

**DETECCIÓN DE ANOMALÍAS EN SEÑALES ELECTROCARDIOGRAFICAS
USANDO MINERÍA DE FLUJO DE DATOS**

**LIDIS MAYERLY RONDÓN SUAREZ
GERMAN EDUARDO CORDOBA CARREÑO**

**UNIVERSIDAD INDUSTRIAL DE SANTANDER
FACULTAD DE INGENIERÍAS FISICOMECÁNICAS
ESCUELA DE INGENIERÍA DE SISTEMAS E INFORMÁTICA
BUCARAMANGA**

2013

**DETECCIÓN DE ANOMALÍAS EN SEÑALES ELECTROCARDIOGRAFICAS
USANDO MINERÍA DE FLUJO DE DATOS**

**LIDIS MAYERLY RONDÓN SUÁREZ
GERMAN EDUARDO CORDOBA CARREÑO**

**Trabajo de grado para optar al título de
Ingeniero de Sistemas**

**Directora:
MSc. Lola Xiomara Bautista Rozo**

**UNIVERSIDAD INDUSTRIAL DE SANTANDER
FACULTAD DE INGENIERÍAS FISICOMECÁNICAS
ESCUELA DE INGENIERÍA DE SISTEMAS E INFORMÁTICA
BUCARAMANGA**

2013

A mis padres Ramiro Rondón y Josefina Suárez, por todos sus esfuerzos y apoyo brindado en esta etapa de mi vida. A mis hermanas Janneth, Karen y Marce por su paciencia y comprensión, también gracias a mi pequeña sobrina Alejandra que con sus sonrisas alegra mi existencia. A mis tías, Flor Ángela, María Elena y Elizabeth, por su ayuda incondicional y buenos consejos. A mis amigos Mario, Edinson, Camilo y Pedro, que con sus palabras y experiencias de vida lograron ganarse mi respeto y admiración. A mi compañero de proyecto Germán Córdoba por su trabajo y dedicación.

Lidis Mayerly Rondón Suárez

Dedico este libro a Dios, por todas las cosas que me han pasado en la vida hasta el momento, A mis padres Ricardo y Graciela por su cariño, dedicación, amor y apoyo incondicional, a mis hermanas Claudia Patricia y Luz Stella por su apoyo y cariño, A mi compañera de proyecto Lidis por su esfuerzo, dedicación y persistencia.

A la directora de proyecto Lola y nuestro colaborador Nelson por su acompañamiento, por su apoyo, al los integrantes del grupo de investigación GIB que mas que compañeros son amigos e ingenieros con los cuales compartimos momentos especiales, a los profesores y docentes de la escuela de Ingeniería de Sistemas UIS quienes con su experiencia y labor contribuyeron a mi formación.

A la Coca-Cola por apoyar mi carrera universitaria, a Ecopetrol, mis jefes, compañeros de oficina y amigos, por abrirme las puertas de la empresa y acompañarme en mi crecimiento tanto laboral como personal.

A mis amigos incondicionales, compañeros de aventura y a toda la gente que contribuyo a que este logro se alcanzara, y por construir tantas historias para ser contadas.

A las personas quienes me han dedicado su tiempo y un momento de sus vidas no solo a escucharme sino brindando sus valiosos consejos, compañía y sonrisas.

Mil gracias a todas las personas con las que he compartido en la vida.

Germán Eduardo Córdoba Carreño

AGRADECIMIENTOS

A las personas que caminaron a nuestro lado en el desarrollo de este trabajo, especialmente a la profesora Lola Xiomara Bautista y Nelson Enrique León, por compartir sus conocimientos sobre la temática que aborda este trabajo.

A la Universidad industria de Santander, por contribuir con el crecimiento a la formación académica con sentido social.

Al grupo de Investigación de Ingeniería Biomédica por apoyarnos en el proceso de investigación.

TABLA DE CONTENIDO

INTRODUCCIÓN	18
1. DESCRIPCIÓN DEL PROYECTO	20
1.1 PLANTEAMIENTO DEL PROBLEMA.....	20
2. JUSTIFICACIÓN	22
3. DESCRIPCIÓN DE LA SOLUCIÓN	23
3.1 OBJETIVOS	23
3.1.1 Objetivo General	23
3.1.2 Objetivos específicos.....	23
3.2 METODOLOGÍA	24
3.2.1 Investigación-Acción.....	25
3.2.2 Metodología Adaptativa.....	26
4. MARCO TEORICO Y ESTADO DEL ARTE	28
4.1 MARCO TEORICO	28
4.1.1 Electrocardiograma	28
4.1.1.1 Derivaciones electrocardiográficas	29
4.1.1.2 Ondas, intervalos y segmentos del electrocardiograma.....	31
4.1.2 Minería de Datos	34
4.1.3 Flujo de datos.....	35
4.1.4 Proceso general de minería de flujo de datos	36
4.1.5 Requerimientos de flujo de datos	38
4.1.6 Técnicas de flujo de datos.....	39
4.1.7 Anomalía	40
4.1.7.1 Tipos de anomalías.....	42
4.1.8 Procesamiento digital de señales.....	42
4.1.8.1 La Transformada de wavelet.....	43
4.1.8.2 Transformada Discreta Wavelet (DWT)	44

4.1.8.3 Etapas del Procesamiento Digital de señales	46
4.2 ESTADO DEL ARTE	47
4.2.1 Descripción de técnicas estudiadas	47
4.2.2 Métodos de agrupamiento (Clustering)	55
4.2.2.1 Método de Particional	55
4.2.2.2 Método Jerárquico	56
5. DESARROLLO DE LA SOLUCIÓN	57
5.1 DISEÑO DEL ALGORITMO	57
5.2 ADQUISICIÓN DE LA SEÑAL	62
5.1.1 Montaje de Adquisición, Preprocesamiento y Segmentación	63
5.3 IMPLEMENTACIÓN Y DESARROLLO DE LA INTERFAZ GRAFICA	68
5.3.1 Diagramas de casos de uso	69
5.3.2 Diagrama de clases	70
5.3.4 Demostración de la herramienta Software	71
5.4 PRUEBA DE LA HERRAMIENTA	77
5.4.1 Prueba de Sensibilidad y Especificidad	77
6. CONCLUSIONES Y RECOMENDACIONES	84
6.1 CONCLUSIONES	84
Fuente: Autores del Proyecto	85
6.2 RECOMENDACIONES Y TRABAJO FUTURO	85
7. PRODUCCIÓN INTELECTUAL	86
ANEXOS	92
A. MANUAL DE USUARIO DIAGECG	92

LISTA DE TABLAS

Tabla 1. Prueba de herramienta con ventana de 1000 datos	79
Tabla 2. Cantidad de flujo de datos en R.....	80
Tabla 3. Prueba de herramienta con ventana de 1000 datos	81
Tabla 4. Cantidad de flujo de datos en R.....	83
Tabla 5. Revisión de cumplimiento de objetivos Específicos.....	85

LISTA DE FIGURAS

Figura 1. Representación gráfica del planteamiento del problema.	21
Figura 2. Metodología utilizada	25
Figura 3. Metodología de la Investigación: Investigación-Acción.....	26
Figura 4. Metodología de Desarrollo: Metodología Adaptativa	27
Figura 5. Representación esquemática de un ciclo de ECG normal.....	28
Figura 6. Derivaciones Frontales	30
Figura 7. Derivaciones Precordiales	31
Figura 8. Representación de ondas y complejos en Un ECG	32
Figura 9. Señal de ECG.....	36
Figura 10. Proceso general de minería de datos	37
Figura 11. Clasificación de algoritmos de flujo de datos	40
Figura 12. Anomalía presente en una señal de E.C.G.....	41
Figura 13. Componentes clave asociadas con una técnica de detección de anomalías	41
Figura 14. Proceso de descomposición (Análisis)	44
Figura 15. Proceso de descomposición y reconstrucción.....	45
Figura 16. Algoritmo de Etapas para el procesamiento digital de señales.....	58
Figura 17. Pseudocódigo del algoritmo de k-medias	60
Figura 18. Imagen que muestra un paciente conectado a los 10 electrodos necesarios para un ECG de 12 derivaciones.....	63

Figura 19. Elementos empleados para la adquisición de la señal de ECG.....	64
Figura 20. Montaje para la adquisición de la señal de ECG	64
Figura 21. Secuencia de operaciones algoritmo de Caracterización.	65
Figura 22. Señal Original	66
Figura 23. Señal Filtrada.....	66
Figura 24. Señal filtrada, comparada con Haar para hallar R	67
Figura 25. Detección del complejo QRS con Daubechies y Haar	67
Figura 26. Diagrama de Casos de Uso.....	69
Figura 27. Diagrama de clases	70
Figura 28. Toma de Pantalla de la interfaz inicial de DiagECG	71
Figura 29. Selección del directorio de la señal a cargar	72
Figura 30. Grafica Electrocardiograma herramienta	73
Figura 31. Tabla de datos que resulta de interfaz gráfica	74
Figura 32. Tabla producto de la exportación de la herramienta	74
Figura 33. Tabla de dato ajustable y análisis	75
Figura 34. Grafica que muestra estadísticas de la clasificación	76
Figura 35. Ventana que aparece con la opción analizar	76
Figura 36. Señal analizada por el especialista.....	78
Figura 37. Promedio pacientes ventana de 1000 datos.....	79
Figura 38. Frecuencia promedio	80
Figura 39. Cantidad de R detectadas en flujo.....	81
Figura 40. Promedio pacientes ventana de 1500 datos.....	82
Figura 41. Promedio Frecuencia.....	82
Figura 42. Cantidad de R detectadas en flujo.....	83

Figura A1. Botones de la aplicación DiagECG	92
Figura A2. Botones de Limpiar y Exportar	93
Figura A3. Categorías de clasificación.....	94
Figura A4. Vista de los datos cargados y su respectiva grafica.....	94
Figura A5. Datos ajustables.....	95
Figura A6. Guardar la imagen de la señal	95
Figura A7. Menú desplegable archivo.....	96
Figura A8. Menú Activar y desactivar malla.....	97
Figura A9. Menú Mostrar Puntos	97
Figura A10. Menú ocultar Puntos	97
Figura A11. Menú Agregar.....	98
Figura A12. Menú Eliminar	98
Figura A13. Menú Archivo	98
Figura A14. Menú ayuda.....	99

RESUMEN

TÍTULO:

DETECCIÓN DE ANOMALÍAS EN SEÑALES ELECTROCARDIOGRAFICAS USANDO MINERÍA DE FLUJO DE DATOS*

AUTORES:

LIDIS MAYERLY RONDÓN SUÁREZ

GERMAN EDUARDO CORDOBA CARREÑO**

PALABRAS CLAVES: ECG, flujo de datos, anomalía, minería de datos

DESCRIPCIÓN

El presente trabajo tiene como fin proporcionar una herramienta que permita detectar anomalías en señales electrocardiográficas (ECG), que sirva de soporte al experto y contribuya con el cuidado de la salud cardiovascular. Esta investigación parte de estudios realizados que muestran que las principales causas de muerte en el mundo se presentan por enfermedades cardiovasculares. Una de las características de este trabajo es mostrar los procedimientos que se realizaron para el análisis de las señales ECG a través de técnicas de minería de flujos de datos (Data Stream Mining en inglés), con el fin de detectar cualquier posibilidad de anomalía en las señales. Para esto se hizo una revisión bibliográfica con el fin de investigar técnicas de minería de flujo de datos que sirvieran para la extracción de información relevante en señales de ECG. Posteriormente se procedió a escoger la técnica de minería de flujo de datos que se ajustara de manera eficiente a la solución del problema planteado; una vez seleccionada la técnica se diseñó un algoritmo donde se plantea el proceso general para el tratamiento de la señal y la implementación del algoritmo con la técnica de minería de flujo de datos seleccionada. Basados en las investigaciones realizadas se diseñó una interfaz gráfica que permitiera la detección de anomalías en señales de ECG. Para obtener los resultados se llevó a cabo primero un análisis de sensibilidad donde se muestra la probabilidad que tiene el algoritmo de no detectar anomalías en la señal, y de especificidad que es la probabilidad que tiene el algoritmo de detectar anomalías en la señal. Estas pruebas se hicieron con el fin de comparar el desempeño de la herramienta comparada con la lectura dada por un experto.

* TRABAJO DE GRADO

** FACULTAD DE INGENIERÍAS FISICOMECAÑICAS. ESCUELA DE INGENIERÍA DE SISTEMAS E INFORMÁTICA. DIRECTOR: LOLA XIOMARA BAUTISTA.

SUMMARY

TITLE:

ANOMALY DETECTION ELECTROCARDIOGRAPHIC SIGNALS USING DATA MININGFLOW*

AUTHORS:

LIDIS MAYERLY RONDÓN SUÁREZ

GERMAN EDUARDO CORDOBA CARREÑO**

KEYWORDS: ECG, data flow, anomaly, data mining.

DESCRIPTION

This paper aims at providing a tool to detect abnormalities in electrocardiographic signals (ECG), to serve as expert support and help with cardiovascular health care. This part of research studies that show that the main causes of death in the world are due to cardiovascular diseases. One feature of this paper is to show the procedures carried out for the analysis of ECG signals through mining techniques of data streams (Stream Data Mining in English), to detect any possible abnormality in the signals. For this literature review was to investigate mining techniques serve dataflow to extract relevant information from ECG signals. Then we proceeded to choose the technique of mining data stream efficiently adjust to the solution of the problem, once the technique selected was designed an algorithm which raises the overall process for signal processing and implementation algorithm with the mining technique selected data stream. Based on investigations designed a graphical interface that allows the detection of abnormalities in ECG signals. To obtain the results was carried out first a sensitivity analysis showing the probability of the algorithm does not detect signal anomalies, and specificity is the probability of the algorithm to detect signal anomalies. These tests were conducted to compare the performance of the tool as compared with the reading given by an expert

* WORD DEGREE

** PHYSICOMECHANICAL ENGINEERINGS FACULTY. SYSTEM ENGINEERING SCHOOL.

DIRECTOR: LOLA XIOMARA BAUTISTA

INTRODUCCIÓN

El carácter multifacético de las enfermedades, combinados con una amplia variedad de resultados y las relaciones complejas con otras enfermedades, han hecho del diagnóstico de las enfermedades cardiovasculares una tarea altamente compleja e importante, incluso para los cardiólogos experimentados. No obstante, la gran cantidad de datos que se generan a cada momento [1] y la necesidad de producir alertas tempranas sobre posibles anomalías presentes en personas con problemas cardiovasculares, cuyas señales cardiacas están siendo monitoreadas por medio de sensores [2], bien sea desde su casa o siendo trasladado a un centro asistencial y donde dicha señal se está enviando sin un análisis previo hace que el diagnóstico de una posible anomalía se realice con un tiempo de retardo muy largo.

En este documento se describen conceptos teóricos sobre la minería de datos haciendo énfasis en identificar una técnica que se adapte de manera adecuada a los requerimientos del problema. La minería de flujo de datos en este contexto, sirve para crear un modelo, donde el algoritmo analiza primero los datos proporcionados, en busca de tipos específicos de patrones o tendencias. A su vez este algoritmo usa los resultados de este análisis para definir los parámetros óptimos para la creación del modelo¹.

El tratamiento digital de señales es otro de las características que se presentan en este documento, que consta de unos procedimientos matemáticos que se hacen a la señal con el fin de eliminar ruido en ella y caracterizar sus componentes. De esta manera se obtienen mejores resultados sin perder datos relevantes. El enfoque principal del actual proyecto se basa en el análisis de señales de ECG simulando una entrada de flujo de datos en tiempo real.

El presente documento está distribuido de la siguiente forma: en el capítulo uno se hace la descripción del problema planteado, en el capítulo dos se presenta

¹Analysis services: Minería de datos, <http://msdn.microsoft.com/es-es/library/ms175595.aspx>

la justificación que va ligado con la validez que tiene el tema de investigación y el porqué es factible hacerlo. Posteriormente en el capítulo tres se describe la solución del problema planteado, para el cual se construyó un objetivo general y cinco objetivos específicos, así como la metodología que se siguió para el desarrollo del trabajo. En el capítulo cuatro se explican conceptos teóricos relacionados con la temática que se trata en este trabajo y también se muestra la revisión del estado del arte. Seguido a esto y teniendo en cuenta los tópicos mencionados anteriormente, en el capítulo cinco se hace la descripción del desarrollo de la solución, que es donde se hace el diseño del algoritmo, se describe el proceso de adquisición de la señal, para posteriormente procesarla. En este capítulo también se encuentra la explicación del desarrollo de la interfaz gráfica y las pruebas que se hicieron a la herramienta. Para terminar en el capítulo seis se encuentran las respectivas conclusiones y recomendaciones que se sacaron de la culminación del desarrollo del presente trabajo.

1. DESCRIPCIÓN DEL PROYECTO

1.1 PLANTEAMIENTO DEL PROBLEMA

Estudios han mostrado que una de las principales causas de muerte en Latinoamérica es el infarto; en 2002 en Colombia murieron 31.289 personas debido a problemas cardíacos [3]. La fundación colombiana del corazón, actualmente trabaja con un aval para planes de promoción de la salud cardiovascular [4], debido a que este problema está generando gastos elevados por cada persona con enfermedades cardiovasculares, lo cual genera gran impacto en la economía del país.

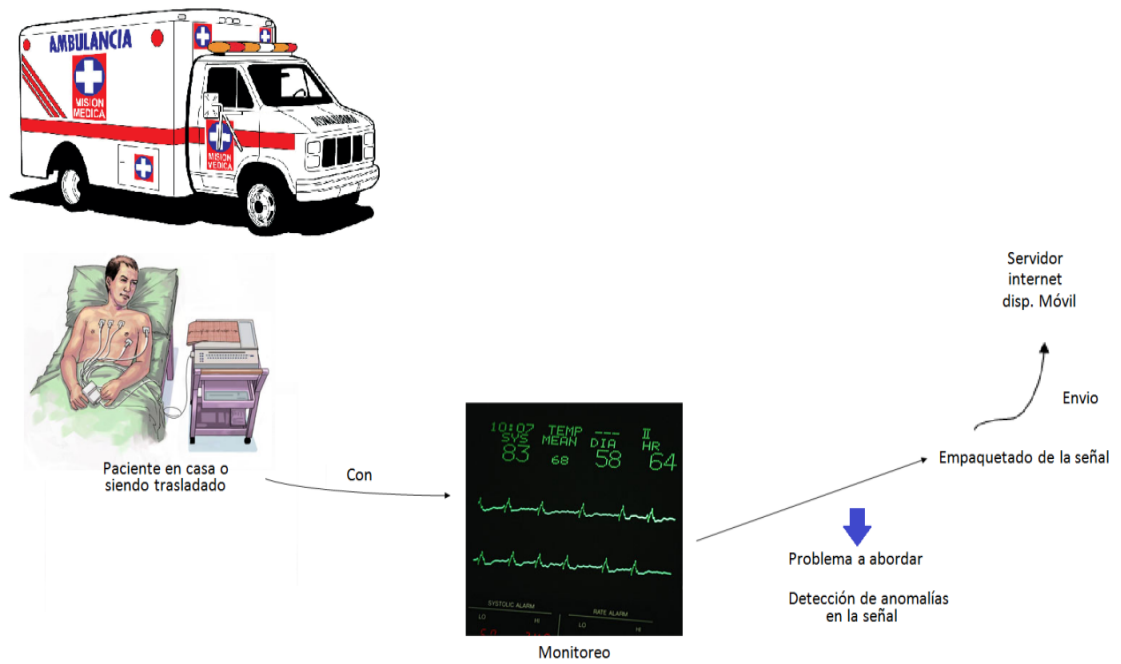
La cantidad de latidos a examinar por parte del equipo médico en un intervalo dado de tiempo puede llegar a ser de varios miles, lo cual convierte la inspección visual de una señal de este tipo en una tarea laboriosa y compleja. Por ejemplo, los monitores Holter, que son dispositivos electrónicos que registran de manera ambulatoria, latidos cardíacos en un período de 24 ó 48 horas, que posteriormente se transmiten al equipo de análisis a una velocidad en un intervalo entre 60 y 120 veces la velocidad de adquisición. Después se procede a la extracción de los eventos significativos muchas veces de forma manual.[5]

Otras situaciones que se presentan con frecuencia son: personas que viven en zonas alejadas y no cuentan con especialista en cardiología, lo cual dificulta el proceso de intervención y diagnóstico del paciente. Existe además situaciones en las cuales la persona está siendo monitoreada desde su casa o está siendo trasladada a un centro asistencial en ambulancia, y en estos casos las señales que se adquieren del paciente no se están analizando previamente además los procedimientos y tiempo que acarrearán estas situaciones son bastante complejos y tardíos.

