.



Después del análisis previo, se evidencia que el Kernel polinomial presenta resultados mejor asimilados en relación con las tres características escogidas para el entrenamiento de la SVM. Del mismo modo, el Kernel gaussiano, presenta resultados similares para todas las validaciones cruzadas (79.75% de precisión y 79.75% de exactitud), pero son métricas que demuestran poca linealidad entre los patrones y se llega a un sobreajuste de las necesidades objetivas de este proyecto, lo cual no es favorable en cuanto a resultados esperados.

Esto demuestra que en general, los datos presentan una complejidad para ser discernidos y clasificados, pero el Kernel RBF muestra resultados más equilibrados, por lo que, es un punto de partida para la evaluación de señales electrocardiográficas, cuando se está en la búsqueda de identificar señales enfermas que tiendan a posibles MI.

El F1-Score es una métrica que indica un rendimiento aceptable del modelo de clasificación, dado que combina la precisión con el radio de verdaderos positivos. Se encuentra un rendimiento máximo del 90.65% en el kernel RBF durante la validacion cruzada 1 y un rendimiento bajo máximo del 32.91% en la validación cruzada 3, con el kernel polinomial.

7. CONCLUSIONES

El preprocesamiento de las señales ECG es una tarea esencial en la clasificación de arritmias producidas por un infarto de miocardio. El ruido en las señales puede afectar la identificación de diferentes ondas y segmentos, lo que dificulta la extracción de características.

El algoritmo pan Tompkins es una herramienta útil para la correcta identificación de los complejos QRS, lo que ayuda en la búsqueda de otras ondas de interés en un electrocardiograma y ayuda también en la definición de características para la SVM.

Las características de amplitud que se utilizan para encontrar evidencias de infarto de miocardio no tienen patrones lineales y no se pueden capturar bien con un Kernel lineal o gaussiano y el Kernel polinomial podría necesitar de un ajuste de los hiperparámetros. Hay mayor adaptabilidad a las características cuando se utiliza el Kernel RBF.

Para la extracción de características de una arritmia producida por un infarto de miocardio, es importante utilizar una gran cantidad de datos. Esto puede ayudar a obtener resultados más precisos y confiables. Además, se debe validar el tipo de anomalía a investigar, ya que se evidencian resultados variables al utilizar señales con daños y antecedentes en diferentes partes del corazón.

Las características utilizadas para la extracción de características muestran que pueden llegar a ser muy similares para el discernimiento entre clases, a pesar de que se acondicionan las señales para su posterior uso en clasificación con Máquinas de Soporte Vectorial (SVM).

El kernel gaussiano, el cual es similar al kernel lineal, es muy poco útil para lograr una clasificación de señales con y sin arritmia, ya que tiende a esperar datos con características más separables y las comúnmente utilizadas para validar si un paciente tiene o no una arritmia, tienden a ser muy similares entre sí.

Al ser señales con una dispersión poco clara por la misma similitud entre las características, el F1-Score combina las medidas de precisión y de exactitud, lo cual permite asumir que importa de igual forma los resultados positivos, sean valores predictivos negativos o positivos, en señales ECG.

8. RECOMENDACIONES

Si bien se hace uso el algoritmo de Pan-Tompkins para identificar los complejos QRS, existen otros algoritmos que pueden ser utilizados para mejorar la calidad de la señal y reducir el ruido, lo que puede aumentar la precisión del modelo desde el preprocesamiento de las señales. Esto dependerá de las ondas y complejos que aspire el programador en encontrar.

La cantidad de datos de entrenamiento es importante para mejorar la precisión y sensibilidad del modelo de clasificación. Sería recomendable recolectar más señales de ECG de pacientes con diferentes tipos de antecedentes de infarto, intentar validar las Máquinas de Vectores de Soporte con cada tipo y contrastar con aquellas señales de control saludable para aumentar el conjunto de datos y mejorar el desempeño del modelo. Asimismo, validar la información con diferentes tipos de aprendizaje automático que permitan diferenciar entre clases similares.

