



Evaluación del desempeño de un clasificador de electrocardiogramas basado en máquina de
soporte de vectores

Danna Marcela Melo Hidalgo

PROHIBIDA SU COPIA

Corporación Universitaria Autónoma De Nariño

Facultad De Ingeniería

Programa De Ingeniería Electrónica

San Juan De Pasto

2019



Evaluación del desempeño de un clasificador de electrocardiogramas basado en máquina de soporte de vectores

Danna Marcela Melo Hidalgo

Asesor:

MG. Oscar Rodríguez Bastidas

Ingeniero Electrónico

Asesora Metodológica:

Esp. Martha Romero Rodríguez

Corporación Universitaria Autónoma De Nariño

Facultad De Ingeniería

Programa De Ingeniería Electrónica

San Juan De Pasto

2019



NOTA DE ACEPTACIÓN

Firma del presidente del Jurado

Firma del jurado

Firma del jurado

PROHIBIDA SU COPIA



DEDICATORIA

Dedico este gran logro a mis padres y familia que han estado apoyándome en estos momentos tan incondicionales de la vida, para lograr culminar mis metas.

PROHIBIDA SU COPIA



Contenido

Contenido	5
Lista de tablas	8
Introducción	1
Título	3
Planteamiento del problema	4
Descripción del problema	4
Formulación del problema	4
Justificación	5
Objetivos	6
Objetivo general	6
Objetivos específicos	6
Marco referencial	7
Marco contextual	7
Marco teórico	7
Funciones kernel	10
Marco conceptual	14
Marco legal	16
Marco tecnológico	17
Diseño de aspecto metodológico	18
Línea de investigación.	18
Enfoque	19



Tipo de investigación	19
Análisis de resultados	20
Patologías relacionadas con el corazón.	20
Cardiopatía isquémica	20
Insuficiencia cardiaca	20
Arritmias	21
Base de datos	21
Preprocesado datos	23
Métricas de desempeño	23
Sensibilidad	24
Especificidad	25
Exactitud	25
Precisión	26
Score F1	26
Metodología experimental	26
Desarrollo del script en lenguaje M	27
Discusión de resultados	28
Experimentación con la clase 1 y 2	28
Experimentación con la clase 1 y 3	30
Experimentación clase 1 y 4	31
Experimentación clase 2 y todas las clases	33
Experimentación utilizando todas las clases	34
Conclusiones	38



PROHIBIDA SU COPIA



Lista de tablas

Tabla 1. Etiquetas de la base de datos.	22
Tabla 2. Matriz de confusión	24
Tabla 3. Resultados obtenidos basado en primeras características de clase 1 y 2.	28
Tabla 4. Resultados obtenidos basado en segundas características de clase 1 y 2.	29
Tabla 5. Resultados obtenidos basado en todas las características de clase 1 y 2.	29
Tabla 6. Resultados obtenidos basado en primeras características de clase 1 y 3.	30
Tabla 7. Resultados obtenidos basado en segundas características de clase 1 y 3.	30
Tabla 8. Resultados obtenidos basado en todas las características de clase 1 y 3.	31
Tabla 9. Resultados obtenidos basado en primeras características de clase 1 y 4.	32
Tabla 10. Resultados obtenidos basado en segundas características de clase 1 y 4.	32
Tabla 11. Resultados obtenidos basado en todas las características de clase 1 y 4.	33
Tabla 12. Resultados obtenidos basado en primeras características de clase 2 y todas	33
Tabla 13. Resultados obtenidos basado en segundas características de clase 2 y todas.	34
Tabla 14. Matriz de confusión todas las clases (lineal)	34
Tabla 15. Matriz de confusión todas las clases (Cuadrático).	35
Tabla 16. Matriz de confusión todas las clases (Cúbico).	36
Tabla 17. Matriz de confusión todas las clases (Gaussiano fino)	37



Lista de figuras

Figura 1. Señal ECG.	9
Figura 2. Clasificación de datos en Maquinas de soporte de vectores	10
Figura 3. MSV lineal	11
Figura 4. Datos clase Rojo y Azul en R2	12
Figura 5. Datos clase Rojo y Azul en R3	12
Figura 6. Diagrama de bloques script	27

PROHIBIDA SU COPIA



Resumen analítico de estudio – rae

Programa académico: Ingeniería Electrónica
Fecha de elaboración del rae: octubre 2019
Autor: Danna Marcela Melo Hidalgo

Asesor técnico: Ing. Oscar Rodríguez Bastidas.
Asesor metodológico: Esp. Martha Romero Rodríguez.

Título: Evaluación del desempeño de un clasificador de electrocardiogramas basado en máquina de soporte de vectores

Palabras claves

Máquinas de soporte de vectores: son técnicas de aprendizaje de máquina por medio de clasificación, está basado en la idea de minimización de riesgo estructural. Conocida como SVM que significa en inglés (support Vector Machines). Se busca con esta técnica de máquina de soporte de vectores hacer un entrenamiento con respecto a una base de datos para poder Extraer conocimiento que pueda ser usado con algún propósito, como clasificar o agrupar datos distintos.

Electrocardiógrafo: aparato electrónico que capta y amplía la actividad eléctrica del corazón a través de electrodos colocados en las 4 extremidades y en 6 posiciones precordiales. El registro de dicha actividad se llama electrocardiograma (ECG).

Electrocardiograma: es la representación gráfica de la actividad eléctrica del corazón que se obtiene, desde la superficie corporal, en el pecho, Es el instrumento principal de la electrofisiología cardíaca y tiene una función de diagnóstico de las enfermedades cardiovasculares.

Electrodo: es un conductor eléctrico utilizado para hacer contacto con una parte no metálica de un circuito, los electrodos poseen una celda electroquímica que se refiere a ánodo y cátodo. El

ánodo es definido como el electrodo donde los electrones salen de la celda y ocurre la oxidación, el cátodo es el cual los electrones entran a la celda y hay reducción. Los electrodos pueden convertirse en ánodo o cátodo.

Anatomía y fisiología del corazón: El corazón es uno de los órganos más importantes del ser humano, este es un órgano muscular contráctil y hueco donde su principal función es la circulación sanguínea, se encuentra entre los dos pulmones, encima del diafragma, delante de la columna vertebral detrás los cartílagos costales y el esternón. Compuesto por dos aurículas y dos ventrículos.

Activación eléctrica del corazón: la actividad eléctrica se genera en el corazón debido a ciertas células especializadas que tienen las características de ser autoexcitables y de funcionamiento periódico, lo cual nos permite decir que son capaces de generar potenciales de acción por si mismas con cierta repetición.

Derivaciones electrocardiográficas: son las encargadas de recoger la diferencia de potencial generada por la actividad eléctrica del corazón, se realiza por medio de electrodos, hay diferentes tipos de derivaciones.

Descripción

La Organización Mundial de la Salud (OMS) en sus notas e informes menciona que cada año hay un aumento en las muertes por enfermedades relacionadas con el corazón. En las cifras reveladas por este ente, calcula que el 30% del total de muertes registradas en el mundo están asociadas a patologías de enfermedades cardiovasculares, también afirma que más del 70 % de las defunciones se producen en países con ingresos bajos y medios.

Por otro lado, el desarrollo de la electrónica aplicada ha incursionado en las áreas médicas. Uno de sus grandes exponentes es el registro de las señales eléctricas del corazón denominado electrocardiograma (ECG), procedimiento no invasivo, que permite obtener información sobre su funcionamiento. El ECG se ha convertido en la principal herramienta de diagnóstico en patologías cardíacas.

También hay que resaltar la importancia que en los últimos años ha ganado la inteligencia artificial. Diversos grupos de investigación han centrado sus esfuerzos en resolver diversos problemas del mundo real, hay que mencionar además que ha llegado hasta la medicina.

Con base en lo anterior expuesto, en este trabajo se plantea el estudio del desempeño de técnicas de clasificación de datos aplicado a registros ECG, específicamente se utiliza el método de máquina de soporte vectorial.

Contenido

El desarrollo de este trabajo se describe a continuación: en la Sección 1 se describe el planteamiento del problema de investigación, la justificación y los antecedentes a este. En la Sección 2 se definen los objetivos generales y específicos. En la Sección 3 se contextualiza al lector y se presenta conceptos básicos que son primordiales para el desarrollo de este trabajo. En la Sección 4 se presenta el diseño metodológico. En la Sección 5 se presenta los principales resultados obtenidos en esta investigación. Finalmente, una discusión y conclusiones respecto a los resultados obtenidos en el trabajo.