Por lo mencionado anteriormente nace la idea de construir una herramienta que permita simular el análisis de señales de ECG en tiempo real usando

técnicas de minería de flujo de datos, que permitan monitorizar y extraer la información relevante en las señales de ECG.

Figura 1. Representación gráfica del planteamiento del problema.



Fuente: Autores del proyecto

2. JUSTIFICACIÓN

En el GIB (Grupo de Investigación en Ingeniería Biomédica), se desarrolló un proyecto de investigación de pregrado que implementó el estándar SCP-ECG para el envío de la señal electrocardiográfica a través de Internet haciendo uso de dispositivos móviles [6]. De esto nació la idea de realizar el procedimiento de análisis de dicha señal de manera automática en el momento de adquisición, de manera que al momento de que ésta sea enviada por la red, esté depurada de ruido y haya sido caracterizada de manera correcta para la identificación de posibles anomalías que esté sufriendo el paciente.

Una de las motivaciones de los autores para la realización de este proyecto es el impacto que se visualiza a nivel social en el campo del cuidado médico, ya que se pueden detectar alertas tempranas para la intervención oportuna por parte del especialista a las personas con problemas cardiacos. Además, este trabajo contribuirá con el crecimiento de la línea de investigación de ingeniería biomédica, en particular en el GIB para posteriormente difundir en el ámbito médico e informático y con ellos mostrar la investigación y alcances que está teniendo el grupo de investigación. Para la realización de este trabajo el grupo de investigación GIB cuenta con las herramientas necesarias para la culminación de este proyecto, lo cual representa un ahorro y aprovechamiento de los recursos disponibles.

3. DESCRIPCIÓN DE LA SOLUCIÓN

Teniendo en cuenta la problemática mencionada, además que es factible el desarrollo de este trabajo, y partiendo de que estos problemas descritos en el capítulo uno hacen parte de la vida real, se plantearon los siguientes objetivos para dar solución al problema.

3.1 OBJETIVOS

3.1.1 Objetivo General

Diseñar una herramienta para el análisis de flujo de datos de señales electrocardiográficas enfocado a la detección de anomalías.

3.1.2 Objetivos específicos

- Investigar las técnicas de minería de flujo de datos para detección de anomalías.
- Seleccionar la técnica de minería de flujo de datos más adecuada para la detección de anomalías en señales electrocardiográficas.
- Implementar un algoritmo con la técnica de minería de flujo de datos que permita la detección de anomalías.
- Desarrollar una interfaz para el detector de anomalías en señales electrocardiografías.
- Verificar el desempeño del detector de anomalías comparándolo contra la lectura dada por el médico.

Una vez plantados estos objetivos se inició a dar cumplimiento a cada uno de ellos. Inicialmente se hizo la investigación y se encontró que la minería de datos ataca el problema de clasificación en la detección de anomalías en señales de ECG, apoyado en el procesamiento digital de señales para la depuración de la señal. Una vez se seleccionó la técnica de minería de datos

que se adaptara a este trabajo, se hizo la implementación de los algoritmos para luego hacer la interfaz gráfica.

La implementación de la aplicación para la detección de anomalías en señales electrográficas fue desarrollada con NetBeans IDE 7.1.2 que es un entorno de desarrollo integrado libre hecho principalmente para JAVA. Esto con el fin de facilitar la depuración y ejecución de la aplicación en este lenguaje de programación. También se utilizó MATLAB® para la fase de preprocesamiento y filtrado de la señal.

A continuación se describe la metodología que se utilizó para el desarrollo de este proyecto.

3.2 METODOLOGÍA

Para llevar a cabo el cumplimiento de los objetivos propuesto en este proyecto, se siguió la metodología de Investigación Acción (IA), debido a que se ajusta con el trabajo realizado. Esta metodología tiene dos finalidades: a) genera beneficios al investigador tales como la identificación de las fuerzas sociales y las relaciones que están detrás de la experiencia humana, además, permite la generación de nuevos conocimientos al investigador y a los grupos involucrados haciendo énfasis en el empleo de los recursos disponibles. Las experiencias que resultan en el campo investigativo proporcionan las informaciones acerca de los procesos históricos de investigaciones hechas anteriormente y al mismo tiempo b) genera conocimientos de investigación notables.

En la figura 2 se muestra el proceso de retroalimentación que se presenta en la metodología de investigación acción la cual pretende ligar el enfoque experimental de la ciencia con programas de acción implantados en la comunidad.

Figura 2. Metodología utilizada



Fuente: Autores del proyecto

3.2.1 Investigación-Acción

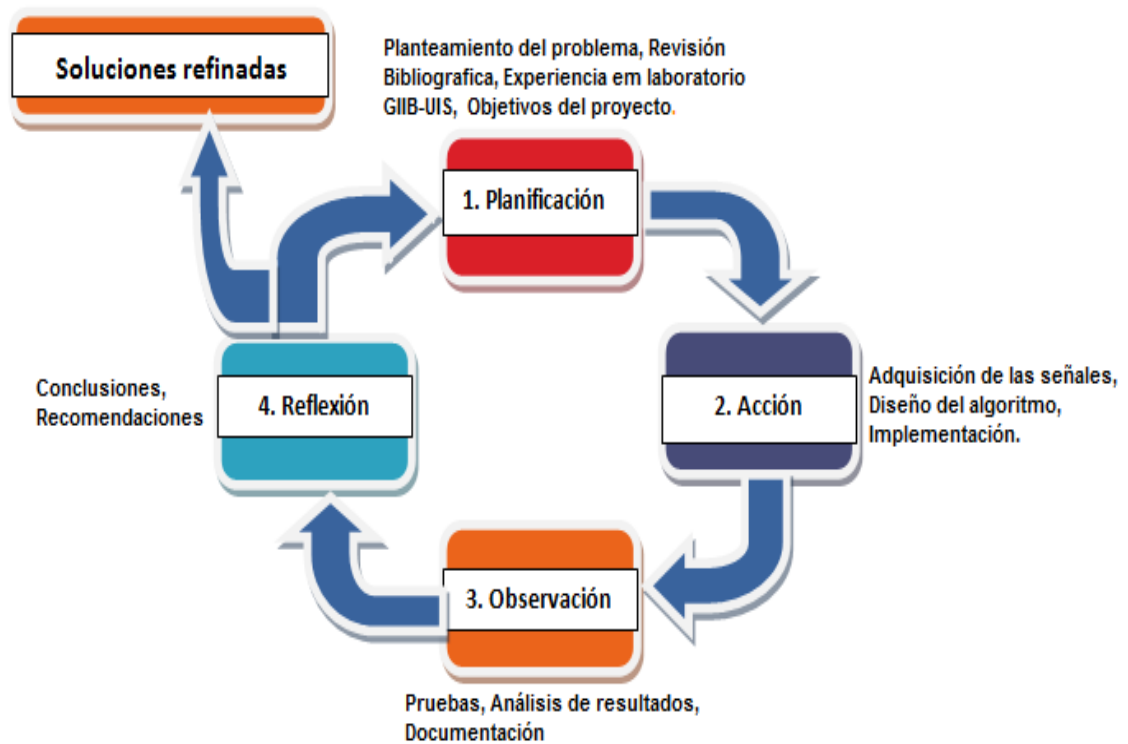
La investigación–acción no se enfoca en un método en concreto, por el contrario abarca varios métodos que tienen en común características tales como: orientación a la acción y el cambio, focalización en un problema, un modelo de proceso “orgánico” que engloba etapas sistemáticas y algunas veces iterativas y colaboración entre los participantes. En la Figura 2 se muestra un modelo que representa el ciclo de actividades propuesto para un caso específico orientado al software, ajustado a las necesidades del presente proyecto.

En el presente proyecto se identificaron cuatro tipos de roles:

- El investigador: los autores del proyecto.
- El objeto investigado: Detección de anomalías en señales de ECG.
- El grupo crítico de referencia: Especialistas en cardiología.
- El beneficiario: Médicos generales, cardiólogos, personal paramédico.

En la figura 3 se muestra gráficamente los pasos y acciones que se llevaron a cabo en el desarrollo del presente proyecto.

Figura 3. Metodología de la Investigación: Investigación-Acción

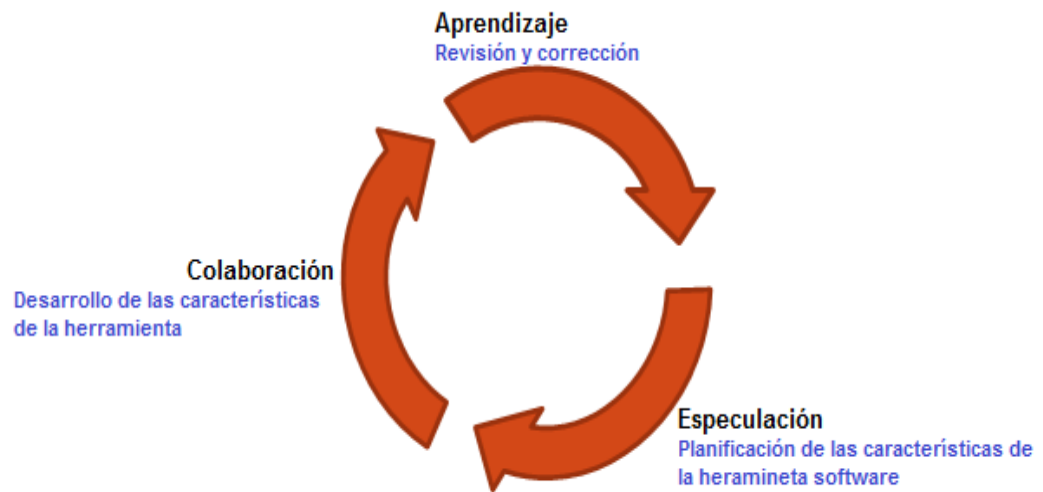


Fuente: Francisco Ruiz, Macario Polo, Mario Piattini [7].

3.2.2 Metodología Adaptativa

Teniendo en cuenta que el proceso desarrollo es variable, se seleccionó la metodología Adaptativa ASD (Adaptive Software Development) [8]. Comprende un ciclo iterativo donde se espera que por cada iteración se produzcan cambios, estos cambios se representan gráficamente en la figura 4. Dichos cambios se representan en este trabajo cuando se hace la evaluación y el ajuste de los parámetros de los algoritmos que se construyeron.

Figura 4. Metodología de Desarrollo: Metodología Adaptativa



Fuente: Jim Hignsmith[9]

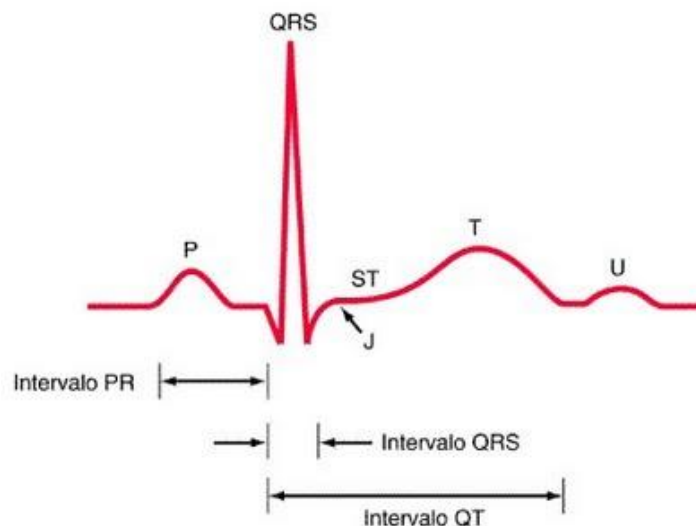
4. MARCO TEORICO Y ESTADO DEL ARTE

4.1 MARCO TEORICO

4.1.1 Electrocardiograma

Es un registro de la actividad eléctrica sobre la piel (pecho) inducida por las ondas de despolarización del musculo del corazón. La frecuencia de las ondas de despolarización de un corazón sano en reposo es del orden de 1 Hz (o ligeramente superior) mientras que la amplitud del pulso eléctrico es grabada del orden de 1 mV. El ciclo normal del corazón tiene una aspecto muy peculiar como se muestra en la Figura 5 .El ciclo comienza por una protuberancia conocida como la onda P, una rápida despolarización característica más prominente del ciclo, conocido como complejo QRS, seguido por otro golpe conocido como onda T. Una estimación fiable de la frecuencia cardíaca es proporcionada por el intervalo RR distancia de dos complejos QRS posteriores [10]. En la Figura 5 se observa que el complejo QRS aparece como el próximo más prominente [11].

Figura 5. Representación esquemática de un ciclo de ECG normal.



Fuente:<http://emergenciasuni.wikispaces.com/ELECTROCARDIOGRAMA+NORMAL>

4.1.1.1 Derivaciones electrocardiográficas

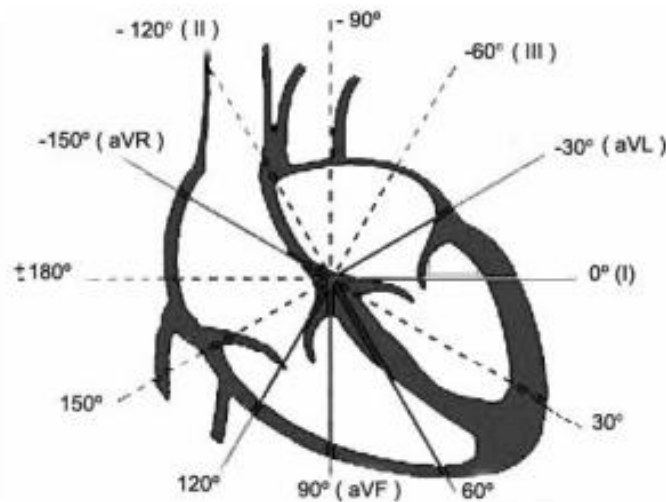
Cuando se toma un ECG, la posición en la que se ubican los electrodos en el cuerpo del paciente se denomina derivaciones; un electrocardiograma completo está compuesto por 12 derivaciones que se clasifican de acuerdo a su posición en el cuerpo, está a su vez están basadas en el triángulo de Einthoven, tres de ellas son bipolares y se conocen con los nombres de D1, D2 y D3; las otras 9 son unipolares y se denominan, por el orden en que se toman, VR, VL y VF, V1, V2, V3, V4, V5 y V6. A continuación se describe cada una de ellas.

1. Derivaciones del plano frontal

Derivaciones bipolares: cada una de ellas corresponde a dos electrodos aproximadamente equidistantes al corazón. Fueron desarrolladas por Willem Einthoven. En la figura 6 se muestra gráficamente estas derivaciones.

- Derivación I (DI): diferencia de potencial entre el brazo derecho (polo negativo) y el izquierdo (polo positivo). El eje de la derivación es 0° .
- Derivación II (DII): diferencia de potencial entre el brazo derecho (polo negativo) y la pierna izquierda (polo positivo). El eje de la derivación es $+60^\circ$.
- Derivación III (DIII): diferencia de potencial entre el brazo izquierdo (polo negativo) y la pierna izquierda (polo positivo). El eje de la derivación es $+120^\circ$ o -60° .

Figura 6. Derivaciones Frontales



Fuente: Carmona R. Juan, Navarro J. Manuel, Carmona R. Luis [12]

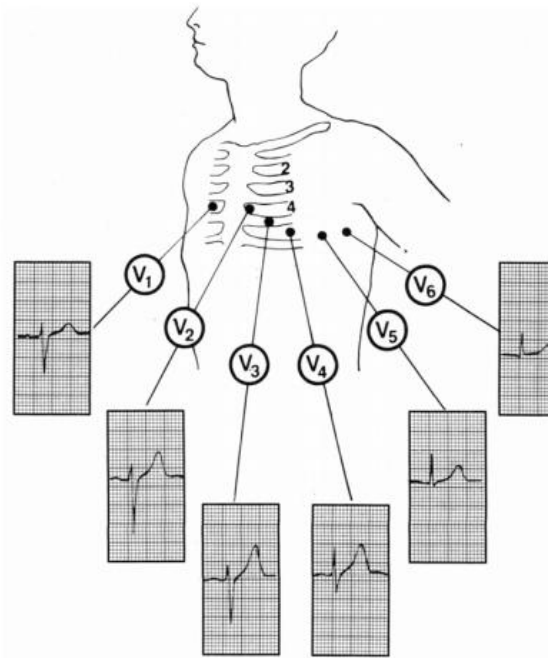
2. Derivaciones del plano horizontal:

Derivaciones precordiales: estas derivaciones son positivas y usan el centro del campo eléctrico del corazón como punto negativo. El electrodo se coloca en la superficie anterior del tórax, encima del corazón, como polo positivo, mientras el polo negativo, se conecta mediante resistencias eléctricas provenientes de los electrodos en las extremidades. Estas derivaciones están representadas en la figura [7] y son:

- V1: cuarto espacio intercostal. Borde derecho del esternón.
- V2: cuarto espacio intercostal. Borde izquierdo del esternón.
- V3: quinto espacio intercostal. Línea media clavicular.
- V4: entre V3 y V5
- V5: quinto espacio intercostal. Línea axilar anterior.
- V6: quinto espacio intercostal. Línea media axilar.
- V7, V8 y V9: son de utilidad ante el cuadro clínico del síndrome isquémico agudo. Estas derivaciones se encuentran al mismo nivel de V4, pero sobre la línea axilar posterior, el omóplato y el borde izquierdo de la columna vertebral respectivamente.

- V3R-V9R: sobre el hemitórax derecho en la misma posición de las derivaciones izquierdas. V2R es la misma V1.