Debido a los patrones no lineales que se evidencian al evaluar características de amplitud para el entrenamiento de una SVM para detección de infartos de miocardio, se recomienda evaluar las características con otros modelos de clasificación actuales.

La colaboración con expertos en cardiología y electrofisiología puede ser beneficiosa para mejorar la comprensión de las señales de ECG y la definición de características relevantes para la clasificación de las arritmias que pueden llegar a conducir a un infarto miocárdico. Es de vital importancia recalcar que no todas las arritmias están relacionadas con el infarto de miocardio, y que existen otras causas y factores de riesgo que pueden desencadenar este tipo de trastornos del ritmo cardíaco.

BIBLIOGRAFÍA

------. Categorías de arritmias. The Texas Heart Institute [página web]. [Consultado el 11, marzo, 2023]. Disponible en Internet: <<https://www.texasheart.org/heart-health/heart-information-center/topics/categorias-de-arritmias/>>.

¿QUÉ ES una arritmia? [Anónimo]. NHLBI, NIH [página web]. (24, marzo, 2022). [Consultado el 3, marzo, 2023]. Disponible en Internet: <<https://www.nhlbi.nih.gov/es/salud/arritmias>>.

ALGORITMO DE Pan Tompkins (QRS) [Anónimo]. Google Docs [página web]. [Consultado el 1, abril, 2023]. Disponible en Internet: <https://docs.google.com/document/d/1Wxa7u0gtGcF_A-ayon7RUWa0mI0yvqcFW05APOE9vD8/preview?hgd=1>.

BETANCOURT, GUSTAVO A. LAS MÁQUINAS DE SOPORTE VECTORIAL (SVMs). En: Scientia Et Technica [en línea]. 2005. vol. 1, no. 27 [consultado el 10, mayo, 2022], p. 6. Disponible en Internet: <<https://doi.org/10.22517/23447214.6895>>. ISSN 0122-1701.

DATA SCIENCE TEAM. Validación cruzada K-Fold. Data science [página web]. (2020). [Consultado el 26, marzo, 2023]. Disponible en Internet: <<https://datascience.eu/es/aprendizaje-automatico/validacion-cruzada-de-k-fold/>>.

EDITORIAL TEAM. How to evaluate you model using the Confusion Matrix. Towards AI [página web]. (23, febrero, 2021). [Consultado el 18, marzo, 2023]. Disponible en Internet: <<https://towardsai.net/p/data-science/how-to-evaluate-you-model-using-the-confusion-matrix>>.

ENFERMEDADES CARDIOVASCULARES [Anónimo]. World Health Organization (WHO) [página web]. [Consultado el 10, mayo, 2023]. Disponible en Internet: <https://www.who.int/es/health-topics/cardiovascular-diseases#tab=tab_1>.

ESTERHUIZEN, Dewald. C# how to: calculating gaussian kernels. Software by Default [página web]. (8, junio, 2013). [Consultado el 11, abril, 2023]. Disponible en Internet: <<https://softwarebydefault.com/2013/06/08/calculating-gaussian-kernels/>>.

GAJARE, Amogh y DEY, Hrishikesh. MATLAB-based ECG R-peak Detection and Signal Classification using Deep Learning Approach. En: 2021 IEEE Bombay Section Signature Conference (IBSSC) [en línea] (18-20, noviembre, 2021: Gwalior, India). [s.l.]: IEEE, 2021 [consultado el 3, marzo, 2023]. Disponible en Internet: <<https://doi.org/10.1109/ibssc53889.2021.9673301>>.

GOLDBERGER, A., et al. PTB Diagnostic ECG Database. PhysioNet [página web]. (25, septiembre, 2004). Disponible en Internet: <<https://physionet.org/content/ptbdb/1.0.0/>>.

