

UNIVERSIDAD AUTÓNOMA DE BAJA
CALIFORNIA
FACULTAD DE CIENCIAS DE LA INGENIERÍA Y TECNOLOGÍA



CLASIFICACIÓN AUTOMATIZADA DE SEÑALES
ELECTROCARDIOGRÁFICAS EN MUJERES EN EL
RANGO DE EDAD ENTRE 40 Y 60 AÑOS

TESIS

Que para cubrir parcialmente los requisitos para obtener el grado de:

BIOINGENIERA

Presenta:

Joana Florescano Olgún

Directora de Tesis:

Dra. Daniela Mercedes Martínez Plata

Marzo 2025

TESIS DEFENDIDA POR

Joana Florescano Olgún

Y APROBADA POR EL SIGUIENTE COMITÉ:

Dra. Daniela Mercedes Martínez Plata
Directora de tesis

Dra. Norma Alicia Barboza Tello

Miembro del comité

Dra. Irma Uriarte Ramírez

Miembro del comité

Dr. Paul Medina Castro

Miembro del comité

Resumen

En este proyecto de tesis, se propone el desarrollo del algoritmo titulado XCorr-In, un algoritmo clasificador basado en los índices de correlación cruzada y el análisis de características morfológicas y temporales de señales electrocardiográficas (ECG) en una población de pacientes femeninas en el rango de edad de 40 a 60 años. Con el objetivo de reducir la tasa de falsos positivos y falsos negativos en la detección de arritmias cardíacas. Los resultados muestran que la integración de criterios de decisión basados en el análisis paramétrico de los índices de correlación, en comparación con los criterios de decisión basados únicamente en su análisis estadístico descriptivo permite reducir la tasa de falsos positivos y falsos negativos. La estrategia de $\beta_{\hat{p},\gamma=0,01}$ presenta un rendimiento sobresaliente, con una sensibilidad (\hat{S}) de 0.9914 y una especificidad (\hat{E}) de 0.9084. Esta estrategia logra reducir significativamente las tasas de falsos positivos (9,16 %) y falsos negativos (0,86 %), con un error total de 4,61 %. Este rendimiento es especialmente relevante en entornos clínicos, donde una alta sensibilidad garantiza que la mayoría de los pacientes con arritmias sean correctamente diagnosticados y tratados. Además, el algoritmo se distingue por su eficiencia computacional, al no requerir grandes volúmenes de datos para su entrenamiento, lo que lo convierte en una herramienta accesible y eficaz para la clasificación rápida de señales ECG.

*A todas las mujeres científicas que han existido a lo largo de la historia,
a aquellas que con su esfuerzo y pasión desafiaron los límites impuestos,
a las que nunca recibieron el reconocimiento que merecían,
a las que tuvieron que luchar el doble para demostrar su valía,
y, sobre todo, a aquellas que soñaron con hacer ciencia
pero a quienes la desigualdad les arrebató la oportunidad.
Gracias por abrir camino con su valentía,
por no rendirse a pesar de los obstáculos,
por sembrar la curiosidad y el conocimiento para que hoy podamos estar
aquí, siendo parte de la ciencia en libertad.*

AGRADECIMIENTOS

Al principio, no tenía claro cómo decidir a quién nombrar en mis agradecimientos ni cómo empezar. Pero, al reflexionar sobre ello, me di cuenta de que esa misma duda es un reflejo de lo afortunada que soy: tengo a tantas personas a quienes agradecer, y eso es un privilegio inmenso.

En primera instancia, quiero agradecer a mi mamá, quien siempre me ha hecho tener confianza en mí y en cada paso que doy siempre está ella sosteniéndome y mi hermana Alexa, quien siempre me da su apoyo incondicional.

A mi directora de tesis, la Dra. Daniela Martínez, por quien siento una gran admiración. A lo largo de los años en que trabajamos juntas, me brindó no solo un profundo conocimiento académico, sino también un gran ejemplo de cómo ser una excelente maestra, científica y persona.

A mi extensa familia quienes siempre están a mi lado brindándome su apoyo en todas mis metas y sueños: a mis primas Oksana e Irina y mis tíos Armando y Vera. Mis tíos Guadalupe y Lucina y mi primo Rubén, mi emotional support Cristina y Jaime, mis padrinos Mónica y Alejandro y mis primos Natalee y Matthew.

Y por último, pero no menos importante, a mis amistades quienes me han acompañado en este proceso, a José quien siempre me escucha, me apoya y juntos alcanzamos nuestras metas, a mis amigas y amigos: Ingrid, Yashojara, Marineé, Fernanda, Juan Camilo y Lua que me acompañaron por mi camino en la universidad.

Índice general

1. Introducción	1
1.1. Justificación	3
1.2. Elementos que dirigen el proyecto de investigación	9
1.2.1. Objetivos	9
1.2.2. Hipótesis	9
1.2.3. Pregunta de investigación	10
1.3. Organización del documento	10
2. Marco teórico	11
2.1. Fundamentos fisiológicos del corazón	11
2.1.1. Fisiología cardíaca	11
2.1.2. Sistema de conducción del corazón	12
2.2. El electrocardiograma	13
2.2.1. Componentes e intervalos del ECG	13
2.2.2. Objetivos de la realización de un electrocardiograma	15
2.2.3. Interpretación del electrocardiograma	16
2.2.4. Diferencias fundamentales en el ECG entre hombres y mujeres	19
2.3. Herramientas para el análisis automatizado de señales electrocardiográficas	21
2.3.1. Herramientas para el análisis y clasificación de señales ECG de acuerdo a la aplicación	22
2.3.2. Herramientas para el análisis y clasificación de señales ECG de acuerdo a la técnica	24
2.3.3. Análisis del ECG: Comparación entre la interpretación humana y la computarizada	28

3. XCorr-In: Clasificador de señales electrocardiográficas basado en los índices de correlación cruzada	31
3.1. Identificación y caracterización de la base de datos	32
3.1.1. Descripción de los datos	33
3.1.2. Integración de los datos	34
3.1.3. Extracción de grupo categórico	35
3.2. Diseño de la señal de referencia	35
3.2.1. Cálculo del tamaño de muestra	37
3.2.2. Selección de la muestra	38
3.2.3. Localización de ondas R	39
3.2.4. Segmentación de la señal	40
3.2.5. Sincronización de las señales segmentadas	42
3.2.6. Promediado de las señales sincronizadas	43
3.3. Algoritmo XCorr-In	44
3.3.1. Correlador	46
3.3.2. Módulo de análisis paramétrico	49
3.3.3. Comparador	59
3.3.4. Integrador de decisiones: Clasificador	63
3.4. Evaluación	64
3.4.1. Modelos para el cálculo de las métricas de efectividad	65
4. Análisis de resultados	68
4.1. Propuesta de clasificación 1: Medidas estadísticas absolutas	69
4.2. Propuesta de clasificación 2: Medidas estadísticas relativas	77
4.3. Discusión de los resultados	79
5. Conclusiones y trabajo a futuro	83
5.1. Conclusiones y contribuciones	83
5.1.1. Contribuciones académicas	88
5.2. Trabajo a futuro	89
Appendices	91

Índice de figuras

1.1.	10 principales causas de muerte, según sexo. [2]	2
2.1.	Representación esquemática de los potenciales de acción registrados en diversas zonas cardíacas y su correlación con el electrocardiograma.[21]	12
2.2.	Representación de dos latidos cardíacos consecutivos en el electrocardiograma. [20]	13
2.3.	Diferentes ondas del electrocardiograma [20]	14
2.4.	Electrocardiograma normal de una mujer de 48 años. [22]	16
2.5.	Papel de ECG. [24]	18
3.1.	Diagrama de flujo de la metodología utilizada para el diseño del algoritmo XCorr-In.	32
3.2.	Diagrama de flujo de la metodología utilizada para el diseño y construcción de la señal ECG de referencia	37
3.3.	Histograma de las edades de la población de pacientes obtenida a partir de la categorización de la base de datos original.	39
3.4.	Representación de la segmentación de la señal ECG en diferentes puntos	41
3.5.	Segmentación de 500 muestras de la señal ECG tomando como referencia la onda R	42
3.6.	Familia de señales ECG para la construcción de la señal de referencia	43
3.7.	Señal ECG de referencia para el algoritmo de clasificación	44
3.8.	Diagrama de proceso para el flujo de información considerado en el diseño del algoritmo XCorr-In	45
3.9.	Representación gráfica del proceso para la obtención de la función de correlación cruzada.	47

3.10. Función de correlación cruzada entre una señal clasificada con ritmo sinusal (azul) y una señal clasificada con taquicardia (rojo).	48
3.11. Distribución de los índices de correlación entre la señal $x_{ref}(t)$ y cuatro señales ECG con ritmo sinusal.	50
3.12. Distribución de los índices de correlación entre la señal $x_{ref}(t)$ y cuatro señales ECG con taquicardia sinusal.	51
3.13. Distribución de la media de los índices de correlación para las 45 señales de la muestra con la señal $x_{ref}(t)$	52
3.14. Metodología para calcular el tiempo de duración del complejo QRS.	61
3.15. Metodología para calcular el tiempo de duración de la onda P.	62
3.16. Metodología para calcular el tiempo de duración del intervalo PR.	62
4.1. Métricas de calidad para el algoritmo 1 en función del valor de α , considerando la estrategia β_μ	72
4.2. Análisis del comportamiento de la sensibilidad y la especificidad en función de los deciles, considerando que $\alpha = 0,6$	74
4.3. Análisis del comportamiento de la sensibilidad y la especificidad en función de los deciles, considerando que $\alpha = 0,7$	74
4.4. Análisis del comportamiento de la sensibilidad y la especificidad en función de los deciles, considerando que $\alpha = 0,8$	75
4.5. Métricas de calidad para el algoritmo 1 en función del valor de α , considerando los valores correspondientes a β_{d_1}	76
4.6. Comparación de las métricas de calidad para el algoritmo 1 en función del valor de α , considerando el valor de β_μ y de β_{d_1}	76
4.7. Comparación de \hat{E} y \hat{S} para las evaluaciones con los diferentes niveles de significancia.	80
4.8. Comparación de las métricas de calidad para el algoritmo 1 en función del valor de α , considerando el valor de β_μ , de β_{d_1} y $\beta_{\hat{p},\gamma=0,01}$	81

Índice de tablas

2.1. Valores normales para las duraciones de las ondas y los intervalos en mujeres [22]	21
3.1. Parámetros de adquisición para cada señal ECG en la base de datos	35
3.2. Base de datos categorizada para el grupo de pacientes femeninos de entre 40 a 60 años	36
3.3. Valores considerados para determinar el tamaño de muestra necesario para el diseño de $x_{ref}(t)$	38
3.4. Medidas descriptivas de los vectores $r[m]$ al correlacionar las 45 señales de la muestra n con la señal $x_{ref}(t)$	53
3.5. Características fisiológicas para extraer de la señal $x(t)$ y sus valores de referencia para el enfoque de medidas estadísticas absolutas.	59
3.6. Características fisiológicas para extraer de la señal $x(t)$ y sus valores de referencia para el enfoque de medidas estadísticas relativas.	60
4.1. Correspondencia de los valores para β_μ obtenidos para cada valor de α en el intervalo evaluado.	71
4.2. Deciles obtenidos para los distintos valores de α , a partir del vector de índices de correlación generado para las 45 señales ECG seleccionadas en la muestra definida en la sección 3.2.2.	73
4.3. Valor del estimador puntual para la proporción poblacional (\hat{p}) correspondiente a cada valor de α en el intervalo evaluado.	78
4.4. Configuraciones óptimas para los tres criterios de decisión propuestos.	80

1.	Métricas de calidad para el Algoritmo 1 en función del valor de α considerando el valor de β_μ	92
2.	Deciles obtenidos para el valor de $\alpha = 0,5$ y las métricas de calidad considerando el valor de β_{d_i}	93
3.	Deciles obtenidos para el valor de $\alpha = 0,55$ y las métricas de calidad considerando el valor de β_{d_i}	94
4.	Deciles obtenidos para el valor de $\alpha = 0,6$ y las métricas de calidad considerando el valor de β_{d_i}	95
5.	Deciles obtenidos para el valor de $\alpha = 0,65$ y las métricas de calidad considerando el valor de β_{d_i}	96
6.	Deciles obtenidos para el valor de $\alpha = 0,7$ y las métricas de calidad considerando el valor de β_{d_i}	97
7.	Deciles obtenidos para el valor de $\alpha = 0,75$ y las métricas de calidad considerando el valor de β_{d_i}	98
8.	Deciles obtenidos para el valor de $\alpha = 0,8$ y las métricas de calidad considerando el valor de β_{d_i}	99
9.	Deciles obtenidos para el valor de $\alpha = 0,85$ y las métricas de calidad considerando el valor de β_{d_i}	100
10.	Deciles obtenidos para el valor de $\alpha = 0,9$ y las métricas de calidad considerando el valor de β_{d_i}	101
11.	Deciles obtenidos para el valor de α correspondiente y las métricas de calidad considerando el valor de β_{d_i}	102
12.	Deciles obtenidos con valores fijos de β_{d_i} y métricas de calidad	103
13.	Métricas de calidad para el Algoritmo 1 en función del valor de α considerando el valor de β_{d_1}	104
14.	Valores de $\beta_{\hat{p}}$ para distintos α	105
15.	Métricas de calidad para el Algoritmo 1 en función del valor de α considerando el valor de $\beta_{\hat{p},\gamma=0,01}$	106
16.	Métricas de calidad para el Algoritmo 1 en función del valor de α considerando el valor de $\beta_{\hat{p},\gamma=0,05}$	107
17.	Métricas de calidad para el Algoritmo 1 en función del valor de α considerando el valor de $\beta_{\hat{p},\gamma=0,1}$	108
18.	Métricas de calidad para el Algoritmo 1 en función del valor de α considerando el valor de $\beta_{\hat{p},\gamma=0,15}$	109

Capítulo 1

Introducción

Las enfermedades cardiovasculares (ECV) son una de las principales preocupaciones de salud pública a nivel mundial, ya que constituyen la principal causa de morbilidad y mortalidad prevenible. Además, de acuerdo con estudios publicados en donde se analiza la tendencia epidemiológica de estos padecimientos, se espera que su incidencia se incremente durante las siguientes décadas.[1]

En América Latina, las ECV son la principal causa de muerte por enfermedades crónicas no transmisibles y, en México, desde 1990 las ECV se han consolidado como la primera causa de muerte. Más del 60 % de la población adulta en México presenta al menos algún factor de riesgo relacionado con estas enfermedades, tales como tabaquismo, obesidad y sobrepeso, hipertensión, diabetes o dislipidemia. Estos factores de riesgo, junto con el proceso de envejecimiento de la población, genera un incremento de enfermedades asociadas a la edad, entre las que destacan las ECV. Se prevé además que la mortalidad por estas causas incrementará en los próximos años. [1]

De acuerdo con el Instituto Nacional de Estadística y Geografía (INEGI) [2], en el primer semestre del año 2024 se registraron un total de 417,408 defunciones en México, lo que representa un aumento de 12,683 muertes respecto al año 2023. En este respecto, las enfermedades del corazón, la diabetes mellitus y los tumores malignos fueron las primeras tres causas de defunción a nivel nacional tanto para mujeres como para hombres. En la Figura 1.1 se detallan las 10 principales causas de defunción según el sexo, donde las enfermedades del corazón se encuentran en el primer lugar para ambos sexos, con un total de 100,710 defunciones, de las cuales 47,187 defunciones (46.40 %) corresponden a mujeres. Además, se observa un incremento de defunciones

por ECV para ambos sexos con respecto al año 2023, con un aumento particular de 1,143 muertes en mujeres del año 2023 al año 2024.

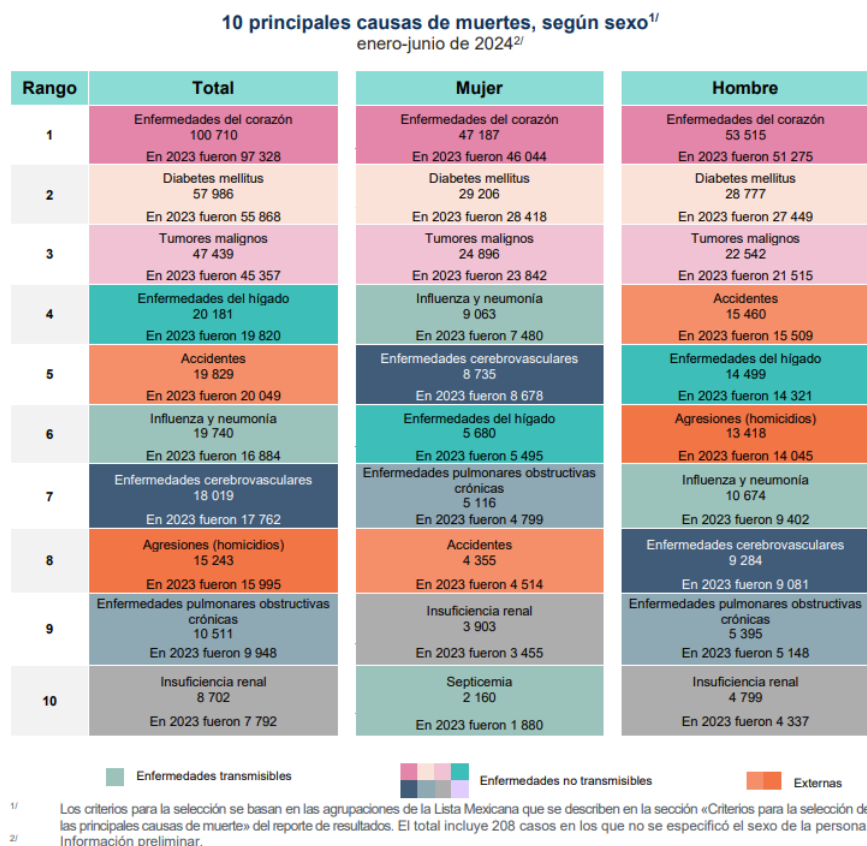


Figura 1.1: 10 principales causas de muerte, según sexo. [2]

La detección oportuna y seguimiento de las ECV impone retos importantes a los sistemas de salud pública, entre los que resalta la sobre-demanda de servicios de atención médica; además, las barreras geográficas, inconvenientes logísticos, costes económicos, dificultades de acceso a las clínicas y centros de salud, entre otros, obstaculizan el correcto seguimiento de los pacientes por parte del personal de salud. [3]

En los últimos años, ha existido una revolución digital acelerando el uso de la telemedicina, la salud digital y la práctica clínica asistida por tecnologías. Estas propuestas se basan en el uso de tecnologías de la información y la comunicación para brindar atención médica personalizada e incluso a la

distancia, permitiendo que los profesionales de la salud se auxilien de dispositivos tecnológicos para evaluar, diagnosticar y tratar a los pacientes. Estos enfoques basados en la telemedicina y salud digital brinda un nuevo panorama para mejorar la asistencia de los pacientes con ECV. [3] Dado que se prevé un aumento en la mortalidad por ECV en los próximos años, es fundamental aprovechar la integración de nuevas tecnologías informáticas y de ciencia de datos para reducir este impacto, priorizando la prevención y la detección de las ECV. Sin embargo, el desarrollo de dispositivos para la detección, clasificación y diagnóstico de ECV enfrenta diversos desafíos entre los que destacan: la confiabilidad de los algoritmos utilizados para los procesos de diagnóstico, el desarrollo de modelos que incorporen la variabilidad fisiológica de las distintas poblaciones, aspectos regulatorios para la validación y certificación de modelos de diagnóstico y tratamiento automatizado, entre otras.

1.1. Justificación

De acuerdo a la Figura 1.1, las ECV son la primera causa de muerte en mujeres; la cifra es tan impresionante que podemos resumir su impacto en un frase: *una mujer muere de ECV cada 5.5 minutos en México*. Entre las ECV más frecuentes para las mujeres se encuentran: la enfermedad del nodo sinusal, la taquicardia sinusal, la taquicardia supraventricular, la taquicardia ventricular idiopática del ventrículo derecho, y el síndrome QT largo congénito y adquirido; en cambio los hombres, presentan prevalencias de otro tipo de arritmias. [4]

Las diferencias entre sexos en cuanto a la incidencia de varios tipos de arritmias cardíacas es una realidad. Si bien las causas subyacentes a estas diferencias siguen siendo en muchos casos desconocidas, se ha propuesto que estas diferencias entre sexo se deben a 2 mecanismos: los efectos hormonales sobre la función de los canales iónicos y las diferencias en el tono autonómico, o bien una combinación de ambos mecanismos.

Riesgo de enfermedades cardiovasculares: Análisis desde una perspectiva de género

Históricamente, la investigación sobre las ECV se ha centrado en poblaciones masculinas, lo que ha traído como consecuencia la falta de conciencia

y atención sobre los riesgos de estas enfermedades en las mujeres, tanto en la investigación como en las políticas sanitarias y la práctica clínica [5]. Un análisis de datos de ClinicalTrials.gov [6], uno de los registros de ensayos más grandes del mundo, encontró que las mujeres representaban menos del 40 % de las personas inscritas en la investigación clínica sobre enfermedades cardíacas y accidentes cerebrovasculares. [7] Sin embargo, en las últimas dos décadas, se han realizado investigaciones más exhaustivas sobre las ECV desde una perspectiva de género. A pesar de ello, la implementación de estas investigaciones en la práctica clínica han avanzado de manera más lenta. Las ECV en las mujeres difieren de los hombres en varios aspectos, como son en los factores de riesgo de la enfermedad, su fisiopatología, el diagnóstico y tratamiento, así como factores sociales. Todos estos factores contribuyen a la morbilidad y mortalidad atribuidas a las ECV en las mujeres. [5]

Una de las barreras para la reducción de los factores de riesgo cardiovascular en las mujeres es una percepción errónea sobre el riesgo de padecer una ECV, siendo la creencia popular que los hombres son el grupo de mayor riesgo de morbilidad y mortalidad, y también la creencia de que el cáncer de mama es el principal riesgo para las mujeres. Esta percepción errónea limita la adopción de medidas preventivas y la atención a los factores de riesgo cardiovascular en las mujeres. Debido a esto, se debe aumentar la conciencia sobre la importancia de prevenir las ECV y reconocer las diferencias de género en la manifestación de estas enfermedades. [8]

Como se mencionó anteriormente, existen diferencias en la manifestación y el impacto de las ECV entre hombres y mujeres. Los factores de riesgo se dividen en dos grupos: los factores de riesgo tradicionales, comunes para ambos géneros, y factores de riesgo femeninos. Adicionalmente se pueden añadir factores de riesgo sociales y de comportamiento.

Factores de riesgo tradicionales

- Diabetes Mellitus.
- Hipertensión: Las mujeres desarrollan hipertensión más tarde que los hombres, volviéndose más prevalente en mujeres mayores en comparación con los hombres.
- Tabaquismo: Aunque la prevalencia del tabaquismo es mayor en hombres, el tabaquismo es un factor de riesgo más fuerte para las ECV en mujeres.

- **Inflamación:** La inflamación es un factor clave de las enfermedades macro y microvasculares en mujeres y hombres, y contribuye a la fisiopatología de la aterosclerosis, el desarrollo de síndromes cardiovasculares e insuficiencia cardíaca.
- **Depresión y estrés psicológico:** Mientras que la depresión es frecuente tanto entre las mujeres como entre los hombres ha demostrado tener una mayor contribución al desarrollo de ECV en mujeres.

Factores de riesgo femeninos

Además de los factores de riesgo comunes, las mujeres enfrentan factores adicionales que aumentan su vulnerabilidad a las ECV:

- Factores hormonales.
- Hipertensión y trastornos metabólicos de embarazo.
- Lactancia.
- Factores relacionados con la terapia del cáncer de mama.
- Enfermedad auto-inmune.

Factores sociales y de comportamiento

Los factores conductuales y sociales también contribuyen a las disparidades de género observadas en el diagnóstico y tratamiento de las ECV. De acuerdo con el estudio [9] la conciencia sobre los factores de riesgo de las ECV y su relevancia personal es baja en las mujeres. Esta falta de conciencia puede dificultar que las mujeres prioricen su salud cardiovascular o que realicen un monitoreo de sus factores de riesgo. Algunos factores que pueden ser un obstáculo para esta conciencia y prioridad y que influyen en estas diferencias de género pueden ser las responsabilidades familiares y la falta de tiempo. [5].

Asimismo, estudios como [10] sugieren que un nivel socioeconómico más bajo es otro factor de riesgo para el desarrollo de ECV en mujeres, lo que resalta la importancia de considerar el contexto socioeconómico al abordar la prevención y el tratamiento de las ECV en mujeres.

Diferencias de edad y su impacto en el desarrollo de enfermedades cardiovasculares

Se ha demostrado que los factores de riesgo para el desarrollo de ECV varían entre mujeres y hombres en diferentes grupos de edad. Aunque la edad es un factor de riesgo importante para ambos sexos, las mujeres tienden a desarrollar ECV aproximadamente 10 años después que los hombres. Algunos estudios han determinado que esta diferencia se deba a los cambios hormonales que ocurren en las mujeres después de la menopausia, cuando los niveles de estrógeno disminuyen, lo que puede aumentar el riesgo de desarrollar problemas cardiovasculares. [8]

Además, la edad a la que aparecen algunos factores de riesgo también es diferente entre sexos. A medida que envejecen, la cantidad de pacientes que desarrollan hipertensión es mayor en mujeres que en hombres, esta disparidad aumenta especialmente después de los 45 años. Los niveles totales de colesterol en las mujeres alcanzan su punto máximo entre 55 y 65 años, aproximadamente una década más tarde que en los hombres. La prevalencia de la diabetes aumenta para ambos, sin embargo, el riesgo de mortalidad por cardiopatía coronaria (CHD) asociado a la diabetes es mayor en las mujeres. [8]

Vulnerabilidad de las mujeres de 40 a 60 años

La poca representación de las mujeres en investigaciones y la falta de conciencia de las mujeres sobre las ECV, hace a las mujeres de entre 40 a 60 años un grupo vulnerable, ya que en este periodo convergen varios factores de riesgo. Por estas razones, este grupo de edad constituye el principal interés de este estudio, ya que es una etapa crucial para implementar estrategias preventivas y terapéuticas destinadas a disminuir el impacto de las ECV, así como identificar y abordar los factores de riesgo en este periodo puede representar una oportunidad clave para reducir la incidencia y la mortalidad asociadas a estas enfermedades.

En diversos estudios electrocardiográficos realizados hasta el momento se han encontrado diferencias entre ambos sexos respecto a la frecuencia cardíaca basal, la variabilidad de la frecuencia cardíaca, el intervalo QT y la duración y el voltaje del complejo QRS. Estas diferencias entre sexos tienen algunas implicaciones clínicas, sobre todo en lo que corresponde a las consideraciones terapéuticas y al tratamiento clínico de las arritmias en las

mujeres [11].

Estas diferencias en las características electrocardiográficas entre hombres y mujeres destacan la necesidad de desarrollar herramientas personalizadas que consideren las particularidades fisiológicas de cada sexo. En particular, los algoritmos de análisis y clasificación de señales electrocardiográficas deben adaptarse para garantizar diagnósticos precisos y efectivos en poblaciones específicas, como las mujeres en el rango de edad de 40 a 60 años.

La formación médica y la dependencia de la interpretación computarizada del ECG

El electrocardiograma (ECG), como técnica no invasiva, rápida y económica, es fundamental para diagnosticar ECV en contextos clínicos, urgencias y emergencias. Por ello, su interpretación forma parte esencial de la formación médica, incluyendo el conocimiento básico de la fisiopatología de las anomalías electrocardiográficas. [12]

No obstante, pocos estudios evalúan el nivel de conocimiento en interpretación de ECG por parte de los médicos. Un estudio con médicos residentes evidenció que consideraban insuficiente su formación académica y carecían de confianza en la interpretación de esta técnica. Debido a estas deficiencias, se han planteado distintas estrategias de mejora, como la interconsulta con Cardiología o el manejo de software de interpretación. Además, se ha demostrado la mejora en la interpretación a través de la experiencia en los primeros años de práctica clínica. [12]

En las últimas décadas, el avance en computación y procesamiento de datos ha impulsado el uso de inteligencia artificial (IA) en el ECG, desarrollando algoritmos para evaluar riesgos, realizar diagnósticos y apoyar la interpretación clínica. [13]. Investigaciones han demostrado que algunos algoritmos son capaces de identificar el ritmo [14] e incluso realizar una interpretación completa de 12 derivaciones [15].