Figura 7. Derivaciones Precordiales



Fuente: Carmona R. Juan, Navarro J. Manuel, Carmona R. Luis [12]

4.1.1.2 Ondas, intervalos y segmentos del electrocardiograma

1. ONDAS

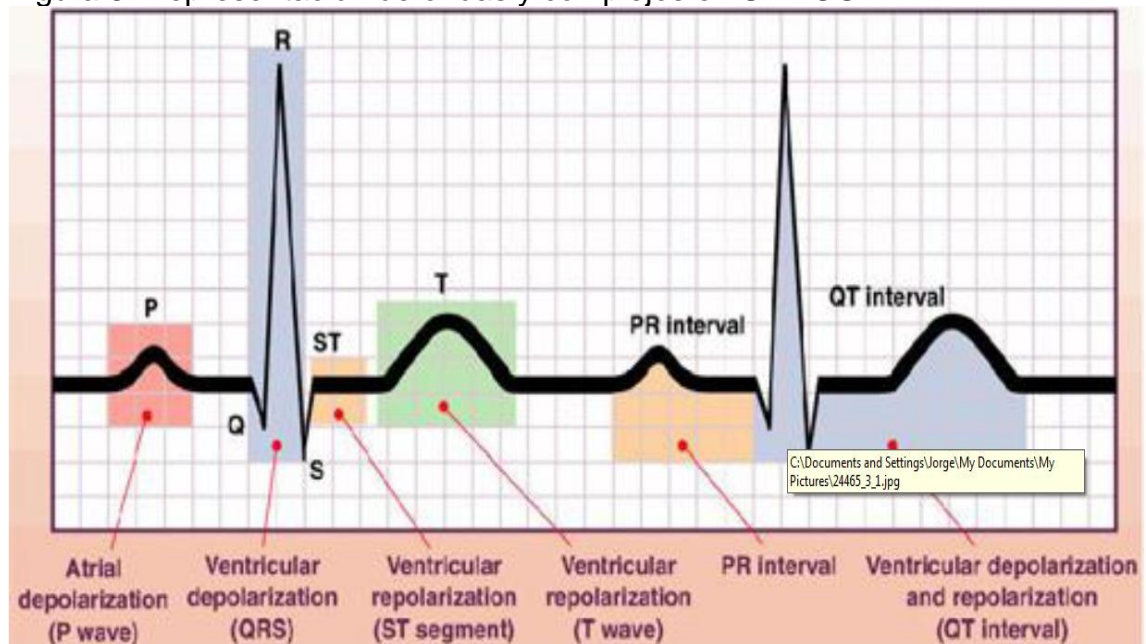
- a) **P:** despolarización auricular, dura entre 0,09 y 0,11 seg. , tiene una altura de hasta 2,5 mV, es de morfología roma o con una pequeña muesca. La primera porción corresponde a la activación de la aurícula derecha y la porción terminal de la aurícula izquierda.
- b) **QRS:** despolarización ventricular, su duración normal es de 0.06 a 0.1 seg. siendo:
 - a. Q: primera onda negativa antes de la primera onda positiva.

- b. R: toda onda positiva. Si existe una segunda onda positiva la llamamos R´.
 - c. .S: toda onda negativa después de una onda positiva.
 - d. QS: complejo totalmente negativo.
- c) **T**: de despolarización ventricular.
- d) **U**: pequeña onda que sigue a la onda T, de significado incierto.

En la figura 8 se hace una descripción grafica de las ondas intervalos y segmentos característicos de un ECG

La re-polarización auricular cae dentro del QRS. La onda T es inicialmente de la misma polaridad que el QRS, siendo habitualmente negativa en aVR y positiva en el resto, aunque puede ser negativa en V1 y en DIII sin que indique patología. En los niños la onda T suele ser negativa de V1 a V4, hasta los 10-15 años en que pasa a positiva.

Figura 8. Representación de ondas y complejos en Un ECG



Fuente: Cruz del Valle W. [13]

2. INTERVALOS:

a) Intervalo PR o PQ: desde el inicio de la onda P al inicio del complejo QRS. Lo forman la onda P y el segmento PR. Su duración normal es menor de 0.2 seg y mayor de 0.12 seg. Corresponde al período que va desde el comienzo de la despolarización auricular, hasta el comienzo de la de la activación ventricular, por lo tanto representa fundamentalmente el retraso fisiológico de la conducción que se lleva a cabo en el nodo AV (segmento PR).

b) Intervalo *QT*: desde el inicio del QRS hasta el final de la onda T. Es proporcional a la frecuencia cardíaca, acortándose al aumentar ésta, ya que al aumentar la frecuencia cardíaca se acelera la repolarización, por lo que se acorta el *QT*. El *QT* ó *QT* corregido en base a la frecuencia cardíaca se calcula por la siguiente fórmula: $QT_c = QT \text{ medio (seg)} / \sqrt{\text{intervalo RR previo}}$.

El QTc normal debe ser < 0.44 , que equivale a que el *QT* sea menor que la mitad del intervalo RR previo cuando la frecuencia cardíaca se encuentra entre 60 y 90.

3. SEGMENTOS

a) Segmento ST:

Comprende desde el fin del complejo QRS (punto J) hasta el inicio de la onda T. Se debe tener en cuenta su relación con la línea de base, ya que con respecto a esta puede estar supra desnivelado, infra desnivelado o ser isoelectrico. Tiene valor patológico si hay desniveles mayores a 1 mm. El punto J, corresponde a la unión entre el fin de la onda S y el inicio del segmento ST. Debido a que no hay mayor cambio de potencial durante esta fase, el segmento ST suele ser isoelectrico en los electrocardiogramas normales.

b) Onda T: Corresponde con la re-polarización ventricular y aparece al final del segmento ST. Su polaridad suele ser positiva en todas las derivaciones

excepto en aVR y ocasionalmente en DIII. La forma de la onda T es redondeada pero asimétrica, por lo general la rama ascendente de la T suele ser de inscripción lenta mientras que la descendente busca la isoeletrica de forma rápida. La amplitud de la onda T es bastante variable de unas a otras derivaciones, aunque nunca debe exceder de 0.6 mV.

c) Intervalo QT: El intervalo QT se mide desde el comienzo del complejo QRS hasta el final de la onda T , se relaciona, dentro de ciertos límites con la duración de la despolarización y de la repolarización. El intervalo QT a veces no indica con precisión el tiempo de recuperación de los ventrículos, su duración habitualmente es de 0.38 a 0.44 s. Sin embargo, varía de acuerdo a la frecuencia cardiaca. Bazett propuso una fórmula que permite calcular el intervalo QT corregido en cuanto a la frecuencia cardíaca: $QT_{c} = \frac{QT}{\sqrt{RR}}$.

4.1.2 Minería de Datos

En esencia la minería de datos se basa en "la extracción no trivial de información implícita, previamente desconocida y potencialmente útil a partir de datos" [14]. Hace uso de técnicas como máquinas de aprendizaje, estadísticas y de visualización para descubrir y presentar los conocimientos en una forma que sea fácilmente. Básicamente, la minería de datos surge para intentar ayudar a comprender el contenido de un repositorio de datos. Con este fin, hace uso de prácticas estadísticas y en algunos casos, de algoritmos de búsqueda próximos a la Inteligencia Artificial y a las redes neuronales. Por lo tanto, la minería de datos abarca los métodos basados en computadora que los patrones de extracción de datos o información aún requiere sólo la intervención humana limitada. La mayoría de estos métodos son relativamente recientes y se basan en el área de inteligencia artificial.

Hay dos objetivos principales en minería de datos: la predicción y la descripción. Predicción se refiere a menudo como minería de datos de

supervisión, mientras que la minería de datos descriptiva incluye los aspectos de supervisión y visualización de datos.

La minería de datos presenta algunas ventajas [15] que son descritas a continuación

Ventajas

- La minería de datos facilita el procesamiento de datos y puede poner de relieve nuevos hechos y contextos.
- Enormes bases de datos pueden ser analizadas mediante la tecnología de la minería de datos. Estas bases de datos pueden ser enormes tanto en largo como en ancho. Por ejemplo, para cada cliente se puede tener cientos de atributos que contienen información detallada y además tener miles de registros de clientes.
- La minería de datos descubre información que no se esperaba obtener. Como muchos modelos diferentes son validados, algunos resultados inesperados tienden a aparecer. En muchos estudios, se ha descubierto que combinaciones particulares de factores entregan efectos inesperados que entregan valor a la compañía.
- Los modelos son confiables. El modelo es probado y comprobado usando técnicas estadísticas antes de ser usado, luego las predicciones que se obtienen por el modelo son válidas y confiables.
- Los modelos se construyen de manera rápida. La minería de datos permite construir y generar modelos en sólo uno minutos u horas. El modelado se torna mucho más fácil puesto que muchos algoritmos son probados y sólo el mejor modelo es entregado al usuario.

4.1.3 Flujo de datos

Un flujo de datos es una secuencia ordenada de puntos que pueden ser leídos sólo una vez o un pequeño número de veces. Formalmente, un flujo de datos

es una secuencia de puntos $X_1, X_2, \dots, X_i, \dots$ que se leen en orden creciente de los índices i . El rendimiento de un algoritmo que opera sobre flujos de datos se mide por el número de pasadas que el algoritmo debe hacer sobre el flujo, limitada en términos de memoria disponible, además de las medidas más convencionales. El modelo de flujo de datos está motivado por la aplicación emergente que implica a grandes conjuntos de datos, por ejemplo, registros telefónicos, grandes conjuntos de páginas web, multimedia, datos y conjuntos de transacciones de la cadena de venta al por menor pueden ser modelados como los flujos de datos [16]. La siguiente imagen muestra un ejemplo de flujo de datos.

Figura 9. Señal de ECG



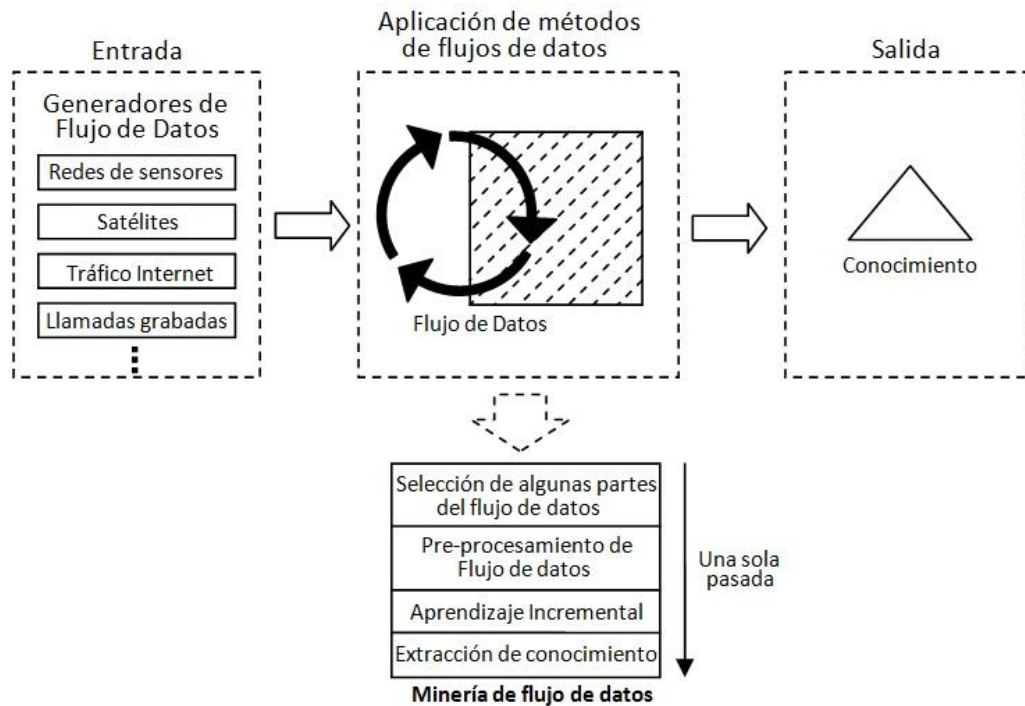
Fuente: Tomado de: <http://www.cardionics.be/Spanish/cardioplug.html>

4.1.4 Proceso general de minería de flujo de datos

Independientemente de los intereses de los usuarios, el proceso de minería de datos puede revelar un gran número de reglas, pero sólo algunas de las

normas descubiertos son relevantes para los usuarios. En la figura 10 se hace una descripción grafica del proceso general de minería de dato.

Figura 10. Proceso general de minería de datos



Fuente: Kholghi, M.; Hassanzadeh, H.; Keyvanpour, M [16]

Aunque en la minería de datos cada caso concreto puede ser radicalmente distinto al anterior, el proceso común a todos ellos se suele componer de cuatro etapas principales:

Determinación de los objetivos. Trata de la delimitación de los objetivos que el cliente desea bajo la orientación del especialista en minería de datos.

Pre-procesamiento de los datos. Se refiere a la selección, la limpieza, el enriquecimiento, la reducción y la transformación de las bases de datos. Esta etapa consume generalmente alrededor del 70% del tiempo total de un proyecto de minería de datos.

Determinación del modelo. Se comienza realizando unos análisis estadísticos de los datos, y después se lleva a cabo una visualización gráfica de los mismos para tener una primera aproximación. Según los objetivos planteados y la tarea que debe llevarse a cabo, pueden utilizarse algoritmos desarrollados en diferentes áreas de la Inteligencia Artificial.

Análisis de los resultados. Verifica si los resultados obtenidos son coherentes y los coteja con los obtenidos por los análisis estadísticos y de visualización gráfica. El cliente determina si son novedosos y si le aportan un nuevo conocimiento que le permita considerar sus decisiones.

4.1.5 Requerimientos de flujo de datos

Requisito 1: Procese una muestra a la vez, y compruebe que esta no aparece solo una vez. Cada muestra debe ser aceptada; una vez inspeccionada o ignorada, una muestra se descarta, sin posibilidad de recuperarla de nuevo.

Requisito 2: Utilizar una cantidad limitada de memoria. La principal motivación para emplear el modelo de flujo de datos es que permite el procesamiento de datos que es muchas veces mayor que la memoria de trabajo disponible.

Requisito 3: Trabajo en una cantidad limitada de tiempo. Un algoritmo debe escalar cómodamente a cualquier número de muestras. La complejidad de tiempo de ejecución debe ser lineal en el número de muestras.

Requisito 4: Esté preparado para predecir en cualquier momento. Un algoritmo ideal debería ser capaz de producir el mejor modelo que puede a partir de los datos que ha observado después de ver cualquier cantidad de ejemplos. En la práctica es probable que haya períodos en los que el modelo se mantiene constante, por ejemplo, cuando un algoritmo basado en lotes es el almacenamiento de hasta el siguiente lote [15].

4.1.6 Técnicas de flujo de datos

Ampliación: Modificar los algoritmos para que acepten grandes cantidades de datos.

Empaquetamiento (wrapper): Gestión de la memoria de un sistema de envoltura sólo puede llevarse a cabo en función de cada modelo, donde la memoria puede ser liberada por olvidar algunos de los modelos que fueron inducidos previamente.

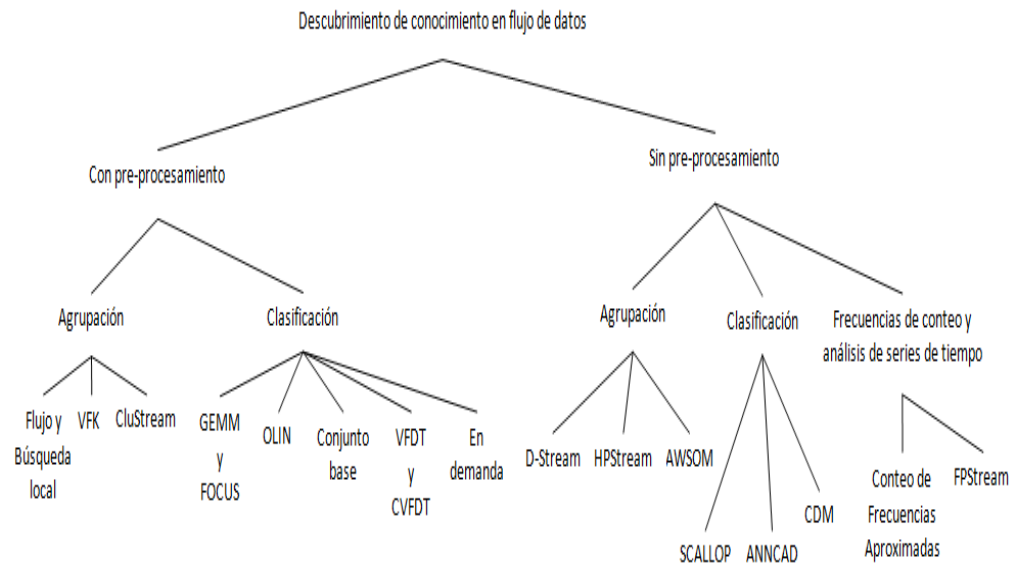
Reglas: un conjunto de reglas puede ser más flexible que la jerarquía de un árbol. Las reglas tienen la ventaja de que cada regla es un componente de la desunión de los modelos que se pueden analizar por separado y se retira del modelo sin grandes trastornos, en comparación con el costo de los árboles de decisión de reestructuración.

Vecinos cercanos: Esta clase de método es descrito como perezoso porque en la configuración de aprendizaje por lotes no se realiza trabajo durante el entrenamiento, pero todo el esfuerzo en los ejemplos de clasificación se retrasa hasta que las predicciones son obligatorias.

Máquinas de soporte vectorial / Redes neuronales: Las redes neuronales son relativamente fáciles de entrenar en un flujo de datos. Una aplicación en el mundo real utilizando redes neuronales está dada por Gama y Rodríguez.

La prueba de CUSUM: La suma acumulada (algoritmo CUSUM), propuso por primera vez, es un algoritmo de detección de cambio que emite una alarma cuando la media de los datos de entrada es significativamente distinta de cero.

Figura 11. Clasificación de algoritmos de flujo de datos



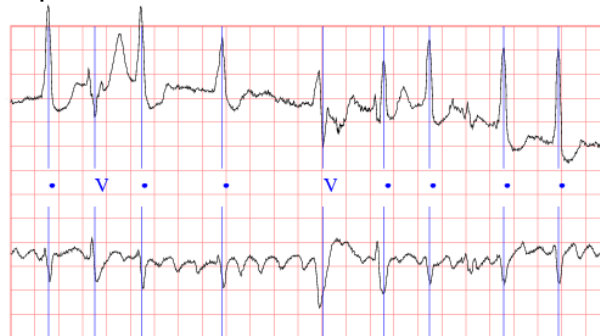
Fuente: hmad Khoureich Ka, Dimitri Petritis [17].

4.1.7 Anomalía

Las anomalías son patrones en los datos que no se ajustan a un concepto bien definido de comportamiento normal. En el ámbito médico una pequeña desviación de lo normal (por ejemplo, fluctuaciones de la temperatura corporal) podría ser una anomalía, mientras que la desviación similar en el ámbito del mercado de valores (por ejemplo, fluctuaciones en el valor de una acción) puede ser considerada como normal [18].

En la Figura 12 se presenta una anomalía debido a que hay una variación en los anchos de los complejos es decir, una anomalía en la formación de señales eléctricas.