HERNÁNDEZ CASILLAS, Andrea. Estudio de la detección y clasificación de la transmuralidad del infarto de miocardio mediante análisis de imágenes de resonancia magnética cardíaca y algoritmos de aprendizaje máquina [en línea]. Tesis de Máster. Valencia: Universitat Politècnica de València, 2020 [consultado el 10, mayo, 2022]. 118 p. Disponible en Repositorio institucional UPV: <<http://hdl.handle.net/10251/174516>>.

IBM DOCUMENTATION [Anónimo]. IBM - Deutschland | IBM [página web]. (17, agosto, 2021). [Consultado el 18, marzo, 2023]. Disponible en Internet: <<https://www.ibm.com/docs/es/spss-modeler/saas?topic=networks-neural-model>>.

INTERVALOS Y Segmentos del Electrocardiograma [Anónimo]. My EKG, La web del Electrocardiograma [página web]. [Consultado el 19, mayo, 2022]. Disponible en Internet: <<https://www.my-ekg.com/generalidades-ekg/intervalos-segmentos-ekg.html>>.

KATHLEEN, Gaines. How To Read An Electrocardiogram (EKG/ECG). Nurse.org [página web]. (9, abril, 2022). [Consultado el 4, marzo, 2023]. Disponible en Internet: <<https://nurse.org/articles/how-to-read-an-ECG-or-EKG-electrocardiogram/>>.

KELDENICH, Tom. Recall, precision, F1 score - simple metric explanation machine learning. Inside Machine Learning [página web]. (2, septiembre, 2001). [Consultado el 11, marzo, 2023]. Disponible en Internet: <<https://inside-machinelearning.com/en/recall-precision-f1-score-simple-metric-explanation-machine-learning/>>.

KOHLI, N. y VERMA, NK. Arrhythmia classification using SVM with selected features. En: International Journal of Engineering, Science and Technology [en línea]. 17, agosto, 2012. vol. 3, no. 8 [consultado el 3, marzo, 2023]. Disponible en Internet: <<https://doi.org/10.4314/ijest.v3i8.10>>. ISSN 2141-2839.

LI, Yaoguang y CUI, Wei. Identifying the mislabeled training samples of ECG signals using machine learning. En: Biomedical Signal Processing and Control [en línea]. Enero, 2019. vol. 47 [consultado el 4, marzo, 2023], p. 168-176. Disponible en Internet: <<https://doi.org/10.1016/j.bspc.2018.08.026>>. ISSN 1746-8094.

LIU, Feifei, *et al.* Performance analysis of ten common QRS detectors on different ECG application cases. En: Journal of Healthcare Engineering [en línea]. 2018. vol. 2018 [consultado el 11, marzo, 2023], p. 1-8. Disponible en Internet: <<https://doi.org/10.1155/2018/9050812>>. ISSN 2040-2309.

LORENZO ROMERO, José. Análisis de señales electrocardiográficas usando técnicas de procesamiento digital [en línea]. Trabajo de grado. [s.l.]: Universitat Oberta de Catalunya, 2015 [consultado el 3, marzo, 2023]. 84 p. Disponible en Repositori Institucional (O2): <<http://hdl.handle.net/10609/40186>>.

MINISTERIO DE SALUD Y PROTECCIÓN SOCIAL DIRECCIÓN DE EPIDEMIOLOGÍA Y DEMOGRAFÍA. Mortalidad en Colombia. Periodo 2020 a 2021: Medición de la Mortalidad por todas las causas y Covid-19 [en línea]. Bogotá: [s.n.], 2022. 45 p. Informe. Disponible en Internet: <<https://www.minsalud.gov.co/sites/rid/Lists/BibliotecaDigital/RIDE/VS/ED/GCFI/mortalidad-colombia-periodo-2020-2021.pdf>>.

NICHOLSON, Chris V. Evaluation metrics for machine learning - accuracy, precision, recall, and F1 defined. Pathmind [página web]. [Consultado el 11, marzo, 2023]. Disponible en Internet: <<https://wiki.pathmind.com/accuracy-precision-recall-f1>>.