Metodología

Línea de investigación

En la Corporación Universitaria Autónoma de Nariño, se manejan actualmente tres áreas de investigación que son: Las Telecomunicaciones, la Electro medicina y Telemedicina y la Automatización Industrial, áreas que se proponen de acuerdo a la profundización que puede obtener cada estudiante mediante la selección de las materias electivas profesionales que se dictan en los semestres octavo, noveno y décimo, disponibles en la Facultad. Dentro de estas tres áreas de investigación se ha contemplado el desarrollo de proyectos, para el presente se utilizará la siguiente línea de investigación que es el electro.

Enfoque

En la Corporación Universitaria Autónoma de Nariño, se manejan actualmente tres áreas de investigación que son: Las Telecomunicaciones, la Electro medicina y Telemedicina y la Automatización Industrial, áreas que se proponen de acuerdo a la profundización que puede obtener cada estudiante mediante la selección de las materias electivas profesionales que se dictan en los semestres octavo, noveno y décimo, disponibles en la Facultad. Dentro de estas tres áreas de investigación se ha contemplado el desarrollo de proyectos, para el presente se utilizará la siguiente línea de investigación que es el electro.

Tipo de investigación

Para el desarrollo de esta investigación se utilizó un tipo de investigación experimental debido a que se basa en la manipulación de variables teniendo el control total sobre ellas, donde se utiliza un fenómeno que para este caso es un análisis de corazón ECG y máquinas de soporte vectorial para su clasificación. Se plantea experimentos con el fin de obtener los resultados que permita validar o negar la hipótesis.

Método

Para el estudio del presente proyecto se utilizó el método inductivo, la cual nos permitió obtener información precisa y veraz que sirvió como sustento teórico y técnico del proyecto de investigación.

“La característica de este método es que utiliza el razonamiento para obtener conclusiones que parten de hechos particulares, aceptados como válidos para llegar a conclusiones cuya aplicación es de carácter general”. Este trabajo se utiliza dicho método debido a que inicia en conclusiones generales para finalizar en las particulares.

Conclusiones

Desde la perspectiva médica existen diferentes patologías que tiene relación con el corazón, en su mayoría estas se ven reflejadas en la alteración de las señales eléctricas del corazón. También es de considerar que, para algunos casos pruebas como ECG no son concluyentes en el diagnóstico, motivo por el cual un clasificador basado en ECG no podría determinar la patología.

Los clasificadores biclase utilizados en el desarrollo de este trabajo, presentan mejores resultados que los multiclase. La media armónica más alta es presentada por un kernel cubico con un valor de 0,958.

Los sistemas multiclase utilizando máquinas de soporte vectorial con kernel gaussiano fino, medio y grueso sus errores son altos, es decir, para las clases 3 en adelante no clasifica ningún valor. Para la clase 1, de 245 registros este clasifico 243 que son positivos verdaderos, pero para también clasifico en esta misma clase registros que no perteneces a esta clase que serían los falsos positivos.

Bibliografía

Alberto, H., & Ortega, C. (2016). *El corazón y sus ruidos cardíacos normales y agregados*. 59, 49–55.

Bernal, O., cardiología, C. M.-R. española de, & 2006, undefined. (n.d.). Arritmias cardíacas en la mujer. *Elsevier*. Retrieved from <https://www.sciencedirect.com/science/article/pii/S0300893206746887>

D. E Rivas-Cisneros, C. Posadas-Castillo, D. A. D.-R. (2014). *Estudio y Diagnóstico de Patologías Cardiacas Utilizando Teoría de Nudos DE PATOLOG ´ IAS CARDIACAS UTILIZANDO TEOR ´*. (May 2016).

DANE. (2017). *Boletín técnico Estadísticas vitales Nacimientos y Defunciones*. (23), 1–4.

Ecg, E. (2005). *Extracción de características de ECG basadas en transformaciones no lineales y wavelets Extracting ECG signal characteristics based on non-linear*

transformations and wavelets. 25(3), 39–48.

Expósito, C., Ávila, R., & Ii, Á. (2008). *Aplicaciones de la inteligencia artificial en la Medicina : perspectivas y problemas Artificial intelligence applied to medicine : prospects and problems*. 1–12.

Fernández Palomeque, C., Bardají Mayor, J. L., Concha Ruiz, M., Cordo Mollar, J. C., Cosín Aguilar, J., Magriñá Ballara, J., & Melgares Moreno, R. (2000). Guías de práctica clínica de la Sociedad Española de Cardiología en la angina estable. *Revista Española de Cardiología*, 53(7), 967–996. <https://doi.org/10.1157/10487>

Ferreira-González, I. (2014). Epidemiología de la enfermedad coronaria. *Revista Espanola de Cardiologia*, 67(2), 139–144. <https://doi.org/10.1016/j.recesp.2013.10.003>

Issn, P. (2005). LAS MÁQUINAS DE SOPORTE VECTORIAL (SVMs). (27), 67–72.

Matheus, D., Villazana, S., & Seijas, C. (2010). *Deteccion de Arritmias en Señales de ECG Basado en Máquinas de Vectores de Soporte Support Vector Machines Based Arrhythmia Detection from ECG Signal*.

Mosquera, R. E. (2017). Diseño de Experimento: Aproximación metodología para el desarrollo de una aplicación interactiva de realidad virtual Inmersiva, para recrear la terapia espejo en la rehabilitación de pacientes amputados de miembros inferiores con dolor de miembro fantasma. *Memorias de Congresos UTP*, 0(0), 224–228. Retrieved from <http://revistas.utp.ac.pa/index.php/memoutp/article/view/1499/2188>

Picarzo, F. J. P. (2006). *Guía rápida para la lectura sistemática del ECG pediátrico*. 127(319), 319–326.

Rodríguez-Artalejo, F., Banegas Banegas, J. R., & Guallar-Castillón, P. (2004). Epidemiología de la insuficiencia cardíaca. *Revista Española de Cardiología*, 57(2), 163–



170. [https://doi.org/10.1016/S0300-8932\(04\)77080-3](https://doi.org/10.1016/S0300-8932(04)77080-3)

Santibañez, T. (2015). *Validez predictiva de un clasificador basado en máquina de soporte vectorial para éxito o fracaso en la extubación de pacientes conectados a ventilación mecánica invasiva en una un ...* (October).

<https://doi.org/10.13140/RG.2.1.4939.2082>

PROHIBIDA SU COPIA

Introducción

La Organización Mundial de la Salud (OMS) en sus notas e informes menciona que cada año hay un aumento en las muertes por enfermedades relacionadas con el corazón. En las cifras reveladas por este ente, calcula que el 30% del total de muertes registradas en el mundo están asociadas a patologías de enfermedades cardiovasculares, también afirma que más del 70 % de las defunciones se producen en países con ingresos bajos y medios.

Por otro lado, el desarrollo de la electrónica aplicada ha incursionado en las áreas médicas. Uno de sus grandes exponentes es el registro de las señales eléctricas del corazón denominado electrocardiograma (ECG), procedimiento no invasivo, que permite obtener información sobre su funcionamiento. El ECG se ha convertido en la principal herramienta de diagnóstico en patologías cardiacas.

También hay que resaltar la importancia que en los últimos años ha ganado la inteligencia artificial. Diversos grupos de investigación han centrado sus esfuerzos en resolver diversos problemas del mundo real, hay que mencionar además que ha llegado hasta la medicina.

Con base en lo anterior expuesto, en este trabajo se plantea el estudio del desempeño de técnicas de clasificación de datos aplicado a registros ECG, específicamente se utiliza el método de máquina de soporte vectorial.

El desarrollo de este trabajo se describe a continuación: en la Sección 1 se describe el planteamiento del problema de investigación, la justificación y los antecedentes a este. En la Sección 2 se definen los objetivos generales y específicos. En la Sección 3 se contextualiza al lector y se presenta conceptos básicos que son primordiales para el desarrollo de este trabajo. En la Sección 4 se presenta el diseño metodológico. En la Sección 5 se presenta los principales



resultados obtenidos en esta investigación. Finalmente, una discusión y conclusiones respecto a los resultados obtenidos en el trabajo.

PROHIBIDA SU COPIA



Título

Evaluación del desempeño de un clasificador de electrocardiogramas basado en máquina de soporte de vectores

PROHIBIDA SU COPIA

Planteamiento del problema

Descripción del problema

Como lo ha expuesto la organización mundial de la salud OMS, las enfermedades asociadas al corazón son las principales causas de muerte a nivel mundial. En Colombia las enfermedades isquémicas del corazón son las principales causas de muerte tanto en hombres como en mujeres (DANE, 2017). Como se puede evidenciar las patologías relacionadas con el corazón son el principal tema médico, además, el diagnóstico de estas enfermedades por lo general las realiza un especialista, médico cardiólogo, que se encuentra en hospitales de segundo y tercer nivel.