Numerosos estudios han demostrado importantes inexactitudes y limitaciones de los programas informáticos de interpretación del ECG. Este es un problema importante, dada la dependencia de los médicos hacia estas herramientas y el impacto que puede tener en la atención al paciente. A pesar de los avances para mejorar la precisión de los algoritmos computarizados de interpretación del ECG, la interpretación final del ECG sigue requiriendo una revisión exhaustiva por parte del médico. [16] Además, incluso con los esfuerzos para estandarizar la interpretación del ECG en los Estados Unidos

y Europa, todavía no ha habido una norma internacionalmente aceptada que abarque todo para la interpretación computarizada del ECG. [17].

Aunque los algoritmos de interpretación computarizada del ECG se mejoran continuamente, su capacidad diagnóstica es limitada, la interpretación computarizada del ECG puede contribuir a las inexactitudes diagnósticas, que no solo afectan la precisión del diagnóstico, sino que también pueden generar un retraso en el inicio del tratamiento adecuado. [13] Por ejemplo, un estudio encontró una alta tasa de falsos negativos en el diagnóstico por ordenador de infarto de miocardio con elevación del segmento ST, lo que desaconseja la interpretación computarizada del ECG para tal uso. [18]

Un estudio con 2112 ECG seleccionados aleatoriamente mostró que el valor predictivo positivo de la interpretación computarizada del ritmo sinusal fue del 95 %, mientras que en ritmos no sinusales cayó al 53.5 % y resultó no interpretable en el 2 % de los casos, esto significa que, en ritmos no sinusales, casi la mitad de los diagnósticos generados por el algoritmo pueden ser incorrectos, lo cual pone en riesgo la toma de decisiones clínicas en escenarios complejos. [17] Esto subraya la necesidad de contar con profesionales capacitados para revisar y validar los resultados generados por los sistemas computarizados.

En este trabajo se presenta un panorama general de las principales tecnologías y técnicas utilizadas en la creación de algoritmos de clasificación de señales electrocardiográficas, con el objetivo de identificar una alternativa viable que coadyuve de manera confiable en los procesos de diagnóstico y tratamiento de ECV, particularmente arritmias cardíacas. A partir del análisis de diversas propuestas existentes, se propone XCorr-In, un clasificador de señales electrocardiográficas basado en los índices de correlación cruzada, así como en características morfológicas y temporales de los registros electrocardiográficos, como una alternativa viable para los procesos de diagnóstico inicial de arritmias, el cual no depende de la disponibilidad de datos para el entrenamiento de modelos de IA, ya que se basa en técnicas de procesamiento de señales y análisis estadístico.

Además, con el objetivo de contribuir en la investigación de ECV con perspectiva de género, el algoritmo propuesto en este trabajo está dirigido a pacientes femeninas de entre 40 y 60 años de edad, permitiendo una mejor caracterización del grupo y optimizando la exactitud en la detección de arritmias cardíacas.

1.2. Elementos que dirigen el proyecto de investigación

1.2.1. Objetivos

Objetivo general

Reducir la tasa de falsos positivos y falsos negativos en la detección de arritmias cardíacas al implementar un clasificador basado en el análisis de características morfológicas y temporales en pacientes femeninas en el rango de 40 a 60 años de edad.

Objetivos específicos

- Analizar las características fisiológicas del corazón y los patrones electrocardiográficos típicos de pacientes femeninos de 40 a 60 años para identificar los parámetros de referencia necesarios para la detección de anomalías.
- Seleccionar una base de datos que cumpla con las características suficientes y adecuadas para el desarrollo y evaluación de un algoritmo para la clasificación de señales electrocardiográficas.
- Diseñar e implementar un algoritmo de clasificación de señales electrocardiográficas basado en el análisis de características morfológicas y temporales.
- Evaluar y seleccionar el criterio óptimo de clasificación que permita reducir las tasas de falsos positivos y falsos negativos.

1.2.2. Hipótesis

La integración de criterios de decisión basados en el análisis paramétrico de los índices de correlación cruzada permite reducir la tasa de falsos positivos y falsos negativos en la detección de arritmias cardíacas de un clasificador basado en las características morfológicas y temporales de las señales electrocardiográficas, en comparación con los criterios de decisión basados solamente en su análisis estadístico descriptivo.

1.2.3. Pregunta de investigación

¿Cuál es el criterio de decisión adecuado para los índices de correlación que permita reducir la tasa de falsos positivos y falsos negativos en un clasificador basado en el análisis de características morfológicas y temporales en pacientes femeninas en el rango de 40 a 60 años de edad?

1.3. Organización del documento

El documento está dividido en cinco capítulos:

- Capítulo 1: Introducción. Se presenta el contexto general en el que se sitúa la problemática que se aborda en este proyecto de tesis. Adicionalmente, se presentan los elementos que dirigen del proyecto de investigación.
- Capítulo 2: Marco teórico. Se desarrollan los fundamentos teóricos de las señales electrocardiográficas desde la perspectiva fisiológica para sentar las bases que permitirán su análisis adecuado, así como identificar las características que deberán ser incorporadas en el diseño del algoritmo clasificador.
- Capítulo 3: XCorr-In: Clasificador de señales electrocardiográficas basado en los índices de correlación cruzada. Se describe con todo detalle el proceso de diseño e implementación del algoritmo clasificador propuesto, así como las métricas consideradas para su evaluación.
- Capítulo 4: Análisis de resultados. Se presentan y discuten las curvas de especificidad y sensibilidad obtenidas para los tres enfoques considerados para la implementación del algoritmo.
- Capítulo 5: Conclusiones y trabajo a futuro. Se exponen las contribuciones del proyecto de tesis y se plasma la reflexión a partir del análisis de los resultados presentados en el Capítulo 4. También se presenta una sección en la que se discuten las limitaciones de la propuesta y se definen rutas para el trabajo futuro

Capítulo 2

Marco teórico

2.1. Fundamentos fisiológicos del corazón

2.1.1. Fisiología cardíaca

El corazón es un órgano vital que se encuentra en el centro del sistema circulatorio. Este sistema se compone, además del corazón, de una red de arterias, venas y capilares, y su tarea es transportar la sangre desde y hacia todas las partes del cuerpo. La sangre transporta el oxígeno y los nutrientes que los órganos necesitan para funcionar correctamente y también transporta dióxido de carbono a los pulmones para que pueda ser exhalado [19].

El corazón se compone de dos aurículas y dos ventrículos. Los ventrículos son los que se encargan de impulsar la sangre que llega al corazón y las aurículas contribuyen al relleno óptimo de los ventrículos en cada latido. El movimiento de aurículas y ventrículos se hace de forma ordenada y coordinada en un ciclo que se repite con cada latido. A este ciclo se le denomina ciclo cardíaco.

El ciclo cardíaco presenta dos fases: diástole y sístole. La diástole es el período del ciclo en el cual los ventrículos están relajados y se están llenando de la sangre que luego tendrán que impulsar. La sístole es el período del ciclo en el cual los ventrículos se contraen y provocan la eyección de la sangre que contienen. Para que el corazón pueda cumplir su función debe poder tanto relajarse, para permitir su llenado, como contraerse, para provocar la eyección de la sangre. La contracción se produce gracias al sistema de conducción cardíaco, el cual permite que la orden de contracción llegue a todas las células musculares en la debida secuencia [20].

2.1.2. Sistema de conducción del corazón

El bombeo cardíaco de sangre a los distintos tejidos del organismo depende de la capacidad del corazón para generar rítmicamente impulsos eléctricos que se propagan de forma ordenada a todo el miocardio. La misión del sistema es funcionar ininterrumpidamente, con una fuerza y una frecuencia ¹ adecuadas a las necesidades del organismo. En condiciones fisiológicas, los impulsos cardíacos se originan en el nodo sinoauricular (SA), que genera entre 60 a 90 potenciales de acción por minuto. [21]

El nodo SA decide cuando debe contraerse el corazón. Llegado el momento, envía un impulso eléctrico a las aurículas y al nódulo auriculoventricular (AV). Este nodo actúa como filtro cuando se reciben más impulsos eléctricos de los necesarios; o envía sus propios impulsos eléctricos si no llega ningún impulso del nodo SA. Los impulsos que salen del nodo AV pasan a una red de conducción que distribuye el impulso eléctrico por los dos ventrículos: el haz de His y el sistema de Purkinje, los cuales a su vez lo distribuyen por los ventrículos. Los ventrículos responden a la onda de propagación contrayéndose de forma sincrónica. Estos elementos se muestran en la Figura 2.1. Todo este proceso no lleva más de 0.3 a 0.4 segundos [20].

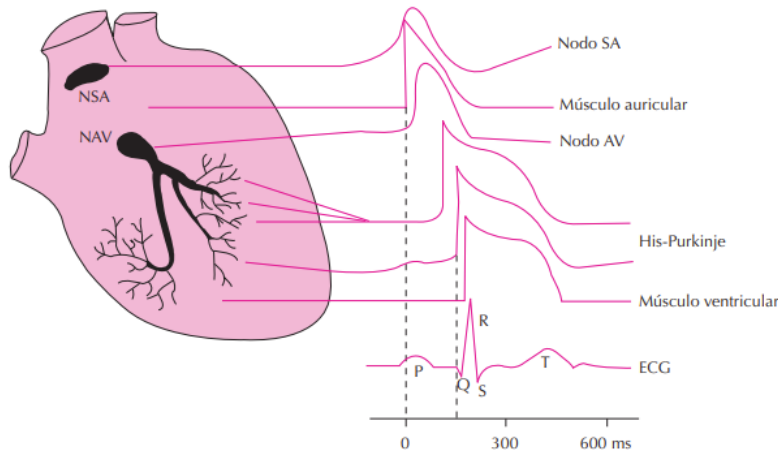


Figura 2.1: Representación esquemática de los potenciales de acción registrados en diversas zonas cardíacas y su correlación con el electrocardiograma.[21]

¹En este contexto se entiende por frecuencia al número de contracciones cardíacas por minuto.

2.2. El electrocardiograma

El flujo de las diversas corrientes eléctricas que producen el funcionamiento del corazón se detecta y representa mediante el electrocardiograma (ECG). El ECG es, entonces, el registro de las diferencias de potencial que ocurren entre varios puntos de la superficie del músculo cardíaco producidas por los fenómenos eléctricos que acompañan al latido cardíaco. [21]

El ECG presenta como referencia eléctrica la denominada línea isoeletrica o línea basal. La línea basal se identifica gráficamente como el trazo horizontal presente entre cada latido, como se muestra en la Figura 2.2 [20].

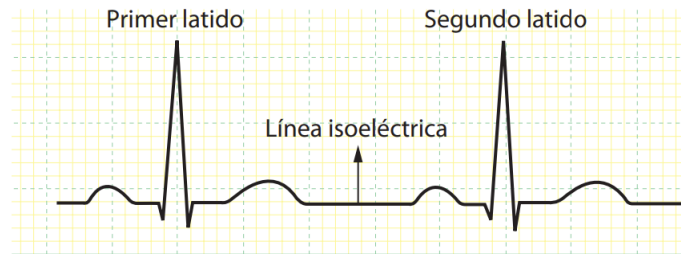


Figura 2.2: Representación de dos latidos cardíacos consecutivos en el electrocardiograma. [20]

Los latidos cardíacos quedan representados en el ECG por las oscilaciones respecto a la línea basal en forma de inflexiones, segmentos, ondas e intervalos, constituyendo una imagen característica que se repite con una determinada frecuencia. El eje vertical representa la magnitud de la diferencia de potencial, en milivolts (mV), en cada momento durante el latido cardíaco; mientras que el eje horizontal representa el tiempo, en segundos (s). La Figura 2.3 ilustra las denominaciones asignadas a las diversas ondas, intervalos y segmentos que componen un latido cardíaco [20].

2.2.1. Componentes e intervalos del ECG

En un ECG normal², cada complejo de ondas consta de una serie de inflexiones que alternan con la línea basal. Realizando la lectura de izquierda a derecha en el ECG, se distinguen la onda P, el segmento PR, el complejo

²En este contexto, se entiende por un ECG normal al registro correspondiente a la actividad eléctrica de un corazón sano.

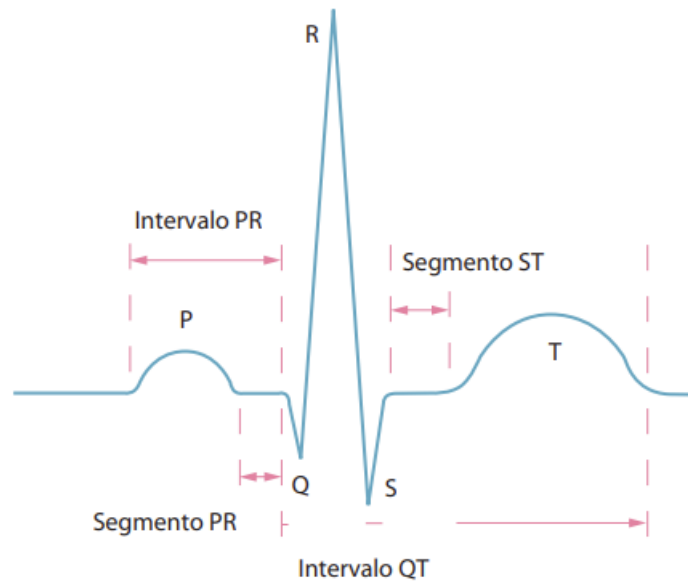


Figura 2.3: Diferentes ondas del electrocardiograma [20]

QRS, el segmento ST y finalmente la onda T, esto se representan en La Figura 2.3. A continuación se presenta una breve descripción de las componentes y segmentos del ECG [20].

Onda P

Es la primera deflexión hacia arriba que aparece en el ECG. Representa el momento en que las aurículas se están contrayendo y enviando sangre hacia los ventrículos.

Segmento PR

Es el tramo de la línea basal que se encuentra entre el final de la onda P y la siguiente deflexión que puede ser hacia arriba (positiva) o hacia abajo (negativa). Durante este período, las aurículas terminan de vaciarse y se produce una relativa desaceleración en la transmisión de la corriente eléctrica a través del corazón, justo antes del inicio de la contracción de los ventrículos.

Complejo QRS

Corresponde con el momento en que los ventrículos se contraen y expulsan su contenido sanguíneo. Consta de las ondas Q,R y S. La onda Q no siempre está presente. Se identifica por ser la primera deflexión negativa presente después del segmento PR. Toda deflexión positiva que aparezca después del segmento PR corresponde a la onda R. La onda S se observa como continuación directa de la onda R y comienza a partir del punto en que esta última, en su fase decreciente, se hace negativa.

Segmento ST

Es el trazado de la línea basal que se encuentra entre el final de la onda S y el comienzo de la onda T.

Onda T

Se inscribe a continuación del segmento ST. Consiste en una deflexión, normalmente positiva, que representa el momento en que el corazón se encuentra en un período de relajación, una vez que ha expulsado la sangre que se hallaba en los ventrículos.

2.2.2. Objetivos de la realización de un electrocardiograma

El ECG es una prueba diagnóstica asequible y segura que proporciona información sobre el estado del corazón. El ECG de una persona sana tiene un trazado característico y los cambios que se producen en el patrón de este trazado característico suelen asociarse con enfermedades cardíacas. Principalmente se utiliza para detectar trastornos del ritmo cardíaco y en el diagnóstico de situaciones que causan un aporte insuficiente de sangre al corazón, por ejemplo infarto de miocardio y angina de pecho. El ECG permite diferenciar el ritmo normal del corazón, denominado ritmo sinusal, de cualquier tipo de arritmia; en este sentido, es posible definir a una arritmia como cualquier ritmo diferente del ritmo cardíaco sinusal. Esto incluye los ritmos en los que el corazón late demasiado rápido, clasificados como taquicardia, o demasiado lento, denominados como bradicardia, así como los resultantes de una interrupción en la transmisión normal del impulso eléctrico a lo largo del corazón, denominados como bloqueos [20].

2.2. EL ELECTROCARDIOGRAMA

El procedimiento para realizar un ECG es relativamente sencillo. Se necesita un electrocardiógrafo, electrodos y un sistema de cables que transmitirán las corrientes detectadas por los electrodos hacia el electrocardiógrafo. El ECG habitual consta de 12 derivaciones: seis obtenidas en el plano frontal y seis en el plano horizontal. Tres de las derivaciones del plano frontal, identificadas como derivaciones estándar I, II, III, son derivaciones bipolares. Las otras tres derivaciones del plan frontal, también nombradas como derivaciones de miembros aVR, aVL y aVF, así como las seis derivaciones precordiales del plano horizontal, identificadas como V1 a V6, son monopolares [21]. La Figura 2.4 muestra las 12 derivaciones de un ECG normal.

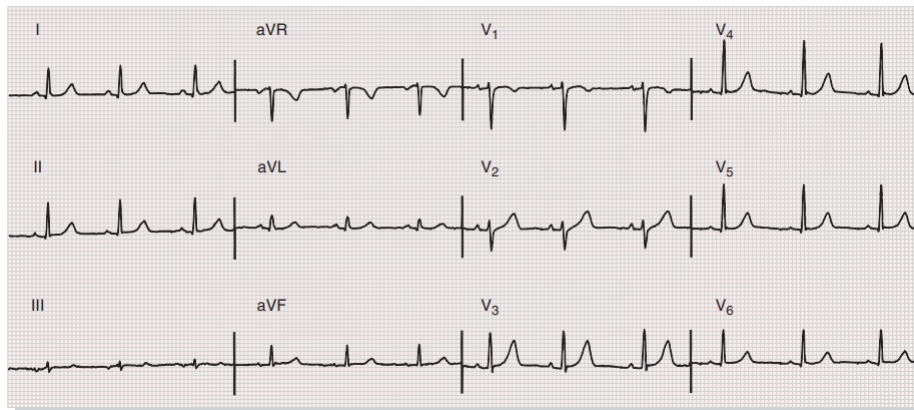


Figura 2.4: Electrocardiograma normal de una mujer de 48 años. [22]

2.2.3. Interpretación del electrocardiograma

Para interpretar correctamente la información que proporciona un ECG es necesario seguir una secuencia ordenada de pasos, en los que se examinan cinco elementos básicos:

- Ritmo
- Frecuencia cardíaca
- Eje cardíaco
- Análisis de ondas, segmentos e intervalos

- Anormalidades

En algunos casos, estos elementos se definen con base en constructos fisiológicos, es decir, cómo debería comportarse eléctricamente el corazón en condiciones normales. Sin embargo, el rango de los valores considerados como normales para estos elementos debe considerar la variabilidad interindividual, ya que lo que se considera normal puede variar significativamente entre personas a causa de factores como la edad, el sexo, el tipo de cuerpo, la orientación del corazón y la fisiología, diferencias demográficas, entre otros. Esto sugiere que un solo rango de valores normales para todos los sujetos puede no ser adecuado, ya que podría dar lugar a diagnósticos erróneos [21].

Papel de ECG

Los ECG se imprimen sobre papel milimetrado, en el cual el eje horizontal representa el tiempo y el eje vertical representa al voltaje. En la cuadrícula del papel, cada 5 mm las líneas de la cuadrícula se hacen más gruesas, quedando así marcados cuadrados grandes de 5 mm x 5 mm y cuadrados chicos de 1 mm x 1 mm como se muestra en la Figura 2.5.

Horizontalmente, el papel de registro corre a una velocidad constante de 25 mm/seg, por lo que cada milímetro equivale a 0.04 segundos o 40 milisegundos, y un cuadrado grande a 0.20 segundos o 20 milisegundos. Verticalmente el voltaje se representa mediante la calibración convencional, la cual establece que 1 cm de amplitud equivale a 1 mV, por lo que 1 mm es igual a 0.1 mV y un cuadrado grande de 5 mm a 0.5 mV. La calibración anteriormente descrita es la más utilizada; sin embargo se puede calibrar con valores diferentes según las necesidades del estudio [23].

Una descripción detallada para la interpretación del ECG haciendo uso del papel milimetrado y considerando los cinco elementos antes mencionados se presenta en [23].

Desafíos y oportunidades en la interpretación clínica de ECG

El análisis manual de los ECG, aunque eficaz en manos de especialistas, está sujeto a limitaciones como la variabilidad en la interpretación, la fatiga humana y la dependencia de la experiencia clínica. El diagnóstico basado en ECG puede verse afectado por la variabilidad interobservador e intra-observador, lo que significa que diferentes cardiólogos pueden interpretar el

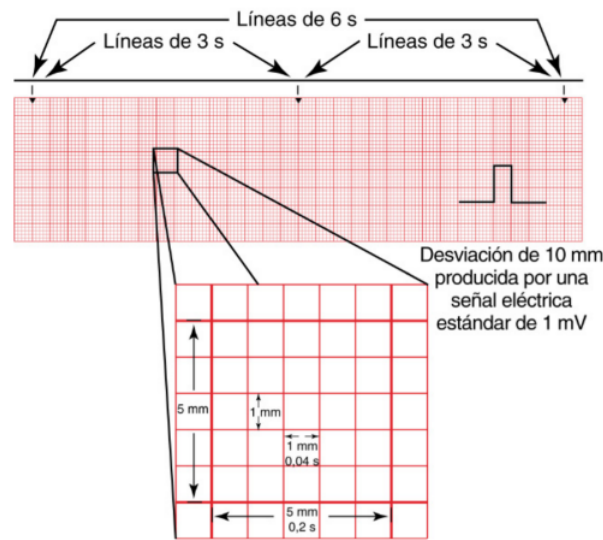


FIGURA 2.1 Papel de ECG.

Figura 2.5: Papel de ECG. [24]

mismo trazo de ECG de manera distinta, especialmente en casos complejos o fronterizos. También es posible que un mismo especialista proporcione diagnósticos ligeramente diferentes al evaluar el mismo ECG en momentos distintos, debido a factores como el contexto clínico o la fatiga. Esto ocurre porque la interpretación del ECG no siempre se basa en criterios absolutos, sino que puede depender de la experiencia del médico, su conocimiento de antecedentes clínicos del paciente y su familiaridad con ciertas patologías.

Por otro lado, los cardiólogos y especialistas en electrofisiología analizan una gran cantidad de ECG diariamente, lo que los puede llevar a un punto de fatiga visual y cognitiva, incrementando con ello la probabilidad de errores en la identificación de patrones anómalos. Algunos estudios han demostrado que la fatiga acumulada en turnos largos reduce la capacidad de detectar anomalías sutiles [25]. Adicionalmente, un alto volumen de pacientes puede generar retrasos en la evaluación de los ECG, afectando la rapidez con la que se toman decisiones médicas críticas.

El diagnóstico por ECG es una habilidad que mejora con la experiencia, lo que provoca que los médicos en entrenamiento experimenten mayor dificultad para identificar ciertos tipos de arritmias o anomalías sutiles en la morfología del ECG. Además, en muchas regiones geográficas el acceso

a cardiólogos capacitados es limitado, lo que puede generar diagnósticos incorrectos o retrasados cuando la interpretación queda en manos de médicos generales con menos experiencia en electrocardiografía.

En los últimos años se han propuesto algoritmos computacionales basados en el análisis de datos, inteligencia artificial y aprendizaje automático que, mediante el procesamiento de grandes volúmenes de datos con alta precisión, facilitan la identificación de patrones complejos asociados con diversas condiciones cardíacas. La clasificación automática de señales ECG mediante redes neuronales, árboles de decisión o modelos basados en aprendizaje profundo ha demostrado ser una herramienta valiosa para la detección de arritmias, isquemias y otros trastornos del ritmo cardíaco. Además de los modelos basados en inteligencia artificial y aprendizaje automático, existen diversas estrategias para la clasificación automática de señales ECG basadas en técnicas de procesamiento digital de señales. Estas estrategias incluyen métodos deterministas y estadísticos que permiten extraer características relevantes de la señal y tomar decisiones en función de reglas predefinidas.

Sin embargo, el desarrollo e implementación de estos sistemas plantea desafíos importantes. La variabilidad interindividual en las señales ECG, las diferencias en la morfología del trazado entre hombres y mujeres, así como la necesidad de bases de datos amplias y representativas, requieren el diseño de modelos robustos y adaptativos. Es fundamental garantizar la interpretabilidad y confiabilidad de los algoritmos para que puedan ser integrados de manera efectiva en la práctica clínica.

2.2.4. Diferencias fundamentales en el ECG entre hombres y mujeres

Hasta ahora, ya se ha mencionado que el ECG es una herramienta fundamental para la evaluación de la actividad eléctrica del corazón y su interpretación precisa es clave para el diagnóstico de diversas patologías cardiovasculares. Sin embargo, la morfología y los parámetros de las señales ECG pueden variar significativamente entre hombres y mujeres debido a diferencias fisiológicas, hormonales y anatómicas. Estas variaciones no sólo influyen en la identificación de valores de referencia para una lectura clínica adecuada, sino que también representan un desafío para la interpretación y el diagnóstico de anomalías cardíacas.

Factores como la frecuencia cardíaca, la duración de los intervalos PR,

QRS y QT, y la amplitud de las ondas P, R y T han demostrado ser distintos entre ambos sexos, lo que sugiere la necesidad de adaptar tanto los criterios clínicos como los modelos computacionales para una evaluación más precisa. Además, las hormonas sexuales, en particular los estrógenos y la testosterona, desempeñan un papel importante en la modulación de la actividad eléctrica cardíaca, afectando la repolarización ventricular y la duración de los intervalos del ECG.

Desde una perspectiva computacional, estas diferencias resaltan la importancia de desarrollar algoritmos de clasificación que incorporen parámetros específicos por sexo para mejorar la precisión del diagnóstico automático. Ignorar estas distinciones podría llevar a una tasa más alta de falsos positivos o negativos en la detección de anomalías cardíacas, especialmente en aplicaciones basadas en inteligencia artificial y aprendizaje automático.

A continuación, se listan las diferencias fundamentales en los ECG entre hombre y mujeres:

- Las mujeres tienden a presentar una frecuencia cardíaca media superior a la de los hombres, con una diferencia de entre 3 y 5 latidos por minuto. Esto se debe a un menor tamaño del corazón y un mayor tono vagal.
- En las mujeres, el intervalo QT suele ser más prolongado que en los hombres, incluso después de la corrección por frecuencia cardíaca (QTc). Esto puede aumentar el riesgo de arritmias como la torsades de pointes. Las mujeres muestran un intervalo QT de 10 a 20 ms superior que el de los hombres en condiciones basales. En especial, esta duración superior se hace más patente en el período menstrual, estas diferencias en la duración del QT están mediadas por el efecto de las hormonas femeninas.
- Generalmente, los hombres presentan intervalos PR y QRS más largos, lo que se asocia con una mayor masa ventricular.
- Las mujeres tienden a presentar ondas R de menor amplitud y ondas T más altas en las derivaciones precordiales.
- Los hombres pueden presentar mayor voltaje en las ondas QRS, lo que se relaciona con un mayor volumen del ventrículo izquierdo.

2.3. HERRAMIENTAS PARA EL ANÁLISIS AUTOMATIZADO DE SEÑALES ELECTROCARDIOGRÁFICAS

- Las hormonas sexuales, en especial los estrógenos, influyen en la repolarización ventricular y en la duración del intervalo QT. Durante la menopausia, algunas de estas diferencias tienden a disminuir.