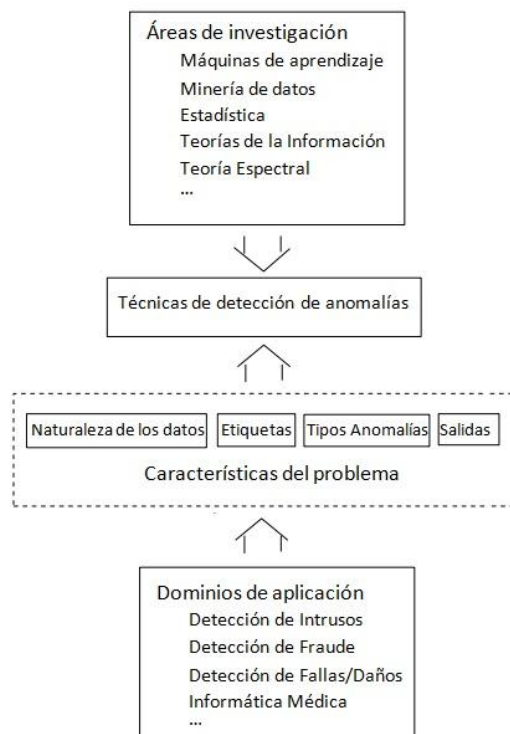
Figura 12. Anomalía presente en una señal de E.C.G



Fuente: Ahmad Khoureich Ka, Dimitri Petritis [19].

En la Figura 13, se describe de manera general las áreas de investigación en las cuales se aborda las técnicas para detección de anomalías teniendo en cuenta las características del problema y el dominio de la aplicación,

Figura 13. Componentes clave asociadas con una técnica de detección de anomalías



Fuente: Ovidiu Apostu, Bogdan Hagiú, Sever Paşca[20].

4.1.7.1 Tipos de anomalías

Anomalías de Punto

Si una instancia individual de un conjunto de datos puede ser considerada como anómala con respecto al resto de los datos, la instancia se denomina como una anomalía de punto. Este es el tipo más simple de anomalía y es el foco de la mayoría de la investigación sobre la detección de anomalías.

Anomalías de Contexto

Si una instancia de datos es anómala en un contexto específico (pero no de otra manera), entonces se denomina como una anomalía contextual. Un ejemplo de anomalía contextual sería una temperatura de 35 °C se presenta en verano es normal, si se presenta en invierno en un solo día se coincidiría como anómalo.

Anomalías Colectivas

Si un conjunto de instancias de datos relacionados es anómalo con respecto al conjunto de datos completo se denomina como una anomalía colectiva. Las instancias de datos individuales en una anomalía colectiva no podrán ser anomalías por sí mismos, pero su aparición juntos en una colección no es normal.

4.1.8 Procesamiento digital de señales

El procesamiento digital de señales se basa en la caracterización, el análisis y la síntesis de las señales y los sistemas que las manipulan juegan papel fundamental en la ingeniería de las tecnologías de la información. El tratamiento de señales por medio de sistemas basados en procesadores

digitales tiene un gran interés debido a su versatilidad y a la capacidad de manejar simultáneamente señales de muy diversos orígenes [21].

A continuación se describe la transformada de wavelet que facilita la extracción de características en señales.

4.1.8.1 La Transformada de wavelet

De manera muy general, la Transformada Wavelet de una función $f(t)$ es la descomposición de $f(t)$ en un conjunto de funciones $\psi_{s,\tau}(t)$, que forman una base y son llamadas las “Wavelets[22].

La Transformada Wavelet se define como:

$$W_f(s, \tau) = \int f(t) \psi_{s,\tau}^*(t) dt.$$

Las Wavelets son generadas a partir de la traslación y cambio de escala de una misma función wavelet $\psi(t)$, llamada la “Wavelet madre”, y se define como:

$$\psi_{s,\tau}(t) = \frac{1}{\sqrt{s}} \psi\left(\frac{t - \tau}{s}\right),$$

Las wavelets $\psi_{s,\tau}(t)$ generadas de la misma función wavelet madre $\psi(t)$ tienen diferente escala s y ubicación τ , pero tienen todas la misma forma. Se utilizan siempre factores de escala $s > 0$. Las Wavelets son dilatadas cuando la escala $s > 1$, y son contraídas cuando $s < 1$. Así, cambiando el valor de s se cubren rangos diferentes de frecuencias. Valores grandes del parámetro s corresponden a frecuencias de menor rango, o una escala grande de $\psi_{s,\tau}(t)$.

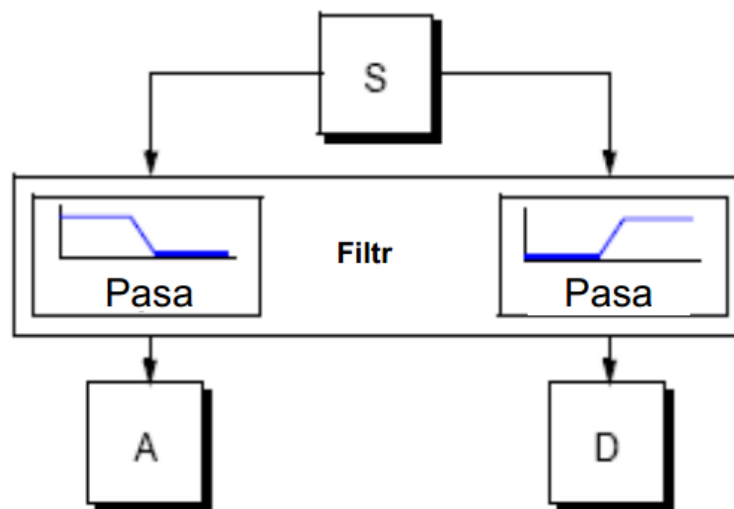
Valores pequeños de s corresponden a frecuencias de menor rango o una escala muy pequeña de $\psi_s, \tau(t)$ [23].

4.1.8.2 Transformada Discreta Wavelet (DWT)

El cálculo de la transformada wavelet para todas las posibles escalas supone una gran cantidad de información. Escoger solo aquellas escalas y posiciones que resulten interesantes para ciertos estudios es una tarea difícil. Si se escogen aquellas escalas y posiciones basadas en potencias de dos, los resultados serán más eficaces. Este análisis se denomina DWT [24].

Es importante resaltar que para muchas señales la información más importante se encuentra en las frecuencias bajas, mientras que en las altas frecuencias se encuentran los detalles o matices de la señal. El análisis wavelet permite descomponer la señal en aproximaciones y detalles, a éste proceso se le conoce con el nombre de análisis. Este filtrado nos proporciona el doble de datos de los que son necesarios, este problema se soluciona con la operación de downsampling.

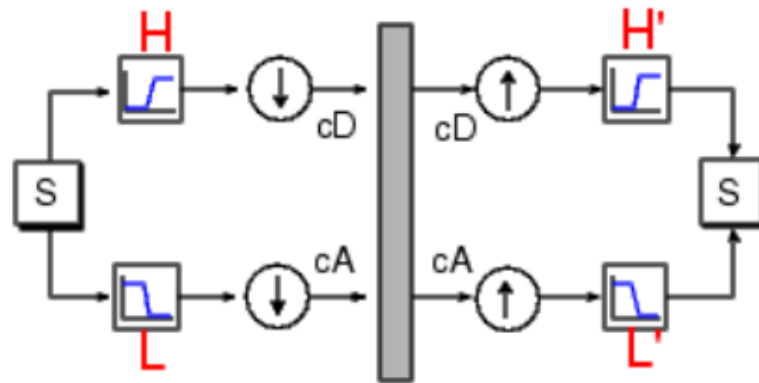
Figura 14. Proceso de descomposición (Análisis)



Fuente: Toolbox Matlab wavelet

El proceso de reconstrucción, también denominado síntesis, se encarga de la obtención de la señal a partir de los detalles y aproximaciones. Éste proceso se lleva a cabo con la transformada wavelet discreta inversa.

Figura 15. Proceso de descomposición y reconstrucción.



Fuente: Toolbox Matlab Wavelet

La elección de las wavelets analizadoras juega un papel muy importante en los resultados finales. Entre las características más importantes a tener en cuenta se encuentran: soporte compacto, simetría, etc. Las wavelets bi-ortogonales son las más eficientes para un posterior proceso de compresión, y en particular, aquellas con pocos coeficientes, ya que el coste de obtención de los coeficientes se incrementa con su número.

El ejemplo más sencillo de transformada wavelet es la transformada de Haar. En este sencillo caso, un análisis (ó convolución) con las funciones base de Haar, consistentes en el cálculo de la media y la diferencia entre cada dos píxeles vecinos, y dichos resultados son submuestreados de manera recursiva.

4.1.8.3 Etapas del Procesamiento Digital de señales

El procesamiento digital de señales, consta de una serie de procedimientos en los cuales se quieren establecer patrones, que hagan que una determinada señal presente transformaciones y a partir de esto, se pueden realizar operaciones para caracterizar, analizar y generar una síntesis de la misma.

Los procedimientos que se siguieron principalmente para el desarrollo de este proyecto se basan en la Adquisición de la señal, Preprocesamiento (filtrado) y Segmentación

Adquisición: Esta etapa comprende la obtención de la señales de ECG por medio de un electrocardiógrafo, que es hardware de adquisición de datos, electrodos que son conductores eléctricos y cables. El electrocardiógrafo, y los electrodos, deben ser capaces de obtener la señal de ECG de una persona, después se procede a guardar ese conjunto de datos en un dispositivo de almacenamiento, para posteriormente analizarlos si aso se quiere.

Preprocesamiento: Corresponde a las operaciones numéricas que realizan mejoras en la señal. En este proceso se encuentran las dimensiones del espacio de reconstrucción que cuantifica apropiadamente las características de la señal. El tamaño del este espacio determina la longitud N de una ventana de tiempo que se desliza sobre toda la serie. Tamaño que corresponde a la dimensión del espacio de reconstrucción.

Segmentación: Comprende la división de la señal en varias partes. El objetivo principal en esta etapa es la simplificación o cambio de la representación de una señal en otras más significativas y más fáciles de analizar. La segmentación se usa tanto para localizar objetos como para encontrar los límites de estos dentro de una señal. El resultado de la segmentación de una señal es un conjunto de segmentos que cubren en conjunto a toda señal para extraer las características más representativas de la misma.

4.2 ESTADO DEL ARTE

El estudio y revisión bibliográfica para el desarrollo de este proyecto de investigación se hizo con el fin de determinar el procedimiento más adecuado para el desarrollo e implementación de los algoritmos que nos permitieran la detección de anomalías con técnicas de minería de flujo de datos de manera más eficiente en las señales de ECG.

A continuación se hace una descripción de las técnicas estudiadas, para posteriormente seleccionar la más adecuada y acorde con nuestro trabajo.

4.2.1 Descripción de técnicas estudiadas

a. Técnicas de agrupamiento Clustering

Definición:

Es una colección de métodos estadísticos que permite agrupar casos sobre los cuales se miden diferentes variables o características. Uno de los problemas del análisis de Clustering es que no existe una definición precisa de esta; lo cual ha originado el desarrollo de una gran cantidad de métodos. Es así que los dos grandes grupos de métodos de Clustering son: los jerárquicos y los no jerárquicos o particionales. En los primeros la pertenencia a un grupo o clúster en un nivel de jerarquía condiciona la pertenencia a grupos de un nivel superior. Los métodos particionales obtienen una única partición de los datos mediante la optimización de alguna función adecuada.

Aplicación:

En el artículo “Un sistema basado en la agrupación para la detección inmediata de anomalías cardíacas a partir de comprimidos de ECG” [25] se encontró que

para llevar a cabo este trabajo se está agregando un módulo de minería de datos (situada en un determinado centro asistencial) para la identificación de las enfermedades cardiovasculares alteración del ECG comprimido, enviado por el paciente, utilizando técnicas de agrupamiento (Clustering). Además la investigación muestra que se utilizó por primera vez un método de selección de atributos que selecciona sólo uno y algunas de sus funciones a partir del ECG comprimido.

Datos y Resultados:

Se tomaron muestras de diferentes segmentos de ECG y se puede observar que cierto grupo de personas tienen diferentes bandas de frecuencia para ECGs normales y anormales. Además muestra la diferencia en los ECG normales y anormales para los 48 atributos seleccionados. Este estudio se mostró por primera vez en Sufiy Khalil (2009) y se amplió el trabajo actual para demostrar la viabilidad de un mecanismo automatizado de alerta basado en técnicas de minería de datos y el ECG comprimido diseñados para salvar las vidas de los pacientes monitorizados.

b. Máquinas de soporte vectorial y Wavelet

Definición:

1. *Máquinas de soporte vectorial*: Son un conjunto de métodos de aprendizaje supervisado para la clasificación y regresión. Viendo los datos de entrada como conjuntos de vectores en un espacio n-dimensional, una SVM construirá un hiper plano de separación en ese espacio, que maximiza el margen entre los conjuntos de datos.

Se hicieron populares por su éxito en reconocimiento de dígitos manuscrito. Una Máquina de Soporte Vectorial (SVM) aprende la superficie de decisión de dos clases distintas de los puntos de entrada. Como un clasificador de una sola clase, la descripción dada por los datos de los vectores de soporte es capaz de

formar una frontera de decisión alrededor del dominio de los datos de aprendizaje con muy poco o ningún conocimiento de los datos fuera de esta frontera.

2. La transformada wavelet: Es un tipo especial de transformada de Fourier que representa una señal en términos de versiones trasladadas y dilatadas de una onda finita (denominada ondula madre).

La teoría de ondículas está relacionada con campos muy variados. Todas las transformaciones de ondículas pueden ser consideradas formas de representación en tiempo-frecuencia y, por tanto, están relacionadas con el análisis armónico.

Aplicación:

Máquinas de soporte vectorial: El sistema inalámbrico expuesto en el artículo titulado "La electrocardiografía ambulatoria inalámbrica a Internet: Nuevos conceptos y estadísticas" [26], basado en internet para ECG es único para el registro y análisis de datos que se encuentran ubicados a grandes distancias y que son tratados a largo plazo. Los acuerdos de colaboración consisten en utilizar la tecnología para permitir que un diagnóstico médico pueda ser observado y seguidamente poder discutir los síntomas con otro médico cuyos pacientes están muy lejos. El análisis mostró algunos resultados prometedores en la predicción de estas dos enfermedades (fibrilación aurícula, arritmias ventriculares), donde los métodos convencionales tienen sólo un valor mal pronosticado.

La transformada wavelet: Para mejorar la precisión de la clasificación de ECG en términos de identificar los diferentes tipos de ritmos cardíacos anormales [27], los autores de este trabajo hicieron una investigación sobre la teoría de la Máquina de Soporte Vectorial (SVM), que ha demostrado ser capaz de reducir al mínimo la probabilidad de hacer una mala clasificación vista por patrones. Según informes estadísticos la plataforma MASN es muy útil para la vigilancia de las aplicaciones de las prácticas médicas, además es una Plataforma de

telemedicina basado en interconexión de diminutos sensores de ECG, llamados médicos ad hoc (redes de sensores MASN).

Datos y Resultados:

Máquinas de soporte vectorial: Los estudios clínicos se realizaron en grupos de diferentes pacientes para predecir la recurrencia de la fibrilación auricular paroxística y la ocurrencia de arritmias ventriculares malignas y / o muerte súbita cardíaca. El análisis mostró algunos resultados prometedores en la predicción de estas dos enfermedades, donde los métodos convencionales tienen sólo pronósticos muy pobres. Además usa los resultados de diversos riesgos como valores de predicción de la telemedicina y de gestión de ECG que se diseñó. Los efectos beneficiosos de ECG-internet-vigilancia en la reducción de la frecuente de imprevistos de re-ingreso en pacientes con CHF pueden persistir a largo plazo y están asociados con la prolongación de la supervivencia, pero estos no pueden ser interpretados a partir de los datos tomados a corto plazo.

La transformada wavelet: La población a tratar fue un grupo de médicos donde se monitoreo la aplicación de las prácticas médicas con pacientes con problemas cardiovasculares. Esta aplicación cuenta con extracción de características de ECG; técnicas de clasificación que se aplican al paciente, los datos y los puntos característicos de intereses extraído. Estos datos proporcionan información significativa para el diagnóstico de posibles enfermedades cardiovasculares. Esto es especialmente útil para extendido grabaciones de ECG que indica cuando el tratamiento humano no sólo consume tiempo, sino también es propenso a errores. Además de estas funciones, el sistema está diseñado para proporcionar medidas de seguridad también contra ataques maliciosos y el robo de información de pacientes.

c. Árboles de decisión

Definición:

Un árbol de decisión es una forma gráfica y analítica de representar todos los eventos (sucesos) que pueden surgir a partir de una decisión asumida en cierto momento. Nos ayudan a tomar la decisión “más acertada”, desde un punto de vista probabilístico, ante un abanico de posibles decisiones. Permite desplegar visualmente un problema y organizar el trabajo de cálculos que deben realizarse.

Aplicación:

El artículo que se menciona en esta sección habla de un sistema de Monitorización de pacientes crónicos con aplicaciones en el cuidado de los pacientes con enfermedades cardiovasculares [28]. El funcionamiento de este sistema, podría ayudar a los diagnósticos generales y enfermedades crónicas. Basado en la arquitectura diseñada, lleva a cabo un sistema de análisis y predicción de enfermedades crónicas a través de ECG. Este sistema tiene en cuenta los signos vitales de un ECG, el sistema puede predecir con un árbol de clasificación e informar a los médicos a tomar acciones en caso de cualquier anomalía que pudiera suceder. Una serie de experimentos sobre los datos mostraron que el PAF es un sistema estable que puede predecir la anomalía de pacientes; ECG datos sin codificación de las normas médicas como se hace en otros enfoques existentes.

Datos y resultados:

Esta estructura ha tenido en cuenta factores médicos con diversas bio-señales. Por lo tanto, el funcionamiento de este sistema no se adaptó a un tipo de señal solamente. Los integrados de múltiples clasificadores podría mejorar la flexibilidad de los requisitos de datos. En base experimentos realizados a los datos reales, se observó que el sistema podría tener una cierta precisión y recordar a la enfermedad. Para publicación de datos de prueba PAF, el sistema podría mantener estable sensibilidad y el resultado no estaba lejos de la logro

de la aprobación médica. Esto muestra que el sistema tenía resultados estables en el análisis de bio-señal y predicción de enfermedades. Esta investigación ha concluido un diseño del sistema y requisito para los pacientes crónicos con bio-señales y técnicas de minería de datos y valida la posibilidad de datos reales.

d. DWT (Distorsión dinámica temporal)

Definición:

La técnica de la distorsión dinámica temporal (DTW) se utiliza para calcular la mejor distorsión de alineamiento posible ϕ , entre T y R , junto con la distorsión asociada $D(T, R)$ un alineamiento óptimo entre las secuencias de longitud variable $T = \{t_1, \dots, t_N\}$ y $R = \{r_1, \dots, r_M\}$. La distorsión total $D(T, R)$ se basa en una suma de distancias locales entre elementos $d(t_i, r_j)$. Una distorsión de alineamiento especial ϕ , alinea T y R mediante un mapeo de punto a punto $\phi = (\phi_t, \phi_r)$, de longitud $K\phi$.