Por otra parte, el médico especialista utiliza como apoyo diagnóstico un electrocardiograma con el fin de determinar la patología cardíaca, y así definir cuál es el tratamiento a seguir. No obstante, algunos pacientes no alcanzan a llegar a ser valorados por el profesional experto, esto se presenta por la congestión en el sistema de salud.

Otro aspecto importante, es la incorporación de la inteligencia artificial en el área médica. En este campo son varias las investigaciones que se han realizado (Expósito, Ávila, & Ii, 2008), sin embargo el problema aún sigue estando abierto. Para desarrollar un sistema experto que pueda realizar una clasificación, la primera fase es evaluar el comportamiento del algoritmo seleccionado para este tipo, esto se realiza evaluando las métricas de desempeño.

Formulación del problema

¿Cuál es la efectividad en el reconocimiento de patologías cardíacas de un clasificador de electrocardiogramas basado en máquinas de soporte de vectores?

Justificación

El electrocardiograma (ECG) es una prueba que registra las señales eléctricas producidas por el corazón. Es un test que se realiza como ayuda diagnóstica para determinar patologías relacionadas con el corazón; la interpretación la realiza un médico especialista (médico cardiólogo) y define cuál es la patología y tratamiento a seguir.

Por otro lado, las máquinas de soporte vectorial son un tipo de algoritmos supervisados, es decir, para realizar una clasificación se debe tener un conocimiento del comportamiento de los datos. También con la evolución de la computación ha surgido la tendencia a diseñar sistemas expertos en la toma de decisiones.

Como es evidente los planes de estratégicos expuestos por el presidente, están orientados a trabajos automatizados en todas las áreas. Entonces, en este trabajo se estudia el comportamiento que tiene las máquinas de soporte vectorial en la clasificación de patologías basado en el estudio de ECG.



Objetivos

Objetivo general

Evaluar el desempeño de un clasificador de electrocardiogramas basado en máquina de soporte a vectores.

Objetivos específicos

- Determinar una patología la cual afecte las señales electrocardiogramas y facilite elegir características que permitan separar las clases.
- Seleccionar las métricas de desempeño que permita evaluar la técnica de máquinas de soporte vectorial.
- Diseñar un protocolo de pruebas con el fin de evaluar el rendimiento de la técnica de máquinas de soporte vectorial.

Marco referencial

Marco contextual

Esta investigación está centrada en la evaluación de máquinas de soporte vectorial aplicadas a la clasificación de patologías utilizando ECG. Las enfermedades seleccionadas deben estar reflejadas en la alteración de las señales eléctricas del corazón, esto con el fin de que se pueda realizar una clasificación.

Como herramienta para el desarrollo de las técnicas de máquinas de soporte vectorial se utiliza MatLab, el cual está instalado en un computador portátil. También es relevante dejar claro que los datos de ECG no son analizados en tiempo real, es decir, se toma una base de datos que contenga registros de ECG ya clasificados. Luego se le hace preproceso con el fin de eliminar datos que puedan presentar algún conflicto.

Marco teórico

Corazón

El corazón es el órgano encargado de irrigar la sangre a diferentes lugares del cuerpo humano, es decir, se puede considerar como una bomba que mueve la sangre. Este órgano es una visera que se encuentra ubicado dentro de la cavidad torácica, de manera esquemática se puede asemejar a una pirámide recostada en sus lados (Alberto & Ortega, 2016).

Electrocardiograma (ECG)

El electrocardiograma es una prueba que registra la actividad eléctrica del corazón, está considerado como una técnica no invasiva. Las ondas del electrocardiograma son: onda P,

complejo QRS, onda T, onda U. Las porciones del electrocardiograma se llaman segmentos y la distancia entre las ondas intervalos (Picarzo, 2006).

Onda P: En condiciones normales es la primera marca reconocible en el ECG. Corresponde a la llegada de la señal de activación a las aurículas. Su duración es menor de 100ms y su voltaje no excede los 2,5mV.

Intervalo PR: Muestra el período de inactividad eléctrica correspondiente al retraso fisiológico que sufre el estímulo en el nodo aurícula ventricular. Su duración debe estar comprendida entre los 120 y 200ms.

Complejo QRS: Es la marca más característica de la señal electrocardiográfica. Representa la llegada de la señal de activación a ambos ventrículos. Su duración es de 80 a 100ms.

Segmento ST: Comprende desde el final del complejo QRS hasta el inicio de la onda T.

Onda T: Corresponde a la repolarización ventricular, aparece al final del segmento ST.

Intervalo QT: Comprende desde el inicio del complejo QRS hasta el final de la onda T y representa la despolarización y repolarización ventricular. Su duración estará entre 320 y 400 ms.

Amplitudes: Onda P: 0.25 mV

Onda R: 1.60 mV

Onda Q: 25% de la onda R

Onda T: 0.2 a 0.5 Mv

Duraciones: intervalo P-R 0.12 a 20 segundos

Intervalo Q-T 0.35 a 0.44 segundos

Segmento S-T 0.05 A 0.15 segundos

Onda P 0.11 segundos

Intervalos QRS 0.09 segundos

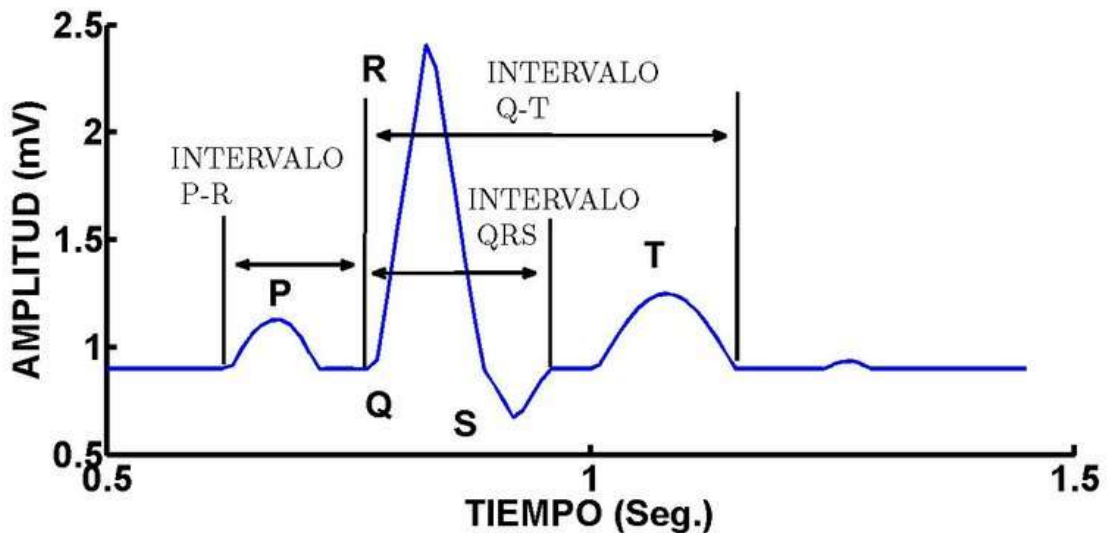


Figura 1. Señal ECG. Fuente: (D. E Rivas-Cisneros, C. Posadas-Castillo, 2014)

Máquina de soporte vectorial (svm)

La máquina de soporte vectorial es una técnica desarrollada por Vladimir Vapnik, son métodos utilizados en clasificación y regresión, además, están catalogadas como aprendizajes supervisado. Inicialmente, las SVM (del inglés *Support Vector Machines*) mapea los datos de

entrada a una dimensión mayor, y busca un hiperplano que los separe con un margen máximo entre las clases(Issn, 2005).

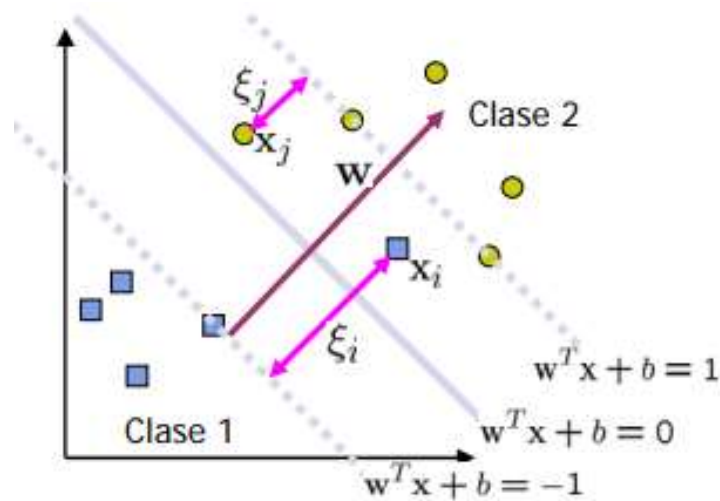


Figura 2. Clasificación de datos en Maquinas de soporte de vectores

Funciones kernel

Dado un conjunto de datos en dos dimensiones que pertenecen a dos clases diferentes, una manera más simple de realizar la agrupación en cada una de sus clases, será utilizar una línea recta. Como se puede observar en la figura 3. Donde las clases están definidas por el color azul y rojo.