El desconocimiento de estas diferencias entre sexos en la interpretación del ECG afecta de forma adversa la validez de los criterios diagnósticos en el ECG de las mujeres. Asimismo, no admitir estas diferencias en cuanto a la duración y el voltaje del QRS lleva a que los criterios electrocardiográficos de hipertrofia ventricular sean más específicos, pero menos sensibles en mujeres. [11]. La Tabla 2.1 presenta los valores utilizados típicamente en la electrocardiografía clínica ajustando los parámetros para la interpretación del ECG en mujeres. [22]

Onda o intervalo	Duración (ms)
Duración de la onda <i>P</i>	< 120
Intervalo <i>PR</i>	< 200
Duración del intervalo <i>QRS</i>	< 110 – 120
Intervalo QT	≤ 440 – 450

Tabla 2.1: Valores normales para las duraciones de las ondas y los intervalos en mujeres [22]

2.3. Herramientas para el análisis automatizado de señales electrocardiográficas

Además de mejorar la precisión diagnóstica en ciertos escenarios, la integración de tecnologías computacionales en los procesos de interpretación clínica de las señales ECG ofrece ventajas significativas en términos de accesibilidad y eficiencia. La implementación de sistemas automatizados en dispositivos portátiles o en plataformas de telemedicina permite un monitoreo continuo y en tiempo real, lo que favorece la detección temprana de anomalías y la toma de decisiones clínicas más oportunas. Esto es especialmente

2.3. HERRAMIENTAS PARA EL ANÁLISIS AUTOMATIZADO DE SEÑALES ELECTROCARDIOGRÁFICAS

relevante en regiones con acceso limitado a cardiólogos especializados, donde el diagnóstico asistido por computadora puede ayudar a reducir la carga de trabajo del personal médico y mejorar la calidad de la atención. La integración de tecnologías computacionales en el análisis y clasificación de señales electrocardiográficas representa un avance significativo en la detección y manejo de enfermedades cardiovasculares. Su desarrollo continuo no sólo tiene el potencial de optimizar el diagnóstico, sino también de transformar la manera en que se monitorea y previene la enfermedad cardíaca, permitiendo una medicina más personalizada, eficiente y accesible para la población.

A continuación se presenta la descripción de distintas herramientas existentes para el análisis automático de señales ECG, clasificadas en función de la aplicación y de los principios de análisis utilizados.

2.3.1. Herramientas para el análisis y clasificación de señales ECG de acuerdo a la aplicación

Esta clasificación basada en la aplicación permite diferenciar las herramientas según su contexto de uso. Mientras que los dispositivos vestibles y los sistemas de telemedicina están orientados a la accesibilidad y monitoreo continuo, los instrumentos médicos en entornos clínicos buscan precisión diagnóstica. Por otro lado, las herramientas de investigación permiten el desarrollo de nuevas tecnologías para mejorar la interpretación de las señales ECG.

Algoritmos de clasificación para sistemas de telemedicina

Estas herramientas están diseñadas para la adquisición, transmisión, almacenamiento y análisis remoto de señales ECG. Suelen integrarse en plataformas de salud digital que permiten el monitoreo continuo de pacientes y la toma de decisiones a distancia. Para que un algoritmo clasificador de señales ECG pueda ser utilizado en aplicaciones de telemedicina, debe cumplir con varias características clave que garanticen su precisión, eficiencia y capacidad de integración en sistemas de salud digital, tomando en consideración las limitaciones propias de los dispositivos vestibles o de monitoreo remoto [26]. Estas características incluyen:

- Alta precisión y robustez en la clasificación. Esto significa que el algoritmo debe minimizar tanto los falsos positivos como los falsos ne-

2.3. HERRAMIENTAS PARA EL ANÁLISIS AUTOMATIZADO DE SEÑALES ELECTROCARDIOGRÁFICAS

gativos, especialmente en la detección de arritmias como la fibrilación auricular, taquicardias ventriculares o bloqueos auriculoventriculares. Además, debe ser entrenado y validado con bases de datos de ECG diversas que incluyan variaciones debidas a edad, género, comorbilidades y condiciones fisiológicas.

- Capacidad de procesamiento en tiempo real o casi real. Para aplicaciones en telemedicina, el algoritmo debe ser capaz de analizar y clasificar el ECG en tiempo real o en cuestión de segundos. Así mismo, debe optimizar el uso de recursos, especialmente si se implementa en dispositivos con capacidad de cómputo limitada, como smartwatches o monitores portátiles.
- Adaptabilidad a señales con ruido y variabilidad. Los ECG obtenidos en entornos ambulatorios suelen tener artefactos de movimiento, interferencias electromagnéticas y ruido muscular. El algoritmo debe integrar técnicas de preprocesamiento como filtros adaptativos, transformada de *wavelet* o técnicas de eliminación de artefactos. Adicionalmente, se requieren algoritmos que puedan recalibrarse o adaptarse a distintas calidades de señal (variaciones en el contacto de electrodos o ruido inducido por la actividad del paciente).
- Capacidad de integración con plataformas de telemedicina. Es importante que algoritmo proporcione resultados exportables en formatos compatibles con sistemas de salud digital como HL7, DICOM-ECG o bases de datos en la nube. Otro aspecto importante es su interoperabilidad, es decir, su capacidad de comunicarse con diferentes dispositivos y plataformas, permitiendo el almacenamiento y visualización remota del ECG.
- Interpretabilidad y explicabilidad. Es deseable que el algoritmo no solo proporcione una clasificación binaria (e.g. "normal" o "anormal"), sino que incluya explicaciones o métricas de confianza en el diagnóstico. En aplicaciones clínicas, también es útil que el sistema muestre los segmentos específicos del ECG que motivaron la clasificación, facilitando la validación por parte del médico.
- Seguridad y cumplimiento normativo. El algoritmo debe cumplir con regulaciones como HIPAA (en EE.UU.) o GDPR (en Europa) para la

2.3. HERRAMIENTAS PARA EL ANÁLISIS AUTOMATIZADO DE SEÑALES ELECTROCARDIOGRÁFICAS

protección de datos de pacientes. Para su uso en entornos médicos, el algoritmo debe haber sido validado con estudios clínicos y aprobado por organismos reguladores como la FDA (EE.UU.) o la CE (Europa).

Algoritmos de clasificación para instrumentos médicos en entornos clínicos

Los dispositivos médicos utilizados en hospitales y clínicas incorporan procesamiento digital de señales para mejorar la calidad de las mediciones y facilitar la interpretación de los ECG. Si bien, se espera que en este tipo de aplicaciones se reduzcan las limitaciones en relación con la capacidad de recursos computacionales en comparación con los dispositivos vestibles o de monitoreo remoto, sigue habiendo requerimientos específicos para que un algoritmo clasificador de señales ECG sea adecuado para su uso en instrumentos médicos en entornos clínicos, sobre todo por temas de compatibilidad con los sistemas hospitalarios.

En este sentido, además de los requerimientos relacionados con la precisión, confiabilidad y robustez previamente discutidos en las aplicaciones de telemedicina, los algoritmos clasificadores que son integrados en instrumentos médicos para ser utilizados entornos clínicos deben ser compatibles con protocolos médicos y exportar sus resultados en formatos estándar como DICOM-ECG, HL7 para su integración con historiales clínicos electrónicos. Además, debe permitir la comunicación con monitores de la Unidad de Cuidados Intensivos y sistemas de gestión hospitalaria, y dependiendo de la infraestructura del hospital, debe ser capaz de ejecutarse localmente en el dispositivo o enviar datos a servidores centralizados.

Finalmente, en temas de regulación, los algoritmos clasificadores deben tener en consideración los lineamientos de sistemas de certificación de dispositivos médicos, como la norma ISO 13485.

2.3.2. Herramientas para el análisis y clasificación de señales ECG de acuerdo a la técnica

Actualmente, existe una diversidad de técnicas de procesamiento digital de señales y de inteligencia artificial para desarrollar algoritmos

2.3. HERRAMIENTAS PARA EL ANÁLISIS AUTOMATIZADO DE SEÑALES ELECTROCARDIOGRÁFICAS

clasificadores de señales ECG. En este apartado se discutirán algunas de estas y al finalizar, se realiza una reflexión en torno a las ventajas y desventajas de cada uno de los enfoques.

Algoritmos clasificadores basados en técnicas de inteligencia artificial

Los algoritmos basados en técnicas de inteligencia artificial son un conjunto de reglas y procesos que las máquinas utilizan para realizar tareas que normalmente requieren inteligencia humana. Estos algoritmos aprovechan grandes cantidades de datos y cálculos para aprender, tomar decisiones y resolver problemas. En el caso de la clasificación de señales ECG, las técnicas de inteligencia artificial proveen la flexibilidad necesaria para adaptar las decisiones a la diversidad de características previamente discutidas para las señales ECG, no obstante, esto requiere una gran cantidad de información para los procesos de entrenamiento de los algoritmos. En esta categoría podemos encontrar algoritmos de clasificación con base en el aprendizaje automático, el aprendizaje profundo y las redes neuronales.

- Técnicas de aprendizaje automático supervisado. Estos algoritmos requieren un conjunto de datos etiquetados (es decir, señales ECG con sus correspondientes diagnósticos) para entrenar el modelo. Son los más comunes en la clasificación de ECG debido a su capacidad para detectar patrones complejos. Desde el punto de vista clínico, los modelos de bosques aleatorios y técnicas lineales destacan por qué sus resultados son interpretables clínicamente, lo que brinda la oportunidad de descubrir nuevos biomarcadores clave en la discriminación de diferentes tipos de latidos cardíacos. Por otro lado, las redes neuronales y los modelos bayesianos ofrecen la ventaja de analizar los datos del ECG directamente, sin necesidad de pre-procesar la señal ni de seleccionar características específicas previamente.
- Técnicas de aprendizaje automático no supervisado. Estos algoritmos no requieren etiquetas en los datos y buscan patrones o grupos dentro de las señales ECG sin conocimiento previo de las clases. Son útiles para explorar datos y descubrir nuevos patrones.

2.3. HERRAMIENTAS PARA EL ANÁLISIS AUTOMATIZADO DE SEÑALES ELECTROCARDIOGRÁFICAS

Algunos algoritmos utilizados que caen dentro de esta categoría son los de tipo *clustering* como el *K-means* o DBSCAN, que pueden agrupar señales ECG similares entre sí y que a menudo se usan en estudios exploratorios o cuando no se tienen etiquetas claras. Otro ejemplo de esta clase de algoritmos son los mapas auto-organizados, que aprenden a clasificar datos de alta dimensión en mapas bidimensionales, resultando de utilidad para clasificar tipos de ECG con variabilidad.

- Redes neuronales. Estos métodos están inspirados en la arquitectura del cerebro humano y son especialmente útiles para procesar grandes cantidades de datos debido a su capacidad para aprender representaciones jerárquicas complejas. Son ideales para tareas de clasificación de señales con patrones espaciales complejos y pueden aprender automáticamente características de la señal sin necesidad de extracción manual. Además las redes neuronales recurrentes utilizan conexiones recurrentes para modelar secuencias temporales, lo que las hace ideales para analizar secuencias temporales en donde las anomalías no se presentan con un patrón periódico.

Algoritmos clasificadores basados en técnicas de procesamiento digital de señales

Además de los modelos basados en inteligencia artificial y aprendizaje automático, existen diversas estrategias para la clasificación automática de señales ECG basadas en técnicas de procesamiento digital de señales. Estas estrategias incluyen métodos deterministas y estadísticos, tanto en el dominio del tiempo como en el dominio de la frecuencia que permiten extraer características relevantes de la señal y tomar decisiones sin la necesidad de un entrenamiento previo. Algunas de las técnicas utilizadas dentro de esta categoría se encuentran los basados en transformación y análisis en el dominio de la frecuencia, los basados en la extracción de características y los basados en el filtrado y modelado matemático.

- Técnicas basadas en la transformación y análisis en el dominio de la frecuencia. En esta categoría encontramos el uso de la trans-

2.3. HERRAMIENTAS PARA EL ANÁLISIS AUTOMATIZADO DE SEÑALES ELECTROCARDIOGRÁFICAS

formada de Fourier para analizar la distribución espectral de la señal ECG y detectar anomalías en ciertas bandas de frecuencia asociadas con arritmias o alteraciones cardíacas; la transformada wavelet que permite descomponer la señal ECG en diferentes niveles de resolución para identificar componentes de baja y alta frecuencia, lo que es útil para la detección de eventos transitorios como las arritmias; y la transformada de Hilbert, que se emplea para obtener la envolvente de la señal y analizar la modulación instantánea de frecuencia y amplitud, útil en la identificación de variaciones en la morfología del ECG.

- Técnicas basadas en la extracción de características. Se extraen características como la duración del complejo QRS, el intervalo RR, la amplitud de la onda P y T, entre otros. Estas características pueden ser utilizadas en sistemas de clasificación basados en reglas o en modelos estadísticos; también es posible emplear métodos como el análisis en el dominio del tiempo (media y desviación estándar de los intervalos RR) y el análisis en el dominio de la frecuencia (potencia en bandas de baja y alta frecuencia) para evaluar la actividad autonómica del corazón.
- Técnicas basadas en el filtrado y modelado matemático. Dentro de esta categoría podemos encontrar a los modelos autorregresivos, por ejemplo el modelo ARMA (por sus siglas en inglés, *AutoRegressive Moving Average*), que se utilizan para modelar la señal ECG y predecir su comportamiento en función de valores previos. Esto es útil para la detección de irregularidades en el ritmo cardíaco. Otra técnica dentro de esta categoría es el análisis de correlación cruzada, el cual se basa en comparar una señal ECG con patrones previamente definidos y detectar similitudes con distintos tipos de arritmias. Como es el caso del algoritmo XCorr-In desarrollado en este proyecto de tesis. Finalmente, existen técnicas de filtrado adaptativo, como los filtros de Kalman o filtros adaptativos LMS (por sus siglas en inglés, *Least Mean Squares*) que resultan de gran utilidad para mejorar la calidad de la señal y detectar patrones anómalos en la morfología del ECG.

En conclusión, es notable como la IA ha revolucionado el análisis del electrocardiograma, proporcionando herramientas avanzadas para la

2.3. HERRAMIENTAS PARA EL ANÁLISIS AUTOMATIZADO DE SEÑALES ELECTROCARDIOGRÁFICAS

detección y predicción de enfermedades cardíacas y no cardíacas. A través de algoritmos de aprendizaje automático, la inteligencia artificial puede identificar patrones sutiles y subclínicos en grandes volúmenes de datos, superando las limitaciones de la interpretación convencional del ECG. Los algoritmos de aprendizaje supervisado han demostrado eficacia en la detección de disfunción ventricular izquierda, fibrilación auricular episódica, y enfermedades estructuras y valvulares; además la inteligencia artificial ha ampliado el uso del ECG más allá del ámbito cardíaco, permitiendo la detección de enfermedades no cardíacas.[27]

No obstante, aun cuando la inteligencia artificial demuestra un gran potencial en el diagnóstico y monitoreo de enfermedades cardiovasculares, su implementación enfrente varios desafíos. Debido a que los modelos de IA requieren una gran cantidad de datos para su entrenamiento, plantea preocupaciones sobre la protección de la información del paciente y el cumplimiento de normativas como el GDPR [28] y la HIPAA [29]. Los modelos pueden presentar un sesgo en los datos, los modelos entrenados en bases de datos específicas pueden no ser generalizables a diferentes poblaciones, lo que podría generar errores diagnósticos en pacientes con características distintas. Para superar estas barreras se requiere de un enfoque multidisciplinario que involucre a ingenieros, médicos, reguladores y expertos en ética. [27].

Por otro lado, las técnicas de procesamiento digital de señales ofrecen alternativas sólidas y complementarias a los modelos de aprendizaje automático para la clasificación de señales ECG. En muchos casos, una combinación de técnicas DSP con modelos estadísticos o sistemas expertos puede proporcionar resultados altamente precisos sin necesidad de entrenar modelos de IA complejos. Estas estrategias son especialmente útiles en aplicaciones con restricciones computacionales o cuando se busca garantizar interpretabilidad en la clasificación de las señales.

2.3.3. Análisis del ECG: Comparación entre la interpretación humana y la computarizada

El análisis del ECG se realiza mediante la interpretación por profesionales de la salud como una interpretación computarizada, ambas interpretaciones presentan ventajas y limitaciones que influyen el

2.3. HERRAMIENTAS PARA EL ANÁLISIS AUTOMATIZADO DE SEÑALES ELECTROCARDIOGRÁFICAS

diagnóstico. La interpretación humana se basa en la experiencia y el conocimiento clínico del paciente. Esto influye factores como antecedentes médicos, síntomas actuales o comorbilidades, lo cual proporciona una visión integral de la interpretación. Sin embargo, la interpretación humana también puede verse limitada por la fatiga, la subjetividad y la dificultad para detectar patrones sutiles en grandes volúmenes de datos.

La interpretación computarizada del ECG se apoya en los diversos métodos mencionados anteriormente, especialmente los que están basados en inteligencia artificial y aprendizaje automático, que pueden procesar grandes cantidades de datos rápidamente. Estos sistemas pueden identificar patrones ocultos en el ECG que son difíciles de detectar por el ojo humano. Sin embargo, los métodos computarizados enfrentan desafíos, por ejemplo, los errores de grabación como la frecuencia de muestreo o la colocación incorrecta de los electrodos, pueden afectar la precisión de los diagnósticos computarizados. Además, las computadoras no tienen la capacidad de integrar el contexto clínico o de considerar factores como el estado emocional del paciente, lo que puede influir en la interpretación clínica de casos más complejos. Los algoritmos también pueden tener dificultades para adaptarse a variaciones interindividuales en la morfología del ECG. [30]

A pesar de los desafíos, la computación avanzada demuestra gran potencial en la detección de arritmias, como la fibrilación auricular o la hipertrofia ventricular, cualquier diagnostico obtenido como salida de una interpretación computarizada debe ser validado en entornos clínicos reales antes de ser implementado en la práctica diaria. El aprendizaje automático ha demostrado su capacidad para mejorar la precisión de los diagnósticos, especialmente en la identificación de patrones no reconocibles por el ojo humano, pero siempre es fundamental contar con la supervisión de expertos médicos. Un desafío importante radica en los algoritmos de decisión rígidos utilizados por muchos sistemas de IA en el análisis de ECGs, que pueden dar lugar a errores si los límites de los parámetros no se ajustan adecuadamente, una leve variación en los parámetros del ECG puede alterar drásticamente la interpretación computarizada, aunque no haya una diferencia clínica significativa. [31]

La creciente demanda de atención cardíaca, combinada con la escasez de especialistas en cardiología, hace que los modelos computarizados

2.3. HERRAMIENTAS PARA EL ANÁLISIS AUTOMATIZADO DE SEÑALES ELECTROCARDIOGRÁFICAS

de interpretación de ECG sean cada vez más cruciales. A medida que aumenta la dependencia de estos sistemas automáticos, también lo hace la necesidad de que los algoritmos sean más precisos, sensibles y adaptables a la diversidad de las poblaciones, teniendo en cuenta las diferencias interindividuales.

Es por ello que la representación adecuada de las mujeres, especialmente en los rangos de edad más vulnerables, es crucial. Garantizar que estos modelos de interpretación incluyan una mayor diversidad de datos permitirá obtener resultados más precisos y fiables, lo que facilitará a los médicos realizar diagnósticos más certeros y adaptados a las características específicas de este grupo poblacional.

Capítulo 3

XCorr-In: Clasificador de señales electrocardiográficas basado en los índices de correlación cruzada

En este capítulo se detallará la metodología utilizada para desarrollar un algoritmo que permite clasificar las señales electrocardiográficas (ECG) de acuerdo a su morfología y periodicidad en dos tipos: ECG con ritmo sinusal y ECG con arritmia. El algoritmo, nombrado como XCorr-In, se fundamenta en técnicas de procesamiento estadístico de señales, en particular, el análisis del comportamiento de los índices de correlación cruzada entre una ECG recibida y un pulso ECG de referencia, de tal forma que sea posible cuantificar y analizar las similitudes morfológicas y temporales entre ambas.

La metodología empleada para el diseño y validación del algoritmo propuesto se ilustra en la figura 3.1. En la primera etapa de la metodología se identificó y caracterizó una base de datos de señales ECG que cumpliera con los requerimientos específicos para la aplicación del algoritmo; posteriormente, una vez que se identificó la base de datos adecuada, se procedió al diseño y selección de la muestra, así como su posterior procesamiento para la construcción de la señal de referencia. Con la señal de referencia lista, se diseñó e implementó la rutina de

3.1. IDENTIFICACIÓN Y CARACTERIZACIÓN DE LA BASE DE DATOS

clasificación basada en pruebas paramétricas y comparadores de factores de forma. Finalmente, se evaluó el clasificador para cuantificar su exactitud, así como determinar los compromisos existentes entre las principales métricas de desempeño: sensibilidad y especificidad. A continuación se describe con mayor detalle cada etapa de la metodología.

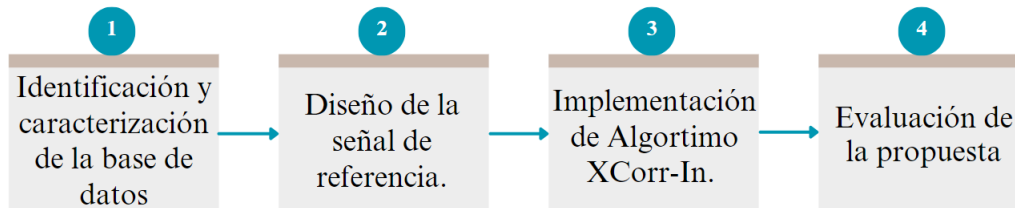


Figura 3.1: Diagrama de flujo de la metodología utilizada para el diseño del algoritmo XCorr-In.

3.1. Identificación y caracterización de la base de datos

En la actualidad existe un gran número de bases de datos de señales ECG que son de libre acceso, las cuales pueden ser aprovechadas para la evaluación y validación de propuestas tecnológicas orientadas al diagnóstico de padecimientos. Un caso particular es el sitio web PhysioBank, alojado por la organización PhysioNet [32], el cual contiene bases de datos de señales ECG recolectadas a partir de diversos estudios, por ejemplo, el estudio clínico publicado en [33] que busca comparar la respuesta electrofisiológica a diversos fármacos a través del análisis de comportamiento de las señales ECG. No obstante, aun cuando las bases de datos disponibles en la página PhysioBank son diversas, estas presentan algunos inconvenientes que les restan utilidad para este proyecto de tesis, entre los que destacan:

- Contienen un número reducido de registros de señales ECG, lo cual limita la confiabilidad de la propuesta del proyecto.

3.1. IDENTIFICACIÓN Y CARACTERIZACIÓN DE LA BASE DE DATOS

- Los registros ECG presentan condiciones específicas que impiden analizar la variabilidad en las características morfológicas y temporales entre señales ECG correspondientes a corazones sanos y señales ECG que presenten algún trastorno del ritmo o alguna patología.
- Las señales ECG corresponden a registros de distintas derivaciones, para efectos de este estudio, la derivación más adecuada es la II, que es la que presenta las ondas y complejos de mayor utilidad para la clasificación.
- Los grupos etarios de los pacientes son diversos, pero no incorporan suficientes muestras para el grupo etario de interés para este estudio.

En este sentido, después de una búsqueda exhaustiva, se identificó la base de datos construida para el estudio titulado *A 12-lead electrocardiogram database for arrhythmia research covering more than 10,000 patients* [34], la cual presenta las características suficientes y adecuadas para el desarrollo y evaluación del algoritmo XCross-In, tanto en un sentido cualitativo como desde la perspectiva cuantitativa. A continuación, se describen con mayor detalle las características de la base de datos seleccionada para este proyecto de tesis.

3.1.1. Descripción de los datos

El estudio desarrollado por [34] recibió la aprobación del comité de ética institucional del Shaoxing People's Hospital, además, se obtuvo el consentimiento informado de los pacientes para el uso públicos de sus datos anónimos.

La base de datos contiene un total de 10,646 registros de señales ECGs, correspondientes a 5,956 pacientes masculinos y 4,690 pacientes femeninos. Del total de los registros, el 17% de las señales ECG presentan ritmo sinusal y el 83% presenta, al menos, una anormalidad cardíaca. Esta base de datos incluye 11 ritmos cardíacos y 56 tipos de condiciones cardiovasculares diagnosticadas por profesionales de la salud. Algunas de las clasificaciones incluidas en la base de datos son: bradicardia, ritmo sinusal, fibrilación auricular, taquicardia, flutter auricular, ritmo

3.1. IDENTIFICACIÓN Y CARACTERIZACIÓN DE LA BASE DE DATOS

irregular, taquicardia supraventricular, taquicardia auricular, taquicardia por re-entrada nodal auriculoventricular, taquicardia por re-entrada auriculoventricular, entre otras.

Un aspecto de particular interés es que la base de datos incluye los registros eléctricos de las 12 derivaciones del ECG completo para cada paciente; así como la descripción total del proceso de adquisición y caracterización de las señales. Esta información resulta particularmente valiosa para los efectos de evaluación de nuestra propuesta.

3.1.2. Integración de los datos

De acuerdo a la descripción brindada por los autores de la base de datos, la integración de los datos se realizó en cuatro etapas. La primera etapa fue para la adquisición de las señales. En esta etapa cada paciente se sometió a un ECG de 12 derivaciones en estado de reposo; esta prueba se tomó durante un periodo de 10 segundos. Los datos adquiridos durante esta etapa se almacenaron en el sistema GE MUSE ECG. En la segunda etapa, las señales fueron categorizadas en función del ritmo y de acuerdo a la posible presencia de trastornos del ritmo o patologías. En esta etapa se integró un comité médico, del cual uno de los integrantes realizó una primera clasificación de la señal con base en el ritmo y otras condiciones cardíacas. Posteriormente, un segundo médico realizó una validación secundaria respecto a la condición cardíaca identificada en el ECG por el primer médico. En caso de haber un desacuerdo entre el primer y segundo médico, un tercer médico toma la decisión final. En la tercera etapa, los datos de la señal ECG y la información de diagnóstico correspondiente a cada señal se exportaron desde el sistema GE MUSE a archivos XML codificados con conversión de nombres específicos definidos por General Electric (GE) [35]. Por último, en la cuarta etapa, se realizó la conversión del archivo XML al formato CVS, este archivo CVS contiene la información de la señal ECG y la información del diagnóstico.

Para el proceso de adquisición de la señal ECG, en el estudio se utilizó un convertidor analógico-digital (ADC, por su sigla en inglés *Analog-to-Digital Converter*) con una resolución de 32 bits, la unidad de amplitud de las mediciones se estableció en microvolts (μV). Los límites

3.2. DISEÑO DE LA SEÑAL DE REFERENCIA

de los valores establecidos para la etapa de cuantización del ADC se definieron entre 32,767 y -32,768, con una frecuencia de muestreo de 500 Hz. La tabla 3.1 presenta los parámetros de adquisición para cada señal ECG incorporada en la base de datos.

Parámetro	Rango o valor
Resolución	32 bits
Frecuencia de muestreo	500 Hz
Tiempo	10 s
Número de muestras	5,000
Número de latidos	10

Tabla 3.1: Parámetros de adquisición para cada señal ECG en la base de datos

3.1.3. Extracción de grupo categórico

El grupo de interés en este trabajo son los pacientes de sexo femenino en un rango de edad de 40 a 60 años. Debido a esto, de la base de datos descrita previamente, se seleccionaron las señales ECG que cumplieran con las siguientes condiciones: sexo femenino, edad de entre 40-60 años. Al seleccionar y extraer los registros que cumplen con las características descritas, se obtiene una nueva base de datos categorizada que presenta la distribución que se muestra en la tabla 3.2.

3.2. Diseño de la señal de referencia

La señal de referencia, denotada por $x_{ref}(t)$ se refiere a una señal ECG que presenta a las características temporales y morfológicas esperadas

3.2. DISEÑO DE LA SEÑAL DE REFERENCIA

Señal ECG	Registros	Porcentaje
Sanas	382	45.2 %
Arritmias	463	54.8 %

Tabla 3.2: Base de datos categorizada para el grupo de pacientes femeninos de entre 40 a 60 años

en una persona femenina sana en el rango de edad de 40 a 60 años. Más específicamente, $x_{ref}(t)$ corresponde a la señal eléctrica generada para un sólo latido. Esta señal es de suma importancia para el correcto funcionamiento del algoritmo XCorr-In, por ello, la selección, procesamiento e interpretación de las muestras utilizadas para la construcción de $x_{ref}(t)$ son un factor clave en el desarrollo del algoritmo.

La señal $x_{ref}(t)$ funcionará como un patrón, el cual será utilizado para identificar la presencia de ritmo sinusal o ritmo ectópico en una señal ECG dada, a la que denotaremos por $x(t)$, mediante el cálculo y análisis de la función de correlación cruzada, $r(\tau)$. Como se describió en el capítulo anterior, $r(\tau)$ representa la fuerza y sentido de correlación entre dos señales que varían en el tiempo. En este sentido, el análisis temporal y morfológico se realizará calculando y analizando $r(\tau)$ para $x_{ref}(t)$ y $x(t)$.