Aplicación: El algoritmo 1 utilizado en este trabajo de investigación titulado "Los métodos de longitud variable para la detección de patrones de anomalías en series de tiempo" [29] está basado en una regresión cuadrática a la serie segmento de tiempo, y obtiene el rango de los patrones de longitud. Algoritmo 2 utiliza DTW (Distorsión dinámica temporal) y los métodos de variables para calcular la similitud de los patrones de forma dinámica, detecta patrones de anomalías en una serie de tiempo dada de forma automática.

Datos y Resultados: Este trabajo utiliza k-distancia de un patrón y mediana para definir factor de anomalía en ECGs, el grado de anomalía, presenta la definición de patrón de anomalías basado en los algoritmos planteados. Los primeros resultados experimentales demuestran el prometedor rendimiento de ellos. La regresión cuadrática el umbral de error determina la longitud de los patrones en una serie temporal, y el umbral de similitud de adyacentes

“patrones”, no autónomo es decir el umbral del partido, que influye en la velocidad de algoritmo 2 en gran medida.

e. Lógica Fuzzy

Definición: La lógica fuzzy, se basa en lo relativo de lo observado como posición diferencial. Este tipo de lógica toma dos valores aleatorios, pero contextualizados y referidos entre sí. Así, por ejemplo, una persona que mida 2 metros es claramente una persona alta, si previamente se ha tomado el valor de persona baja y se ha establecido en 1 metro. Ambos valores están contextualizados a personas y referidos a una medida métrica lineal.

Aplicación: En el presente trabajo se propone mostrar una metodología para la creación automática de sistemas expertos difusos, aplicado en la clasificación de isquemias y +arrítmica " [30]. La metodología propuesta crea automáticamente un sistema experto difuso a partir de un conjunto de datos de entrenamiento inicial. El método consta de tres etapas: (a) extracción de un conjunto clásico de reglas a partir de un árbol de decisión inducida a partir de la formación de datos, (b) transformación del conjunto clásico de reglas en un modelo difuso, y (c) la optimización del modelo difuso de parámetros usando optimización global.

Datos y Resultados:

La metodología propuesta se ha evaluado en la detección de los latidos cardíacos isquémicos en ECG grabaciones con datos de la base de datos ESC ST-T. Además, se ha evaluado en clasificación del latido arrítmico, utilizando los datos de la base de datos MIT-B IH arritmia. En ambos casos clasificaron alto resultados se obtuvo, la precisión (Acc) es del 92% y el 96% para la FES isquémicos y arrítmico, respectivamente.

f. ACO (Optimización por colonia de hormigas)

Definición:

Optimización de colonia de hormigas (antcolonyoptimization o ACO) está inspirado en el rastro y seguimiento de feromonas realizado por las hormigas como medio de comunicación. La optimización de colonias de hormigas (ACO) es un método basado en una población para resolver problemas de optimización combinatoria inspirado en el comportamiento de las hormigas (Dorigo & Socha 2006). Los algoritmos de ACO se consideran como parte de inteligencia de enjambres (swarm intelligence), que es el campo de investigación que estudia algoritmos inspirados por la observación del comportamiento de enjambres. Estos algoritmos utilizan individuos que cooperan a través de auto-organización (sin ningún control central que actúe sobre los miembros del enjambre).

Aplicación:

ACO (Ant Colony Optimization) algoritmo[31]; Se basa en técnicas de clustering basados en ACO y algunas mejoras se han aplicado con el fin de alcanzar grupo final establece más rápidamente. Este algoritmo de kNN (K-Nearest Neighbor) se integra con ACO como un mecanismo de búsqueda local. En este trabajo, la sensibilidad es considerada como el factor de éxito más crítico.

Datos y Resultados:

Las señales ECG son tomadas de MIT-BIH base de datos de ECGs, que se utilizan para clasificar 6 arritmias diferentes. La frecuencia de muestreo es de $f = 360$ Hz. Estos son el ritmo sinusal normal (N), contracción ventricular prematura (PVC), contracción auricular prematura (APC), bloqueo de rama derecha (BRD), la fusión del ventrículo izquierdo (F) y fusión (f). Estos registros han sido tomados de 32 pacientes diferentes. La distribución de los registros utilizados para probar la corrección del algoritmo se muestra en la Tabla 1 [31] En primer lugar conjuntos de entrenamiento diferentes se ha construido a partir

de estos conjuntos de registros mediante un muestreo aleatorio. Después del entrenamiento se pasa a establecer el resto de muestras que se etiquetan como muestras de prueba para cada clase de arritmia.

4.2.2 Métodos de agrupamiento (Clustering)

4.2.2.1 Método de Particional

Este método parte de una base de datos que tiene una longitud de n objetos, donde de una partición se pueden obtener k particiones de dichos datos, y cada partición representa un cluster además k puede ser menor o igual a n . Dicho esto es preciso que cada partición o grupo contenga por lo menos un objeto y cada objeto debe pertenecer a exactamente un grupo [32].

Dado k , el número de particiones a construir, se crea una partición inicial. A continuación, se utiliza una técnica iterativa para la reubicación de los intentos de mejorar la compartimentación por objetos en movimiento a partir de un grupo a otro. Para saber si una partición es buena se deben clasificar los objetos de la misma categoría, en relación el uno del otro. Hay varios tipos de criterios para juzgar la calidad de las particiones. Para lograr la optimalidad global en el reparto basada en la agrupación requeriría la enumeración exhaustiva de todas las posibles particiones. En cambio, la mayoría de aplicaciones adoptan uno de los métodos heurísticos poco populares, tales como (1) el algoritmo k -means, donde cada grupo está representado por el valor medio de los objetos en el conglomerado, y (2) el algoritmo de k -medoides, donde se representa cada grupo por uno de los objetos situado cerca del centro de la agrupación.

Estos métodos de agrupamiento heurísticas funcionan bien para encontrar esférico en forma de cluster en bases de datos de pequeño y mediano tamaño.

Para encontrar grupos con formas complejas y para agrupar conjuntos de datos muy grandes, la partición basada en métodos deben ser extendidos.

La descripción del método jerárquico se hará a continuación.

4.2.2.2 Método Jerárquico

Un método de agrupación jerárquica funciona mediante la agrupación de objetos de datos en un grupo de árboles. Los métodos de agrupación jerárquica pueden ser clasificados ya sea como aglomeración o de división, dependiendo de si la descomposición jerárquica se forma de abajo hacia arriba (fusión) o de arriba hacia abajo (división). La calidad de una agrupación del método jerárquico puro sufre de su incapacidad para realizar el ajuste una vez que una decisión de fusión o división ha sido ejecutado. Es decir, si toma una decisión particular (división) y más tarde resulta haber sido una mala elección, el método no puede dar marcha atrás y corregirlo [32].

Teniendo en cuenta, el planteamiento del problema, los objetivos planteados y el marco teórico, se dio paso a al desarrollo del de la solución del problema plantado.

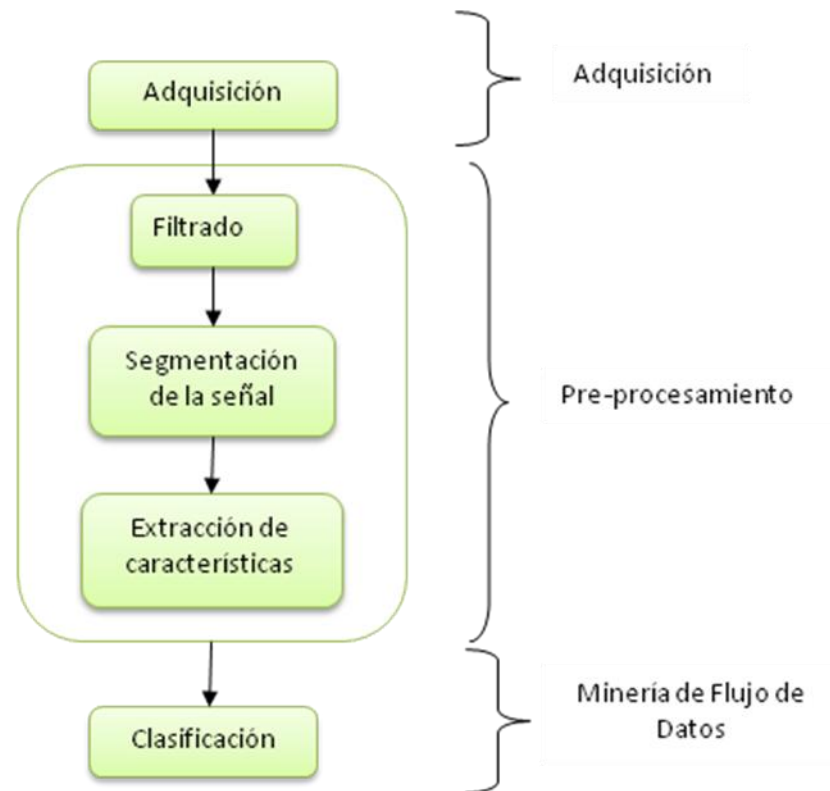
5. DESARROLLO DE LA SOLUCIÓN

Para dar inicio a la solución del problema planteado en el presente trabajo, se hizo énfasis especialmente en las etapas de adquisición preprocesamiento segmentación e implementación de los algoritmos plantados para el análisis de señales de ECG. Dado que el grupo de investigación de ingeniería biomédica (GIIB), cuenta con los equipos necesarios para la toma de señales de ECG, se optó por hacer nuestra propia base de datos, que cuenta con alrededor de 40 señales; siendo los estudiantes de la materia de ingeniería biomédica y persona del Hospital Universitario de Santander las personas a quienes se les tomo el ECG. A continuación se describen los procedimientos que se llevaron a cabo para la adquisición, preprocesamiento y segmentación de las señales de ECG.

5.1 DISEÑO DEL ALGORITMO

Teniendo en cuenta las técnicas resaltadas en el estudio del estado del arte, se planteó un algoritmo que tiene como objetivo realizar la detección de anomalías en señales de ECG , este algoritmo cuenta con una etapa de clasificación que es donde se aplica la técnica de minería de datos. En la Figura 16 se presenta a modelo general el proceso seguido.

Figura 16. Algoritmo de Etapas para el procesamiento digital de señales



Fuente: Autores del Proyecto

Para la parte de clasificación que se muestra en el algoritmo de la Figura 16, se detalla el proceso de minería de datos; iniciamos haciendo la descripción de del proceso de minería de datos para implementar esta parte del algoritmo, este proceso compone de:

- Toma de datos: Comprende todo lo referente a la toma de señales de ECG.
- Limpieza de los datos: Esta parte comprende la eliminación de ruidos y de datos incoherentes presentes en la señal de ECG.

- Selección de datos: La selección de datos consiste en coger los datos ya filtrados para posteriormente realizar las transformaciones necesarias.
- Transformación de datos: En esta etapa se consolidan los datos o se transforman, es decir se realizan las operaciones necesarias para la transformación de la señal, aquí es donde se hace la caracterización de la señal lo cual se realizó con la ayuda de la Transformada Discreta de Wavelet. Después se realiza la clasificación con las técnicas de minería de datos.
- La minería de datos: Para llevar a cabo este proceso esencial en el desarrollo de este proyecto; se seleccionó un método inteligente con el fin de extraer patrones de datos que permitan la detección de anomalías en la señal de ECG.

A continuación se hace la descripción de la técnica de minería de datos seleccionada; esto se hizo teniendo en cuenta las técnicas estudiadas y las características de la señal de ECG. Clustering es la técnica seleccionada porque se enfoca en la clasificación o agrupamiento de objetos similares (clusters), que a su vez enfatiza en descubrir un nuevo conjunto de categorías o grupos de interés en sí mismo y hacer una evaluación intrínseca [33]. En esta oportunidad utilizamos la clasificación como método de aprendizaje supervisado, y el agrupamiento para la clasificación en señales de ECG.

Clustering se compone de dos métodos el particional y el jerárquico; como el método particional es el que más se adapta a las condiciones del actual trabajo, a continuación se hace una descripción de la implementación de esta técnica teniendo como base el pseudocódigo de K-medias, que hace parte de los algoritmos de clustering.

Figura 17. Pseudocódigo del algoritmo de k-medias

```
Entrada

    K: número de clúster
    D: Conjunto de datos que contiene n objetos

Salida

    Conjunto de K grupos (cluster)

Método

1. elegir arbitrariamente k objetos de D como los centros
de conglomerados iniciales;

2. Repetir

3. (re) asignar a cada objeto a la agrupación a la que el
objeto es el más similar, basado en el valor medio de los
objetos en el conglomerado;

4. actualizar los medios de clúster, es decir, calcular el
valor medio de los objetos para cada cluster;

5. Hasta que no hayan cambios
```

Fuente: Molina J.M, García J[34]

Los procedimientos que se llevaron a cabo para el desarrollo e implantación del algoritmo son los siguientes:

1. Ventana de datos de 1000 (o de la longitud que decida el usuario); donde entran de 60 a 100 ventanas por minuto. Esto nos da un promedio de 60000 datos por minuto. (Con mil datos se dibuja un pulso)
2. Filtrar los datos en paquetes o ventanas de a 1000 datos. (db4 Nivel 4).
3. Extraer características de los 1000 datos y sacar 5 datos (amplitudes); esto se hace con los siguientes pasos:

Con Haar nivel 1 hallamos el punto C,

Con el punto C hallamos el punto R

Con el punto R hallamos Q y S,
Con estos puntos hallamos P y T
Si no aparece el complejo QRS es tomado la señal como ruido
Si el punto P no aparece es buscando en la ventana anterior
Esto también se compara con las duraciones normales de los
segmentos P QRS ST T.

Con Clustering

Tomamos al azar K clusterr iniciales, en este caso $K=2$ a los cuales se les da el nombre de:

- Normal
 - Requiere revisión
1. Con los 5 Datos de altura P Q R S T, si cada uno existe se marca como “Esta” o “Ruido” del pulso.
 2. Si “Esta”, envía los 9 valores, 5 alturas P Q R S T y 4 duraciones P QRS ST y T al método de clustering. (R9)
 3. Si el primer pulso está en los parámetros normales se toma este como centroide del Cluster “Normal”, y se toma como centroide de “Revisar” un punto alejado de este. (Cada Cluster en ese momento ya tiene un conjunto de datos perteneciente a cada Cluster).
 4. Se calcula la distancia Euclidiana del siguiente pulso a cada uno de estos centroide, al que este más cercano se agrega a este clustering.
 5. Si llega al cluster “Revisar” se compara cada parámetro contra los valores estándar de altura y duración y se decide agregarlo al cluster “Normal” o definitivamente al cluster “Revisar”
 6. Esta agregación implica que se agrega el punto y se re-cálculo el centroide (Se promedian todos los puntos que hasta el momento tiene el clustering, con el fin de establecer el nuevo centroide.

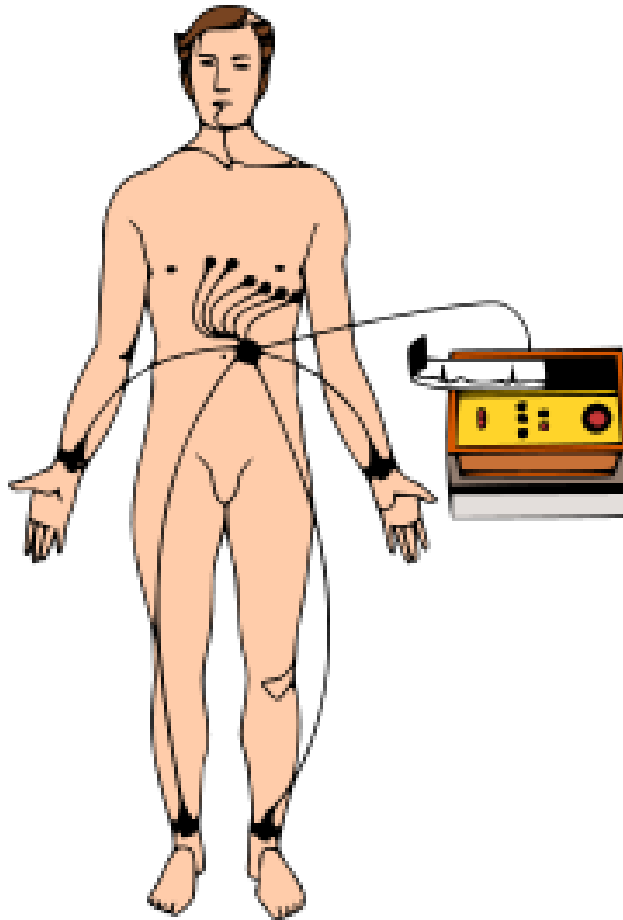
7. Se toma el siguiente pulso y se sigue el paso 4.

Una de las razones por las cuales se trabajó con este método es que no presenta desventajas similares a las del método jerárquico, el cual presenta una incapacidad para escalar bien y además la complejidad de tiempo de algoritmos jerárquicos se encuentra en $O(m^2)$ (Donde m es el número total de casos), que es no lineal con el número de objetos. Por otra parte los métodos jerárquicos no pueden deshacer lo que se hizo anteriormente. Es decir, no tiene capacidad de back-tracking. Además el algoritmo que se implementó del método particional, es decir, k-means es un algoritmo sencillo y uno de los más usados para atacar problemas de esta magnitud.

5.2 ADQUISICIÓN DE LA SEÑAL

En esta etapa se obtuvieron las señales de ECG con ayuda del BIOPAC mp30B-CE, que es un hardware de adquisición de datos, en este caso señales de ECG. Para adquirir la señal, los electrodos se colocan en la piel del paciente, en localizaciones predeterminadas de manera universal, de modo que nos permite obtener registros comparables entre sí. Con los cables correctamente colocados podemos obtener 12 derivaciones (ver Figura.18). Una vez conectado todo de forma adecuada la señal se transfiere por USB a una PC cuyo software visualiza la señal, y la guarda en un dispositivo de almacenamiento. En este trabajo solo se adquirió la segunda derivada. Es interesante observar que cuando se está haciendo la toma de ECG, se pueden generar interferencias por otras actividades que realiza el cuerpo simultáneamente.

Figura 18. Imagen que muestra un paciente conectado a los 10 electrodos necesarios para un ECG de 12 derivaciones



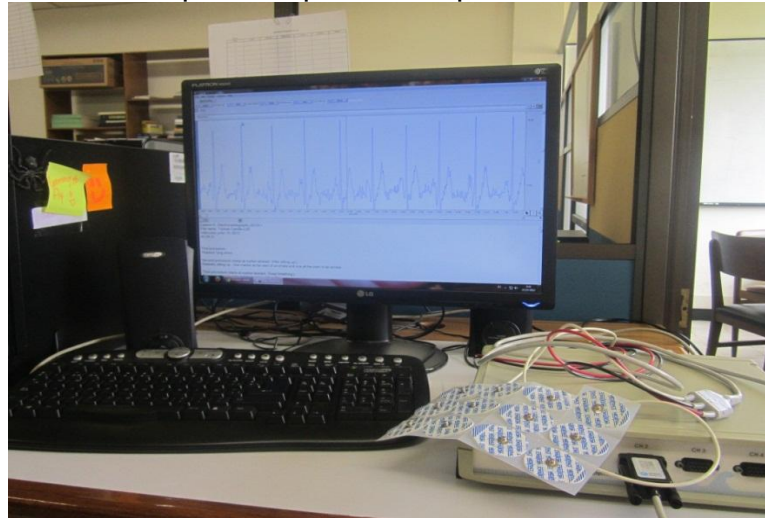
Fuente:<http://commons.wikimedia.org/w/index.php?title=File:ECGcolor.svg&page=1>

5.1.1 Montaje de Adquisición, Preprocesamiento y Segmentación

Montaje de Adquisición: El montaje de adquisición se llevó a cabo dentro del GIIB, mediante los dispositivos con los que cuenta el grupo, ver Figura 19. Se utilizó un BIOPAC mp30B-CE, electrodos, Cables troncales y de paciente para electrocardiografía y un computador con las siguientes características: Procesador Intel(R) Core™2 Quad CPU Q8200 @2.333Hz 2.33Hz con sistema operativo de 32 Bits. Los principales problemas que se presentaron al momento de la adquisición de señal se dieron debido a los movimientos que realizaba el

paciente, incluyendo la forma como respiraba, aunque estos problemas son comunes, se realizaron varias tomas durante el desarrollo del proyecto con el fin de encontrar señales con las cuales se pudieran detectar intervalos bien definidos. Para la toma de las demás señales siempre se usó el mismo montaje.