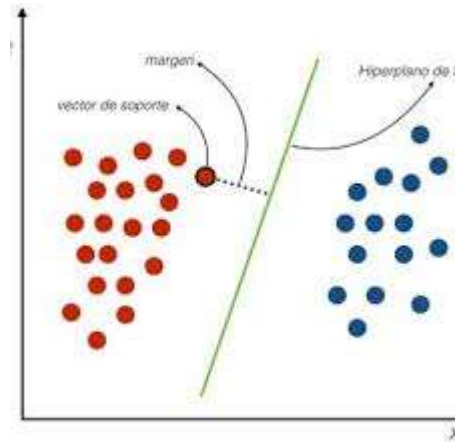


Figura 3. MSV lineal Fuente: (Santibañez, 2015)

Consideremos el problema de clasificación binario expresado como $\{(x_i, y_i = \pm 1)\}$. Se desea separar utilizando un clasificador lineal mostrado en la siguiente expresión.

$$f(x) = w^T x + b = \langle w, x \rangle + b$$

Los vectores de soporte se expresan como:

$$w^T x_j + b = \pm 1$$

Ahora bien, este tipo de estrategias no siempre se pueden aplicar a todos los datos encontrados en el mundo real. Para resolver este tipo de problemas se suele mapear los datos a una dimensión mayor con el fin de que sea fácil separar los datos. En la figura 4 muestra dos clases las cuales están representadas por el color rojo y azul, también se observa que no es posible separar utilizando una línea recta. No obstante, la figura 5 representa los datos de la figura 4, pero ahora se encuentran en una dimensión mayor, en la que se observa que son separables utilizando una línea recta.

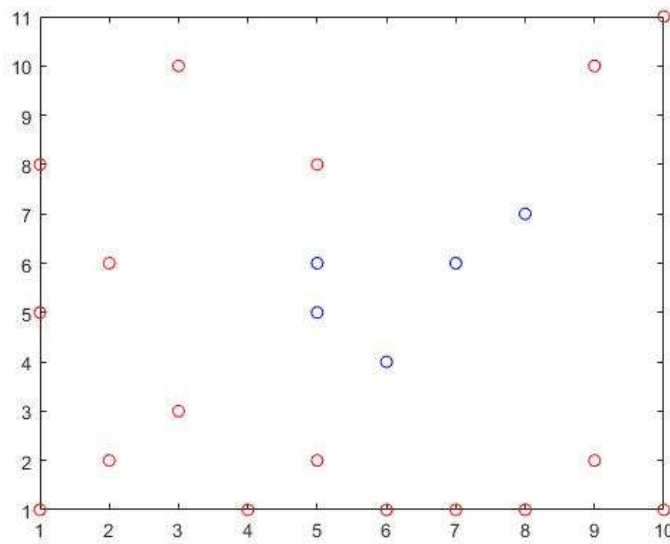


Figura 4. Datos clase Rojo y Azul en R2

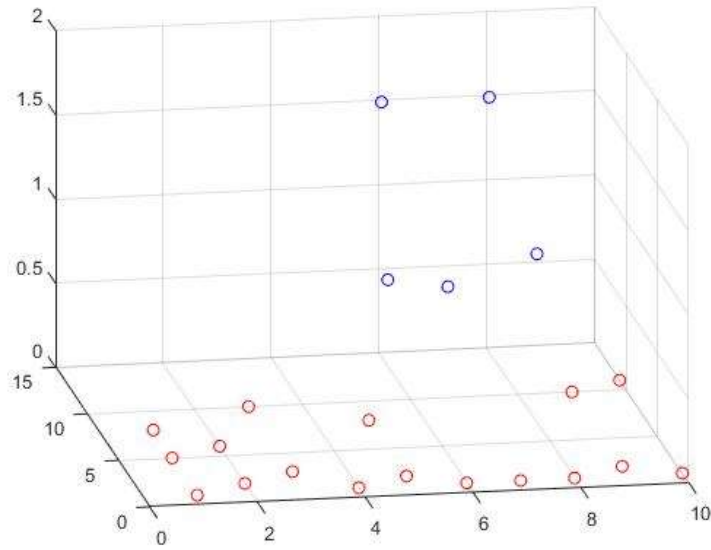


Figura 5. Datos clase Rojo y Azul en R3

Como se puede observar es posible separar los datos que no son lineales utilizando funciones kernels, algunas de ellas se expresan a continuación:

Kernel polinomio se expresa con la siguiente ecuación

$$k(x_i, x_j) = (x_i \cdot x_j + 1)^d$$

Donde d es el grado del polinomio, para el caso cuadrático es 2 y cubico 3.

Kernel cuadrático

$$k(x_i, x_j) = (x_i \cdot x_j + 1)^2$$

Kernel cubico

$$k(x_i, x_j) = (x_i \cdot x_j + 1)^3$$

Kernel Gaussiano

$$k(x, y) = \exp\left(\frac{\|x - y\|^2}{2\sigma^2}\right)$$

Para el caso de gaussiano fino, medio y grueso se debe modificar sigma.

Antecedentes

Los sistemas expertos buscan elegir o clasificar un conjunto de datos. Las máquinas de soporte vectorial han sido usadas para clasificar ECG. En el trabajo presentado en (Matheus, Villazana, & Seijas, 2010) su objetivo es clasificar tres tipos de arritmias diferentes, usa características seleccionadas por el algoritmo Lempel-Ziv y entropía de Shannon. El clasificador diseñado tiene una precisión del 96%.

Por otro lado, la extracción de características de ECG es importante en la clasificación. En este documento (Ecg, 2005) utilizan técnicas de máquina de soporte vectorial para el análisis de la extracción de características de un ECG, analizan dos formas de obtener las características principales; una de ellas es usando transformaciones no lineales y otra usando wavelets. Como conclusión determinan que hay ambigüedad para determinar la técnica más efectiva.

Marco conceptual

Máquinas de soporte de vectores: son técnicas de aprendizaje de máquina por medio de clasificación, está basado en la idea de minimización de riesgo estructural. Conocida como SVM que significa en inglés (support Vector Machines). Se busca con esta técnica de máquina de soporte de vectores hacer un entrenamiento con respecto a una base de datos para poder Extraer conocimiento que pueda ser usado con algún propósito, como clasificar o agrupar datos distintos.

Electrocardiógrafo: aparato electrónico que capta y amplía la actividad eléctrica del corazón a través de electrodos colocados en las 4 extremidades y en 6 posiciones precordiales. El registro de dicha actividad se llama electrocardiograma (ECG) [12].

Electrocardiograma: es la representación gráfica de la actividad eléctrica del corazón que se obtiene, desde la superficie corporal, en el pecho, Es el instrumento principal de la electrofisiología cardíaca y tiene una función de diagnóstico de las enfermedades cardiovasculares [13].

Electrodo: es un conductor eléctrico utilizado para hacer contacto con una parte no metálica de un circuito, los electrodos poseen una celda electroquímica que se refiere a ánodo y cátodo. El ánodo es definido como el electrodo donde los electrones salen de la celda y ocurre la oxidación, el cátodo es el cual los electrones entran a la celda y hay reducción. Los electrodos pueden convertirse en ánodo o cátodo [14].

Anatomía y fisiología del corazón: El corazón es uno de los órganos más importantes del ser humano, este es un órgano muscular contráctil y hueco donde su principal función es la circulación sanguínea, se encuentra entre los dos pulmones, encima del diafragma, delante de la columna vertebral detrás los cartílagos costales y el esternón. Compuesto por dos aurículas y dos ventrículos.

Activación eléctrica del corazón: la actividad eléctrica se genera en el corazón debido a ciertas células especializadas que tienen las características de ser autoexcitables y de funcionamiento periódico, lo cual nos permite decir que son capaces de generar potenciales de acción por si mismas con cierta repetición.

Derivaciones electrocardiográficas: son las encargadas de recoger la diferencia de potencial generada por la actividad eléctrica del corazón, se realiza por medio de electrodos, hay diferentes tipos de derivaciones.