Determinar la forma de onda y duración de $x_{ref}(t)$ requiere un claro entendimiento de la variabilidad implícita en las amplitudes y características temporales de las señales ECG. Como se describió en el capítulo anterior de este documento, la forma de onda de una señal ECG adquirida por un sistema de electrodos siempre incorporará una componente de ruido aleatorio. En este sentido, es posible considerar que tanto la amplitud, como la duración, de las distintas ondas que componen el ECG de cada latido son variables aleatorias. De acuerdo a lo anterior, la construcción de $x_{ref}(t)$ requiere incorporar información suficiente para asegurar la representatividad de las variaciones más significativas en el ECG.

En este trabajo se propone la metodología presentada en la Figura 3.2 para el diseño y construcción de $x_{ref}(t)$. A continuación se describen

con mayor detalle las etapas propuestas en la metodología.

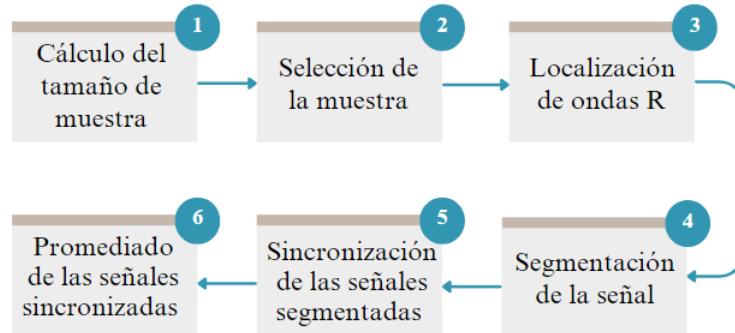


Figura 3.2: Diagrama de flujo de la metodología utilizada para el diseño y construcción de la señal ECG de referencia

3.2.1. Cálculo del tamaño de muestra

Una muestra es un subconjunto de elementos seleccionados dentro una población de interés. Para que las inferencias que se realizan sobre la población a partir de la muestra sean válidas, se debe de obtener una muestra que sea representativa de la población. [36].

De acuerdo a los datos presentados en la subsección 3.1.3, la población de señales de ECG sanas consideradas en este trabajo se compone de 382 registros. En este sentido, para determinar el tamaño de la muestra y asegurar la representatividad de la población, se utiliza la ecuación (3.1) [37] :

$$n = \frac{N \cdot Z^2 \cdot p \cdot q}{d^2 \cdot (N - 1) + Z^2 \cdot p \cdot q} \quad (3.1)$$

En donde:

- n : Tamaño de la muestra.
- N : Tamaño de la población.
- Z : Valor de Z asociado al nivel de confianza.

3.2. DISEÑO DE LA SEÑAL DE REFERENCIA

- p : Probabilidad de éxito (Nivel de confianza).
- q : Probabilidad de fracaso (Significancia).
- d^2 : Precisión (error máximo admisible en términos de proporción).

Los valores utilizados en este trabajo para la determinación del tamaño de la muestra se presentan en la Tabla 3.3. Sustituyendo los valores de la Tabla 3.3 en la ecuación (3.1), se obtiene el siguiente tamaño de muestra:

$$n = \frac{382 \cdot 1,64^2 \cdot 0,95 \cdot 0,05}{0,05^2 \cdot (382 - 1) + 1,64^2 \cdot 0,95 \cdot 0,05} = 45,17 \approx 45$$

Por tanto, se determina que, para construir una señal de referencia que incorpore las características necesarias para ser estadísticamente representativa de la población de señales ECG de mujeres sanas de 40 a 60 años, es necesario seleccionar aleatoriamente 45 registros de señales ECG de la categoría identificada con ritmo sinusal.

Variable	Valor
N	382
Z	1.64
p	0.95
q	0.05
d^2	0.05

Tabla 3.3: Valores considerados para determinar el tamaño de muestra necesario para el diseño de $x_{ref}(t)$

3.2.2. Selección de la muestra

Para eliminar el riesgo de sesgo por sobremuestreo de características es recomendable utilizar un método de muestreo aleatorio, lo cual significa

que los registros que formarán parte de la muestra serán seleccionados al azar y de manera independiente.

Un vez que se obtuvo el valor de n mediante la ecuación 3.1, los n registros que integrarán la muestra se seleccionan de manera aleatoria dentro de la población de 382 señales. Esta selección aleatoria se realiza aplicando un muestreo simple utilizando una rutina desarrollada en Python, con la función `random.sample()` de la biblioteca `random`. En la técnica de muestreo simple todos los elementos de la población tienen la misma probabilidad de ser seleccionados. En este caso, el muestreo simple es adecuado debido a que la distribución de edades en la población pacientes del sexo femenino es aproximadamente uniforme, como se puede constatar en la Figura 3.3

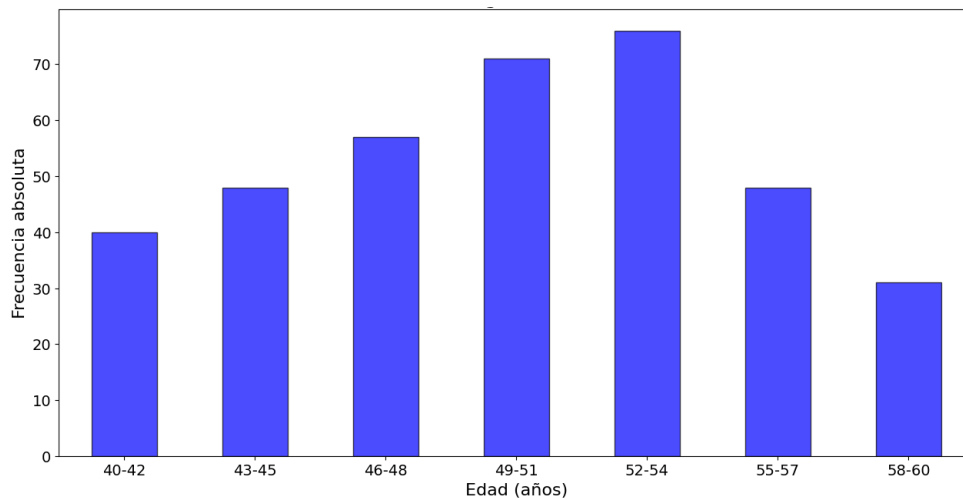


Figura 3.3: Histograma de las edades de la población de pacientes obtenida a partir de la categorización de la base de datos original.

3.2.3. Localización de ondas R

Una vez que se han seleccionado los 45 registros de señales ECG que integrarán la muestra necesaria para construir $x_{ref}(t)$, el siguiente paso es localizar los máximos locales correspondientes a las ondas R en todos los latidos que integran el registro de cada una de las 45 señales ECG

seleccionadas. Para esto se utiliza el toolbox BioSppy [38], el cual agrupa diversos métodos de procesamiento y reconocimiento de patrones orientados al análisis de bioseñales.

Para localizar las ondas R, se utiliza el módulo `biosppy.signals.ecg`, el cual provee métodos y funciones para el análisis específico de señales ECG. La función para la localización de estas ondas es `rpeaks`, la cual proporciona el índice temporal en el que se localiza el valor máximo de la onda R.

3.2.4. Segmentación de la señal

Como se mencionó en la tabla 3.1, cada registro de señal ECG de la base de datos está compuesto por un total de 5,000 muestras de amplitud de la señal eléctrica, las cuales fueron recolectadas durante 10 segundos. Dado que la señal $x_{ref}(t)$ busca representar la forma de onda de un sólo latido cardíaco sano, no es necesario utilizar el registro completo de la señal ECG. En este sentido, se requiere elegir un único latido cardíaco para cada una de las 45 señales ECG seleccionadas para construir $x_{ref}(t)$. Para lograr esto, es necesario segmentar la señal. La segmentación de una señal se refiere al proceso de extraer un subconjunto de muestras consecutivas o no consecutivas, de tal forma que formen una nueva señal (segmentada) que posea las características de interés. [39]

Para segmentar las señales ECG, primero se debe calcular el número de muestras necesarias para representar cada latido cardíaco. El número de muestras por latido, denotado por n_B , se calcula con la expresión:

$$n_B = \frac{n_s}{B} \quad (3.2)$$

Donde n_s corresponde al número total de muestras que integra el registro de la señal ECG y B corresponde al número de latidos presentes en el registro analizado. Dado que la duración promedio de un latido cardíaco en un corazón sano es de 1 s y que los registros que integran la base de datos tienen una duración de 10 s, entonces es posible determinar que cada registro incorpora aproximadamente 10 latidos. En este sentido, de acuerdo a la ecuación 3.2, se requiere extraer 500

3.2. DISEÑO DE LA SEÑAL DE REFERENCIA

muestras para representar un latido cardíaco. Adicionalmente, estas 500 muestras deben seleccionarse tal que asegure que se encuentren todos los elementos correspondientes a un latido cardíaco (Onda P, complejo QRS y onda T). La figura 3.4 representa la segmentación de 500 muestras en diferentes puntos de la señal.

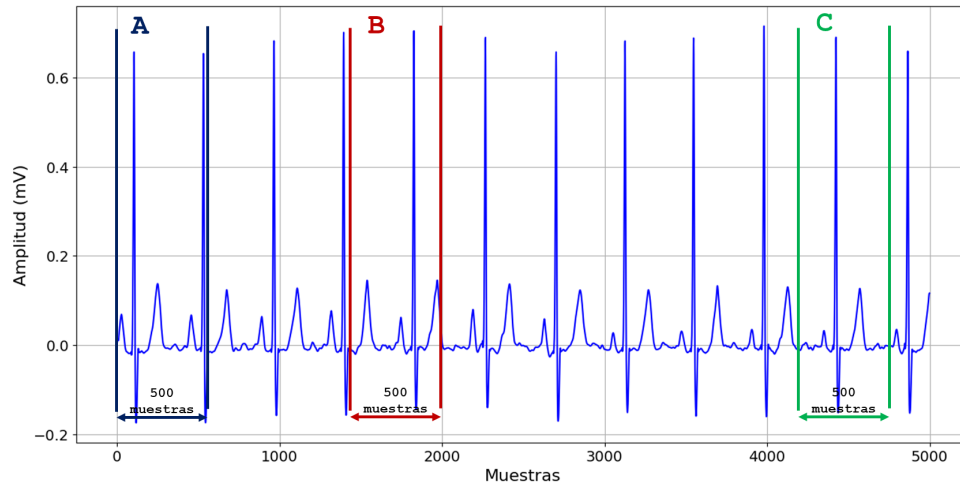


Figura 3.4: Representación de la segmentación de la señal ECG en diferentes puntos

En la Figura 3.3 se aprecia que si la segmentación de las 500 muestras se realiza en un intervalo arbitrario de la señal, se corre el riesgo de que no se incluyan todos los elementos del latido. Esto se representa gráficamente por los segmentos A y B. En el segmento A se observa que no se incluye completamente el complejo QRS, en el segmento B se observa que no se incluye completamente la onda T. Por otro lado, en el segmento C se observa una correcta integración de todos los elementos necesarios para representar el latido cardíaco.

Para asegurar una correcta segmentación de cada latido, el criterio de selección del latido se basa en identificar el latido con la amplitud máxima de la onda R. Una vez que se ha localizado, se seleccionan 250 muestras a la derecha y 250 muestras a la izquierda de este punto, con esto es posible asegurar que se incluyan las onda P y T, como lo muestra la figura 3.5.

Este procedimiento se repite para cada una de las 45 señales ECG

3.2. DISEÑO DE LA SEÑAL DE REFERENCIA

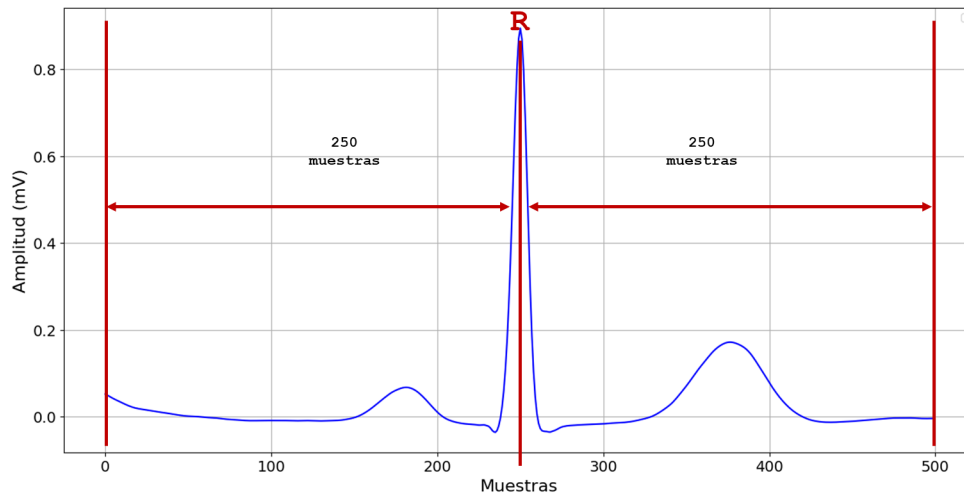


Figura 3.5: Segmentación de 500 muestras de la señal ECG tomando como referencia la onda R

seleccionadas para construir $x_{ref}(t)$, obteniendo como resultado 45 latidos cardíacos, denotados por $x_{ref,i}(t)$, todos con una longitud de 500 muestras, equivalentes a una duración de 1 s.

3.2.5. Sincronización de las señales segmentadas

La sincronización de señales implica la coordinación temporal entre dos o más señales, de manera que estas puedan ser comparadas, combinadas, transmitidas o procesadas sin pérdida de información o errores. En el procesamiento de señales, la sincronización es fundamental para que las señales se encuentren alineadas en el tiempo, frecuencia o fase. [39]

Una vez que se tienen los 45 latidos cardíacos, éstos se sincronizarán temporalmente en función de la onda R. Este proceso es necesario para poder construir el latido de referencia aplicando el promedio de amplitudes. El conjunto de señales sincronizadas se muestra en la figura 3.6.

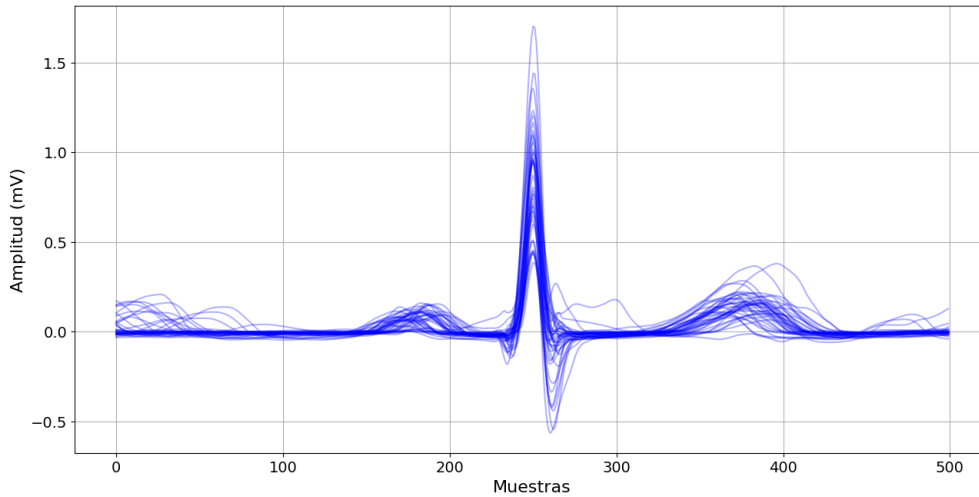


Figura 3.6: Familia de señales ECG para la construcción de la señal de referencia

3.2.6. Promediado de las señales sincronizadas

Finalmente, después de la sincronización, la señal $x_{ref}(t)$ se obtiene promediando las amplitudes de los 45 latidos previamente seleccionados, segmentados y sincronizados.

Como se mencionó anteriormente, la señal $x_{ref}(t)$ debe reflejar las características morfológicas y temporales que se esperan en el ECG de una paciente sana y en el rango de edad de 40 a 60 años. Para esto, es necesario incorporar la diversidad y la variabilidad inherente a las señales ECG. Ya se ha mencionado que la amplitud de una señal ECG en un instante dado se considera una variable aleatoria. Adicionalmente, es bien sabido que el primer momento central es la mejor representación para el valor esperado de una variable aleatoria. En este sentido, para estimar el primer momento central de las amplitudes de conjunto de señales ECG se utilizará la media muestral de las amplitudes correspondientes a cada índice temporal en los latidos cardíacos sincronizados.

De acuerdo a lo anterior, el latido de referencia, en su representación en tiempo discreto, se obtiene a partir de:

$$x_{ref}[1, j] = \frac{1}{n} \sum_{i=1}^n \mathbf{x}_s[i, j] \quad (3.3)$$

Para $j = 1 \dots n_B$, donde n representa el número de señales consideradas para obtener la señal promediada, es decir el tamaño de la muestra definido por la ecuación 3.1. La matriz \mathbf{x}_s de tamaño $i \times j$ definida en la ecuación 3.3 contiene las n_B muestras de amplitud seleccionadas durante el proceso de segmentación para cada uno de los n latidos. En este sentido, el promedio se aplica a los conjuntos de valores contenidos en cada una de las j columnas de la matriz.

El resultado de la señal de referencia construida para esta investigación se observa en la figura 3.7

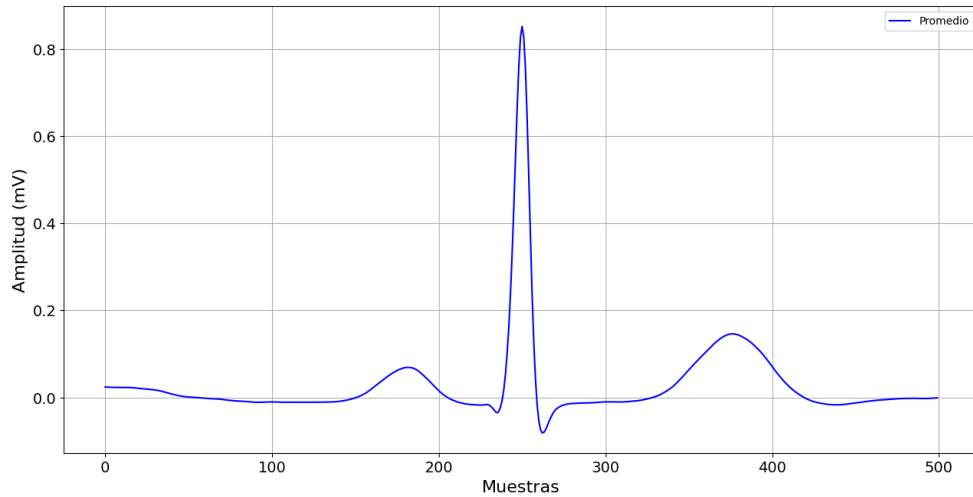


Figura 3.7: Señal ECG de referencia para el algoritmo de clasificación

3.3. Algoritmo XCorr-In

El algoritmo XCorr-In se basa, fundamentalmente, en la comparación de formas de onda para identificar la presencia o ausencia de patrones conocidos. XCorr-in se inspira en la teoría de los filtros acoplados para

la recepción de señales de radio comúnmente utilizadas en sistemas de RaDAR [40].

El diagrama de proceso para el flujo de los datos en el algoritmo XCorr-In se presenta en la Figura 3.8. Una vez recibida una señal ECG, denotada por $x(t)$, esta se ingresa a dos módulos de procesamiento paralelo, un correlador y un comparador. El correlador tiene como entrada también a la señal $x_{ref}(t)$, o su equivalente en tiempo discreto $x_{ref}[1, j]$, que corresponde al latido de referencia descrito en la sección anterior. La función principal del correlador es obtener la función de correlación cruzada entre la señal ECG recibida y el latido de referencia. La función de correlación cruzada será procesada en el módulo de análisis paramétrico para ejecutar una prueba de hipótesis a los estadísticos más relevantes identificados en los índices de correlación que integran la función. Por otro lado, el comparador ejecutará diversas rutinas de procesamiento de bioseñales para extraer características específicas de la señal ECG recibida y compararlas con valores de referencia establecidos. Finalmente, el clasificador consiste en una serie de operaciones lógicas entre las salidas del comparador y el módulo de análisis paramétrico para generar una decisión en torno a la clasificación de la señal ECG recibida.

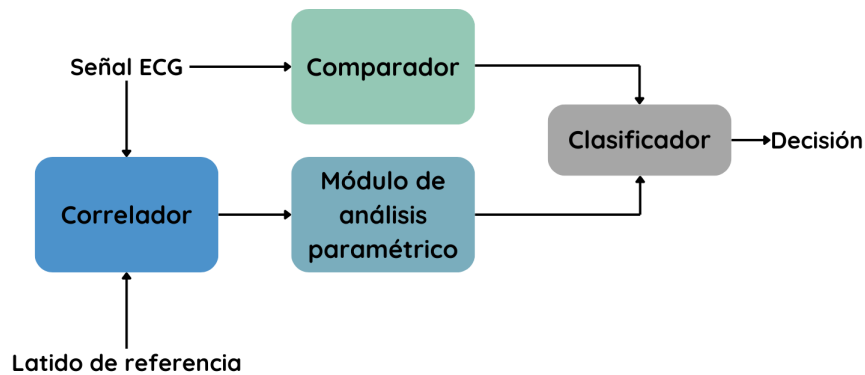


Figura 3.8: Diagrama de proceso para el flujo de información considerado en el diseño del algoritmo XCorr-In

3.3.1. Correlador

Dada una señal ECG recibida, en su representación en tiempo discreto denotada por $x[k]$, la función de correlación cruzada discreta entre $x[k]$ y la señal de referencia $x_{ref}[k]$ se obtiene mediante:

$$r[m] = \frac{1}{2k-1} \sum_{i=1}^k x_{ref}[i + \Delta m] x[i] \quad (3.4)$$

Donde m corresponde al desplazamiento temporal discreto aplicado a $x_{ref}[k]$ y Δ corresponde a un factor de desplazamiento que permite controlar la cantidad de muestras en la función de correlación. Note que el índice temporal de la señal de referencia, definido en la ecuación 3.3 se ha cambiado para mantener consistencia en el modelo para la función de correlación cruzada. No obstante, se sigue haciendo referencia a la señal definida por dicha ecuación.

La representación gráfica del proceso para calcular la correlación cruzada se muestra en la Figura 3.9. En esta figura se presentan cuatro señales: La señal azul corresponde a la señal ECG del paciente, $x[k]$; la señal de color rojo, corresponde a la señal $x_{ref}[k]$; finalmente, las señales en color magenta y naranja son versiones de $x_{ref}[k]$ desplazadas en el tiempo, es decir, $x_{ref}[k - m]$, donde $m = 400$ para la señal magenta y $m = 800$ para la señal naranja.

La ecuación 3.4 implica que el índice de correlación normalizado se calcula en cada desplazamiento de $x_{ref}[k]$. Por lo tanto, a la salida del correlador tendremos un vector de longitud $(1 \times M)$, donde $M = K + n_B - 1$ para $\Delta = 1$. K corresponde al número de muestras de amplitud de la señal $x[k]$, y como se mencionó previamente, n_B corresponde al número de muestras de amplitud en $x_{ref}[k]$. Cada índice de correlación en el vector $r[m]$ indica el nivel de similitud entre $x_{ref}[k]$ y $x[k]$ para el intervalo de muestras definido por $[m, m + n_B]$.

La Figura 3.10 presenta dos ejemplos de funciones de correlación cruzada, considerando dos señales ECG distintas, la curva de color azul corresponde a la función de correlación cruzada entre $x_{ref}(t)$ y una señal ECG clasificada con ritmo sinusal, la curva en color rojo muestra la función de correlación entre $x_{ref}(t)$ y una señal ECG clasificada con

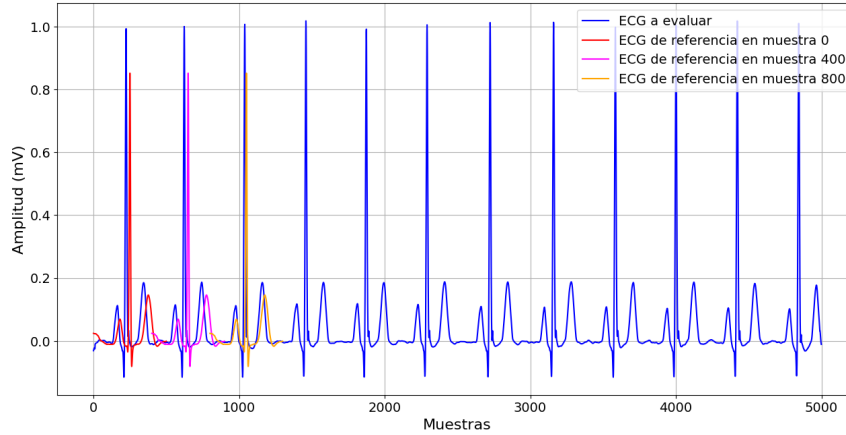


Figura 3.9: Representación gráfica del proceso para la obtención de la función de correlación cruzada.

taquicardia sinusal, correspondiente a un trastorno del ritmo cardíaco que se manifiesta como respuesta fisiológica a estímulos diversos como el estrés, la fiebre, reducción del volumen sanguíneo, la ansiedad, el ejercicio, hipotensión arterial o la insuficiencia cardíaca congestiva [41]. Es posible notar que las magnitudes de los índices de correlación difieren significativamente entre una señal ECG con ritmo sinusal y una señal con una arritmia, aun cuando la señal $x_{ref}(t)$ es constante para todos los procesos de correlación. Esto significa que, a través del análisis de comportamiento de los índices de correlación, es posible identificar la presencia de ritmos específicos en una ECG dada.

Para la curva azul que representa la función de correlación entre $x_{ref}(t)$ y una ECG clasificada con ritmo sinusal, se observa que el vector $r[m]$ presenta valores de amplitud cercanos a 1 en ciertos desplazamientos; no obstante, predominan los valores de índices de correlación en el rango $(-0,2, 0,2)$. Esto se debe a que al calcular la correlación entre la señal $x_{ref}(t)$ y la señal ECG recibida no están sincronizadas en la mayoría de los desplazamientos. Sin embargo, cuando la señal $x_{ref}(t)$ se alinea adecuadamente con un latido cardíaco, los índices de correlación alcanzan valores altos, cercanos a 1, indicando una fuerte similitud entre las dos señales. Estos picos de correlación también presentan una periodicidad, lo que indica similitud entre la señal $x_{ref}(t)$ y la señal ECG clasificada con ritmo sinusal en intervalos regulares, esta periodicidad coincide con

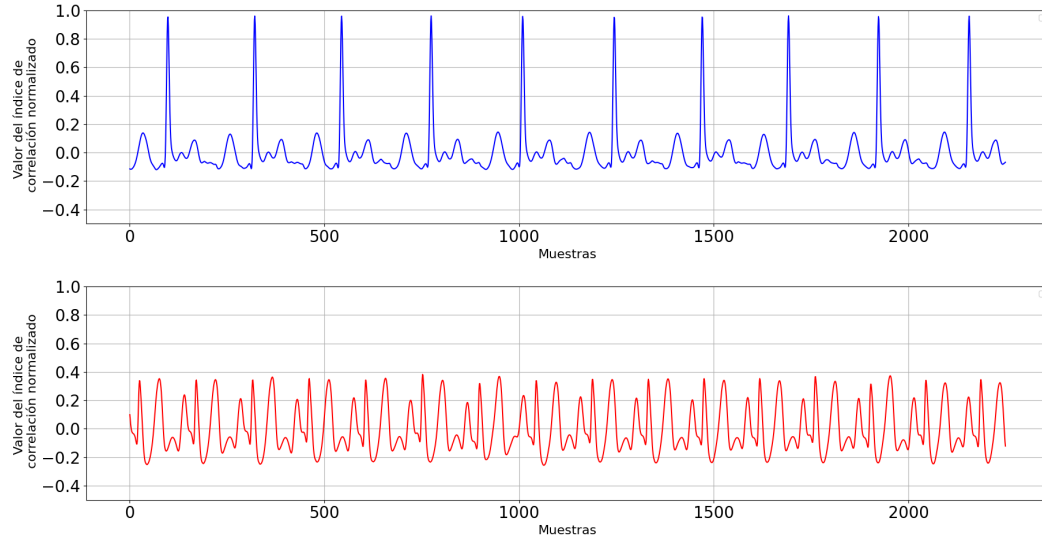


Figura 3.10: Función de correlación cruzada entre una señal clasificada con ritmo sinusal (azul) y una señal clasificada con taquicardia (rojo).

la frecuencia cardíaca. Por otro lado, para la curva roja que representa la señal clasificada con taquicardia, el vector $r[m]$ muestra valores de correlación bajos en todos los desplazamientos de la señal, en este caso los índices de correlación no superan el valor de 0.4, indicando que no existen intervalos de similitud significativa con la señal $x_{ref}(t)$.