PAN, Jiapu y TOMPKINS, Willis J. A real-time QRS detection algorithm. En: IEEE Transactions on Biomedical Engineering [en línea]. Marzo, 1985. BME-32, no. 3 [consultado el 11, marzo, 2023], p. 230-236. Disponible en Internet: <<https://doi.org/10.1109/tbme.1985.325532>>. ISSN 0018-9294.

PARO CARDÍACO súbito [Anónimo]. The Texas Heart Institute [página web]. [Consultado el 19, mayo, 2022]. Disponible en Internet: <<https://www.texasheart.org/heart-health/heart-information-center/topics/paro-cardiaco-subito/>>.

PRIETO-SOLÍS, José A.; BENITO, Natividad y MARTÍN-DURÁN, Rafael. Diagnóstico electrocardiográfico de la obstrucción del tronco coronario izquierdo mediante el análisis vectorial del segmento ST y el complejo QRS. *En*: Revista Española de Cardiología [en línea]. Febrero, 2008. vol. 61, no. 2 [consultado el 4, marzo, 2023], p. 137-145. Disponible en Internet: <<https://doi.org/10.1157/13116201>>. ISSN 0300-8932.

QIAN ZHENG, *et al.* A novel multi-resolution SVM (MR-SVM) algorithm to detect ECG signal anomaly in WE-CARE project. *En*: 2013 ISSNIP Biosignals and Biorobotics Conference: Biosignals and Robotics for Better and Safer Living (BRC) [en línea] (18-20, febrero, 2013: Rio de Janeiro). [s.l.]: IEEE, 2013 [consultado el 3, marzo, 2023]. Disponible en Internet: <<https://doi.org/10.1109/brc.2013.6487453>>. ISBN 9781467330251.

SIDDHARDHAN. 7.3.3. support vector machine - kernels [video]. YouTube. (29, octubre, 2021). [Consultado el 18, marzo, 2023]. 19:51 min. Disponible en Internet: <<https://www.youtube.com/watch?v=WLYFhROVM18>>.

TEIXEIRA, C. A., *et al.* EPILAB: A software package for studies on the prediction of epileptic seizures. *En*: Journal of Neuroscience Methods [en línea]. Septiembre, 2011. vol. 200, no. 2 [consultado el 4, marzo, 2023], p. 257-271. Disponible en Internet: <<https://doi.org/10.1016/j.jneumeth.2011.07.002>>. ISSN 0165-0270.

THE TEXAS HEART INSTITUTE. Anatomía del corazón. The Texas Heart Institute [página web]. [Consultado el 19, marzo, 2023]. Disponible en Internet: <<https://www.texasheart.org/heart-health/heart-information-center/topics/anatomia-del-corazon/>>.

THYGESEN, Kristian, *et al.* Fourth universal definition of myocardial infarction (2018). *En*: Circulation [en línea]. 13, noviembre, 2018. vol. 138, no. 20 [consultado el 5, junio, 2023]. Disponible en Internet: <<https://doi.org/10.1161/cir.0000000000000617>>. ISSN 1524-4539.

VALORACIÓN DEL segmento ST [Anónimo]. My EKG, La web del Electrocardiograma [página web]. [Consultado el 1, mayo, 2023]. Disponible en Internet: <<https://www.my-ekg.com/como-leer-ekg/segmento-st.html>>.

VENKATESAN, C., *et al.* ECG signal preprocessing and SVM classifier-based abnormality detection in remote healthcare applications. En: IEEE Access [en línea]. 2018. vol. 6 [consultado el 11, marzo, 2023], p. 9767-9773. Disponible en Internet: <<https://doi.org/10.1109/access.2018.2794346>>. ISSN 2169-3536.

VILLAMIL-MUNÉVAR, Alejandro; ALVARADO, José y OLAYA-SÁNCHEZ, Alejandro. Onda J, predictor de muerte súbita. En: Revista Colombiana de Cardiología [en línea]. Mayo, 2016. vol. 23, no. 3 [consultado el 11, marzo, 2023], p. 168-171. Disponible en Internet: <<https://doi.org/10.1016/j.rccar.2015.07.007>>. ISSN 0120-5633.