Marco legal

DECRETO NUMERO 4725 DE 2005: 26 DIC 2005: Por el cual se reglamenta el régimen de registros sanitarios, permiso de comercialización y vigilancia sanitaria de los dispositivos médicos para uso humano [16].

DECRETO 1011 DE 2006 sistema obligatorio de garantía de calidad de la atención de salud del sistema general de seguridad social de salud [17].

RESOLUCIÓN NÚMERO 00002003 DE 2014 (28 MAY 2014) Por la cual se definen los procedimientos y condiciones de inscripción de los Prestadores de Servicios de Salud y de habilitación de servicios de salud [18].

Resolución 1043 de 2006 “Por la cual se establecen las condiciones que deben cumplir los Prestadores de Servicios de Salud para habilitar sus servicios e implementar el componente de auditoria para el mejoramiento de la calidad de la atención y se dictan otras disposiciones” [19].

Resolucion-00800- de2015: “el cual da atributos de funciones de requisitos y competencias laborales para el manejo de salud para cumplir las funciones de eficiencia, eficacia y efectividad para cumplir logros al desarrollo, según las funciones que la ley y los reglamentos de salud señalan según el ministerio de salud [20].

Marco tecnológico

Al tratarse de una tarea compleja y de gran importancia para el tratamiento de señales electrocardiográficas se busca encontrar una viabilidad en el desarrollo por medio bases de datos solidas donde se brinde información importante para llevar acabo el estudio.

Actualmente el estudio de la toma de electrocardiogramas se realiza a partir de fuentes convencionales en el área de inteligencia computacional dentro del campo de la electrónica y telecomunicaciones basada en el estudio de las características de cada una de sus fases que son procesamiento digital de señales, sustento matemático, y aplicación de estos, para caracteriza las señales.

Para el estudio de esta propuesta es innegable la complejidad de tecnología que encontramos en todos los ámbitos del ser humano. La medicina no es ajena a todo lo que día a día se vive por implementar nuevas tecnologías que aporten a nuevos conocimientos que sean de manera segura y tengan un grado alto de confiabilidad.

Diseño de aspecto metodológico

Línea de investigación.

En la Corporación Universitaria Autónoma de Nariño, se manejan actualmente tres áreas de investigación que son: Las Telecomunicaciones, la Electro medicina y Telemedicina y la Automatización Industrial, áreas que se proponen de acuerdo a la profundización que puede obtener cada estudiante mediante la selección de las materias electivas profesionales que se dictan en los semestres octavo, noveno y décimo, disponibles en la Facultad. Dentro de estas tres áreas de investigación se ha contemplado el desarrollo de proyectos, para el presente se utilizará la siguiente línea de investigación que es el electro.

Diseño y modelado de sistemas y equipos electrónicos: Uno de los objetivos principales del Ingeniero Electrónico es el diseño de sistemas y equipos electrónicos, que tengan como finalidad la solución de un problema o el mejoramiento de procesos y de la maquinaria existente dentro de los diferentes campos de acción de la electrónica. Para ello es importante pensar en el contexto en el que se desarrolla un proyecto de esta índole, costos, materiales, eficiencia y demás parámetros que llevan a conseguir un excelente producto.

En la línea de investigación de diseño de sistemas y equipos electrónicos, se distinguen dos vertientes diferenciadas que se han venido desarrollando en la Facultad, el diseño digital y el diseño analógico, o bien, la combinación de ellas en algunas implementaciones electrónicas. Para el diseño se tiene como concepto general el uso de las artes aplicadas a un contexto de desarrollo específico como en este caso es la tecnología y específicamente la electrónica y los fundamentos

eléctricos asociados al objeto de estudio que son los dispositivos eléctricos, electrónicos y los sistemas informáticos asociados a ellos.

Enfoque

El enfoque cuantitativo busca probar una hipótesis con ayuda de la estadística, en general analiza datos que son medibles para establecer patrones de comportamiento en una población (Mosquera, 2017). El proyecto se utiliza un enfoque cuantitativo debido a que el estudio está en basado en experimentos que sus resultados son numéricos, estos a su vez se analizan para determinar la efectividad del método.

Tipo de investigación

Para el desarrollo de esta investigación se utilizó un tipo de investigación experimental debido a que se basa en la manipulación de variables teniendo el control total sobre ellas, donde se utiliza un fenómeno que para este caso es un análisis de corazón ECG y máquinas de soporte vectorial para su clasificación. Se plantea experimentos con el fin de obtener los resultados que permita validar o negar la hipótesis.

Análisis de resultados

Patologías relacionadas con el corazón.

Las patologías relacionadas con el corazón son: Cardiopatía isquémica, Insuficiencia cardiaca, Arritmias, entre otras.

Cardiopatía isquémica

Conjunto de alteraciones cardíacas que ocurren por un desequilibrio entre el flujo sanguíneo de las arterias coronarias o flujo coronario y el requerimiento de oxígeno del músculo cardíaco o miocardio. Este desequilibrio produce una isquemia cuyos efectos son metabólicos (aumento del ácido láctico, acidosis, disminución del ATP, disminución de fosfocreatinas), mecánicos (disminución de la contractilidad del corazón, disminución de la distensibilidad de la zona isquémica, y otros) y eléctricos (modificación de potenciales de reposo y acción, inestabilidad eléctrica y los consiguientes trastornos del ritmo). Es la principal causa de muerte en la mayoría de los países desarrollados (Fernández Palomeque et al., 2000).

La principal causa de la enfermedad coronaria es el estrechamiento de las arterias coronarias que irrigan el corazón a causa del aterosclerosis, que básicamente consiste en la acumulación de lípidos en el lumen (ateroma) de una o más arterias coronarias principales en las cuales su revestimiento interno se encuentra inflamado crónicamente. Otras causas incluyen las embolias, la arteritis, la disección, las estenosis ostiales. También se considera el espasmo coronario o angina de Prinzmetal (Ferreira-González, 2014).

Insuficiencia cardiaca

La insuficiencia cardiaca es un problema del corazón, este presenta falencias en la cumplir su función principal de bombear sangre a los demás órganos. Involucrando la sístole y la diástole;

una baja en el gasto cardiaco, falla en la sístole, presenta algunos malestares como es el cansancio, astenia, debilidad frialdad en las extremidades, entre otras (Rodríguez-Artalejo, Banegas Banegas, & Guallar-Castillón, 2004).

Arritmias

El corazón realiza movimientos coordinados que generan señales eléctricas. Las arritmias cardiacas se presentan cuando hay una alteración de los latidos de este órgano, puede ser ritmos rápidos o lentos. Las alteraciones del pueden provenir de una etiología primaria o de causas reversibles, que necesariamente deben ser corregidas (Bernal, cardiología, & 2006). Las principales arritmias son: Cardiopatía isquémica, infarto del miocardio anterior, infarto del miocardio inferior, taquicardia sinusal, bradicardia sinusal, contracción prematura ventricular, contracción supraventricular prematura, bloqueo de rama izquierda y derecha, bloqueo atrioventricular primer grado, segundo y tercero, hipertrofia del ventrículo izquierdo, fibrilación y aleteo auricular.

Base de datos

Cabe aclarar que más que ser una base de datos es un conjunto de datos (dataset). Los datos utilizados para el desarrollo de este trabajo se encuentran en el repositorio UCI, se encuentran almacenadas 16 clases con etiquetas del 1 a 16 que representa una clasificación del tipo de arritmia ver Tabla 1. También cabe aclarar que, en esta colección de datos se incluye 452 pacientes, para cada uno de ellos se tiene 279 características diferentes, estas son asociadas a la

información clínica y a valores de señales adquiridas del corazón. Las características de cada uno de los registros se listan a continuación:

- Edad.
- Sexo
- Talla
- Cintura
- Duración QRS
- Intervalo P-R
- Intervalo Q-T
- Intervalo T
- Intervalo P
- QRS
- T
- P
- QRST
- J
- Ritmo cardiaco
- Canal DI:
- Canal DII
- Canal DIII
- Señales de canales

Tabla 1. Etiquetas de la base de datos.

Etiqueta	Arritmia	Etiqueta	Arritmia
1	Normal	9	Bloqueo de rama izquierda
2	Cardiopatía isquémica	10	Bloqueo de rama derecha
3	Infarto del miocardio anterior	11	Bloqueo atrioventricular primer grado
4	Infarto del miocardio inferior	12	Bloqueo atrioventricular segundo grado
5	Taquicardia sinusal	13	Bloqueo atrioventricular tercero grado
6	Bradicardia sinusal	14	Hipertrofia del ventrículo izquierdo

7	Contracción prematura ventricular	15	Fibrilación y aleteo auricular.
8	Contracción supraventricular prematura	16	Otros

Fuente: esta investigación

Preprocesado datos

En esta sección se realiza una etapa de depuración de la información que se encuentra almacenada en el conjunto de datos. Inicialmente se verifica que todos los datos se encuentren en formato numérico, cuando un elemento de otro tipo se remplaza por un valor numérico 0. A continuación se organiza los datos en un arreglo bidimensional, las filas representan los registros que fueron tomados a pacientes, y columnas contienen los atributos o características de cada registro.