Como se mencionó previamente, en la ecuación 3.4, el parámetro Δm representa el desplazamiento temporal discreto aplicado a $x_{ref}[k]$. En esta investigación se establece que $\Delta = 2$, lo que implica que el tamaño del vector de índices de correlación será de $(1 \times \frac{M}{2})$. Utilizar un valor de $\Delta = 1$ daría como resultado un vector de mayor longitud pero incluiría más valores de correlación cercanos a cero. Por otra parte, utilizar un valor de $\Delta > 2$ representa pérdidas de información al analizar la señal, ya que desplazamientos más amplios provocan que para algunos puntos de la señal recibida no se calcule el valor de correlación, impidiendo su análisis. Un valor muy elevado de Δ podría llevar a omitir el desplazamiento en el cual la señal $x_{ref}[k]$ se alinea con el latido cardíaco de la señal $x[k]$, imposibilitando la identificación de un ritmo sinusal.

Para decidir si una ECG dada, $x(t)$, presenta una similitud fuerte con

el pulso de referencia, $x_{ref}(t)$, es necesario asegurar que en el vector $r[m]$ existen suficientes índices de correlación que superan un umbral de clasificación denotado por α . En este sentido, α representará el nivel de similitud mínimo esperado entre dos latidos cardíacos normales sincronizados, lo que corresponde al análisis morfológico de las señales. De acuerdo a lo anterior, α es un parámetro de configuración en el algoritmo XCorr-In, ya que establece el criterio para clasificar la intensidad de correlación entre las dos señales que se comparan.

Por otro lado, determinar la presencia de ritmo sinusal o ritmo ectópico requiere más que evaluar la presencia de índices de correlación fuerte. También es fundamental establecer la cantidad mínima de índices de correlación que se espera superen el umbral α en el vector $r[m]$. Esto permite asegurar que la similitud observada no sea el resultado de coincidencias aleatorias, sino que se mantiene con cierta periodicidad entre las dos señales. Para llevar a cabo el análisis temporal de la función de correlación, es necesario inspeccionar la distribución de los valores de los índices de correlación existentes en $r[m]$, esto se lleva a cabo en el módulo de análisis paramétrico descrito a continuación.

3.3.2. Módulo de análisis paramétrico

Ya se ha mencionado en diversas ocasiones que las amplitudes de cualquier señal ECG presentan variaciones debido a alteraciones aleatorias provocadas por las diversas fuentes de ruido y las características anatómicas y fisiológicas de cada paciente. En este sentido, es posible imaginar que, aun cuando las amplitudes de la función $r[m]$ presentan tendencias de comportamiento en función de la clasificación de la ECG recibida, los índices de correlación en $r[m]$ deben ser modelados como variables aleatorias.

En este sentido, los índices de correlación obtenidos a la salida del correlador corresponden a observaciones de una variable aleatoria con distribución $f(r)$. Por tanto, resulta imperativo analizar las características de esta distribución para poder diseñar una prueba de hipótesis que permita discernir sobre la clasificación de una señal ECG dada.

Análisis estadístico descriptivo

En esta sección se examina el comportamiento estadístico de los índices de correlación que se obtienen al correlacionar la señal $x_{ref}(t)$ con las n señales seleccionadas de acuerdo a lo descrito en la sección 3.2.2. Este análisis tiene como objetivo principal establecer la distribución $f(r)$ de referencia que permita describir el comportamiento probabilístico de los índices de correlación para señales clasificadas con ritmo sinusal.

Para ilustrar el comportamiento de las distribuciones de los índices de correlación, se elaboraron los histogramas correspondientes a cuatro señales clasificadas con ritmo sinusal, como se observa en la figura 3.11, y a cuatro señales clasificadas con taquicardia sinusal, como se aprecia en la figura 3.12.

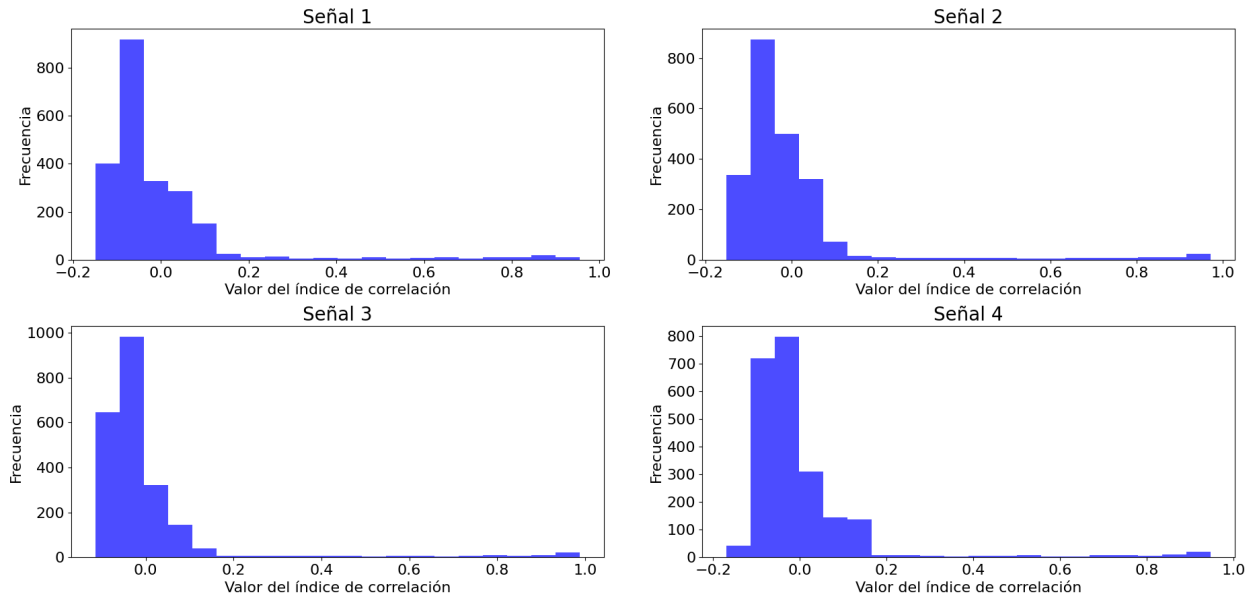


Figura 3.11: Distribución de los índices de correlación entre la señal $x_{ref}(t)$ y cuatro señales ECG con ritmo sinusal.

A partir de los histogramas presentados en las figuras 3.11 y 3.12, se observa que la mayoría de los índices de correlación se concentran en el rango $(-0,2, 0,2)$.

En los histogramas correspondientes a las señales clasificadas con rit-

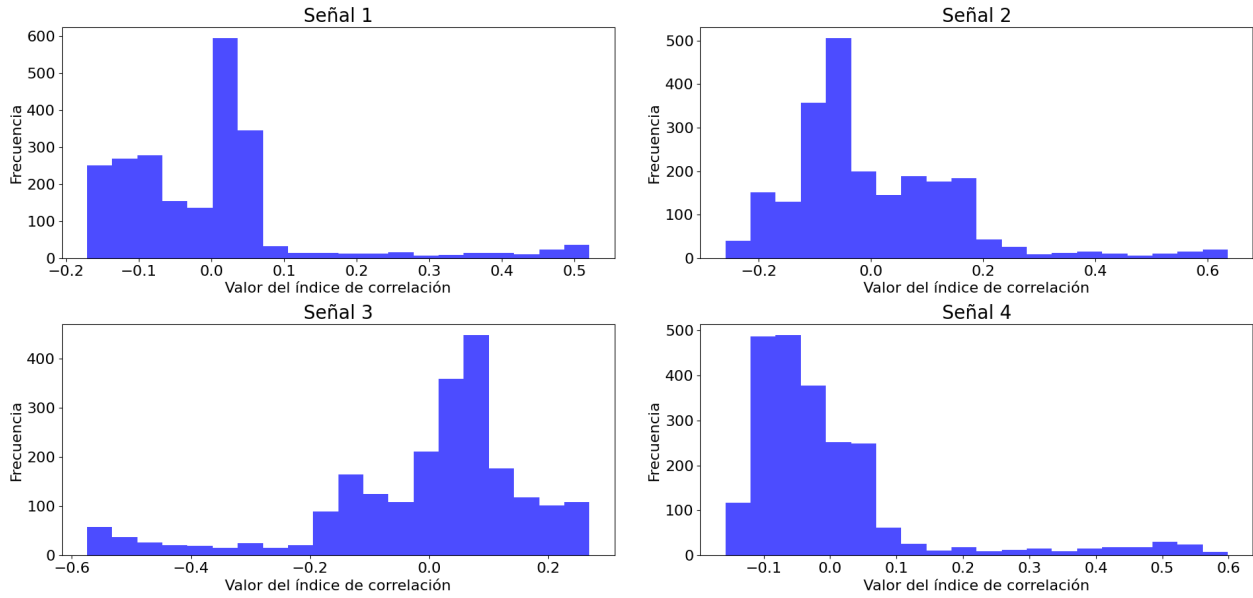


Figura 3.12: Distribución de los índices de correlación entre la señal $x_{ref}(t)$ y cuatro señales ECG con taquicardia sinusal.

mo sinusal en la Figura 3.11, se identifican observaciones distribuidas en el rango $(0,5,1)$. La presencia de estos valores indica que existen desplazamientos en la función de correlación cruzada donde las señales presentan una alta similitud.

Por el contrario, los histogramas de las señales clasificadas con taquicardia sinusal de la Figura 3.12 muestran que los valores de los índices de correlación son, generalmente, menores a 0,6. A partir de esta exploración de los datos muestrales, es posible establecer que los índices de correlación alcanzan valores más altos en señales con ritmo sinusal, en comparación a las señales con arritmia. Esta diferencia en las distribuciones de frecuencia de los índices de correlación establece las bases para construir estrategias de clasificación basadas en los distintos parámetros poblacionales de la distribución $f(r)$.

La Figura 3.13 presenta el histograma para el promedio del valor del índice de correlación cruzada, \bar{r} , para las 45 señales de la muestra con la señal $x_{ref}(t)$. Estos promedios se obtuvieron de acuerdo al siguiente modelo:

$$\bar{r} = \frac{1}{M} \sum_{m=1}^M r[m] \quad (3.5)$$

Visualmente, es posible notar que la distribución de la media de los índices de correlación es aproximadamente normal; esto permite suponer que la distribución del resto de los parámetros que dependen del primer momento central estarán basadas en esta misma distribución.

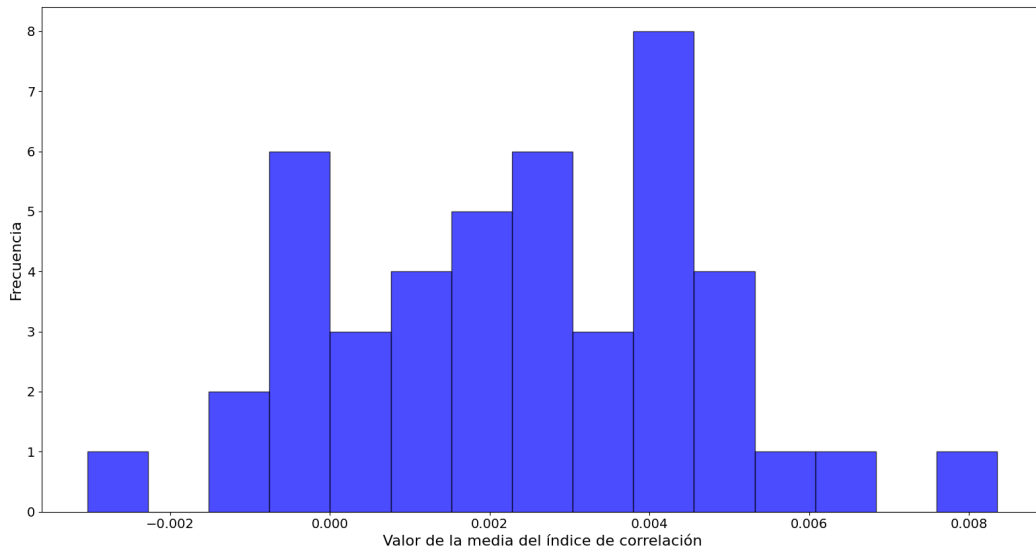


Figura 3.13: Distribución de la media de los índices de correlación para las 45 señales de la muestra con la señal $x_{ref}(t)$.

Finalmente, el comportamiento estadístico de los índices de correlación se describe cuantitativamente empleando tres medidas estadísticas: media, varianza y sesgo. En la tabla 3.4 se presentan las medidas estadísticas calculadas para los 45 vectores $r[m]$ correspondientes a las señales de la muestra n . En primer lugar, se calcularon la media, la varianza y el sesgo de cada uno de los 45 vectores $r[m]$. Posteriormente, se obtuvo el promedio de estas medidas individuales (medias, varianzas y sesgos) para generar métricas globales que representan el comportamiento estadístico general de las 45 señales. A partir de los valores presentados en la Tabla 3.4, podemos resaltar la amplia variabilidad y el sesgo presente

en los índices de correlación, lo cual coincide con las representaciones visuales mostradas en los histogramas. Dada la naturaleza de la función de correlación cruzada empleada para este proyecto, es de esperarse que la media de los índices de correlación sea cercana a cero, dado que sólo una reducida proporción de índices provocan valores elevados de correlación. Esta información resulta de utilidad para determinar que el uso directo de los índices de correlación como variable de decisión para clasificar las señales ECG pudiera limitar la confiabilidad del algoritmo, principalmente por la variabilidad y el sesgo presente en la distribución de $r[m]$. En este sentido, es necesario integrar una variable sujeta a un factor de control que dependa de la fuerza de correlación esperada entre una señal ECG con ritmo sinusal y $x_{ref}(t)$.

Métrica	Valor
Media aritmética	0.002371
Varianza	0.027489
Sesgo	3.519060

Tabla 3.4: Medidas descriptivas de los vectores $r[m]$ al correlacionar las 45 señales de la muestra n con la señal $x_{ref}(t)$

Análisis de parámetros poblacionales

Una vez descrito el comportamiento de los índices de correlación, es necesario profundizar en el análisis de los parámetros poblacionales que se han identificado como relevantes para determinar si la señal $x(t)$ presenta una similitud fuerte con la señal $x_{ref}(t)$. Esto es crucial para clasificar $x(t)$ como una señal con ritmo sinusal o una señal con ritmo ectópico.

Al finalizar el análisis estadístico descriptivo de los índices de correlación se determinó que es necesario definir una nueva variable que estuviera sujeta a un factor de control asociado a la fuerza de correlación. Para esto, se define α como el primer criterio de configuración

para el algoritmo XCorr-In. α corresponde a un umbral de decisión que le permitirá al algoritmo determinar la cantidad de valores en $r[m]$ que superan un determinado valor de correlación.

El segundo criterio de configuración para el algoritmo XCorr-In es el número de desplazamientos esperados en los que $r[m] \geq \alpha$ para una señal clasificada con ritmo sinusal. Este criterio requiere definir un umbral de decisión, β , a partir del cual será posible inferir si una señal ECG presenta ritmo sinusal o una arritmia. Para poder establecer el umbral β es necesario determinar una estrategia basada en el comportamiento de los estadísticos muestrales presentados en la sección previa. De acuerdo al análisis estadístico descriptivo, la distribución de frecuencias de los índices de correlación difiere cuando la ECG presenta ritmo sinusal y cuando presenta alguna alteración del ritmo. En este sentido, se proponen tres estrategias para establecer el valor de β :

1. Promedio (β_μ): El umbral se obtendrá a partir de la estimación puntual de la media de observaciones para las que $r[m] \geq \alpha$.
2. Deciles (β_{d_1}): El umbral se obtendrá mediante un decil de la distribución acumulativa de las observaciones para las que $r[m] \geq \alpha$.
3. Proporción ($\beta_{\hat{p}}$): El umbral se determina a partir de la distribución muestral de las proporciones de observaciones para las que $r[m] \geq \alpha$ con respecto a $|r[m]|$.

De acuerdo a las Figuras 3.11 y 3.12, resultan notables las diferencias entre los valores de los índices de correlación cuando la señal $x_{ref}(t)$ se correlaciona con una señal ECG con ritmo sinusal, en contraste con los obtenidos a partir de la correlación con una ECG que presenta arritmia. Además, dado que las diferentes arritmias causan distintos comportamientos morfológicos y temporales en las señales ECG, resulta más adecuado basarse en la búsqueda de signos de *normalidad* para definir las estrategias de clasificación y, a partir del resultado de esta búsqueda, concluir sobre la presencia de *anormalidades*. En este sentido, las tres estrategias presentadas previamente para establecer el umbral β se basan en el análisis de los índices de correlación obtenidos al correlacionar $x_{ref}(t)$ con las 45 señales utilizadas en la Sección 3.2.2. Para esto, es necesario definir la variable y_i , para $i = 1, \dots, 45$ que representa la

cantidad índices de correlación que cumplen con $r[m] \geq \alpha$ para cada una de las 45 señales seleccionadas.

Para establecer el valor de β_μ se considera la media aritmética definida de acuerdo al siguiente modelo:

$$\beta_\mu = \frac{1}{n} \sum_{i=1}^n y_i \quad (3.6)$$

Esta estrategia se basa en la premisa de que cualquier señal ECG con ritmo sinusal debe tener, al menos, β_μ índices que indiquen un valor de correlación con el pulso de referencia mayor o igual a α . Sin embargo, debido a la variabilidad en la anatomía de los pacientes y los diversos factores que pueden afectar la amplitud y periodicidad de las ECG, existe la posibilidad de que una ECG sana no cumpla con este criterio basado en el estimador puntual de la media poblacional. Incluso si la señal $x(t)$ tiene un valor muy cercano al valor de β_μ , quedaría excluida de la clasificación si el valor no es exactamente igual o superior.

Debido a la fuerte influencia de los valores atípicos en el enfoque basado en β_μ , se propuso un segundo enfoque basado en la segmentación de la distribución acumulativa de los valores de y_i . Para este enfoque, es necesario ordenar de manera ascendente los valores que componen la variable y_i , dando lugar a $y_{asc}[i]$. Posteriormente, localizamos los deciles en $y_{asc}[i]$, los cuales denotaremos como y_{d_l} para $l = 1, \dots, 9$. Para ejemplificar esto, considere que y_{d_1} es el valor de $y_{asc}[i]$ ubicado en la posición $i = d_1 = \frac{n}{10}$. Es importante recordar que si el valor de d_1 no es un número entero, entonces se deben promediar los valores correspondientes a las posiciones más cercanas al valor obtenido. En el caso particular de la muestra analizada en este trabajo, $d_1 = \frac{45}{10} = 4,5$, por lo que $y_{d_1} = (y_4 + y_5)/2$. En este sentido, el valor de β_{d_1} se obtiene de:

$$\beta_{d_l} = y_{d_l} = y_{asc}[d_l] \quad (3.7)$$

Este segundo enfoque se basa en la localización de un valor de referencia para el número de índices que cumplen con $r[m] \geq \alpha$ considerando que el $(100 - 10l)\%$ de las veces, una señal ECG con arritmia sobrepasará

dicha cantidad. No obstante, aun cuando esta segunda estrategia subsana las limitaciones identificadas para β_μ , el valor de β_{d_i} depende de la longitud de la señal ECG analizada. Esto significa, que este valor necesitaría estarse adaptando en función de la duración de la señal ECG analizada, implicando posibles sesgos adicionales durante la estimación del primer decil y con ellos, errores de decisión.

Con el objetivo de incorporar la incertidumbre en los parámetros de forma de la distribución subyacente de $r[m]$ y proponer una estrategia robusta a la duración del registro de la señal ECG que se analiza, se determinó conveniente analizar la proporción de índices de correlación que superan el umbral α en relación con el total de índices de correlación presentes en cada señal, dando lugar a una variable aleatoria que no depende de la duración del registro de la señal ECG recibida, sino de la distribución de probabilidades de la proporción muestral.

Para el conjunto $\mathbf{Y}_n = \{y_1, y_2, \dots, y_n\}$, la proporción \hat{p}_i se expresa como:

$$\hat{p}_i = \frac{y_i}{\frac{M}{2}} \quad (3.8)$$

Donde \hat{p}_i representa la proporción de índices de correlación que superan el umbral α para las n señales y $\frac{M}{2}$ representa la longitud del vector de salida del correlador, es decir, el numero total de indices de correlación calculados en la señal ECG recibida.

La ecuación 3.8 corresponde a un estimador puntual de la proporción poblacional de índices que cumplen con $r[m] \geq \alpha$ esperada en una señal ECG con ritmo sinusal. No obstante, definir este valor directamente como criterio de decisión no nos permite incorporar el factor de incertidumbre en el proceso de toma de decisiones. En este sentido, es necesario diseñar una prueba de hipótesis paramétrica para determinar el valor del umbral $\beta_{\hat{p}}$. De acuerdo con [36], la distribución muestral de \hat{p}_i es aproximadamente normal siempre y cuando se cumpla alguno de los siguientes supuestos: (i) $n \geq 30$, o (ii) $np \geq 5$. En este sentido, es posible afirmar que

$$f(\hat{p}_i) = \frac{1}{\sigma_{\hat{p}_i} \sqrt{2\pi}} \exp - \frac{(\hat{p}_i - \mu_{\hat{p}_i})^2}{2\sigma_{\hat{p}_i}^2} \quad (3.9)$$

Es decir, \hat{p}_i presenta una distribución normal con media:

$$\mu_{\hat{p}_i} = np \quad (3.10)$$

y varianza:

$$\sigma_{\hat{p}_i}^2 = np(1 - p) \quad (3.11)$$

Donde n es el tamaño de la muestra y p es la proporción poblacional, que en este caso indicaría la proporción de los índices de correlación que cumplen con $r[m] \geq \alpha$ en cualquier señal ECG con ritmo sinusal, no sólo para las señales de la muestra analizada en esta investigación. Una vez definido esto, es posible plantear la prueba de hipótesis necesaria para definir el umbral $\beta_{\hat{p}}$. Para esto se definen las siguientes hipótesis estadísticas:

$$H_0 : \hat{p} \geq \beta_{\hat{p}} \quad (3.12)$$

$$H_1 : \hat{p} < \beta_{\hat{p}}$$

La hipótesis H_0 describe el comportamiento esperado para la proporción muestral bajo el supuesto de que la señal recibida presenta ritmo sinusal, mientras que la hipótesis H_1 representa el caso de una señal con arritmia. Este planteamiento permite establecer una prueba de hipótesis unilateral de tal forma que:

$$\begin{array}{c} H_1 \\ Z_{\hat{p}} \leq Z_{\gamma} \\ H_0 \end{array} \quad (3.13)$$

Donde $Z_{\hat{p}}$ corresponde al estadístico de prueba, que a su vez dependerá de \hat{p} , la proporción de índices de la señal recibida, y Z_{γ} corresponde al

valor crítico definido para la prueba de hipótesis, el cual dependerá del nivel de significancia elegido, denotado por γ . $Z_{\hat{p}}$ se define mediante:

$$Z_{\hat{p}} = \frac{\hat{p} - np}{\sqrt{np(1-p)}} \quad (3.14)$$

Mientras que Z_{γ} se obtiene mediante:

$$Z_{\gamma} = Q^{-1}(1 - \gamma) \quad (3.15)$$

Donde $Q^{-1}(\cdot \cdot \cdot)$ representa la función Q inversa, la cual permite obtener los valores críticos para una distribución Normal Estándar. De acuerdo a lo anterior, para poder clasificar una señal ECG como sana, es decir con ritmo sinusal, es necesario que se cumpla con la siguiente condición:

$$Z_{\hat{p}} > Z_{\gamma} \quad (3.16)$$

o equivalentemente,

$$\frac{\hat{p} - np}{\sqrt{np(1-p)}} > Q^{-1}(1 - \gamma) \quad (3.17)$$

Re-acomodando los términos, la hipótesis H_0 de la expresión 3.13 se convierte en:

$$\hat{p} > \frac{\sqrt{np(1-p)}}{np} Q^{-1}(1 - \gamma) \quad (3.18)$$

Por lo que $\beta_{\hat{p}}$ corresponde al término del lado derecho de la desigualdad presentada en la expresión 3.18. Note que en la expresión 3.18 es necesario estimar la proporción poblacional definida previamente. Esta estimación se puede realizar mediante la construcción de un intervalo de confianza y a partir de este seleccionar el valor con el menor margen de error; no obstante, en este trabajo se decidió considerar el estimador puntual de la proporción a partir de la muestra de señales analizada previamente.

3.3.3. Comparador

Una vez establecidas las bases del módulo de análisis paramétrico, el módulo comparador se encarga de analizar características temporales de la señal ECG recibida. En este proyecto se propusieron dos enfoques para el módulo comparador: uno que genera información para acoplar la decisión basada en los enfoques absolutos para el análisis de los índices de correlación presentados en la sección anterior, es decir, con el promedio y los deciles; y otro, con técnicas avanzadas de procesamiento de señales, que trabaja en conjunto con la estrategia basada en los índices de correlación. A continuación, se describen ambas propuestas.

Enfoque 1

Este primer enfoque opera con dos características fisiológicas: la frecuencia cardíaca y la amplitud máxima del complejo QRS. Se define una rutina para extraer esta información de la señal ECG recibida y comparar los valores obtenidos con valores de referencia definidos en la literatura. Los valores de referencia identificados para el proceso de comparación se muestran en la tabla 3.5.

Variable	Valor
Frecuencia cardíaca	60-100 lpm
Amplitud de QRS	0.5 mV

Tabla 3.5: Características fisiológicas para extraer de la señal $x(t)$ y sus valores de referencia para el enfoque de medidas estadísticas absolutas.

Para calcular la frecuencia cardíaca se utilizó el siguiente modelo:

$$\text{lpm} = \frac{60}{t_{R-R}} \quad (3.19)$$

Donde t_{R-R} representa el intervalo de tiempo entre dos ondas R consecutivas, y se obtiene mediante la siguiente ecuación:

$$t_{R-R} = \frac{T}{R_{peaks}} \quad (3.20)$$

R_{peaks} representa el número de ondas R presentes en la señal y se obtiene mediante una función para la detección de máximos y T es el tiempo total de registro de la señal. Finalmente, la amplitud del complejo QRS se estima obteniendo el valor máximo de toda la señal ECG.

Enfoque 2

En este segundo enfoque se consideran características de mayor relevancia clínica con el objetivo de analizar temporalmente cada componente de un latido cardíaco. Al igual que en el enfoque 1, se define una rutina para extraer dicha información de la señal ECG recibida y comparar los valores obtenidos con valores de referencia definidos en la literatura. Las características fisiológicas para extraer y los valores de referencia considerados en este segundo enfoque se muestran en la tabla 3.6.

Variable	Valor
Frecuencia cardíaca	60-100 lpm
Duración del complejo QRS	≤ 110 ms
Duración de la onda P	≤ 120 ms
Duración del intervalo PR	≤ 200 ms

Tabla 3.6: Características fisiológicas para extraer de la señal $x(t)$ y sus valores de referencia para el enfoque de medidas estadísticas relativas.

Para calcular la frecuencia cardíaca se utiliza la función `heart-rate`, del módulo de Python `biosppy.signals.ecg`

La Figura 3.14 presenta el proceso para calcular la duración del complejo QRS en la señal $x(t)$. Primero, se localizan las ondas R a lo largo de toda la señal utilizando la función `rpeaks`. Una vez localizadas las

posiciones de las ondas R, se delimitan las ventanas de búsqueda, estas ventanas se establecieron con una longitud de 60 muestras, lo que corresponde a un intervalo de 120 ms alrededor de cada onda R, abarcando 30 muestras antes y 30 muestras después, permitiendo así capturar el complejo QRS completo. Posteriormente, para localizar el punto Q se identifica el valor mínimo en la ventana izquierda (previa de la onda R), mientras que para el punto S se identifica el valor mínimo de la ventana derecha (posterior a la onda R). Finalmente, la duración del complejo QRS se determina como la diferencia en las posiciones de las onda S y Q. Una vez obtenidas las duraciones individuales del complejo QRS para cada latido, se calcula la duración promedio en milisegundos, permitiendo así una evaluación de la duración del complejo QRS.