Figura 19. Elementos empleados para la adquisición de la señal de ECG



Fuente: Autores del Proyecto

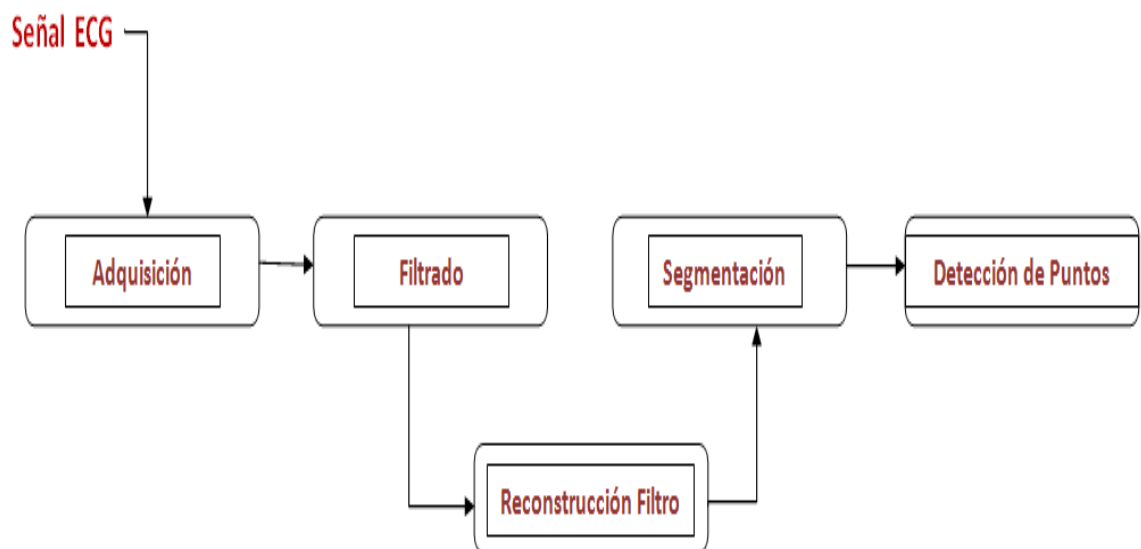
Figura 20. Montaje para la adquisición de la señal de ECG



Fuente: Autores del Proyecto

Preprocesamiento: En esta etapa se simula la entrada de la señal de ECG en tiempo real. Partiendo de esto, se procede a realizar los diferentes análisis numéricos para esclarecer y concretar los datos obtenidos de dicha señal. El ruido presente en las señales electrocardiográficas se debe principalmente al ruido electromiográfico, la respiración, artefactos del movimiento, ruido electroquirúrgico, variaciones de la línea base e interferencias de la línea de potencia. Se han usado filtros lineales para reducir el efecto de estas interferencias. En la Figura 21 se muestra el algoritmo que se construyó para este proceso.

Figura 21. Secuencia de operaciones algoritmo de Caracterización.



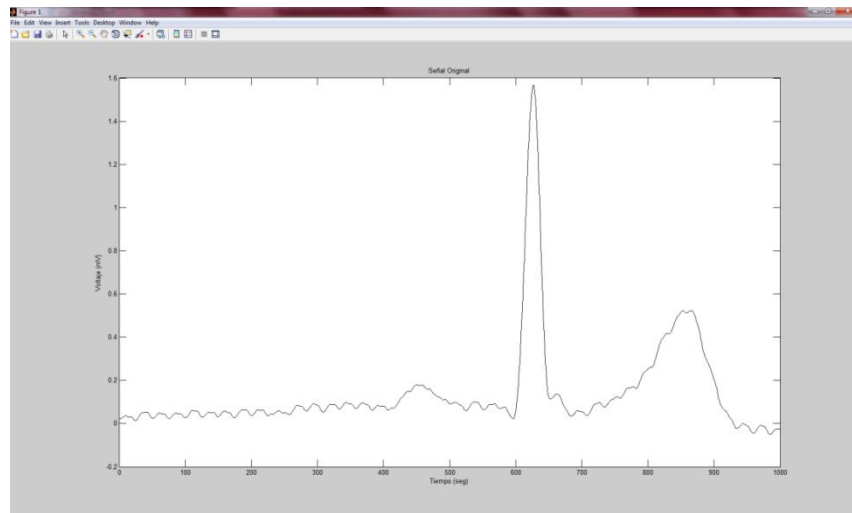
Fuente: Data Mining: Concepts and Techniques, 2nd edition (with Micheline Kamber), The Morgan Kaufmann Series in Data Management Systems

La implementación de este algoritmo se realizó con ayuda de MATLAB®, a continuación se muestran imágenes de la forma como se hizo la caracterización de la señal.

Inicialmente se tomaron 1000 datos porque con esta cantidad se dibuja un complejo completo de una señal de ECG. Además la frecuencia de entrada de datos del electrocardiógrafo utilizado (BIOPAC) es de 1000 datos por segundo.

Partiendo de la señal original obtenida de una de las personas a que se les realizó el ECG, ver Figura 22.

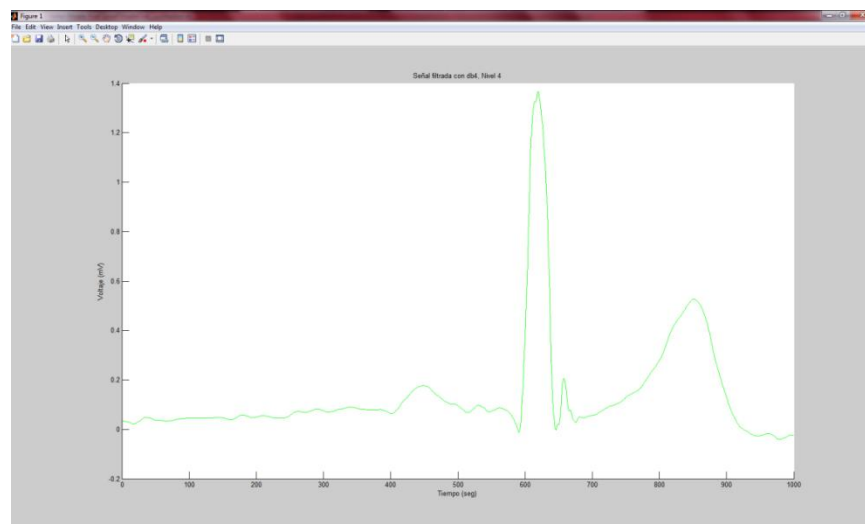
Figura 22. Señal Original



Fuente: Autores del Proyecto

Se inicio el proceso de filtrado con db4, ver Figura. 23, donde se puede observar una señal mucho más definida, con la que posteriormente se realizar la caracterización de la señal.

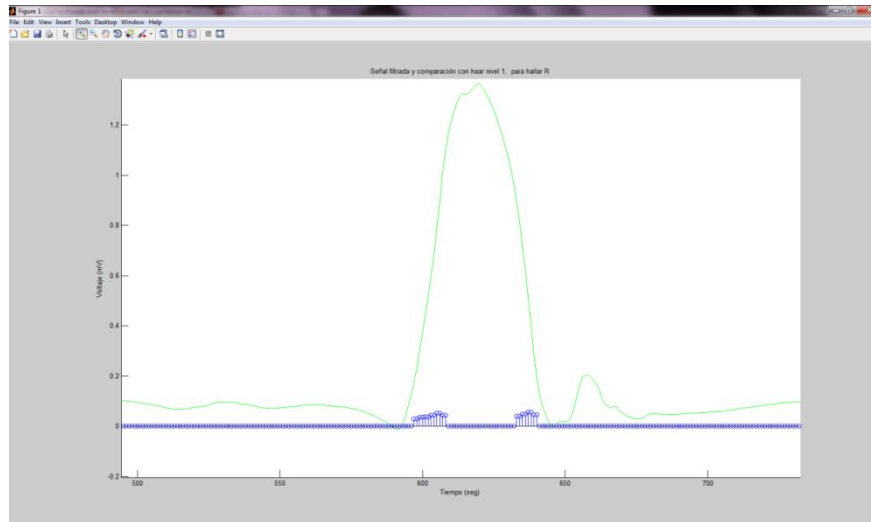
Figura 23. Señal Filtrada



Fuente: Autores del Proyecto

El paso que sigue se realizó usando la Haar para de este modo halla el punto R; esto procedimiento se hizo comparando la señal filtrada con Haar.

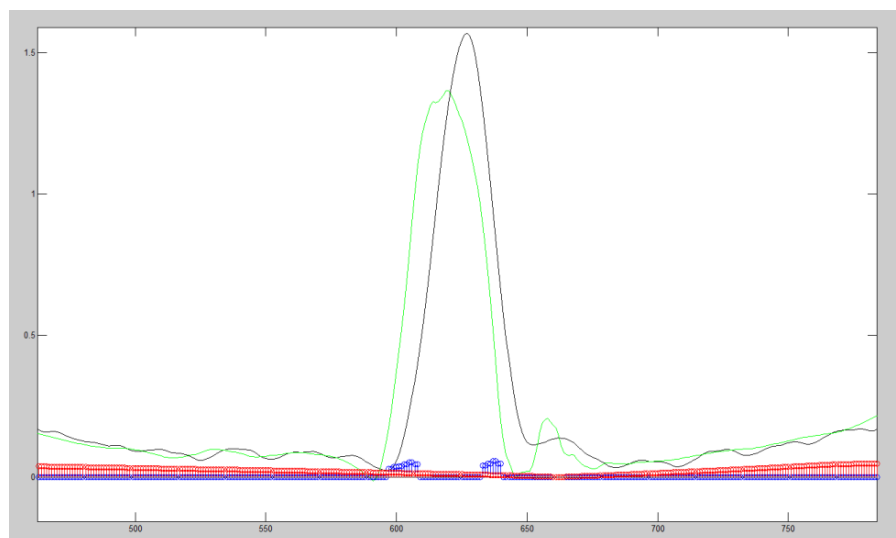
Figura 24. Señal filtrada, comparada con Haar para hallar R



Fuente: Autores del Proyecto

En la Figura 25 se ve que con el uso de las wavelets madre Daubechies y Haar se puede detectar de manera eficiente el complejo QRS.

Figura 25. Detección del complejo QRS con Daubechies y Haar



Fuente: Autores del proyecto

Es importante resaltar que en la detección y caracterización de la señal de ECG se tomó como herramienta base la transformada discreta de wavelet (WTD) y se implementó en MATLAB®, como se mencionado anteriormente.

A continuación se describe el proceso de segmentación también mencionado en el algoritmo de caracterización de la señal.

Segmentación: Esta etapa comprende la división de la señal de ECG por intervalos. Los electrocardiogramas se dividen tradicionalmente en cuatro grandes eventos eléctricos, cada uno reflejando la actividad eléctrica asociada con una fase particular del ciclo cardíaco. Estos cuatro eventos se llaman la onda P, el complejo QRS, el segmento ST y la onda T. Una de las herramientas usadas en esta etapa fue la transformada de Wavelet, que se usó para filtrar y caracterizar la señal por segmentos, debido a que la wavelet está especialmente indicada para señales con pulsos o intermitencias: sucesos que ocurren de manera no periódica y a su vez preserva las características esenciales de la señal. Los procedimientos mencionados anteriormente no alteran los datos iniciales, se mantiene el equilibrio y la precisión y se hace con el fin de hallar una mejor caracterización de la señal. Los inconvenientes que se presentan en estos procedimientos son las interferencias eléctricas que resultan en la toma de la señal lo cual dificulta la detección de las ondas “P, Q, R, S, T”.

5.3 IMPLEMENTACIÓN Y DESARROLLO DE LA INTERFAZ GRAFICA

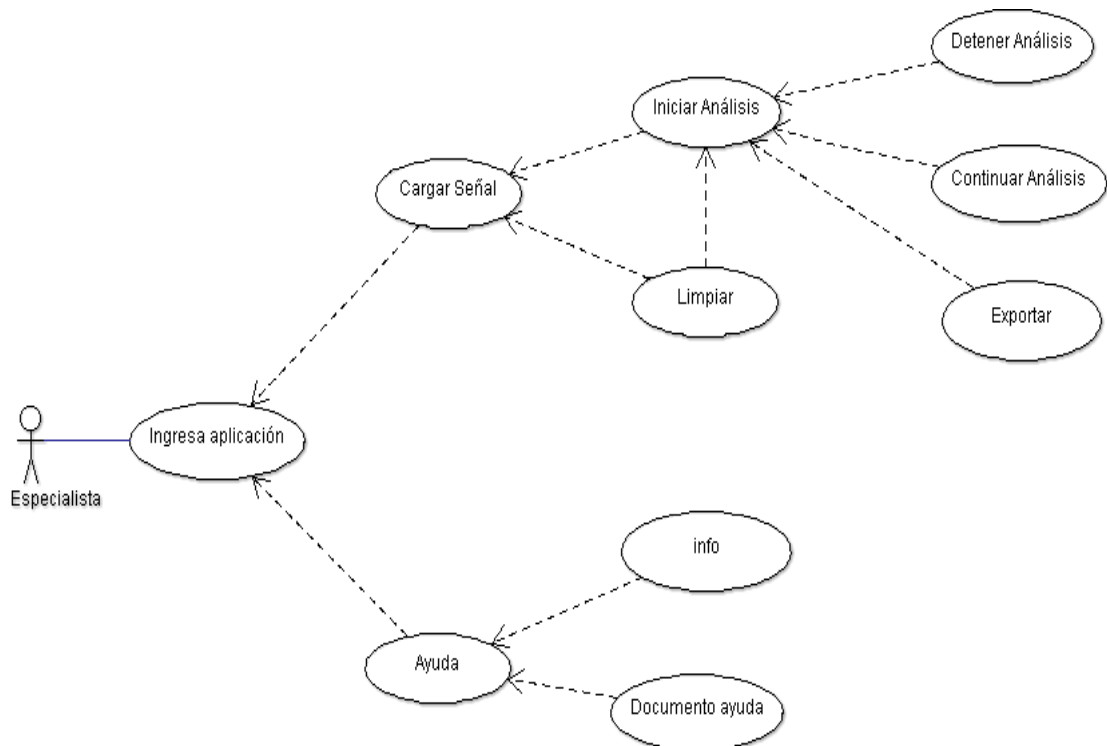
El desarrollo de la aplicación para el actual trabajo, esta implementado sobre el framework de NetBeans IDE 7.1.2 bajo el lenguaje Java. La plataforma NetBeans es un marco genérico para aplicaciones Swing. El proyecto NetBeans consiste en un IDE de código abierto. Este IDE está escrito en Java y puede ejecutarse en cualquier equipo donde esté instalado Java. Es

importante precisar que todas las funcionalidades del IDE se proporcionan por módulos, donde cada módulo proporciona una función bien definida como un apoyo para el lenguaje Java. Esta herramienta fue escogida como entorno de desarrollo para la actual aplicación, debido a que NetBeans cuenta con todos los módulos necesarios para el desarrollo Java en una sola descarga, permitiendo al usuario empezar a trabajar inmediatamente.

5.3.1 Diagramas de casos de uso

En este aparte se muestra la descripción del diagrama de casos de usos que se creó para el desarrollo de la aplicación a la que se le dio el nombre de DiagECC. Aquí se muestran las características generales de la herramienta software.