Métricas de desempeño

Los índices o métricas de desempeño permiten evaluar el comportamiento de los procesos, en el área de la inteligencia artificial se ha definido diferentes medidas con las que se puede concluir la eficiencia de los algoritmos.

Para conceptualizar las métricas, primero observe la tabla 2, esta representa dos clases etiquetadas como positivas y negativas. En el eje horizontal están ubicadas la clase mostradas por el clasificador y en el eje vertical representa las clases verdaderas, la intersección positiva – positivo representa los valores verdaderos, en otras palabras, estos son clasificados como positivos y realmente son positivos. La intersección positivo-negativo son denominados falsos negativos, es decir, estos datos son reales pero el clasificador los etiqueta como negativos. El

cruce entre negativo-positivo su resultado son los falsos positivos, esto es, los valores negativos son etiquetados como positivos, pero en realidad son negativos. Y finalmente, queda los verdaderos negativos que son clasificados como sus valores reales.

Tabla 2. Matriz de confusión

		Predicción	
		Positivo	Negativo
Observación	Positivo	Verdaderos Positivos (TP)	Falsos Negativos (FN)
	Negativo	Falsos Positivos (FP)	Verdaderos Negativos(TN)

Fuente: esta investigación

Con el fin de interpretar la matriz de confusión se determina que las clases están relacionadas con un conjunto de pacientes, algunos de ellos se encuentran enfermos y otros sanos. A pacientes enfermos se les asocia la clase positivo y pacientes sanos negativos. Es decir, un usuario positivo es quien tiene la enfermedad, por el contrario, una persona sana es negativo debido a que no padece la patología. Con base en lo anterior expuesto se puede realizar la definición de cada métrica.

Sensibilidad

Es la capacidad que posee un clasificador para determinar valores positivos. Es decir, detecta la enfermedad en personas que la padece. Este valor puede estar en porcentaje o tener una

cantidad punto flotante entre 0 y 1; cuando el resultado es 1 indica que clasifico todas las etiquetas correctamente, por el contrario, un valor 0 indica que no clasificó ningún valor. Para el cálculo se utiliza la siguiente expresión:

$$Se = \frac{TP}{TP + FN}$$

Especificidad

Es una métrica que indica la capacidad que tiene un estimador para clasificar casos negativos cuando realmente son negativos. El valor se encuentra entre 0 y 1, donde un 1 indica que todos los resultados fueron clasificados exitosamente.

$$Sp = \frac{TN}{TN + FP}$$

Exactitud

Es un índice que muestra las predicciones correctas. Este indicador puede tomar valores entre 0 y 1, la mejor exactitud sería 1 y, por el contrario, 0 sería la peor. Matemáticamente es la suma de las clasificaciones correctas sobre el número total de registros.

$$e = \frac{TP + TN}{TN + TP + FN + FP}$$

Precisión

Es un indicador que proporciona información sobre qué proporción de los pacientes diagnosticados enfermos, realmente tiene la enfermedad.

$$R = \frac{TP}{TP + FP}$$

Score F1

Es un indicador de la media armónica entre la precisión y la sensibilidad.

$$F1 = \frac{2 \cdot R \cdot Se}{R + Se}$$

Metodología experimental

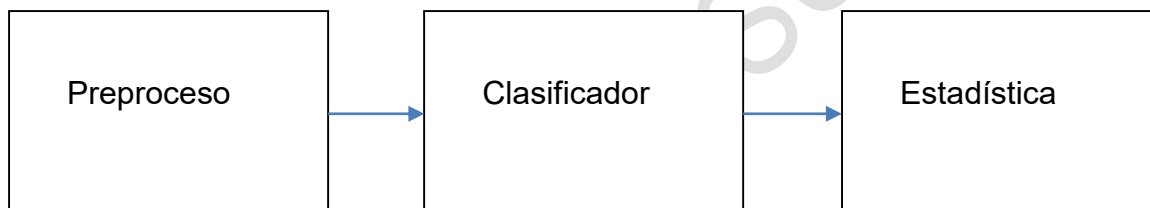
En esta sección se describe el proceso llevado a cabo en el desarrollo de esta investigación. Como el conjunto de datos seleccionados tiene diferentes clases que representan una arritmia particular, se procede inicialmente a realizar pruebas utilizando un clasificador biclase, es decir, que los datos de entrada solo pertenecen a una de dos clases. Luego, se procede a experimentar con un clasificador multiclase, es decir, los datos de entrada al clasificador contienen diferentes clases, en este caso se utiliza todas las etiquetas del dataset.

También en las pruebas realizadas tanto en el clasificador biclase y multiclase se realizó utilizando validación cruzada, consiste en dividir el conjunto de datos de prueba en subconjuntos, con el fin de evitar sesgos en los resultados. Por otro lado, se evalúa cambiando las dimensiones de los valores, en otras palabras, se realiza modificando las características.

Desarrollo del script en lenguaje M

Matlab es una herramienta matemática, que está orientada a trabajar con matrices, usada en diferentes procesos ingenieriles. Este software tiene su propio lenguaje de programación denominado **M**. En este trabajo se utilizó esta herramienta debido a su trabajo con matrices e incursión en algoritmos machine learning. Para cumplir los objetivos se desarrolló un script que permite obtener información pertinente. En la figura 6 se muestra el diagrama de bloques.

Figura 6. Diagrama de bloques script



Fuente: esta investigación

En el bloque denominado preproceso se realiza la lectura de los datos, luego se construye una matriz multidimensional, donde las columnas se asocia con los atributos y las filas con registros, también se verifica que todos los datos sean numéricos. Posterior a esto se realiza la extracción de las características seleccionadas.

En el bloque clasificador está el modelo clasificador, en él se incluye las funciones matemáticas para diversos tipos de kernel lineal, polinómico, gaussiano. Esta función recibe los datos que son entregados por el bloque preproceso y luego realiza la clasificación utilizando todos los kernels mencionados. Y finalmente entrega una matriz de datos para que sean analizados por el siguiente modulo.

Finalmente, se encuentra el modulo denominado estadístico, este se encarga de entregar información estadística de los resultados obtenidos por el clasificador, aquí se encuentran integradas las métricas de desempeño mostradas en la sección anterior.

Discusión de resultados

Experimentación con la clase 1 y 2

La tabla 3 muestra los índices de desempeño de las clases 1 y 2, se utilizó las primeras características. El índice armónico (F1) más alto se obtuvo al utilizar el kernel gaussiano medio, con un valor de 0,89271392. Por otro lado, el menor valor es 0,84775087 que corresponde al kernel gaussiano fino.

Tabla 3. Resultados obtenidos basado en primeras características de clase 1 y 2.

Tipo	Exactitud	Tasa error	Sensibilidad	Especificidad	Precisión	F1
Lineal	0,87889273	0,12110727	0,95918367	0,43181818	0,90384615	0,8911948
Cuadrático	0,8615917	0,1384083	0,93469388	0,45454545	0,90513834	0,88282835
Cubico	0,82352941	0,17647059	0,88571429	0,47727273	0,90416667	0,86196624
Gaussiano Fino	0,84775087	0,15224913	1	0	0,84775087	0,84775087
Gaussiano Medio	0,87889273	0,12110727	0,95510204	0,45454545	0,90697674	0,89271392
Gaussiano Grueso	0,85467128	0,14532872	0,9877551	0,11363636	0,86120996	0,85792816

Fuente: esta investigación

En la tabla 4 se presenta las métricas de las clases 1 y 2, al contrario de la anterior aquí se utiliza características diferentes. Los kernel gaussiano fino y grueso tienen una especificidad igual a 0, es decir, el clasificador no etiqueto ningún dato, en estas dos técnicas presenta un índice armónico menor.

Tabla 4. Resultados obtenidos basado en segundas características de clase 1 y 2.