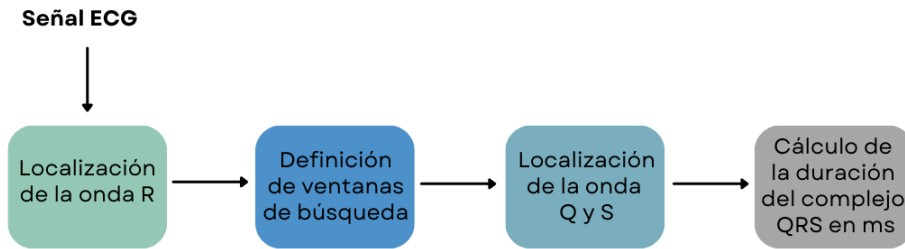


Figura 3.14: Metodología para calcular el tiempo de duración del complejo QRS.

La Figura 3.15 presenta el proceso para calcular la duración de la onda P en la señal $x(t)$. Primero, se localizan las ondas R, con la función `rpeaks`, lo que permite establecer el valor de referencia para las ventanas de búsqueda necesarias para analizar la onda P. Posteriormente, se delimitan las ventanas de búsqueda empleando una longitud de 100 muestras, lo que corresponde a un intervalo de 200 ms previo a cada onda R. En la tercera etapa, dentro de cada ventana se identifica el valor máximo de amplitud, el cual corresponde a la onda P. Para determinar el inicio y el final de la onda P, en la cuarta etapa se seleccionan sub-ventanas alrededor de la onda P, es decir, a la izquierda y a la derecha del punto máximo encontrado. Posteriormente, se localiza el inicio y el final de la onda P en cada una de estas sub-ventanas buscando los valores mínimos, que representan los límites inicial y final de la onda P. Una vez obtenidas las duraciones individuales de la onda P para cada latido, finalmente se calcula la duración promedio en milisegundos.

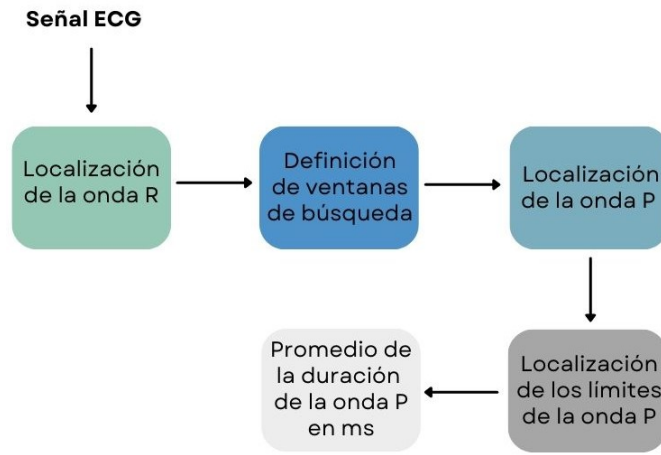


Figura 3.15: Metodología para calcular el tiempo de duración de la onda P.

Finalmente, la Figura 3.16 presenta el proceso para calcular la duración del intervalo PR en la señal $x(t)$, para este proceso, la primera, segunda y tercera etapas son las descritas para el proceso para calcular la duración de la onda P. En la cuarta etapa, la duración del intervalo se calcula como la diferencia entre la onda R y la onda P, una vez que se tienen las duraciones de los intervalos para cada latido, se calcula la duración promedio en milisegundos.

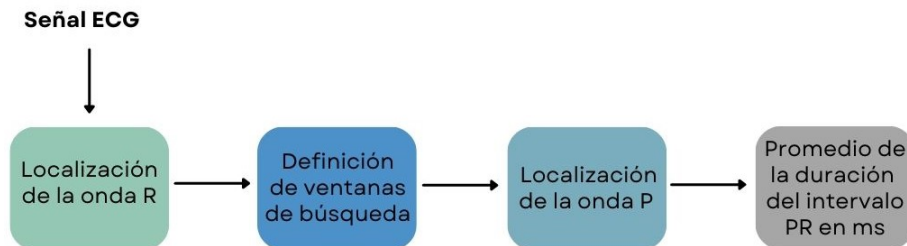


Figura 3.16: Metodología para calcular el tiempo de duración del intervalo PR.

3.3.4. Integrador de decisiones: Clasificador

Una vez que se obtienen las salidas del comparador y del módulo de análisis paramétrico, estas salidas se integran en el módulo clasificador utilizando una serie de operaciones lógicas para obtener una decisión en torno a la clasificación de la señal ECG recibida. En total son dos propuestas de clasificación las que se proponen en este trabajo de tesis, las cuales se basan en la información de los módulos de análisis paramétrico y comparador. A continuación se presentan las propuestas en conjunto con el algoritmo que describe a cada una de ellas.

Propuesta de clasificación 1: Medidas estadísticas absolutas

La propuesta de clasificación 1 se basa en las estrategias 1 y 2 para la definición del umbral β y el enfoque 1 para el análisis temporal de las señales, ambas definidas en las subsecciones previas.

La decisión de clasificación se basa en una regla AND como se describe en el algoritmo 1. Cuando la salida es $class = 0$ indica que la señal ECG recibida puede ser considerada normal; por el contrario cuando la salida es $class = 1$ indica que la señal ECG recibida puede ser considerada anormal.

Propuesta de clasificación 2: Medidas estadísticas relativas

La propuesta de clasificación 2 se basa en la estrategia 3 para la definición del umbral β y el enfoque 2 para el análisis temporal de las señales. En general, la clasificación de las señales también se basa en una regla AND, sólo que al haber una mayor cantidad de elementos de decisión, la estructura del algoritmo se vuelve más compleja. En el algoritmo 2 se describen los pasos a seguir para el proceso de clasificación. Nuevamente, cuando la variable de salida del algoritmo es $class = 0$ indica que la señal ECG recibida puede ser considerada normal; por el contrario cuando la salida es $class = 1$ indica que la señal ECG recibida puede ser considerada anormal.

Algorithm 1 Algoritmo clasificador de ECG

Require: r α β bpm V_{max}

```

1:  $P_n \leftarrow 0$ 
2: for  $k \leftarrow 1$  to  $K$  do
3:   if  $r[k] \geq \alpha$  then
4:      $P_n \leftarrow P_n + 1$ 
5:   end if
6: end for
7:  $class \leftarrow 1$ 
8: if  $P_n \geq \beta$  then
9:   if bpm  $\geq 60$  & bpm  $\leq 100$  then
10:    if  $V_{max} \geq 0,5$  then
11:       $class \leftarrow 0$ 
12:    end if
13:  end if
14: end if

```

3.4. Evaluación

Para evaluar la confiabilidad del algoritmo se propone el uso de dos métricas: la sensibilidad y la especificidad. La sensibilidad se refiere a la capacidad del algoritmo para identificar correctamente la presencia de una arritmia en una señal ECG. Por otro lado, la especificidad corresponde a la capacidad de clasificar como normal una señal ECG que presenta ritmo sinusal. Ambas métricas son importantes, por lo que mantener tanto una alta sensibilidad como una alta especificidad es prioridad en los procesos de diseño y configuración del algoritmo clasificador. No obstante, como se discutirá en lo subsecuente, estas métricas representan compromisos de diseño, lo que implica que incrementar una inevitablemente implicará reducir la otra en una determinada proporción.

Algorithm 2 Algoritmo para la clasificación de señales ECG

Require: \mathbf{r} , α , \hat{p} , n , \hat{Z} , $Z_{\alpha/2}$, bpm, QRS_{ms} , P_{ms} , PR_{ms}

- 1: $P_n \leftarrow 0$, $K \leftarrow \text{length}(\mathbf{r})$
- 2: **for** $k \leftarrow 1$ to K **do**
- 3: **if** $r[k] \geq \alpha$ **then**
- 4: $P_n \leftarrow P_n + 1$
- 5: **end if**
- 6: **end for**
- 7: $p \leftarrow \frac{r_i}{\frac{M}{2}}$, $q \leftarrow 1 - p$
- 8: $\hat{Z}_p \leftarrow \frac{p - \hat{p}}{\sqrt{\frac{pq}{n}}}$
- 9: **if** $\hat{Z}_p \leq -Z_{\alpha/2}$ **then**
- 10: $class \leftarrow 1$
- 11: **end if**
- 12: **if** $\hat{Z}_p \geq -Z_{\alpha/2}$ **then**
- 13: **if** bpm ≥ 60 & bpm ≤ 100 **then**
- 14: **if** $QRS_{ms} \leq 110$ **then**
- 15: **if** $P_{ms} \leq 120$ **then**
- 16: **if** $PR_{ms} \leq 200$ **then**
- 17: $class \leftarrow 0$
- 18: **end if**
- 19: **end if**
- 20: **end if**
- 21: **end if**
- 22: **end if**
- 23: **return** class

3.4.1. Modelos para el cálculo de las métricas de efectividad

La sensibilidad (S) de una prueba diagnóstica se define como la probabilidad condicionada de obtener un resultado positivo, dado que efectivamente existe presencia de la condición detectable. Así mismo, la especificidad (E) de la prueba diagnóstica corresponde a la probabilidad condicionada de obtener un resultado negativo, dado que ciertamente prevalece la ausencia de la condición detectable.

Para plantear estos conceptos en término de las probabilidades correspondientes, es necesario definir los siguientes eventos, que en este caso están contextualizados de acuerdo al problema que se aborda en este proyecto:

- D : El ECG presenta una arritmia cardíaca.
- H : El ECG presenta ritmo sinusal.
- P : El algoritmo decide que hay una arritmia cardíaca.
- N : El algoritmo decide que hay ritmo sinusal.

En este sentido, los modelos de cálculo para (S) y (E) están dados por:

$$S = \Pr(P | D) = \frac{f(x(t); P \cap D)}{f(x(t); D)} \quad (3.21)$$

$$E = \Pr(N | H) = \frac{f(x(t); N \cap H)}{f(x(t); H)} \quad (3.22)$$

Los términos expresados como $f(x(t); \cdot \cdot \cdot)$ en las ecuaciones 3.21 y 3.22 se refieren a las distribuciones de probabilidad de las amplitudes de la señal ECG que se analiza bajo los supuestos especificados en cada función de distribución; por ejemplo, para el supuesto denotado por $P \cap D$ se refiere a la distribución conjunta de la señal ECG cuando hay una arritmia presente y esta es detectada por el algoritmo. La obtención de las funciones de distribución de probabilidad necesarias para derivar las probabilidades teóricas que describen las ecuaciones 3.21 y 3.22 es un proceso complejo, el cual queda fuera de los objetivos de este proyecto de tesis. No obstante, se optó por emplear un enfoque frecuentista para obtener las aproximaciones de (S) y (E) a partir de una simulación tipo Monte Carlo. Para esto, se seleccionaron dos muestras de señales previamente clasificadas con base en el estándar de oro definido en el trabajo [34], un grupo corresponde a señales ECG con ritmo sinusal y el otro corresponde a señales diagnosticadas con arritmia. Con base en esta información preliminar, las señales fueron ingresadas al algoritmo y se registró la clasificación para cada una de ellas. De acuerdo a este procedimiento, se obtuvieron cuatro posibles resultados a la salida del algoritmo:

- Verdadero negativo (TN): Señal ECG con ritmo sinusal clasificada como **negativo a arritmia**. En función de los eventos previamente descritos, este resultado correspondería a la ocurrencia del evento compuesto denotado por $H \cap N$.
- Falso Negativo (FN): Señal ECG con arritmia clasificada como **negativo a arritmia**. Respecto a los eventos previamente descritos, en este caso, el resultado correspondería a la ocurrencia del evento compuesto denotado por $D \cap N$.
- Verdadero positivo (TP): Señal ECG con arritmia clasificada como **positivo a arritmia**. Para este resultado el evento compuesto está dado por $D \cap P$.
- Falso positivo (FP): Señal ECG con ritmo sinusal clasificada como **positivo a arritmia**. Finalmente, el evento compuesto correspondiente sería $H \cap P$.

Así, es posible definir estimadores para la sensibilidad y la especificidad, a los que denotaremos con \hat{S} y \hat{E} , en función de la cardinalidad de los eventos $D \cap P$, $D \cap N$, $H \cap P$ y $H \cap N$ quedando de la siguiente manera:

3.23 y 3.24 respectivamente.

$$\hat{S} = \frac{n(TP)}{n(FN) + n(TP)} \quad (3.23)$$

$$\hat{E} = \frac{n(TN)}{n(FP) + n(TN)} \quad (3.24)$$

Donde $n(A)$ indica la cardinalidad del conjunto A .

Capítulo 4

Análisis de resultados

La evaluación de los algoritmos 1 y 2 se implementó en el entorno de Python, el cual se seleccionó por su versatilidad y variedad de librerías disponibles que facilitan el análisis de datos y procesamiento de las señales. En particular, se utilizaron las siguientes librerías:

- Numpy: Es una librería que ofrece una amplia gama de funciones para operaciones matemáticas, manipulación de arreglos, transformadas de Fourier discretas, álgebra lineal básica, estadísticas, simulación aleatoria y otras tareas computacionales avanzadas. [42]
- Pandas: Es una librería diseñada para la manipulación y análisis de datos estructurados. Proporciona estructuras de datos eficientes, como *dataframes*, que permiten realizar operaciones como filtrado, limpieza, agrupamiento, transformación y análisis de grandes conjuntos de datos con facilidad. [43]
- Matplotlib: Es una librería para la creación de gráficos y visualización de datos. Permite generar gráficos estáticos, interactivos y animados en diversos formatos, como líneas, barras, histogramas y gráficos de dispersión, facilitando el análisis visual y la interpretación de resultados. [44]
- Biosppy: Es una librería diseñada específicamente para el procesamiento y análisis de señales fisiológicas. Ofrece funciones para tareas como el filtrado, la detección de picos, la extracción de

4.1. PROPUESTA DE CLASIFICACIÓN 1: MEDIDAS ESTADÍSTICAS ABSOLUTAS

características y la segmentación de datos, siendo particularmente útil para trabajar con señales como ECG, fotopletomografía (PPG, por sus siglas en inglés *photoplethysmography*) y respiración. [38]

Para la evaluación del algoritmo XCorr-In primero se seleccionaron las señales de la base de datos publicada por [34] que cumplen con el criterio de interés para este trabajo, es decir, señales ECG correspondientes a mujeres en el rango de edad entre 40 y 60 años. El conjunto de señales resultante para la evaluación se compone de 382 señales con ritmo sinusal y 463 señales con arritmia. Una vez seleccionadas las señales, cada una se utiliza como entrada a los algoritmos descritos en el capítulo previo y se registra el resultado de su clasificación, ya sea $class = 1$ (correspondiente a una arritmia) o $class = 0$ (correspondiente a ritmo sinusal). A partir de los resultados registrados, se calculan las métricas de evaluación, sensibilidad (\hat{S}) y especificidad (\hat{E}), utilizando las ecuaciones 3.23 y 3.24 y se producen las gráficas correspondientes.

4.1. Propuesta de clasificación 1: Medidas estadísticas absolutas

De acuerdo a la descripción del algoritmo 1, para su operación se requieren cuatro parámetros de configuración: α , β , bpm y V_{max} . Recordemos que α es un parámetro que determina la fuerza de correlación esperada entre la señal analizada y el pulso de referencia, previamente denotado como $x_{ref}(t)$, y que puede tomar valores entre $[0, 1]$. β representa la ocurrencia esperada de índices de correlación que sobrepasen el umbral α , es decir, $n(r[m] \geq \alpha)$ para una señal ECG con ritmo sinusal. Finalmente, bpm y V_{max} corresponden a variables fisiológicas establecidas en la literatura médica.

Con el propósito de indagar sobre el impacto del umbral α en esta primera propuesta, se analizó el desempeño del algoritmo considerando valores para α en el intervalo definido por $[0,5, 1]$, considerando incrementos de $\Delta\alpha = 0,05$. No se consideran valores de α menores a 0.5 porque los niveles de correlación débiles comprometen la sensibilidad del algoritmo, al incrementar la probabilidad de correspondencia de la

4.1. PROPUESTA DE CLASIFICACIÓN 1: MEDIDAS ESTADÍSTICAS ABSOLUTAS

señal recibida con la señal de referencia, incluso cuando la señal recibida no presente ritmo sinusal. Con respecto a las variables bpm y V_{max} , se consideraron los valores definidos en la Tabla 3.5.

Por otro lado, con respecto al parámetro β , se definieron dos procesos de evaluación independientes para el algoritmo 1. El primero considera la estrategia basada en el promedio del número índices de correlación que cumplen con $r[m] \geq \alpha$, es decir, el valor de β se obtuvo de acuerdo a la ecuación 3.6. El segundo proceso de evaluación utiliza la estrategia basada en los deciles para seleccionar el valor de β , utilizando como fundamento lo expuesto en el modelo definido por la ecuación 3.7.

Es importante mencionar que la selección del valor de β está estrechamente relacionada con el valor seleccionado para α , puesto que, al modificar el valor de α , la distribución $f(r)$ que describe el comportamiento de los índices de correlación cambiará. Esto significa que se debe realizar el proceso de análisis paramétrico para establecer el valor de β cada vez que se modifica el valor de α .

Para ejemplificar lo anterior y de acuerdo al análisis paramétrico presentado en el capítulo previo para el promedio de los índices de correlación, para el valor de $\alpha = 0,8$, el umbral β para señales ECG que presentan las características descritas en la Tabla 3.1 sería $\beta_\mu = 33$. Esto significa que, para que una señal ECG compuesta por 10 latidos y que presenta ritmo sinusal sea clasificada como $class = 0$, se espera que al menos haya 33 índices de correlación que cumplen con $r[m] \geq 0,8$, y que cualquier cantidad inferior a este valor indicará la presencia de una arritmia. Siguiendo este proceso de análisis, se obtuvieron los valores para β_μ correspondientes a los distintos valores de α definidos para la evaluación, lo cuales se presentan en la Tabla 4.1

La Figura 4.1 resume el comportamiento del algoritmo 1 respecto a la sensibilidad y la especificidad para los distintos valores de α evaluados y sus correspondientes valores para β_μ . En la gráfica se presentan dos curvas, una para la sensibilidad y otra para la especificidad. Respecto a la sensibilidad es posible notar que el algoritmo basado en el promedio de los índices de correlación es altamente capaz de identificar la presencia de una señal ECG con arritmia, puesto que para todos los valores de α y β_μ evaluados el algoritmo logra niveles de sensibilidad por encima del 0.95, significando que fue capaz de detectar la presencia

4.1. PROPUESTA DE CLASIFICACIÓN 1: MEDIDAS ESTADÍSTICAS ABSOLUTAS

Valor de α	Valor de β_μ
0,5	70
0,55	64
0,6	57
0,65	52
0,7	46
0,75	39
0,8	33
0,85	26
0,9	16
0,95	6
1	0

Tabla 4.1: Correspondencia de los valores para β_μ obtenidos para cada valor de α en el intervalo evaluado.

de arritmias en el 95 % de las señales ECG que efectivamente presentaban una arritmia. No obstante, esta estrategia no resulta ser efectiva en términos de la especificidad, ya que el valor más alto que se logra es de 0.6282. Esto significa que el algoritmo no resulta eficiente cuando se trata de clasificar una señal con ritmo sinusal como tal, fallando en el 37 % de las señales ECG con ritmo sinusal analizadas.

Con respecto al segundo proceso de evaluación, dado que el valor de β se obtiene a partir de los deciles de y_i , es necesario identificar el decil óptimo que maximiza el desempeño del algoritmo para cada valor de α .

4.1. PROPUESTA DE CLASIFICACIÓN 1: MEDIDAS ESTADÍSTICAS ABSOLUTAS

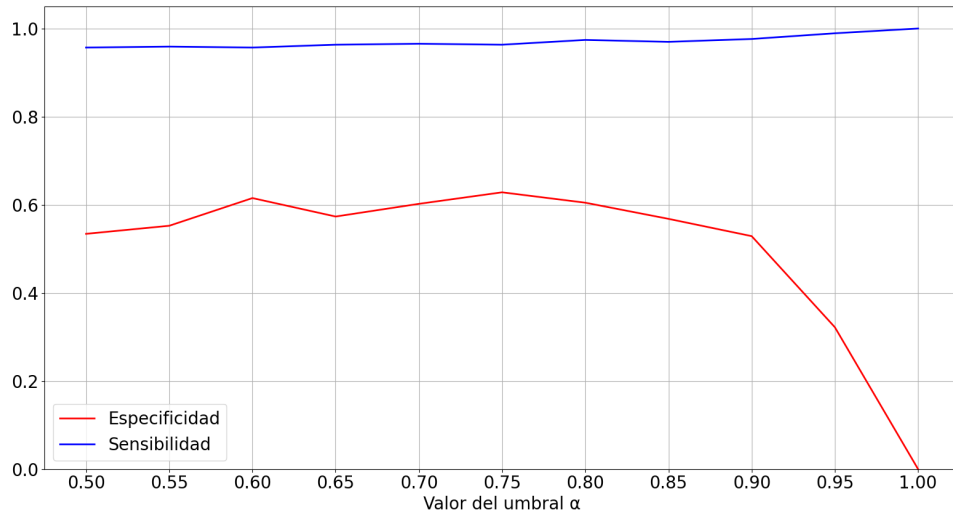


Figura 4.1: Métricas de calidad para el algoritmo 1 en función del valor de α , considerando la estrategia β_μ

Para ello, se evaluó el algoritmo 1 considerando los distintos valores de α y los deciles correspondientes a la distribución acumulativa de y_i , los cuales se obtienen a partir del procedimiento descrito por la ecuación 3.7. Los valores para los deciles con respecto a cada valor de α se presentan en la Tabla 4.2. Para comprender mejor los datos presentados en la Tabla 4.2, tomemos como ejemplo el valor correspondiente al segundo decil, d_2 , para $\alpha = 0,8$, el cual es 31. Este valor indica que, de las 45 señales analizadas para establecer la distribución acumulativa de y_i , el 80% presentó 31 o más índices de correlación que cumplen con $r[m] \geq 0,8$. Note que el porcentaje está asociado al decil, y no al valor de α , así, si elegimos el primer decil, d_1 , el valor 16 indica que el 90% de las 45 señales presentaron 16 o más índices de correlación que cumplen con $r[m] \geq 0,8$.

La especificidad y sensibilidad del algoritmo obtenida cuando el umbral β_{d_i} es igual a los deciles listados en la Tabla 4.2 para $\alpha = 0,6$, $\alpha = 0,7$ y $\alpha = 0,8$ se presentan en las Figuras 4.2, 4.3 y 4.4, respectivamente. A partir de las curvas presentadas es posible notar como al incrementar el valor del decil la especificidad disminuye, mientras que la sensibilidad aumenta. Esto se debe a que los valores más altos de β_{d_i} establecen un

4.1. PROPUESTA DE CLASIFICACIÓN 1: MEDIDAS ESTADÍSTICAS ABSOLUTAS

Decil	Valor de α										
	0,5	0,55	0,6	0,65	0,7	0,75	0,8	0,85	0,9	0,95	1
d_1	56	48	43	41	33	24	16	5	0	0	0
d_2	63	58	52	48	42	38	31	21	4	0	0
d_3	66	60	56	51	46	39	33	24	13	0	0
d_4	69	64	58	54	48	41	35	27	17	0	0
d_5	73	66	60	55	49	43	36	29	20	1	0
d_6	74	68	62	58	50	44	37	30	21	9	0
d_7	76	70	66	59	53	46	39	33	23	11	0
d_8	80	74	67	62	54	48	41	34	24	13	0
d_9	85	78	70	64	57	51	43	36	27	15	0
d_{10}	93	85	70	72	65	55	49	41	29	19	0

Tabla 4.2: Deciles obtenidos para los distintos valores de α , a partir del vector de índices de correlación generado para las 45 señales ECG seleccionadas en la muestra definida en la sección 3.2.2.

criterio más exigente para clasificar las señales como ECG con ritmo sinusal.

A partir del análisis de estos resultados, se identificó al decil d_1 como el que proporciona un mejor balance entre ambas métricas, logrando valores de sensibilidad por encima de 0.92 para todos los valores de α evaluados, mientras que la especificidad se mantiene por encima de 0.71 para los valores de $\alpha \leq 0,9$. En el caso de los valores de $\alpha > 0,9$, la sensibilidad cae por debajo de 0.43 para todos los deciles evaluados. Utilizar

4.1. PROPUESTA DE CLASIFICACIÓN 1: MEDIDAS ESTADÍSTICAS ABSOLUTAS

los deciles más bajos como d_1 , favorece la clasificación de señales con ritmo sinusal pero reduce la sensibilidad, mientras que utilizar deciles más altos como d_5 , disminuyen drásticamente la especificidad.

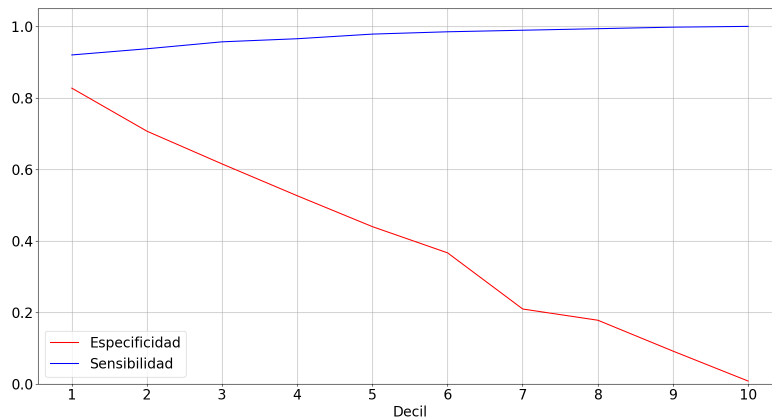


Figura 4.2: Análisis del comportamiento de la sensibilidad y la especificidad en función de los deciles, considerando que $\alpha = 0,6$.

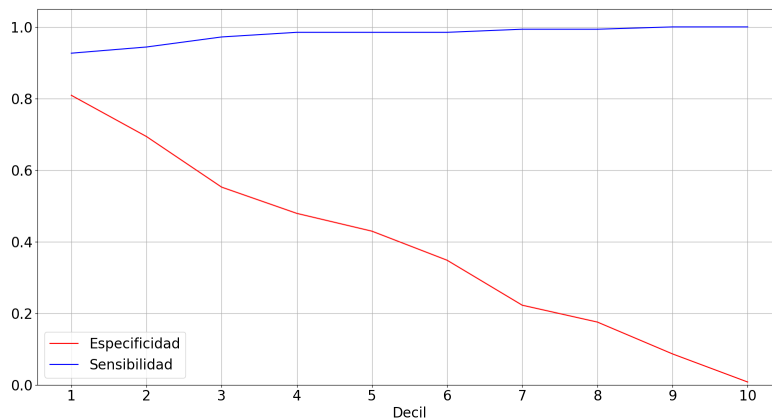


Figura 4.3: Análisis del comportamiento de la sensibilidad y la especificidad en función de los deciles, considerando que $\alpha = 0,7$.

Con objetivo de mostrar con mayor claridad el impacto de α en el desempeño de Algoritmo 1 cuando integra la estrategia de los deciles para la selección de los valores de β , en la Figura 4.5 se presentan los valores

4.1. PROPUESTA DE CLASIFICACIÓN 1: MEDIDAS ESTADÍSTICAS ABSOLUTAS

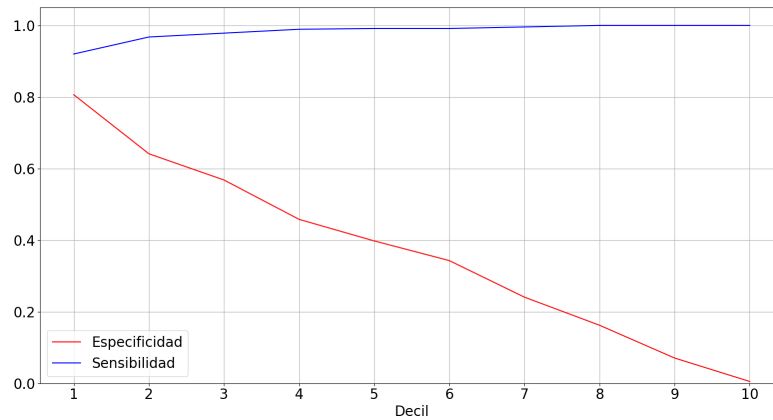


Figura 4.4: Análisis del comportamiento de la sensibilidad y la especificidad en función de los deciles, considerando que $\alpha = 0,8$.

de sensibilidad y especificidad en función de α , considerando β_{d_1} como criterio de decisión. A partir de las curvas presentadas, es posible notar que ambas métricas permanecen prácticamente constantes para valores de α entre 0,5 y 0,85, observando ligeras variaciones en la sensibilidad, mientras que para los valores de $\alpha \geq 0,9$ la especificidad decae significativamente. De esta gráfica podemos concluir que el valor de α que presenta el mejor compromiso es 0,6, dando lugar a una sensibilidad de 0.92 y una especificidad de 0.8272.