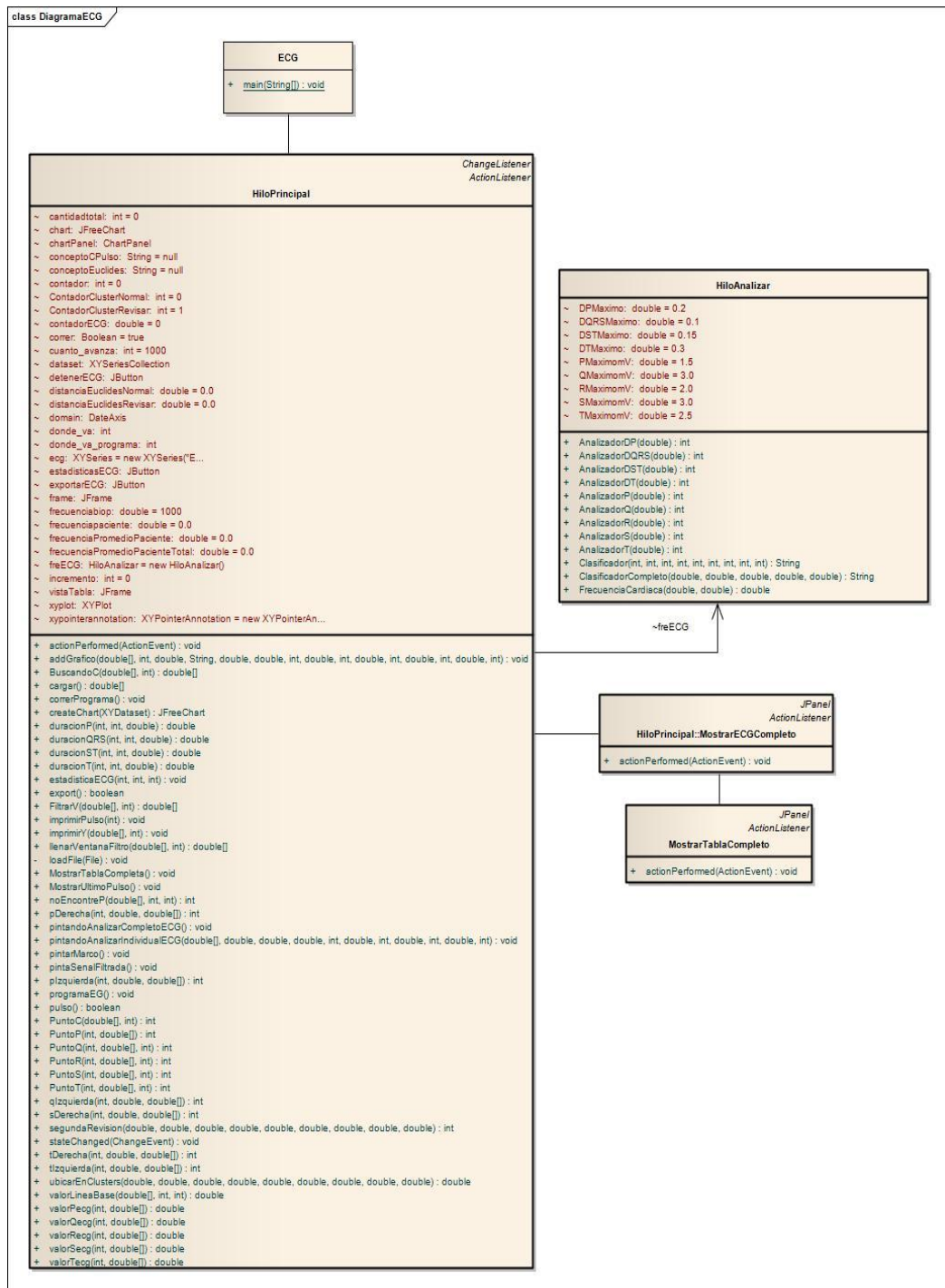
Figura 26. Diagrama de Casos de Uso



Fuente: Autores del Proyecto

5.3.2 Diagrama de clases

Figura 27. Diagrama de clases



Fuente:Autores del proyecto

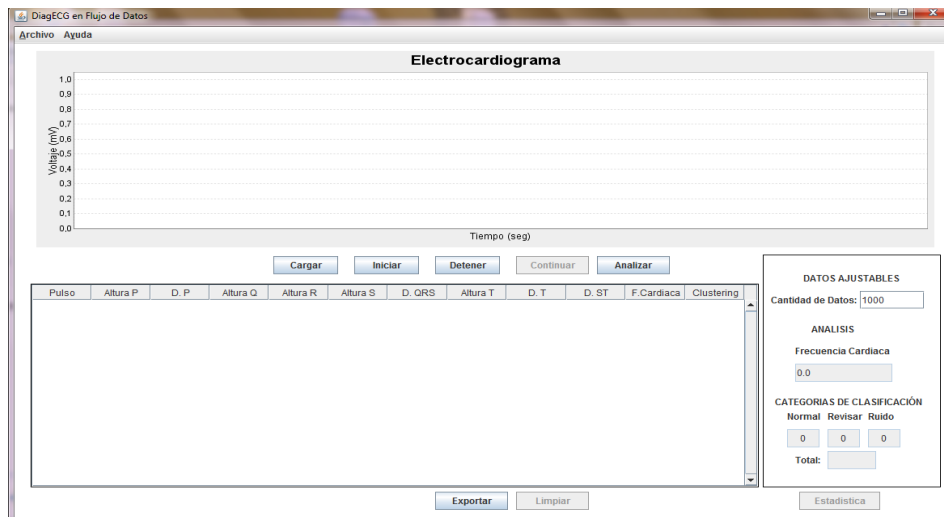
5.3.4 Demostración de la herramienta Software

La implementación de esta herramienta se hizo bajo el lenguaje Java y bajo el paradigma de objetos, con el apoyo de la librería jfreechart, Filtro clase java compiladas de funciones Matlab usando el JBuilder.

A demás para la implementación de los métodos se tuvo en cuenta el diagrama de la Figura26 y se implementaron métodos y la interfaz que se describen a continuación.

Esta interfaz gráfica fue creada pensando en la fácil manipulación por parte del usuario. Consta de una ventana principal, que es donde se realiza el análisis de la señal de ECG. A continuación se hace una descripción grafica de la herramienta software.

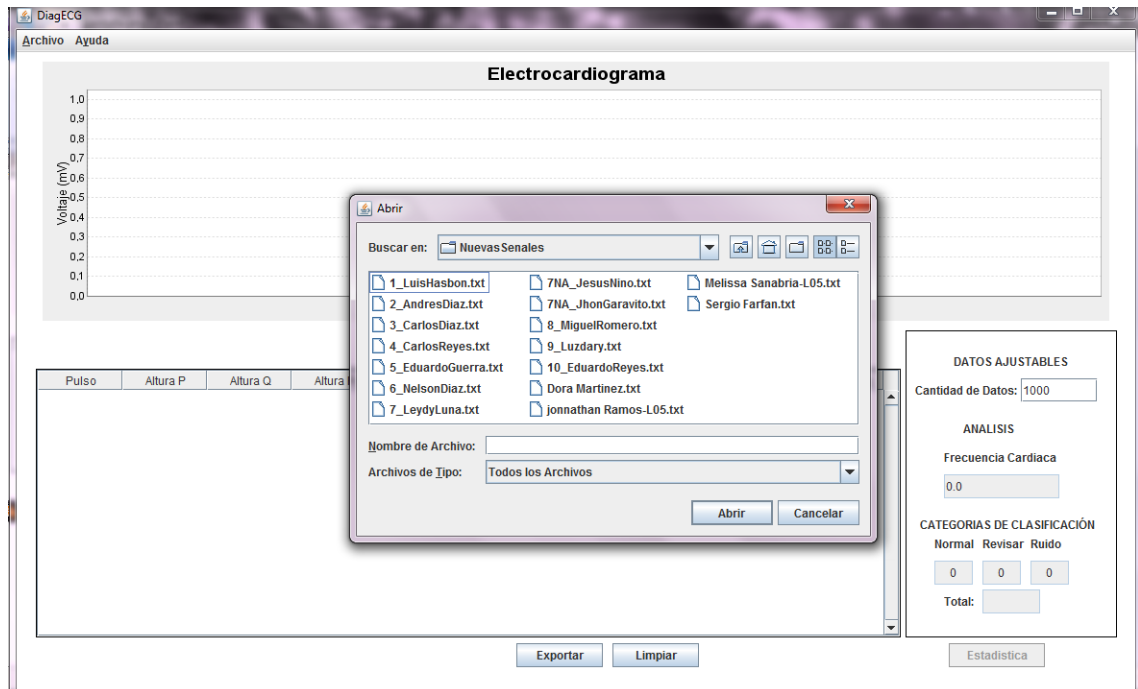
Figura 28. Toma de Pantalla de la interfaz inicial de DiagECG



Fuente: Autores del proyecto

Con la opción Cargar (Figura. 28) se accede al directorio de la señal que se desea analizar. La herramienta soporta los formatos .txt, que es el formato que nos proporciona el Biopac.

Figura 29. Selección del directorio de la señal a cargar

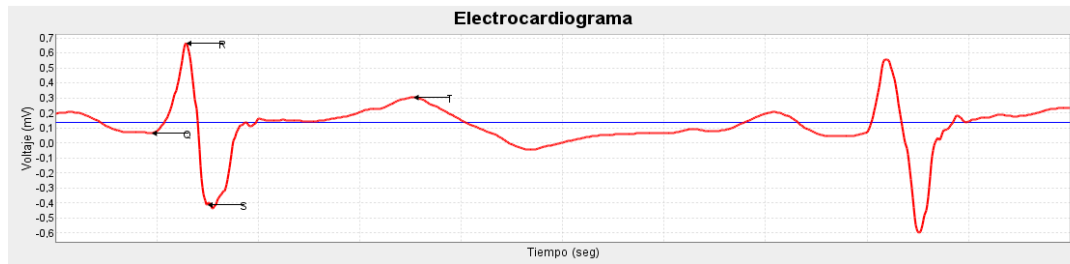


Fuente: Autores del Proyecto

Una vez seleccionada la señal, se activa el botón de Iniciar el cual toma una ventana a realizar procesos con una cantidad de datos especificada en el cuadro de texto Cantidad de Datos.

Dicha ventana realiza el proceso con una Cantidad de Datos por defecto de 1000 datos (modificable por el usuario) y es enviada a la librería Java Filtro compilada de funciones presentes en Matlab con ayuda de JBuilder de MATLAB®, la cual nos aplica el Filtrado con la wavelet db4 de la misma manera como se haría en Matlab, y es graficada en la parte superior de la ventana principal del programa.

Figura 30. Grafica Electrocardiograma herramienta



Fuente: Autores del Proyecto

Posterior a esto la ventana resultante es enviada a la función BuscandoC la cual nos procesa esta ventana con el uso de la librería Filtro y apoyados en la haar nivel 1, con una thr de 0,03 nos ayuda a ubicar un punto C, donde posiblemente esta la onda R.

A partir de este punto C, por análisis de aumento o reducción en el valor en mV de la señal es calculada la posición de la onda R y su valor, seguidamente se encuentra el punto Q y S y sus respectivos valores, luego de esto con la búsqueda de máximos entre el borde izquierdo de la ventana y el punto Q es calculado la posición y el valor de la onda P, y de la misma manera entre S y el borde derecho de la ventana es calculado posición y valor de la onda T.

Todos estos valores extraídos de la señal son agregados por filas en la tabla que se encuentra en la parte inferior de la pantalla principal.

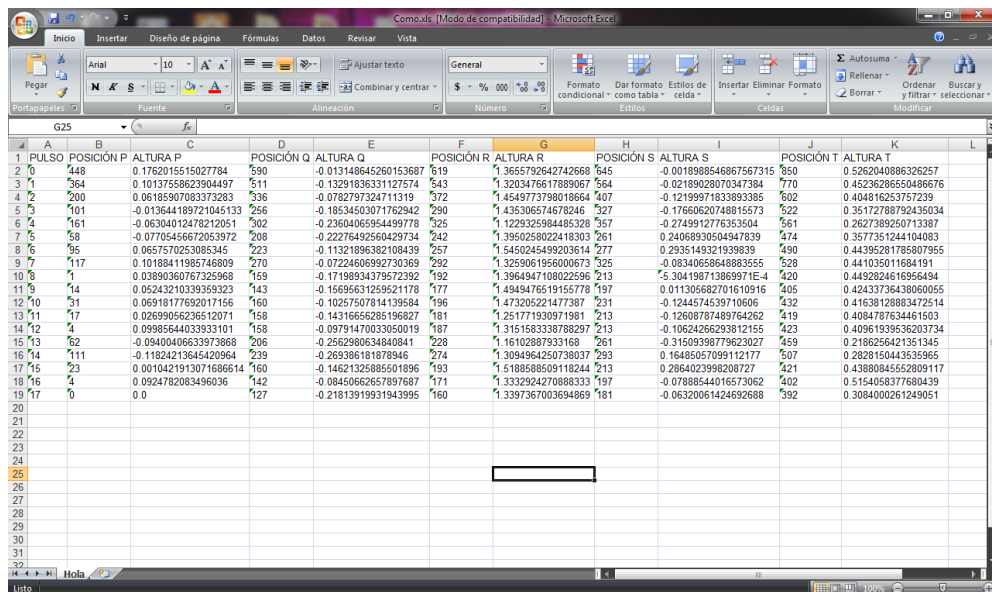
Figura 31. Tabla de datos que resulta de interfaz gráfica

Pulso	Posición P	Altura P	Posición Q	Altura Q	Posición R	Altura R	Posición S	Altura S	Posición T	Altura T
0	448	0.17620155...	590	-0.0131486...	619	1.36557926...	645	-0.0018988...	850	0.52620408...
1	364	0.10137558...	511	-0.1329183...	543	1.32034766...	564	-0.0218902...	770	0.45236286...
2	200	0.06185907...	336	-0.0782797...	372	1.45497737...	407	-0.1219997...	602	0.40481625...
3	101	-0.0136441...	256	-0.1853450...	290	1.43530657...	327	-0.1766062...	522	0.35172788...
4	161	-0.0630401...	302	-0.2360406...	325	1.12293259...	357	-0.2749912...	561	0.26273892...
5	58	-0.0770545...	208	-0.2227649...	242	1.39502580...	261	0.24068930...	474	0.35773512...
6	95	0.06575702...	223	-0.1132189...	257	1.54502454...	277	0.29351493...	490	0.44395281...
7	117	0.10188411...	270	-0.0722460...	292	1.32590619...	325	-0.0834065...	528	0.44103501...
8	1	0.03890360...	159	-0.1719893...	192	1.39649471...	213	-5.3041987...	420	0.44928246...
9	14	0.05243210...	143	-0.1569563...	177	1.49494765...	197	0.01130568...	405	0.42433736...
10	31	0.06918177...	160	-0.1025750...	196	1.47320522...	231	-0.1244574...	432	0.41638128...
11	17	0.02699056...	158	-0.1431665...	181	1.25177193...	213	-0.1260878...	419	0.40847876...
12	4	0.09985644...	158	-0.0979147...	187	1.31515833...	213	-0.1062426...	423	0.40961939...
13	62	-0.0940040...	206	-0.2562980...	228	1.16102887...	261	-0.3150939...	459	0.21862564...
14	111	-0.1182421...	239	-0.2693861...	274	1.30949642...	293	0.16485057...	507	0.28281504...
15	23	0.00104219...	160	-0.1462132...	193	1.51885885...	213	0.28640239...	421	0.43880845...

Fuente: Autores del proyecto.

Al detener el análisis de la señal se activan los botones de Continuar, Exportar y Limpiar los cuales permiten exportar la tabla de datos a un archivo en Excel, para continuar con el análisis de la señal y volver a comenzar con otra señal respectivamente.

Figura 32. Tabla producto de la exportación de la herramienta



	A	B	C	D	E	F	G	H	I	J	K	L
1	PULSO	POSICIÓN P	ALTURA P	POSICIÓN Q	ALTURA Q	POSICIÓN R	ALTURA R	POSICIÓN S	ALTURA S	POSICIÓN T	ALTURA T	
2	0	448	0.1762015515027784	590	-0.013148645260153687	619	1.3655792642742668	645	-0.0018988546867567315	850	0.5262040886326257	
3	1	364	0.10137558623904497	511	-0.13291836331127574	543	1.3203476617889067	564	-0.02189028070347384	770	0.45236286550486676	
4	2	200	0.06185907083373283	336	-0.0782797324711319	372	1.4549773798018664	407	-0.12199971633893385	602	0.404816253757239	
5	3	101	-0.013644189721045133	256	-0.18534503071762942	290	1.435306574678246	327	-0.17660620748815673	522	0.35172788792435034	
6	4	161	-0.06304012478212051	302	-0.23604065954499778	325	1.1229325984485328	357	-0.2749912776353504	561	0.2627389250713387	
7	5	58	-0.07705456672053972	208	-0.22276492560429734	242	1.3950258022418303	261	0.24068930504947839	474	0.3577351244104083	
8	6	95	0.0657570253085345	223	-0.11321896382108439	257	1.5450245499203614	277	0.2935149321939839	490	0.44395281788807955	
9	7	117	0.10188411985746809	270	-0.07224606992730369	292	1.3259061956000673	325	-0.0834065848883555	528	0.441035011684191	
10	8	1	0.03890360767329568	159	-0.17198934379572392	192	1.3964947108022596	213	-5.304198719369971E-4	420	0.4492824616956434	
11	9	14	0.05243210339393323	143	-0.15695631259521178	177	1.4949476519155778	197	0.011305682701610916	405	0.4243373643060055	
12	10	31	0.06918177692017156	160	-0.10257507814139584	196	1.473205221477387	231	-0.1244574539710606	432	0.41638128883472514	
13	11	17	0.02699056236512071	158	-0.14316656285196827	181	1.251771930971981	213	-0.12608787489764262	419	0.4084787634461503	
14	12	4	0.09985644033933101	158	-0.09791470033050019	187	1.315158338788297	213	-0.1062426293812155	423	0.40961939536203734	
15	13	62	-0.09400406633973868	206	-0.2562980634840841	228	1.16102887933168	261	-0.3150939779623027	459	0.2186256421351345	
16	14	111	-0.11824215845420964	239	-0.269386181878945	274	1.3094964250738037	293	0.1648505709912177	507	0.2828150443535965	
17	15	23	0.001042191307168654	160	-0.14621325895501896	193	1.5188588509182244	213	0.2864023952089277	421	0.4388084552809117	
18	16	0	0.0924782083496036	142	-0.08450662657897687	171	1.3332924270888333	197	-0.07888544016573062	402	0.5154058377680439	
19	17	0	0.0	127	-0.21813919931943995	160	1.3397367003694869	181	-0.06320061424692688	392	0.3084000261249051	
20												
21												
22												
23												
24												
25												
26												
27												
28												
29												
30												
31												
32												

Fuente: Autores del proyecto

En la parte inferior derecha de la herramienta se presenta una tabla con un dato ajustable y el valor de la frecuencia cardiaca calculado con base en la distancia entre las onda R; en la parte inferior de esta tabla también se encuentran, 4 cuadros de texto los cuales nos presentan un consolidado de la aplicación de la técnica de minería de datos de clustering y un total de pulsos por señal.

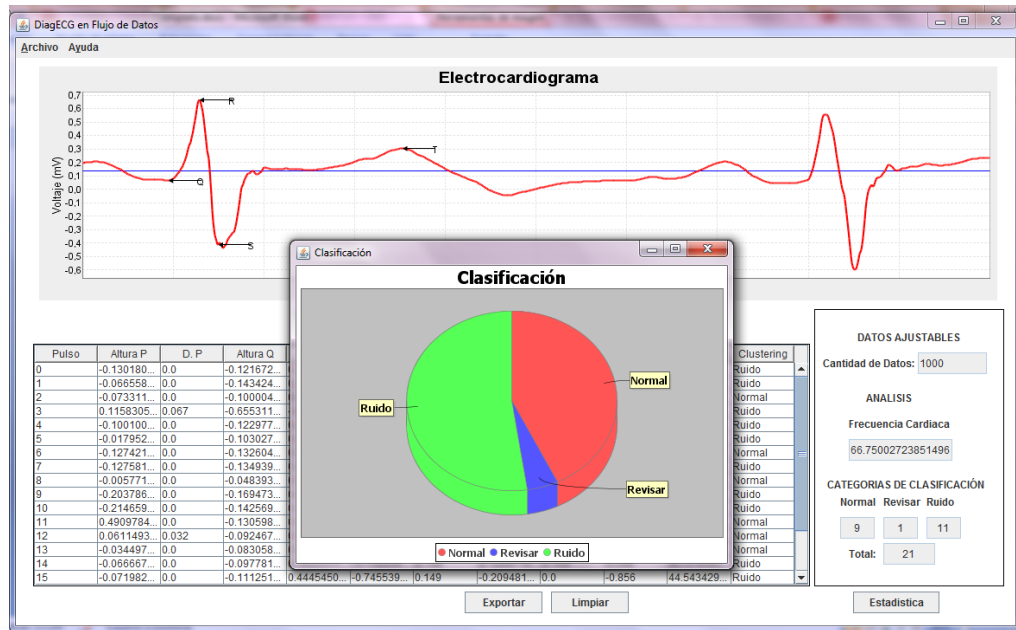
Figura 33. Tabla de dato ajustable y análisis

DATOS AJUSTABLES		
Cantidad de Datos:	1000	
ANALISIS		
Frecuencia Cardiaca		
21.45748987854252		
CATEGORIAS DE CLASIFICACIÓN		
Normal	Revisar	Ruido
108	1	/
Total:		110

Fuente: Autores del Proyecto

Cuando se da click en el botón de estadísticas aparece una grafica como la que se muestra en la Figura 34 donde se muestra el porcentaje de clasificación que se obtuvo del análisis de la señal.