Tipo	Exactitud	Tasa error	Sensibilidad	Especificidad	Precisión	F1
Lineal	0,84429066	0,15570934	0,99183673	0,02272727	0,84965035	0,84696202
Cuadrático	0,85121107	0,14878893	0,95510204	0,27272727	0,87969925	0,86522072
Cubico	0,84429066	0,15570934	0,94693878	0,27272727	0,87878788	0,86119394
Gaussiano Fino	0,84775087	0,15224913	1	0	0,84775087	0,84775087
Gaussiano Medio	0,84775087	0,15224913	0,9877551	0,06818182	0,85512367	0,85142131
Gaussiano Grueso	0,84429066	0,15570934	0,99591837	0	0,84722222	0,8457539

Fuente: esta investigación

En la tabla 5 se visualiza los resultados calculados de las métricas de desempeño. Cabe aclarar que para esta prueba se unieron las características utilizadas en las dos tablas anteriores. F1 tiene un mayor valor cuando se usa un kernel lineal, y su valor es menor cuando la sensibilidad es 0.9918 y especificidad 0.

Tabla 5. Resultados obtenidos basado en todas las características de clase 1 y 2.

Tipo	Exactitud	Tasa error	Sensibilidad	Especificidad	Precisión	F1
Lineal	0,88927336	0,11072664	0,9755102	0,40909091	0,90188679	0,89553566
Cuadrático	0,88927336	0,11072664	0,97142857	0,43181818	0,90494297	0,89703974
Cubico	0,85813149	0,14186851	0,93877551	0,40909091	0,8984375	0,87782207
Gaussiano Fino	0,84775087	0,15224913	1	0	0,84775087	0,84775087
Gaussiano Medio	0,87543253	0,12456747	0,97959184	0,29545455	0,88560886	0,88049129
Gaussiano Grueso	0,84083045	0,15916955	0,99183673	0	0,8466899	0,84375

Fuente: esta investigación

Las tablas 3,4 y 5 muestran valores numéricos de diferentes índices de desempeño, pero en este caso solo se tiene en cuenta el índice F1 o media armónica, esto se debe a que los datos no están balanceados por tal motivo si se analiza otro parámetro tiende a tener sesgo hacia una métrica. El kernel que presenta un mejor desempeño es el lineal cuando se utiliza todas las características.

Experimentación con la clase 1 y 3

Para esta prueba se tuvieron en cuenta las clases 1 y 3. En la tabla 6 son presentados las métricas de desempeño del algoritmo. Se puede observar que con los kernel lineal, gaussiano fino y grueso la especificidad tiene un valor de 0. No obstante, los valores de la media armónica son altos, todos superan el 0.93, pero el valor más alto es presentado por el kernel cubico.

Tabla 6. Resultados obtenidos basado en primeras características de clase 1 y 3.

Tipo	Exactitud	Tasa error	Sensibilidad	Especificidad	Precisión	F1
Lineal	0,93461538	0,06538462	0,99183673	0	0,94186047	0,93822394
Cuadrático	0,95	0,05	0,9877551	0,33333333	0,96031746	0,95513087
Cubico	0,95384615	0,04615385	0,9877551	0,4	0,96414343	0,95896715
Gaussiano Fino	0,94230769	0,05769231	1	0	0,94230769	0,94230769
Gaussiano Medio	0,94230769	0,05769231	0,99591837	0,06666667	0,94573643	0,94401895
Gaussiano Grueso	0,94230769	0,05769231	1	0	0,94230769	0,94230769

Fuente: esta investigación

En la tabla 7, se observa que los kernel gaussiano fino y grueso presentan sensibilidad alta y una especificidad nula. La media armónica se mantiene por encima de 0.94, siendo el kernel lineal el que mejor score tiene.

Tabla 7. Resultados obtenidos basado en segundas características de clase 1 y 3.

Tipo	Exactitud	Tasa error	Sensibilidad	Especificidad	Precisión	F1
Lineal	0,94615385	0,05384615	0,99183673	0,2	0,95294118	0,94953538
Cuadrático	0,93846154	0,06153846	0,97959184	0,26666667	0,9561753	0,94723561
Cubico	0,91923077	0,08076923	0,95102041	0,4	0,96280992	0,9405158
Gaussiano Fino	0,94230769	0,05769231	1	0	0,94230769	0,94230769

Gaussiano Medio	0,93846154	0,06153846	0,99183673	0,06666667	0,94552529	0,94198017
Gaussiano Grueso	0,94230769	0,05769231	1	0	0,94230769	0,94230769

Fuente: esta investigación

En la tabla 8 se muestran los resultados de las métricas de las clases 1 y 3 al utilizar los dos grupos de características. La especificidad tiene un valor de 0 al utilizar kernel gaussiano, y la sensibilidad es 1 en gaussiano fino y grueso.

Tabla 8. Resultados obtenidos basado en todas las características de clase 1 y 3.

Tipo	Exactitud	Tasa error	Sensibilidad	Especificidad	Precisión	F1
Lineal	0,95	0,05	0,99591837	0,2	0,953125	0,95155993
Cuadrático	0,95	0,05	0,97959184	0,46666667	0,96774194	0,9587889
Cubico	0,95769231	0,04230769	0,9877551	0,46666667	0,968	0,96281857
Gaussiano Fino	0,94230769	0,05769231	1	0	0,94230769	0,94230769
Gaussiano Medio	0,93461538	0,06538462	0,99183673	0	0,94186047	0,93822394
Gaussiano Grueso	0,94230769	0,05769231	1	0	0,94230769	0,94230769

Las tablas 6, 7 y 8 evidencian los índices de desempeño cuando se compara las clases 1 y 3.

Las técnicas de gaussiano fino y grueso presentan una especificidad con valor de 0, en otras palabras, el clasificador cuando utiliza este tipo de kernels no puede clasificar ningún paciente. También se encuentra que el valor más alto de la media armónica es 0,958, este se presenta al utilizar gaussiano cubico basado en el primer grupo de características.

Experimentación clase 1 y 4

En la tabla 9 encontramos que la sensibilidad para kernel tipo lineal, gaussiano fino, medio y grueso es igual a 1, para este tipo de técnicas también la especificidad es 0.

Tabla 9. Resultados obtenidos basado en primeras características de clase 1 y 4.

Tipo	Exactitud	Tasa error	Sensibilidad	Especificidad	Precisión	F1
Lineal	0,94230769	0,05769231	1	0	0,94230769	0,94230769
Cuadrático	0,91923077	0,08076923	0,96734694	0,13333333	0,948	0,93339375
Cubico	0,90769231	0,09230769	0,95510204	0,13333333	0,94736842	0,92710607
Gaussiano Fino	0,94230769	0,05769231	1	0	0,94230769	0,94230769
Gaussiano Medio	0,94230769	0,05769231	1	0	0,94230769	0,94230769
Gaussiano Grueso	0,94230769	0,05769231	1	0	0,94230769	0,94230769

Fuente: esta investigación

La tabla 10 muestra los resultados de los indicadores de desempeño cuando se hace uso del segundo grupo de características. Se puede observar que la métrica de sensibilidad es igual a 0 cuando se utiliza kernel lineal, cuadrático y gaussiano con todas sus variaciones, también la sensibilidad del modelo cuadrático es 0,93.

Tabla 10. Resultados obtenidos basado en segundas características de clase 1 y 4.

Tipo	Exactitud	Tasa error	Sensibilidad	Especificidad	Precisión	F1
Lineal	0,94230769	0,05769231	1	0	0,94230769	0,94230769
Cuadrático	0,90384615	0,09615385	0,95918367	0	0,94	0,92156863
Cubico	0,88846154	0,11153846	0,93061224	0,2	0,95	0,91820084
Gaussiano Fino	0,94230769	0,05769231	1	0	0,94230769	0,94230769
Gaussiano Medio	0,94230769	0,05769231	1	0	0,94230769	0,94230769
Gaussiano Grueso	0,94230769	0,05769231	1	0	0,94230769	0,94230769

Fuente: esta investigación

En la tabla 11 la especificidad cuando se utiliza kernel lineal y gaussiano fino, grueso y medio no hay ninguna clasificación para esta clase. Por el contrario, en todas las combinaciones de gaussiano la sensibilidad es 1, es decir, que fueron clasificadas bien todas las muestras.

Tabla 11. Resultados obtenidos basado en todas las características de clase 1 y 4.

Tipo	Exactitud	Tasa error	Sensibilidad	Especificidad	Precisión	F1
Lineal	0,93461538	0,06538462	0,99183673	0	0,94186047	0,93822394
Cuadrático	0,94615385	0,05384615	0,9877551	0,26666667	0,95652174	0,95130954
Cubico	0,93846154	0,06153846	0,97959184	0,26666667	0,9561753	0,94723561
Gaussiano Fino	0,94230769	0,05769231	1	0	0,94230769	0,94230769
Gaussiano Medio	0,94230769	0,05769231	1	0	0,94230769	0,94230769
Gaussiano Grueso	0,94230769	0,05769231	1	0	0,94230769	0,94230769

Fuente: esta investigación

Experimentación clase 2 y todas las clases

En esta prueba sigue siendo un clasificador biclase, la modificación que se realiza es que las clases diferentes a la 2 se las cambio para que pertenezcan a una sola clase.