En la Figura 4.6 se observa la comparación de la sensibilidad y la especificidad del algoritmo 1 para los criterios de decisión basados en β_μ y β_{d_1} . A partir de las curvas podemos confirmar como la integración de la distribución de los índices de correlación como parte del criterio de decisión mejora la clasificación hecha por el algoritmo, reduciendo significativamente la tasa de falsos positivos. Por ejemplo, para $\alpha = 0,6$, el criterio basado en el promedio logra una tasa de falsos positivos igual a 0,3849 y una tasa de falsos negativos igual a 0,0432, dando como resultado un error total igual al 19,86 % de las decisiones, mientras que el criterio basado en el primer decil obtuvo una tasa de falsos positivos igual a 0,1727 y una tasa de falsos negativos igual a 0,0799, dando como resultado un error total igual al 12,18 % de las decisiones, significando una reducción del 7,68 % del error total del algoritmo.

4.1. PROPUESTA DE CLASIFICACIÓN 1: MEDIDAS ESTADÍSTICAS ABSOLUTAS

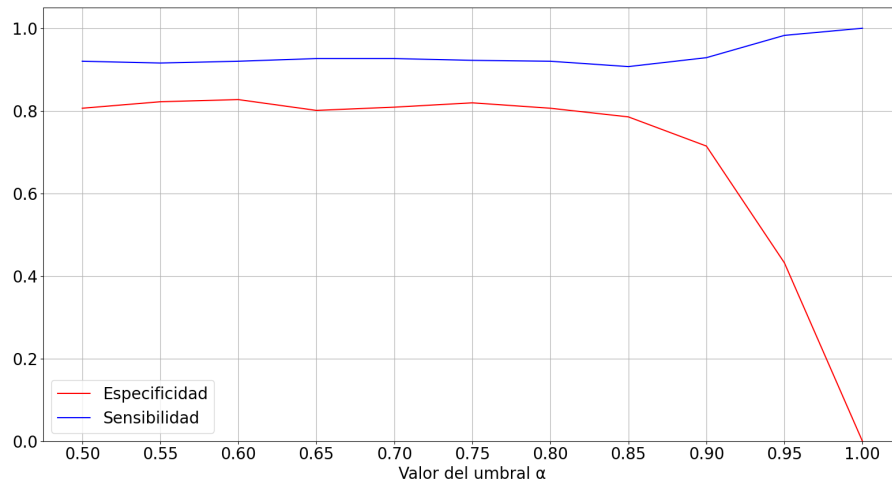


Figura 4.5: Métricas de calidad para el algoritmo 1 en función del valor de α , considerando los valores correspondientes a β_{d_1}

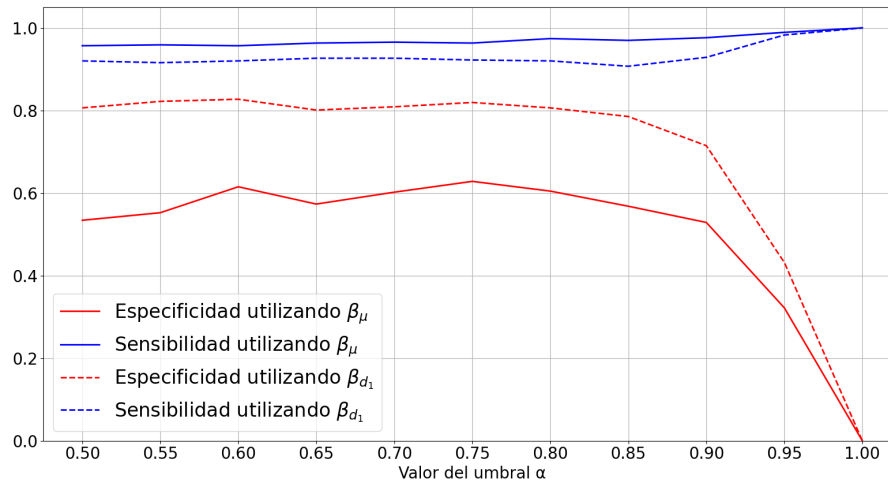


Figura 4.6: Comparación de las métricas de calidad para el algoritmo 1 en función del valor de α , considerando el valor de β_μ y de β_{d_1} .

4.2. Propuesta de clasificación 2: Medidas estadísticas relativas

En esta segunda propuesta de clasificación se integra una mayor cantidad de parámetros de operación del algoritmo con el objetivo de volverlo más robusto a la incertidumbre generada por la variabilidad en la forma de onda de las señales ECG y la duración del registro de la misma. En este sentido, para la operación del Algoritmo 2 se requieren los siguientes parámetros de operación: α , \hat{p} , n , $\beta_{\hat{p}}$, bpm , t_{QRS} , t_P , t_{PR} .

Al igual que en la propuesta 1, α define el grado de similitud esperado entre la señal analizada y el pulso de referencia, \hat{p} representa la estimación puntual de la proporción poblacional de los índices de correlación que cumplen con $r[m] \geq \alpha$ con respecto a la cantidad total de índices de correlación generados para el registro de la señal, $\beta_{\hat{p}}$ representa el criterio de decisión basado en el análisis paramétrico de la distribución de probabilidades para proporción muestral y bpm , t_{QRS} , t_P y t_{PR} corresponden a las variables fisiológicas que se consideran en la literatura médica como críticas para la identificación de trastornos del ritmo cardíaco.

Para evaluar el desempeño del Algoritmo 2, nuevamente se consideraron valores para α en el intervalo definido por $[0,5, 1]$, con incrementos de $\Delta\alpha = 0,05$. Para las variables fisiológicas bpm , t_{QRS} , t_P y t_{PR} se establecieron los valores definidos en la Tabla 3.6. Por otro lado, para definir el criterio de decisión $\beta_{\hat{p}}$ es necesario obtener la estimación puntual de la proporción poblacional, \hat{p} , tal como se describe en la Ecuación 3.18. Para esto, se promediaron las proporciones muestrales para las 45 señales seleccionadas en la Sección 3.2.2 y se considera a este promedio como la estimación puntual \hat{p} . Es importante mencionar que se debe obtener el valor de \hat{p} para cada valor de α evaluado, puesto que, al variar α , varía la cantidad de índices de correlación que superan este valor. Las estimaciones puntuales para cada valor de α se presentan en la Tabla 4.3. Adicionalmente, en la Ecuación 3.18 se especifica que se requiere definir el nivel de significancia (γ) empleado para la ejecución de la prueba de hipótesis a partir de la cual se decide sobre la clasificación de la señal analizada. Para efectos de esta evaluación se están considerando los siguientes niveles de significancia: 0,01,

4.2. PROPUESTA DE CLASIFICACIÓN 2: MEDIDAS ESTADÍSTICAS RELATIVAS

0,05, 0,10 y 0,15. Los niveles de significancia nos permiten diseñar las pruebas de hipótesis a partir de un valor esperado para la ocurrencia de errores tipo I, que en este caso, sería una tasa teórica esperada de falsos positivos.

Valor de α	Valor de \hat{p}
0,5	0.03128
0,55	0.02856
0,6	0.02562
0,65	0.02326
0,7	0.02035
0,75	0.01756
0,8	0.01480
0,85	0.01146
0,9	0.007269
0,95	0.002567
1	0

Tabla 4.3: Valor del estimador puntual para la proporción poblacional (\hat{p}) correspondiente a cada valor de α en el intervalo evaluado.

En la Figura 4.7 se presentan cuatro gráficas que muestran curvas de sensibilidad y especificidad, una gráfica para cada nivel de significancia evaluado. La sensibilidad presenta un comportamiento similar en todas las gráficas con valores aproximadamente constantes y superiores a 0.96 para todos los valores de α evaluados. Esto se puede atribuir a que,

además de cambiar el enfoque para analizar los índices de correlación, se integró el análisis de características fisiológicas de mayor relevancia clínica, lo que permite que, además de analizar las características morfológicas de la señal a través de la función de correlación cruzada, también se analicen las características temporales de los componentes del latido cardíaco. De esta manera, al contar con más criterios de análisis, se facilita la identificación de señales con arritmia.

Por otra parte, se observa que la especificidad presenta un comportamiento decreciente para los cuatro niveles de significancia a medida que el valor de α aumenta. De hecho, el máximo nivel de especificidad obtenido es de 0,9083 cuando $\gamma = 0,01$ y $\alpha = 0,5$, generando un error total de tan solo el 4.61 % de las decisiones. La diferencia entre la tasa real de falsos positivos (0,0917) y la tasa teórica (0,01) se debe al error residual generado por la estimación puntual de la proporción poblacional. No obstante, incluso con este error residual, el criterio de selección basado en la proporción de índices de correlación presenta un error total menor que las estrategias basadas en el promedio y en los deciles.

4.3. Discusión de los resultados

La Figura 4.8 muestra la comparación de los resultados obtenidos para los tres criterios de decisión propuestos en este trabajo de tesis. Las líneas punteadas representan las métricas obtenidas al utilizar el algoritmo 2 con $\beta_{\hat{p}}$ y $\gamma = 0,01$, las líneas discontinuas representan las métricas obtenidas al utilizar el algoritmo 1 con el criterio definido por β_{d_1} y, por último, las líneas sólidas representan las métricas obtenidas al utilizar el algoritmo 1 con el criterio definido por β_{μ} .

En el Apéndice 1 se presentan las Tablas completas con los resultados numéricos de las evaluaciones realizadas. Haciendo un análisis exhaustivo de estos valores, y tomando como base el error total, se tiene que para el criterio de decisión basado en β_{μ} , el valor de α que da lugar al menor error es 0,75, observado un 18.82 % de decisiones erróneas; para el criterio de decisión basado en β_{d_1} , el menor error se obtuvo para $\alpha = 0,6$, correspondiendo al 12.18 % de las decisiones; finalmente, para el criterio basado en $\beta_{\hat{p}}$, el menor se obtuvo para $\gamma = 0,01$ y $\alpha = 0,5$,

4.3. DISCUSIÓN DE LOS RESULTADOS

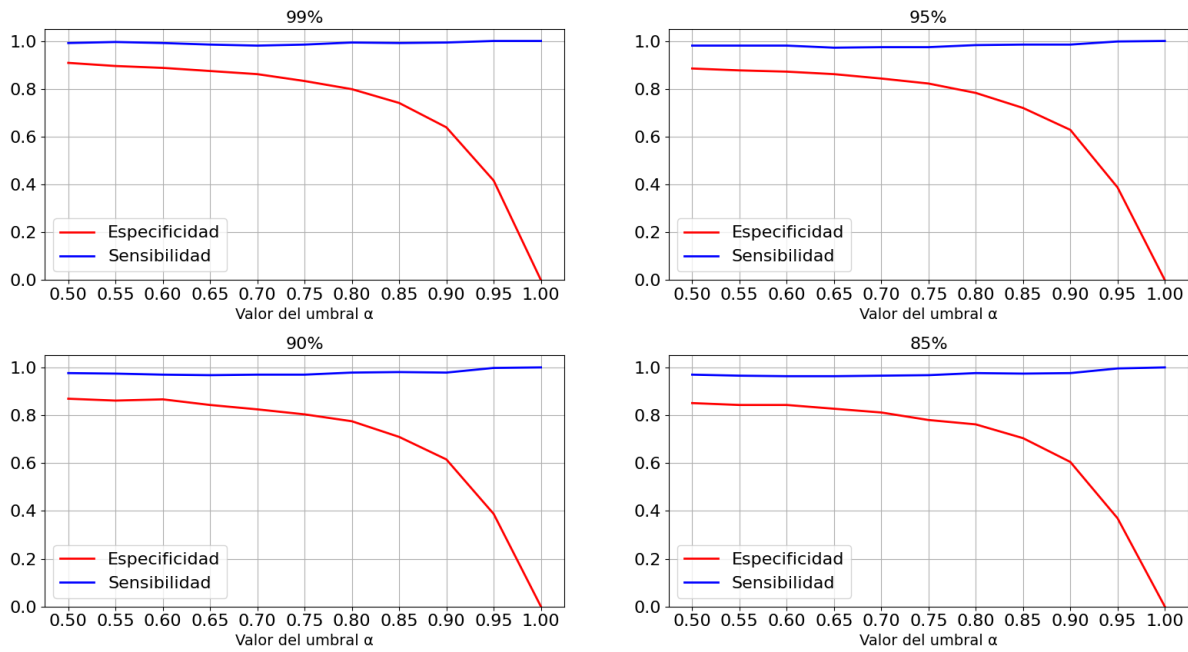


Figura 4.7: Comparación de \hat{E} y \hat{S} para las evaluaciones con los diferentes niveles de significancia.

resultando en el 4.61 % de las decisiones. Esto se resume en la Tabla 4.4.

Criterio de decisión	α	\hat{E}	\hat{S}	FP	FN	Error total
β_{μ}	0.75	0.6282	0.9632	0.3718	0.0368	0.1822
β_{d_1}	0.6	0.8272	0.9201	0.1728	0.0799	0.1219
$\beta_{\hat{p}, \gamma=0,01}$	0.5	0.9084	0.9914	0.0916	0.0086	0.0461

Tabla 4.4: Configuraciones óptimas para los tres criterios de decisión propuestos.

Para el algoritmo XCorr-In se busca tener una baja tasa de falsos positivos y falsos negativos. Un falso positivo indica que el algoritmo ha

4.3. DISCUSIÓN DE LOS RESULTADOS

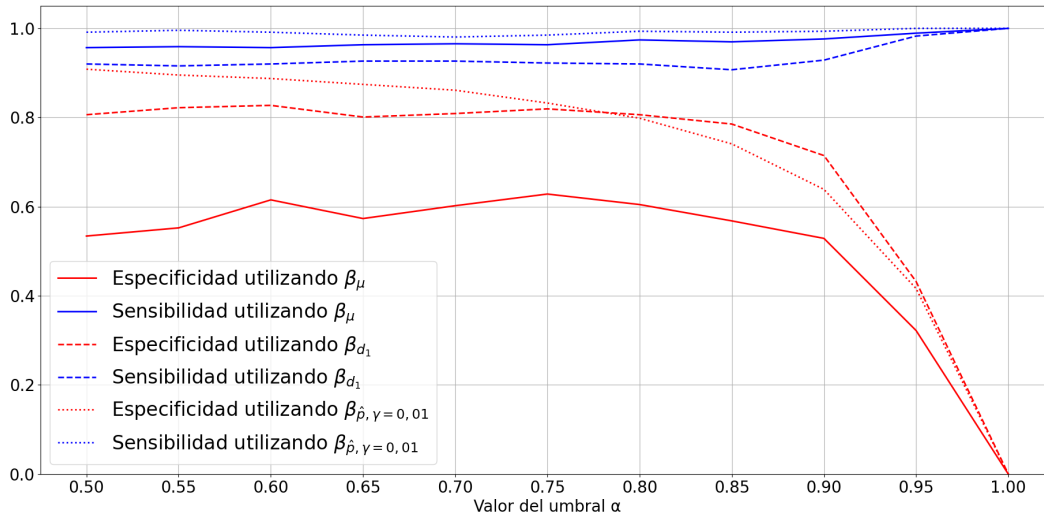


Figura 4.8: Comparación de las métricas de calidad para el algoritmo 1 en función del valor de α , considerando el valor de β_μ , de β_{d_1} y $\beta_{\hat{p}, \gamma=0,01}$

clasificado erróneamente una señal anormal como normal, mientras que un falso negativo indica que el algoritmo ha clasificado erróneamente una señal normal como anormal.

Desde un enfoque clínico, es fundamental priorizar una alta sensibilidad, ya que esto reduce la tasa de falsos negativos. Un falso negativo puede tener consecuencias graves para la salud del paciente, ya que clasificar erróneamente una señal anormal como normal podría impedir que el paciente reciba la atención o el tratamiento necesarios. Por esta razón, la sensibilidad debe tener prioridad sobre la especificidad. Sin embargo, también es importante mantener una especificidad alta. Clasificar una señal normal como anormal, aunque no tendría un impacto tan crítico en la salud del paciente, implicaría una pérdida de recursos, como tiempo del personal de salud para realizar diagnósticos adicionales y el uso de recursos clínicos innecesarios.

Considerando los resultados obtenidos para el algoritmo XCorr-In mostrados en la Figura 4.8, el enfoque basado en medidas estadísticas relativas demuestra un desempeño superior en comparación con los enfoques de medidas estadísticas absolutas. Los resultados indican que el enfoque basado en la proporción, junto con un nivel de significancia igual a 0,01

4.3. DISCUSIÓN DE LOS RESULTADOS

es el más adecuado, maximizando el desempeño de las métricas. Este enfoque no solo permite alcanzar una sensibilidad más alta, una prioridad clave por su relevancia clínica, sino que también logra un equilibrio entre sensibilidad y especificidad. Esto se debe a que, al emplear una proporción en el enfoque de medidas estadísticas relativas, se incorpora la variabilidad de las señales ECG. Este enfoque permite una mejora en la especificidad, ya que se tiene en cuenta un análisis más completo de las señales. Además, al considerar parámetros clínicos adicionales, como la duración del complejo QRS, la duración de la onda P y el intervalo PR, se realiza un análisis más detallado de las características morfológicas y temporales de las señales ECG. Estas consideraciones contribuyen significativamente a la mejora de las métricas observadas.

Capítulo 5

Conclusiones y trabajo a futuro

5.1. Conclusiones y contribuciones

En este trabajo, se propone el algoritmo titulado como XCorr-In, el cual es un algoritmo clasificador de señales electrocardiográficas basado en los índices de correlación cruzada, que incorpora características morfológicas y temporales de pacientes femeninas de entre 40 a 60 años. La evaluación para el algoritmo XCorr-In se realizó mediante su capacidad de identificar la presencia de arritmias en registros electrocardiográficos de pacientes con trastornos del ritmo diagnosticados por un panel de expertos en cardiología, así como su capacidad de identificar el ritmo sinusal en los registros electrocardiográficos de corazones sanos.

Para la implementación del algoritmo se propusieron tres enfoques. El primero se basa en el promedio de los índices de correlación, el segundo se basa en los cuartiles de la distribución empírica de los índices de correlación, y el tercero se basa en la proporción de índices de correlación que cumplen con un determinado criterio. Los primeros dos enfoques se consideran como propuestas de medidas absolutas ya que la clasificación se realiza con base en los valores absolutos de medidas estadísticas; mientras que el tercer enfoque se considera una propuesta de medidas relativas, ya que opera sobre la razón de índices de correlación respecto al total de índices en un determinado registro.

A continuación, se subrayarán los principales hallazgos identificados a partir de la evaluación de las métricas de desempeño del algoritmo Xcorr-In en sus tres enfoques.

Propuesta de clasificación 1: Medidas estadísticas absolutas

Dentro de la propuesta de clasificación 1, se tienen dos enfoques. El primero considera el promedio de los índices de correlación calculado a partir de un conjunto de señales muestra como criterio de clasificación, denotado por β_μ , y el segundo, que considera el primer decil de la distribución empírica de los índices de correlación para el mismo conjunto muestra de señales ECG como criterio de clasificación y que en este documento es referido como β_{d_1} .

De acuerdo a los resultados obtenidos para β_μ , al variar el umbral de correlación, denotado por α , de 0,5 a 1 con un factor de variación $\Delta\alpha = 0,05$, se observa que a medida que α se acerca a 1, el valor de β_μ disminuye. Esto se debe a que α , representa el índice de correlación entre el pulso de referencia, denotado por $x_{ref}(t)$, y la señal ECG que se analiza, a la cual referiremos como $x(t)$. A medida que este índice aumenta, se reducen las coincidencias en la correlación a lo largo de la señal. Esto ocurre debido a la variabilidad de las señales, ya que es muy poco probable que dos señales ECG sean idénticas. Como resultado, exigir un valor más alto de α implica que habrá una menor cantidad de índices de correlación que superarán este valor, generando un menor valor esperado para β_μ .

Esta estrategia resulta restrictiva, ya que la señal $x(t)$ se compara con un valor único de β_μ y no permite integrar la variabilidad inherente en las señales ECG. Esto significa que, aunque la señal analizada $x(t)$ tenga un valor bastante cercano a β_μ , se clasificaría como una señal ECG con arritmia. En esta propuesta de clasificación no se considera la variabilidad que existe entre las señales, ya que incluso las señales clasificadas con ritmo sinusal no son idénticas. Otro aspecto sobre esta propuesta, es que el valor de β_μ depende de la duración de la señal ECG, ya que los valores obtenidos en la Tabla 4.1 son correspondientes al análisis de las n señales obtenidas de la base de datos, con una duración de 5000 muestras, esto quiere decir que si, la señal a analizar $x(t)$ presenta una duración diferente de 5000 muestras, los valores de

β_μ no serían representativos de esa señal, y sería necesario ajustar los valores de la Tabla 4.1 para señales con distintas duración o por otra parte, ajustar la duración de la señal $x(t)$. No obstante, definir este criterio sirve como una referencia para el resto de los enfoques.

Para los resultados obtenidos con esta estrategia, como se observa en la Figura 4.1, la sensibilidad (\hat{S}) es superior al 90 % para todos los valores de α . Por otro lado, la especificidad (\hat{E}) alcanza un valor máximo de 61.51 %. Aunque esta propuesta presenta un alta sensibilidad, su baja especificidad representa un problema en el contexto clínico. Una especificidad reducida implica que un paciente que presente una señal ECG con ritmo sinusal tendría aproximadamente un 40 % de probabilidad de ser clasificado como una señal con arritmia. Esto convella costos económicos para el paciente y perdida de recursos como tiempo y esfuerzo del personal de salud, ya que podría requerir estudios o pruebas adicionales para descartar la presencia de arritmias. Además, la salud del paciente podría verse afectada si se le administra un tratamiento o medicamento innecesario.

De acuerdo a los resultados obtenidos para β_{d_i} , en la Tabla 4.2 se observa el valor de β_{d_i} para los valores de α de 0,5 a 1 con $\Delta\alpha = 0,05$. En esta distribución se observa un comportamiento similar que para β_μ , a medida que el valor de α aumenta el valor de β_μ disminuye, en la Figura 4.2, 4.3 y 4.4, se observa el comportamiento de la especificidad y sensibilidad en función de los deciles, para los valores de α de 0,6, 0,7 y 0,8, en estas figuras se observa que la sensibilidad se mantiene constante y presenta un valor por encima del 90 %, sin embargo, la especificidad disminuye drásticamente, esto indica que en esta estrategia la distribución de los índices de correlación tiene un impacto en la especificidad, ya que el valor que toma β_{d_i} aumenta de acuerdo al número de decil, por lo que, este criterio se hará más elevado cada vez y esto indica que menos señales cumplirán con esta alta exigencia de similitud, dando como resultado una especificidad baja.

Para esta estrategia, se identificó al decil d_1 como el que proporciona un mejor balance entre ambas métricas, logrando una sensibilidad por encima del 92 % para todos los valores de α y una especificidad por encima de 71 % para los valores $\alpha \leq 0,9$. Utilizar deciles más bajos como d_1 , favorece la especificidad pero reduce la sensibilidad, en cambio, utilizar deciles más altos como d_5 , disminuyen drásticamente la

especificidad. Como se menciono anteriormente el impacto de la baja especificidad en un contexto clínico, es importante priorizar ambas métricas.

En la Figura 4.6 se muestra la comparación de la sensibilidad y especificidad para el algoritmo 1 en función del valor de α , considerando el valor de β_μ y β_{d_1} . Se observa que el valor de β_{d_1} presenta un mejor desempeño para la especificidad pero una reducción en la sensibilidad, sin embargo, esta reducción es más pequeña comparada con la mejora de la especificidad. La integración de la distribución de los índices de correlación como parte del criterio de decisión mejora la clasificación del algoritmo 1, la especificidad mejora lo que significa que la tasa falsos positivos se reduce.

Propuesta de clasificación 2: Medidas estadísticas relativas

Para esta segunda propuesta de clasificación, además de la integración de una mayor cantidad de parámetros de operación del algoritmo se requirió integrar un análisis paramétrico con el objetivo de volverlo más robusto a la incertidumbre generada por la variabilidad de la forma de onda de las señales ECG y la duración del registro de la señal. En esta propuesta al utilizar una estimación puntual de la proporción poblacional de los índices de correlación que cumplen con $r[m] \geq \alpha$ con respecto a la cantidad de total de índices de correlación generados para el registro de la señal. $\beta_{\hat{p}}$ representa un criterio de decisión basado en el análisis paramétrico de la distribución de probabilidades para proporción muestral. A diferencia de β_μ y β_{d_1} al utilizar $\beta_{\hat{p}}$ se toma en cuenta la variabilidad de las señales. Además la Figura 4.7 muestran las curvas de sensibilidad y especificidad para cuatro niveles de significancia, donde para los cuatro niveles se observa un comportamiento similar, siendo el nivel de significancia de 99%, el que presenta valores más elevados tanto para la sensibilidad y la especificidad.

En la Figura 4.8 se muestra la comparación de las métricas de calidad en función del valor de α , considerando el valor de β_μ , de β_{d_1} y $\beta_{\hat{p}, \gamma=0,01}$. En esta Figura se confirma la mejora del desempeño del algoritmo en cada estrategia, comenzando con el análisis de β_μ que presenta una especificidad baja, debido a la alta exigencia en el criterio de decisión de las señales, lo que ocasionaba que las señales con ritmo sinusal se

5.1. CONCLUSIONES Y CONTRIBUCIONES

clasificaran con señales con arritmias, posteriormente, se observa una mejora en la especificidad al utilizar β_{d_1} , esto demuestra que al utilizar como estrategia el análisis de la distribución de los índices de correlación mejora la especificidad, sin embargo, se observa una reducción en la sensibilidad. Por último, al utilizar $\beta_{\hat{p},\gamma=0,01}$ se observa una mejora en la especificidad, esto es debido a que al utilizar la proporción se tiene en cuenta la variabilidad de las señales, además el 2 cuenta con más parámetros de decisión como son la duración del complejo QRS, la duración de la onda P y el intervalo PR, esto permite realizar un análisis más detallado de las características morfológicas y temporales de las ECG, lo que da como resultado una mejora en las métricas, sobre todo en la especificidad.

De acuerdo a los resultados presentados en la Tabla 4.4, donde se presentan las configuraciones óptimas para los tres criterios de decisión propuestos, se observa que para cada una de las tres estrategias se presenta un valor óptimo de α diferente. En esta tabla se confirma que al implementar un analisis parametrico se reduce la tasa de falsos positivos, de acuerdo con los resultados presentados en Tabla, para la estrategia β_{μ} se tiene una tasa de falsos positivos de 0.3718, la cual se logra disminuir a 0.0916 al implementar un análisis de la distribución de los índices de correlación mejora la especificidad.