En la tabla 3 gaussiano en todas sus combinaciones tiene una especificidad 0, al igual que el tipo lineal. El valor más alto de la media armónica es 0,91 presentado por el kernel cuadrático.

Tabla 12. Resultados obtenidos basado en primeras características de clase 2 y todas

Tipo	Exactitud	Tasa error	Sensibilidad	Especificidad	Precisión	F1
Lineal	0,89823009	0,10176991	0,99509804	0	0,90222222	0,90022173
Cuadrático	0,89823009	0,10176991	0,96813725	0,25	0,9228972	0,91039658
Cubico	0,88053097	0,11946903	0,94852941	0,25	0,92142857	0,90051566
Gaussiano Fino	0,90265487	0,09734513	1	0	0,90265487	0,90265487
Gaussiano Medio	0,90044248	0,09955752	0,99754902	0	0,90243902	0,90143965
Gaussiano Grueso	0,90265487	0,09734513	1	0	0,90265487	0,90265487

Fuente: esta investigación

En la tabla 13 el valor de la sensibilidad es igual a 1 en gaussiano grueso, y la especificidad tiene valores 0 en tipos lineal, todas las combinaciones de gaussiano.

	8	1	1														
	9	1	5					3									
	10	50															
	14	4															
	15	4	1														
	16	18	3														
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Predicción																	

Fuente: esta investigación

En la tabla 15 se muestra la matriz de confusión para un kernel cuadrático. El clasificador para las clases 6, 7, 8, 11, 12 y 13 no encuentra datos. En las clases 1, 2 y 3 es donde clasifica mayores datos independiente si son verdaderos o falsos, la clase 1 y 2 presenta un mayor número de aciertos.

Tabla 15. Matriz de confusión todas las clases (Cuadrático).

Observaciones	1	226	7	1	2	1				5							3	
	2	21	18		1					3								1
	3	9	3	2													1	
	4	12	1	2														
	5	7	1		1					1				1				2
	6	25																
	7	3																
	8	1	1															
	9		5							2	1				1			
	10	41	1		1	1					5							1
	14	2	1			1												
	15	4	1															
	16	13	4								2					1		2
			1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
	Predicción																	

Fuente: esta investigación

La tabla 16 muestra los resultados de la matriz de confusión cuando se utiliza un kernel cubico. Se puede observar que hay una distribución de datos en toda la tabla, ahora bien, evaluando que tan eficiente es este tipo de estrategia, se puede afirmar que en su mayoría los datos la clasificación entregada por el sistema esta errada, con la excepción de las clases 1 y 2 que presenta mejores resultados que para las otras clases.

Tabla 16. Matriz de confusión todas las clases (Cubico).

Observaciones	1	189	11	4	2	6	11				18					4	
	2	19	18	1	2						3					1	
	3	5	3	4	1						1					1	
	4	8	3	1	1	1					1						
	5	6		1	2						1			1	1	1	
	6	20	1				1				2					1	
	7	3															
	8	1	1														
	9				1						7					1	
	10	32	2	2		1					6					4	
	14	1		1							2						
	15	2									1					1	
	16	7	3				1	1	1		4			1	2	1	
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
	Predicciones																

Fuente: esta investigación

En la tabla 17 se presenta los resultados obtenidos al utilizar gaussiano fino. Para la case 1, de 245 registros verdaderos, el clasificador entrego 242 solo dejando por fuera de esta clase a 3 registros, en cambio para la clase 2, de 44 registros, solo 4 fueron clasificados exitosamente. Para las demás clases no clasifico ningún dato.

Tabla 17. Matriz de confusión todas las clases (Gaussiano fino)

Observaciones	1	242	3															
	2	40	4															
	3	15																
	4	15																
	5	13																
	6	25																
	7	3																
	8	2																
	9	6	3															
	10	50																
	14	4																
	15	5																
	16	22																
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	
	Predicciones																	

Fuente: esta investigación

Como se puede evidenciar en la tabla 17 es pobre la clasificación que realiza este tipo de técnicas, los resultados de técnicas maquinas supervisadas utilizando kernel gaussiano medio y grueso tuvieron resultados similares a los presentados por gaussiano fino.



Conclusiones

Desde la perspectiva médica existen diferentes patologías que tiene relación con el corazón, en su mayoría estas se ven reflejadas en la alteración de las señales eléctricas del corazón.

También es de considerar que, para algunos casos pruebas como ECG no son concluyentes en el diagnóstico, motivo por el cual un clasificador basado en ECG no podría determinar la patología.

Los clasificadores biclase utilizados en el desarrollo de este trabajo, presentan mejores resultados que los multiclase. La media armónica más alta es presentada por un kernel cubico con un valor de 0,958.

Los sistemas multiclase utilizando máquinas de soporte vectorial con kernel gaussiano fino, medio y grueso sus errores son altos, es decir, para las clases 3 en adelante no clasifica ningún valor. Para la clase 1, de 245 registros este clasifico 243 que son positivos verdaderos, pero para también clasifico en esta misma clase registros que no perteneces a esta clase que serían los falsos positivos.



Referencias

- Alberto, H., & Ortega, C. (2016). *El corazón y sus ruidos cardíacos normales y agregados*. 59, 49–55.
- Bernal, O., cardiología, C. M.-R. española de, & 2006, undefined. (n.d.). Arritmias cardiacas en la mujer. *Elsevier*. Retrieved from <https://www.sciencedirect.com/science/article/pii/S0300893206746887>
- D. E Rivas-Cisneros, C. Posadas-Castillo, D. A. D.-R. (2014). *Estudio y Diagnóstico de Patologías Cardiacas Utilizando Teoría de Nudos DE PATOLOGÍAS CARDIACAS UTILIZANDO TEORÍA*. (May 2016).
- DANE. (2017). *Boletín técnico Estadísticas vitales Nacimientos y Defunciones*. (23), 1–4.
- Ecg, E. (2005). *Extracción de características de ECG basadas en transformaciones no lineales y wavelets Extracting ECG signal characteristics based on non-linear transformations and wavelets*. 25(3), 39–48.
- Expósito, C., Ávila, R., & Ii, Á. (2008). *Aplicaciones de la inteligencia artificial en la Medicina : perspectivas y problemas Artificial intelligence applied to medicine : prospects and problems*. 1–12.
- Fernández Palomeque, C., Bardají Mayor, J. L., Concha Ruiz, M., Cordo Mollar, J. C., Cosín Aguilar, J., Magriñá Ballara, J., & Melgares Moreno, R. (2000). Guías de práctica clínica de la Sociedad Española de Cardiología en la angina estable. *Revista Española de Cardiología*, 53(7), 967–996. <https://doi.org/10.1157/10487>
- Ferreira-González, I. (2014). Epidemiología de la enfermedad coronaria. *Revista Espanola de Cardiologia*, 67(2), 139–144. <https://doi.org/10.1016/j.recesp.2013.10.003>



Issn, P. (2005). *LAS MÁQUINAS DE SOPORTE VECTORIAL (SVMs)*. (27), 67–72.

Matheus, D., Villazana, S., & Seijas, C. (2010). *Deteccion de Arritmias en Señales de ECG*

Basado en Máquinas de Vectores de Soporte Support Vector Machines Based Arrhythmia Detection from ECG Signal.

Mosquera, R. E. (2017). Diseño de Experimento: Aproximación metodología para el desarrollo de una aplicación interactiva de realidad virtual Inmersiva, para recrear la terapia espejo en la rehabilitación de pacientes amputados de miembros inferiores con dolor de miembro fantasma. *Memorias de Congresos UTP*, 0(0), 224–228. Retrieved from <http://revistas.utp.ac.pa/index.php/memoutp/article/view/1499/2188>

Picarzo, F. J. P. (2006). *Guía rápida para la lectura sistemática del ECG pediátrico*. 127(319), 319–326.

Rodríguez-Artalejo, F., Banegas Banegas, J. R., & Guallar-Castillón, P. (2004). Epidemiología de la insuficiencia cardíaca. *Revista Española de Cardiología*, 57(2), 163–170. [https://doi.org/10.1016/S0300-8932\(04\)77080-3](https://doi.org/10.1016/S0300-8932(04)77080-3)

Santibañez, T. (2015). *Validez predictiva de un clasificador basado en máquina de soporte vectorial para éxito o fracaso en la extubación de pacientes conectados a ventilación mecánica invasiva en una un ...* (October). <https://doi.org/10.13140/RG.2.1.4939.2082>