La estrategia para la cual el algoritmo XCorr-In presenta un mejor desempeño, es para $\beta_{\hat{p},\gamma=0,01}$, esta estrategia de clasificación presenta una especificidad de 0.9084, una sensibilidad de 0.9914, es esta estrategia la tasa de falsos positivos se logra reducir a 0.0916 y la de falsos negativos a 0.0086, presentando un error total de solo 0.0461. En un ambiente clínico ambas métricas tienen un impacto en la salud del paciente, por una parte, al tener una alta sensibilidad se esta asegurando que el 99.14% de los pacientes que presenten una señal ECG con arritmia serán diagnosticados correctamente, por lo que el personal de salud podrá empezar un diagnostico y tratamiento para el paciente; sería peligroso presentar una sensibilidad baja, ya que esto implica que un paciente que presenta una señal ECG con arritmia se diagnosticaría como ritmo sinusal, para ese paciente no se empezaría un diagnostico ni un tratamiento, por lo que la enfermedad seguirá avanzando. Por otra parte esta estrategia también presenta una especificidad alta, si bien, es menos critico diagnosticar a un paciente con un falso positivo

que con un falso negativo, una baja especificidad representa otro tipo de pérdidas, como tiempo del personal de salud y el uso de recursos clínicos innecesarios. La prioridad del algoritmo XCorr-In es una alta sensibilidad ya que esta tiene un mayor impacto en la salud del paciente, sin embargo, con la estrategia $\beta_{\hat{p},\gamma=0,01}$, se logra tener un equilibrio entre ambas métricas, cuidando de esta manera la salud del paciente, el tiempo del personal de salud y los recursos clínicos.

El algoritmo XCorr-In al ser un algoritmo de decisión basado en los índices de correlación cruzada y parámetros fisiológicos, hace que su demanda computacional sea menor comparada con otras técnicas, por ejemplo, redes neuronales o inteligencia artificial. Se presenta una alta sensibilidad y especificidad sin necesidad de tener un gran volumen de datos para entrenamiento.

5.1.1. Contribuciones académicas

Como parte de los productos generados por este proyecto de investigación se tienen las siguientes contribuciones académicas:

1. Mayo del 2022. Participación en el Primer Concurso de Prototipos de Bioingeniería con el prototipo titulado "Detección y clasificación de la frecuencia cardíaca de una señal ECG con ruido".
2. Noviembre del 2022. Presentación del cartel titulado "Análisis de la Función de Correlación en Señales Electrocardiográficas para la Detección de Arritmias" dentro del marco del 7mo Foro Nacional de Bioingeniería, con el que se obtuvo el primer lugar en el concurso de carteles.
3. Diciembre del 2023. Presentación de la ponencia titulada "Statistical Analysis of Cross-Correlation Index for Identifying Abnormal ECG Signals" en la conferencia *2023 International Conference on Computational Science and Computational Intelligence (CSCI)*, cuyas memorias están publicadas en la base de datos de la IEEEExplore: <https://ieeexplore.ieee.org/document/10590579>.
4. Marzo del 2025. Registro de la propiedad intelectual del software desarrollado como parte de la implementación del algoritmo

XCorr-In ante el Instituto Nacional del Derecho de Autor (En proceso).

5. Junio del 2025. Artículo titulado "XCorr-In: Cross-Correlation Index Analysis for robust ECG Signal Classification" que será enviado a la revista *Annals of Biomedical Engineering* <https://link.springer.com/journal/10439/aims-and-scope>.

5.2. Trabajo a futuro

Los resultados obtenidos demuestran un desempeño adecuado del algoritmo XCorr-In. Sin embargo, el trabajo se presenta algunas limitaciones.

Una de las principales limitaciones es la señal $x_{ref}(t)$, ya que esta fue diseñada para ser representativa de una población con características específicas, en este caso, mujeres en un rango de edad de 40 a 60 años. Por lo tanto, si se desea analizar una población con características diferentes, es necesario hacer un proceso de calibración, el cual consiste en seguir la metodología descrita en la Figura 3.2, para obtener una señal $x_{ref}(t)$ que sea representativa de la población objetivo.

Otra limitación es que el algoritmo XCorr-In realiza una clasificación dicotómica, ya que la clasificación es si presenta un ritmo sinusal o si presenta una arritmia, esto es una limitación al personal de salud, ya que para poder diagnosticar que tipo de arritmia tiene, se deben de hacer estudios y pruebas adicionales. El área de oportunidad más grande para el algoritmo XCorr-In es incluir más parametros para una segunda decisión, después de la primera decisión sobre si la señal ECG presenta una arritmia, la siguiente decisión implicada clasificar esa arritmia, para esto es necesario realizar una investigación más exhaustiva sobre los tipos de arritmias y las características de cada una, para realizar esta segunda clasificación también sería necesario incluir otro tipo de técnicas al algoritmo para la identificación de estas características en cada arritmia. Debido a que el algoritmo XCorr-In realiza esta clasificación dicotómica, es porque se presenta una sensibilidad alta, el algoritmo es capaz de identificar las señales que presentan un ritmo fuera del ritmo sinusal, pero realizar una segunda clasificación representa la integración de muchos más parametros de clasificación y otras técnicas

5.2. TRABAJO A FUTURO

para el algoritmo, por ejemplo, el uso de inteligencia artificial o redes neuronales.

Appendices

α	β_μ	Especificidad (\hat{E})	Sensibilidad (\hat{S})	Error total
0.5	70	0.5340	0.9568	0.2343056
0.55	64	0.5523	0.9589	0.2248832
0.6	57	0.6151	0.9568	0.1976484
0.65	52	0.5732	0.9632	0.21308
0.7	46	0.6020	0.9654	0.1988568
0.75	39	0.6282	0.9632	0.18822
0.8	33	0.6047	0.9740	0.1929236
0.85	26	0.5680	0.9697	0.2118684
0.9	16	0.5287	0.9762	0.22607
0.95	6	0.3219	0.9892	0.3124196
1	0	0.0000	1.0000	0.452

Tabla 1: Métricas de calidad para el Algoritmo 1 en función del valor de α considerando el valor de β_μ

d_l	βd_l	Especificidad (\hat{E})	Sensibilidad (\hat{S})
d_1	56	0.8063	0.9201
d_2	63	0.7042	0.9330
d_3	66	0.6152	0.9460
d_4	69	0.5340	0.9568
d_5	73	0.3796	0.9806
d_6	74	0.3455	0.9827
d_7	76	0.2827	0.9849
d_8	80	0.1545	0.9935
d_9	85	0.0654	0.9978
d_{10}	93	0.0131	1.0000

Tabla 2: Deciles obtenidos para el valor de $\alpha = 0,5$ y las métricas de calidad considerando el valor de β_{d_l}

d_l	βd_l	Especificidad (\hat{E})	Sensibilidad (\hat{S})
d_1	49	0.8220	0.9158
d_2	58	0.7016	0.9352
d_3	60	0.6414	0.9460
d_4	64	0.5026	0.9633
d_5	66	0.4293	0.9741
d_6	68	0.3246	0.9806
d_7	70	0.2565	0.9827
d_8	74	0.1597	0.9957
d_9	78	0.0576	0.9978
d_{10}	85	0.0183	1.0000

Tabla 3: Deciles obtenidos para el valor de $\alpha = 0,55$ y las métricas de calidad considerando el valor de β_{d_l}

d_l	βd_l	Especificidad (\hat{E})	Sensibilidad (\hat{S})
d_1	43	0.8272	0.9201
d_2	52	0.7068	0.9374
d_3	56	0.6152	0.9568
d_4	58	0.5262	0.9654
d_5	60	0.4398	0.9784
d_6	62	0.3665	0.9849
d_7	66	0.2094	0.9892
d_8	67	0.1780	0.9935
d_9	70	0.0916	0.9978
d_{10}	79	0.0079	1.0000

Tabla 4: Deciles obtenidos para el valor de $\alpha = 0,6$ y las métricas de calidad considerando el valor de β_{d_l} .

d_l	βd_l	Especificidad (\hat{E})	Sensibilidad (\hat{S})
d_1	41	0.8010	0.9266
d_2	48	0.6859	0.9438
d_3	51	0.5733	0.9633
d_4	54	0.4660	0.9806
d_5	55	0.4005	0.9827
d_6	58	0.2827	0.9892
d_7	59	0.2408	0.9914
d_8	62	0.1230	0.9957
d_9	64	0.0733	1.0000
d_{10}	72	0.0079	1.0000

Tabla 5: Deciles obtenidos para el valor de $\alpha = 0,65$ y las métricas de calidad considerando el valor de β_{d_l}

d_l	βd_l	Especificidad (\hat{E})	Sensibilidad (\hat{S})
d_1	33	0.8089	0.9266
d_2	42	0.6937	0.9438
d_3	46	0.5524	0.9719
d_4	48	0.4791	0.9849
d_5	49	0.4293	0.9849
d_6	50	0.3482	0.9849
d_7	53	0.2225	0.9935
d_8	54	0.1754	0.9935
d_9	57	0.0864	1.0000
d_{10}	65	0.0079	1.0000

Tabla 6: Deciles obtenidos para el valor de $\alpha = 0,7$ y las métricas de calidad considerando el valor de β_{d_l}

d_l	βd_l	Especificidad (\hat{E})	Sensibilidad (\hat{S})
d_1	24	0.8194	0.9222
d_2	38	0.6283	0.9633
d_3	39	0.5916	0.9719
d_4	41	0.5000	0.9827
d_5	43	0.3822	0.9892
d_6	44	0.3351	0.9892
d_7	46	0.2565	0.9935
d_8	48	0.1702	1.0000
d_9	51	0.0524	1.0000
d_{10}	55	0.0105	1.0000

Tabla 7: Deciles obtenidos para el valor de $\alpha = 0,75$ y las métricas de calidad considerando el valor de β_{d_l}

d_l	βd_l	Especificidad (\hat{E})	Sensibilidad (\hat{S})
d_1	16	0.8063	0.9201
d_2	31	0.6414	0.9676
d_3	33	0.5681	0.9784
d_4	35	0.4581	0.9892
d_5	36	0.3979	0.9914
d_6	37	0.3429	0.9914
d_7	39	0.2408	0.9957
d_8	41	0.1623	1.0000
d_9	43	0.0707	1.0000
d_{10}	49	0.0052	1.0000

Tabla 8: Deciles obtenidos para el valor de $\alpha = 0,8$ y las métricas de calidad considerando el valor de β_{d_l}

d_l	βd_l	Especificidad (\hat{E})	Sensibilidad (\hat{S})
d_1	5	0.7853	0.9071
d_2	21	0.6728	0.9611
d_3	24	0.5890	0.9698
d_4	27	0.4764	0.9849
d_5	29	0.3770	0.9870
d_6	30	0.3246	0.9914
d_7	33	0.1283	0.9978
d_8	33	0.1283	0.9978
d_9	36	0.0445	1.0000
d_{10}	41	0.0000	1.0000

Tabla 9: Deciles obtenidos para el valor de $\alpha = 0,85$ y las métricas de calidad considerando el valor de β_{d_l}

d_l	βd_l	Especificidad (\hat{E})	Sensibilidad (\hat{S})
d_1	0	0.7147	0.9287
d_2	4	0.6675	0.9417
d_3	13	0.5628	0.9762
d_4	17	0.4686	0.9849
d_5	20	0.3613	0.9978
d_6	21	0.3194	1.0000
d_7	23	0.2251	1.0000
d_8	24	0.1440	1.0000
d_9	27	0.0471	1.0000
d_{10}	29	0.0105	1.0000

Tabla 10: Deciles obtenidos para el valor de $\alpha = 0,9$ y las métricas de calidad considerando el valor de β_{d_l}

d_l	βd_l	Especificidad (\hat{E})	Sensibilidad (\hat{S})
d_1	0	0.4319	0.9827
d_2	0	0.4319	0.9827
d_3	0	0.4319	0.9827
d_4	0	0.4319	0.9827
d_5	1	0.4005	0.9849
d_6	9	0.2435	0.9957
d_7	11	0.1806	0.9978
d_8	13	0.1126	0.9978
d_9	15	0.0445	1.0000
d_{10}	19	0.0079	1.0000

Tabla 11: Deciles obtenidos para el valor de α correspondiente y las métricas de calidad considerando el valor de β_{d_l}

d_i	β_{d_i}	Especificidad (\hat{E})	Sensibilidad (\hat{S})
d_1	0	0	1
d_2	0	0	1
d_3	0	0	1
d_4	0	0	1
d_5	0	0	1
d_6	0	0	1
d_7	0	0	1
d_8	0	0	1
d_9	0	0	1
d_{10}	0	0	1

Tabla 12: Deciles obtenidos con valores fijos de β_{d_i} y métricas de calidad

α	β_{d_1}	Especificidad (\hat{E})	Sensibilidad (\hat{S})
0.5	56	0.8063	0.9201
0.55	49	0.8220	0.9158
0.6	43	0.8272	0.9201
0.65	41	0.8010	0.9266
0.7	33	0.8089	0.9266
0.75	24	0.8194	0.9222
0.8	16	0.8063	0.9201
0.85	5	0.7853	0.9071
0.9	0	0.7147	0.9287
0.95	0	0.4319	0.9827
1.0	0	0.0000	1.0000

Tabla 13: Métricas de calidad para el Algoritmo 1 en función del valor de α considerando el valor de β_{d_1}

α	$\beta_{\hat{p}}$
0.5	0.03128
0.55	0.02856
0.6	0.02562
0.65	0.02326
0.7	0.02035
0.75	0.01756
0.8	0.01480
0.85	0.01146
0.9	0.00727
0.95	0.00257
1.0	0.00000

Tabla 14: Valores de $\beta_{\hat{p}}$ para distintos α

α	$\beta_{\hat{p},\gamma=0,01}$	Especificidad (\hat{E})	Sensibilidad (\hat{S})	Error total
0.5	0.03128	0.9084	0.9914	0.0461
0.55	0.02856	0.8953	0.9957	0.0497
0.6	0.02562	0.8874	0.9914	0.0556
0.65	0.02326	0.8743	0.9849	0.0651
0.7	0.02035	0.8613	0.9806	0.0734
0.75	0.01756	0.8325	0.9849	0.0840
0.8	0.01480	0.7984	0.9935	0.0947
0.85	0.01146	0.7408	0.9914	0.1219
0.9	0.00727	0.6387	0.9935	0.1668
0.95	0.00257	0.4162	1.0000	0.2639
1.0	0.00000	0.0000	1.0000	0.4520

Tabla 15: Métricas de calidad para el Algoritmo 1 en función del valor de α considerando el valor de $\beta_{\hat{p},\gamma=0,01}$.

α	$\beta_{\hat{p},\gamma=0,05}$	Especificidad (\hat{E})	Sensibilidad (\hat{S})
0.5	0.03128	0.8848	0.9806
0.55	0.02856	0.8770	0.9806
0.6	0.02562	0.8717	0.9806
0.65	0.02326	0.8613	0.9719
0.7	0.02035	0.8429	0.9741
0.75	0.01756	0.8220	0.9741
0.8	0.01480	0.7827	0.9827
0.85	0.01146	0.7199	0.9849
0.9	0.00727	0.6283	0.9849
0.95	0.00257	0.3874	0.9978
1.0	0.00000	0.0000	1.0000

Tabla 16: Métricas de calidad para el Algoritmo 1 en función del valor de α considerando el valor de $\beta_{\hat{p},\gamma=0,05}$.

α	$\beta_{\hat{p},\gamma=0,1}$	Especificidad (\hat{E})	Sensibilidad (\hat{S})
0.5	0.03128	0.8691	0.9762
0.55	0.02856	0.8613	0.9741
0.6	0.02562	0.8665	0.9698
0.65	0.02326	0.8429	0.9676
0.7	0.02035	0.8246	0.9698
0.75	0.01756	0.8037	0.9698
0.8	0.01480	0.7749	0.9784
0.85	0.01146	0.7094	0.9806
0.9	0.00727	0.6152	0.9784
0.95	0.00257	0.3874	0.9978
1.0	0.00000	0.0000	1.0000

Tabla 17: Métricas de calidad para el Algoritmo 1 en función del valor de α considerando el valor de $\beta_{\hat{p},\gamma=0,1}$.

α	$\beta_{\hat{p},\gamma=0,15}$	Especificidad (\hat{E})	Sensibilidad (\hat{S})
0.5	0.03128	0.8508	0.9698
0.55	0.02856	0.8429	0.9654
0.6	0.02562	0.8429	0.9633
0.65	0.02326	0.8272	0.9633
0.7	0.02035	0.8115	0.9654
0.75	0.01756	0.7801	0.9676
0.8	0.01480	0.7618	0.9762
0.85	0.01146	0.7042	0.9741
0.9	0.00727	0.6047	0.9762
0.95	0.00257	0.3691	0.9957
1.0	0.00000	0.0000	1.0000

Tabla 18: Métricas de calidad para el Algoritmo 1 en función del valor de α considerando el valor de $\beta_{\hat{p},\gamma=0,15}$.

Bibliografía

- [1] C. A. D. Cervantes. «Tendencia e impacto de la mortalidad por enfermedades cardiovasculares en México, 1990-2015.» Accedido: 2024-11-27. (2019), dirección: https://www.scielosp.org/article/ssm/content/raw/?resource_ssm_path=/media/assets/rcsp/v45n4/1561-3127-rcsp-45-04-e1081.pdf.
- [2] Instituto Nacional de Estadística y Geografía (INEGI), *INEGI*, Último acceso: 20-Feb-2025, 2024. dirección: <https://www.inegi.org.mx/default.html>.
- [3] L. Fácila Rubio, C. Lozano-Granero, R. Vidal-Pérez, V. Barrios y R. Freixa-Pamias, «Nuevas tecnologías para el diagnóstico, tratamiento y seguimiento de las enfermedades cardiovasculares,» *Revista Española de Cardiología*, vol. 77, n.º 1, págs. 88-96, 2024, ISSN: 0300-8932. DOI: <https://doi.org/10.1016/j.recesp.2023.07.011>. dirección: <https://www.sciencedirect.com/science/article/pii/S0300893223003238>.
- [4] American Heart Association. «Women Still Underrepresented in Clinical Research, Science, and Medicine—That Could Save Them from Their No. 1 Killer.» Accedido: 2024-11-27. (2024), dirección: <https://newsroom.heart.org/news/women-still-underrepresented-in-clinical-research-science-and-medicine-that-could-save-them-from-their-no-1-killer>.
- [5] L. Geraghty, G. A. Figtree, A. E. Schutte, S. Patel, M. Woodward y C. Arnott, «Cardiovascular Disease in Women: From Pathophysiology to Novel and Emerging Risk Factors,» *Heart, Lung and Circulation*, vol. 30, n.º 1, págs. 9-17, 2021, Women's Heart Health, ISSN: 1443-9506. DOI: <https://doi.org/10.1016/j.hlc.2020>.

- 05.108. dirección: <https://www.sciencedirect.com/science/article/pii/S1443950620302808>.
- [6] U.S. National Library of Medicine, *ClinicalTrials.gov*, Accedido: 2024-11-27, 2024. dirección: <https://clinicaltrials.gov>.
- [7] American Heart Association. «Go Red for Women.» Accedido: 2024-11-27. (2024), dirección: <https://www.goredforwomen.org/es/search#q=citar&sort=relevancy>.
- [8] European Society of Cardiology, «Cardiovascular Disease in Women: A Major Public Health Issue,» *European Heart Journal*, 2006, Accedido: 28 de noviembre de 2024. dirección: <https://www.escardio.org/static-file/Escardio/EU-affairs/CVDinWomen-2006EHJ278994.pdf>.
- [9] L. A. McDonnell, A. L. Pipe, C. Westcott, S. Perron, D. Younger-Lewis, N. Elias et al., «Perceived vs actual knowledge and risk of heart disease in women: findings from a Canadian survey on heart health awareness, attitudes, and lifestyle,» *Canadian Journal of Cardiology*, vol. 30, n.º 7, págs. 827-834, 2014.
- [10] K. Backholer, S. A. E. Peters, M. L. Bots, A. Peeters, R. R. Huxley y M. Woodward, «Sex differences in the relationship between socioeconomic status and cardiovascular disease: a systematic review and meta-analysis,» *Journal of Epidemiology and Community Health*, vol. 71, n.º 6, págs. 550-557, 2017.
- [11] Revista Española de Cardiología. «Arritmias cardíacas en la mujer: ¿un desafío en la práctica clínica?» Accedido: 2024-11-27. (2018), dirección: <https://www.revespcardiol.org/es-arritmias-cardiacas-mujer-articulo-13089748>.
- [12] F. J. Alonso Moreno y et al., «Estudio de fiabilidad en la interpretación del electrocardiograma por médicos de familia y médicos residentes,» *Semergen*, 2017. DOI: 10.1016/j.semerg.2016.12.001. dirección: <http://dx.doi.org/10.1016/j.semerg.2016.12.001>.
- [13] N. Rafie, A. H. Kashou y P. A. Noseworthy, «ECG Interpretation: Clinical Relevance, Challenges, and Advances,» *Hearts*, vol. 2, n.º 4, págs. 505-513, 2021, ISSN: 2673-3846. DOI: 10.3390/

- hearts2040039. dirección: <https://www.mdpi.com/2673-3846/2/4/39>.
- [14] A. H. Kashou, W. Y. Ko, Z. I. Attia, M. S. Cohen, P. A. Friedman y P. A. Noseworthy, «A comprehensive artificial intelligence-enabled electrocardiogram interpretation program,» *Cardiovascular Digital Health Journal*, vol. 1, n.º 2, págs. 62-70, 2020. DOI: 10.1016/j.cvdhj.2020.08.005.
- [15] A. H. Kashou, S. K. Mulpuru, A. J. Deshmukh et al., «An artificial intelligence-enabled ECG algorithm for comprehensive ECG interpretation: Can it pass the 'Turing test'?» *Cardiovascular Digital Health Journal*, vol. 2, n.º 3, págs. 164-170, 2021. DOI: 10.1016/j.cvdhj.2021.04.002.
- [16] P. Martínez-Losas, J. Higuera, J. C. Gómez-Polo et al., «The influence of computerized interpretation of an electrocardiogram reading,» *American Journal of Emergency Medicine*, vol. 34, n.º 10, págs. 2031-2032, oct. de 2016. DOI: 10.1016/j.ajem.2016.07.029.
- [17] J. Schlöpfer y H. J. Wellens, «Computer-Interpreted Electrocardiograms: Benefits and Limitations,» *Journal of the American College of Cardiology*, vol. 70, n.º 9, págs. 1183-1192, ago. de 2017. DOI: 10.1016/j.jacc.2017.07.723.
- [18] J. L. Garvey, J. Zegre-Hemsey, R. Gregg y J. R. Studnek, «Electrocardiographic diagnosis of ST segment elevation myocardial infarction: An evaluation of three automated interpretation algorithms,» *Journal of Electrocardiology*, vol. 49, n.º 5, págs. 728-732, sep. de 2016. DOI: 10.1016/j.jelectrocard.2016.04.010.
- [19] National Heart, Lung and Blood Institute. «Cómo funciona el corazón.» Last accessed: 2025-02-22. (2022), dirección: <https://www.nhlbi.nih.gov/es/salud/corazon>.
- [20] J. C. G. Rubira, «Fisiología cardíaca,» en *Libro de la salud cardiovascular del Hospital Clínico San Carlos y la Fundación BBVA*, Fundación BBVA, 2009, págs. 41-48.

- [21] J. Tamargo y E. Delpón, «Función de bomba del corazón: el ciclo cardíaco,» *Tresguerres JA, Benítez EA, Cachofeiro MV, Cardinali DP, Gil Loyzaga P, Lahera JV, et al. Fisiología humana. 3ra. ed. Madrid: McGraw Hill Interamericana*, págs. 485-506, 2012.
- [22] D. Mirvis y A. Goldberger, «Electrocardiography,» *Braunwald's Heart Disease, a Textbook of Cardiovascular Medicine*, págs. 146-147, ene. de 2011.
- [23] Departamento de Fisiología de la Universidad Nacional Autónoma de México. «Taller de interpretación del electrocardiograma.» Last accessed: 2025-02-22. (s.f.), dirección: <https://fisiologia.facmed.unam.mx/index.php/taller-de-interpretacion-del-electrocardiograma/>.
- [24] K. Wesley, *Huszar. Interpretación del ECG: monitorización y 12 derivaciones: Guía práctica para la interpretación y el tratamiento*. Elsevier Health Sciences, 2017.
- [25] D. S. Tawfik, J. Profit, T. I. Morgenthaler et al., «Physician Burnout, Well-being, and Work Unit Safety Grades in Relationship to Reported Medical Errors,» *Mayo Clinic Proceedings*, vol. 93, n.º 11, págs. 1571-1580, 2018.
- [26] A. Borodin, A. Pogorelov e Y. Zavyalova, «Overview of algorithms for electrocardiograms analysis,» en *2013 13th Conference of Open Innovations Association (FRUCT)*, 2013, págs. 14-19. DOI: 10.23919/FRUCT.2013.8124224.
- [27] K. Siontis, P. Noseworthy, Z. Attia et al., «Artificial intelligence-enhanced electrocardiography in cardiovascular disease management,» *Nature Reviews Cardiology*, vol. 18, págs. 465-478, 2021. DOI: 10.1038/s41569-020-00503-2.
- [28] European Parliament and Council of the European Union, *General Data Protection Regulation (GDPR)*, Accessed: 2025-01-30, 2016. dirección: <https://gdpr-info.eu>.
- [29] U.S. Department of Health and Human Services, *HIPAA for Professionals*, Accessed: 2025-01-30, n.d. dirección: <https://www.hhs.gov/hipaa/index.html>.
- [30] H. Smulyan, «The computerized ECG: friend and foe,» *The American journal of medicine*, vol. 132, n.º 2, págs. 153-160, 2019.

-
- [31] J. L. Semmlow, *Circuits, Systems, and Signals for Bioengineers: A MATLAB-Based Introduction*. Burlington, MA, USA: Academic Press, Elsevier Inc., 2005, ISBN: 0-12-088493-3.
- [32] PhysioNet. «PhysioBank.» (2016), dirección: <https://archive.physionet.org/physiobank/>.
- [33] A. L. Goldberger, L. A. N. Amaral, L. Glass et al., «PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals,» *Circulation*, vol. 101, n.º 23, e215-e220, 2000. DOI: 10.1161/01.CIR.101.23.e215.
- [34] J. Zheng, J. Zhang, S. Danioko et al., «A 12-lead electrocardiogram database for arrhythmia research covering more than 10,000 patients,» *Scientific Data*, vol. 7, pág. 48, 2020. DOI: 10.1038/s41597-020-0386-x. dirección: <https://doi.org/10.1038/s41597-020-0386-x>.
- [35] G. Healthcare. «MUSE v9.» (2024), dirección: <https://www.gehealthcare.es/products/diagnostic-cardiology/muse-v9>.
- [36] R. E. Walpole, R. H. Myers, S. L. Myers y K. Ye, *Probabilidad y estadística para ingeniería y ciencias*, 9.^a ed., trad. por L. E. Pineda Ayala. Naucalpan de Juárez, Estado de México: Pearson Educación de México, S.A. de C.V., 2012, ISBN: 978-607-32-1417-9.
- [37] A. Morillas, «Muestreo en poblaciones finitas,» *Muestreo en Poblaciones Finitas*, 2007, Obtenido de Muestreo en Poblaciones Finitas, 20017.
- [38] C. Carreiras, A. P. Alves, A. Lourenço, F. Canento, H. Silva, A. Fred et al., *BioSPPy: Biosignal Processing in Python*, [Online; accessed {today}], 2015-. dirección: <https://github.com/PIA-Group/BioSPPy/>.
- [39] J. G. Proakis y D. G. Manolakis, *Procesamiento digital de señales*. Prentice Hall, 1996.

- [40] J. Roman, M. Rangaswamy, D. Davis, Q. Zhang, B. Himed y J. Michels, «Parametric adaptive matched filter for airborne radar applications,» *IEEE Transactions on Aerospace and Electronic Systems*, vol. 36, n.º 2, págs. 677-692, 2000. DOI: 10.1109/7.845259.
- [41] J. Loscalzo, A. Fauci, D. Kasper, S. Hauser, D. Longo y J. L. Jameson, *Harrison. Principios de Medicina Interna 21e*. McGraw-Hill, 2022.
- [42] C. R. Harris, K. J. Millman, S. J. van der Walt et al., «Array programming with NumPy,» *Nature*, vol. 585, n.º 7825, págs. 357-362, sep. de 2020. DOI: 10.1038/s41586-020-2649-2. dirección: <https://doi.org/10.1038/s41586-020-2649-2>.
- [43] T. pandas development team, *pandas-dev/pandas: Pandas*, ver. latest, feb. de 2020. DOI: 10.5281/zenodo.3509134. dirección: <https://doi.org/10.5281/zenodo.3509134>.
- [44] J. D. Hunter, «Matplotlib: A 2D graphics environment,» *Computing in Science & Engineering*, vol. 9, n.º 3, págs. 90-95, 2007. DOI: 10.1109/MCSE.2007.55.