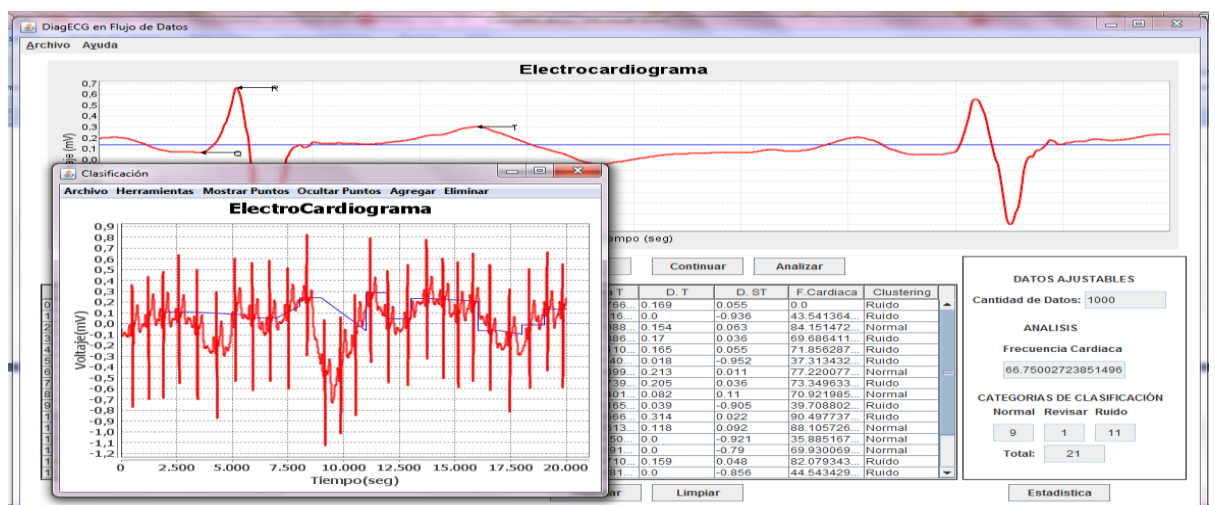
Figura 34. Grafica que muestra estadísticas de la clasificación



Fuente: Autores del Proyecto

Al presionar el botón analizar aparece una venta en la cual se encuentran opciones que el usuario puede seleccionar de acuerdo a lo que desee hacer. En la Figura 35 se muestra gráficamente la ventana que aparece. En esta sección se puede ver la señal junto con su respectiva línea base.

Figura 35. Ventana que aparece con la opción analizar



Fuente: Autores del Proyecto

En esta ventana se pueden realizar modificaciones de acuerdo a los requerimientos del usuario.

5.4 PRUEBA DE LA HERRAMIENTA

5.4.1 Prueba de Sensibilidad y Especificidad

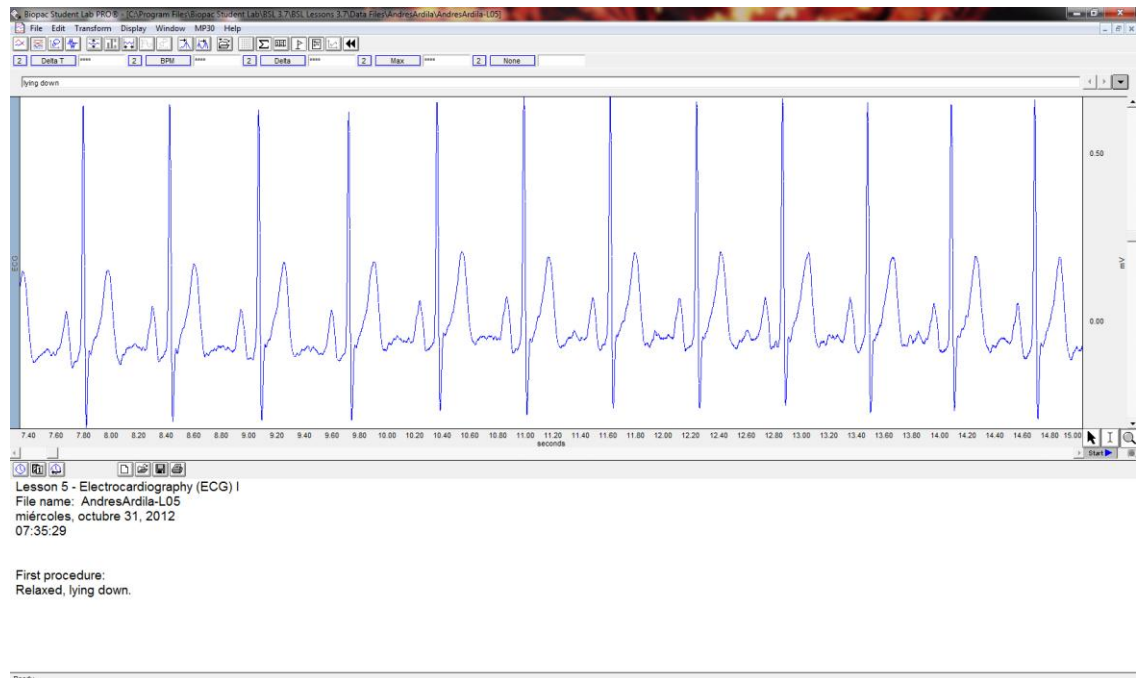
Para verificar la funcionalidad del algoritmo se hicieron pruebas de Especificidad y sensibilidad. Para las pruebas mostradas en este documento se escogieron de manera previa 10 señales de ECG con 2 tipos de ventaneo; estas señales fueron de las que se adquirieron en el grupo de investigación GIIIB. La sensibilidad es la probabilidad que tiene el algoritmo de no detectar anomalías en la señal (Inexistencia de falsos positivos) y la especificidad es la probabilidad que tiene el algoritmo de detectar anomalías en la señal (Existencia de verdaderos positivos).

La se define como: $Sencibilidad = \frac{VP}{VP+FN}$; donde VP son los verdaderos positivos y FN los falsos negativos.

La especificidad se define como: $Especificidad = \frac{VN}{VN+FP}$; donde VN son los verdaderos negativos y FP los falsos positivos.

Para las pruebas realizadas se parte de que el especialista debe realizara los cálculos manualmente. En la Figura 42 se muestra una aproximación de como el especialista ve la señal sin el uso de ventanas adaptativas.

Figura 36. Señal analizada por el especialista



Fuente: Autores del Proyecto

Los resultados que se muestran a continuación son preliminares, ya que actualmente se están realizando más pruebas para mostrar unos resultados más óptimos. Para logra este cometido se cuenta con la colaboración del grupo de investigación del Dr. Oscar Leonel Rueda Ochoa director del (Grupo de electrocardiografía) de la facultad de salud UIS, quien es el encargado de dar validación a la aplicación que se desarrollo. Las muestras que se están recogiendo son a pacientes que presentan diferentes tipos de patologías; esto con el fin de verificar el desempeño del software y dejar una base de datos para pruebas en desarrollos de proyectos futuros.

En la tabla 1 se muestra las pruebas hechas con una ventana de 1000 datos.

Tabla 1. Prueba de herramienta con ventana de 1000 datos

Tamaño Ventana:	1000					
Paciente	Frecuencia Cardiaca Distancia RR	Cantidad de Pulsos			Total Pulsos detectados en flujo	Nro datos
		Normales	Requiere Revisión	Ruido		
Paciente 1	60,229263	120	1	1	122	122000
Paciente 2	65,136728	131	1	0	132	132000
Paciente 3	64,266994	40	1	51	92	92000
Paciente 4	63,407000	130	1	0	131	131000
Paciente 5	54,990886	103	1	28	132	132000
Paciente 6	58,912575	121	1	0	122	122000
Paciente 7	59,020646	81	1	50	132	132000
Paciente 8	65,315834	85	1	37	123	123000
Paciente 9	62,546612	127	1	4	132	132000
Paciente 10	63,560124	120	1	0	121	121000

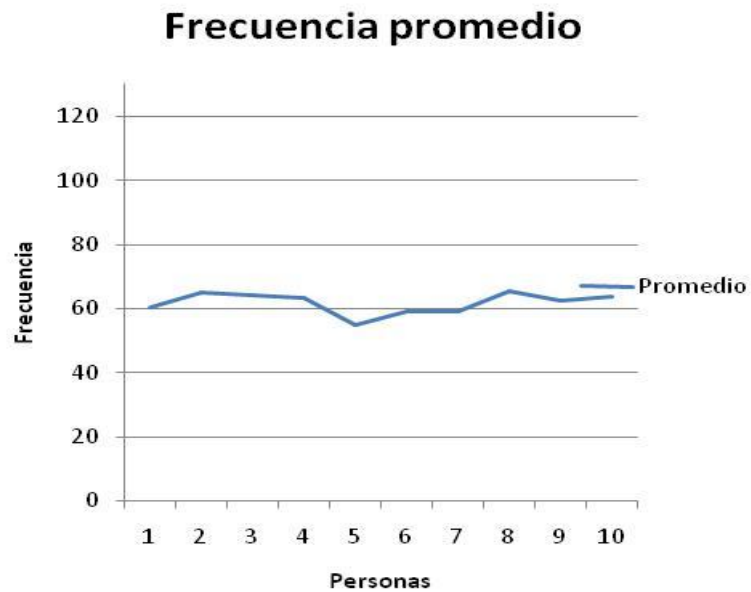
Fuente: Autores del Proyecto

Figura 37. Promedio pacientes ventana de 1000 datos



Fuente: Autores del proyecto

Figura 38. Frecuencia promedio



Fuente: Autores del proyecto

En la tabla 2 se muestra la cantidad de ondas R que nos reconoce la aplicación en flujo de datos con una ventana de tamaño 1000 y el porcentaje de error que produce.

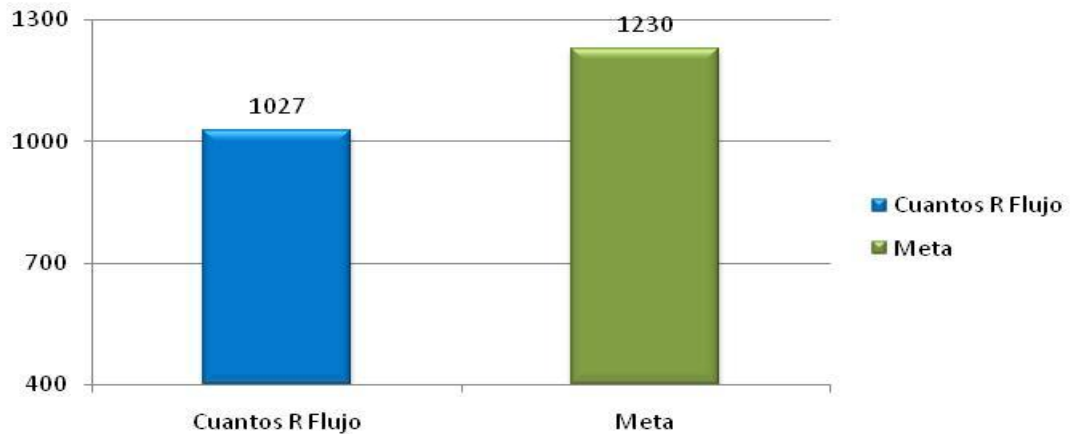
Tabla 2. Cantidad de flujo de datos en R

Cuantos R Flujo	Cuantos R aparecen	Error R
121	122	1%
109	130	16%
67	92	27%
114	130	12%
116	130	11%
94	121	22%
93	131	29%
87	122	29%
122	131	7%
104	121	14%

Fuente: Autores del proyecto

Esta figura muestra la cantidad de onda R que están en los diez ECGs tomados, comparado con la cantidad que nos detecta el flujo con la ventana de 1000 datos.

Figura 39. Cantidad de R detectadas en flujo



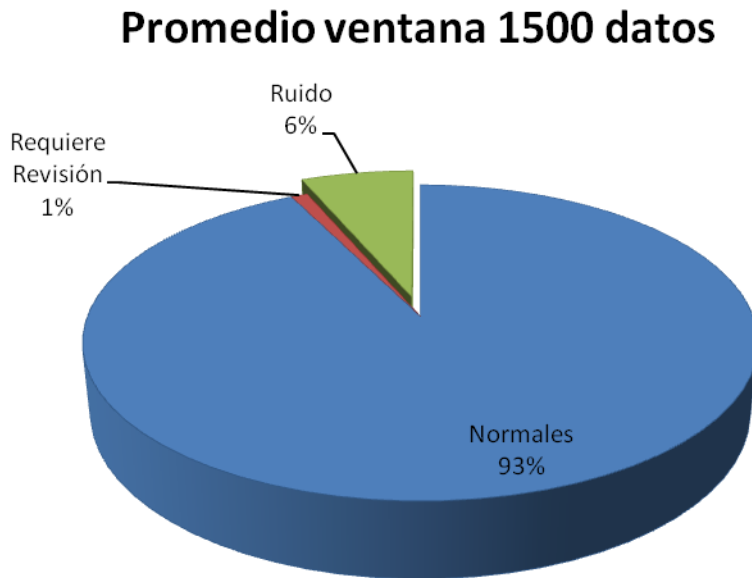
Fuente: Autores del proyecto

Tabla 3. Prueba de herramienta con ventana de 1000 datos

Tamaño Ventana:	1500	Cantidad de Pulsos			Total Pulsos	Nro datos
Paciente	Frecuencia Cardiaca Distancia RR	Normales	Requiere Revisión	Ruido		
Paciente 1	72,142807	80	1	0	81	81000
Paciente 2	48,111469	131	1	0	132	132000
Paciente 3	49,293441	130	1	0	131	131000
Paciente 4	52,358741	130	1	1	132	132000
Paciente 5	75,854488	39	1	2	42	42000
Paciente 6	59,287411	81	1	0	82	82000
Paciente 7	57,289471	129	1	1	131	131000
Paciente 8	54,547580	67	1	63	131	131000
Paciente 9	57,669473	129	1	2	132	132000
Paciente 10	48,275312	80	1	0	81	81000

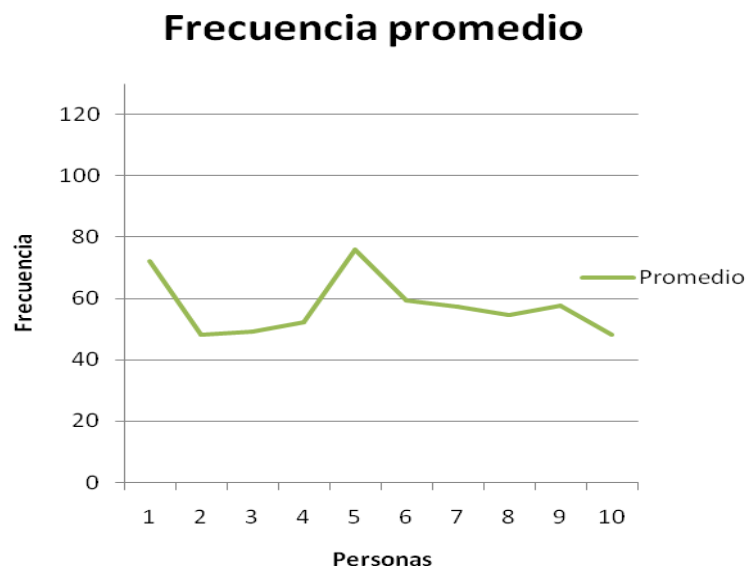
Fuente: Autores del proyecto

Figura 40. Promedio pacientes ventana de 1500 datos



Fuente: Autores del proyecto

Figura 41. Promedio Frecuencia



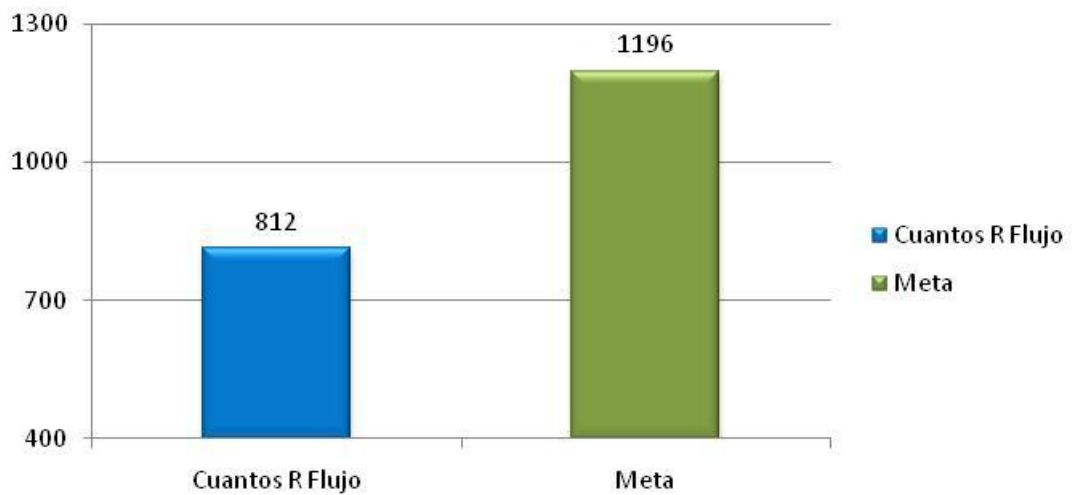
Fuente: Autores del proyecto

Tabla 4. Cantidad de flujo de datos en R

Cuantos R Flujo	Cuantos R aparecen	Error R
81	122	34%
93	130	28%
40	92	57%
113	130	13%
82	130	37%
80	121	34%
84	129	35%
62	131	53%
102	130	22%
75	81	7%

Fuente: Autores del proyecto

Figura 42. Cantidad de R detectadas en flujo



Fuente: Autores del proyecto

6. CONCLUSIONES Y RECOMENDACIONES

6.1 CONCLUSIONES

Este trabajo de investigación de detección de anomalías en señales electrocardiográficas se cumplió con las funciones básicas y esenciales que se plantearon como objetivos durante la planeación del proyecto.

Luego de investigar y analizar los diferentes y posibles filtros a emplear se encontró que la mejor opción fue la del filtro db4, el cual nos redujo el ruido de la señal sin perder datos relevantes de la misma.

En la experiencia de la toma de la señal con el paciente se encontró que muchos de los ruidos encontrados en las señales se deben a las actividades que realiza el paciente paralelamente mientras se está practicando el examen.

La técnica de minería de datos escogida se ajusta de manera adecuada con las características del proyecto debido a que estos métodos de agrupamiento heurístico funcionan bien para bases de datos de pequeño y mediano tamaño.

Con las pruebas realizadas se encontró que la herramienta sirve de base computacional para el área de cardiología para la detección de anomalías en señales de ECG y reduce el tiempo de análisis.

Los comentarios y recomendaciones hecha por los expertos nos ayudaron, para la realización del diseño de la interfaz gráfica. Estas intervenciones y sugerencias contribuyeron de manera positiva para mejorar la aplicación, ya que son ellos los usuarios finales.

En resumen, la siguiente tabla muestra la correspondencia del cumplimiento de los objetivos.

Tabla 5. Revisión de cumplimiento de objetivos Específicos

Objetivo	Cumplimiento
Objetivo 1	Ver Capítulo 4.2: Estado del arte
Objetivo 2	Ver Capítulo 4.2.2: Descripción de Clustering
Objetivo 3	Ver Capítulo 5.1: Diseño e implementación de algoritmo
Objetivo 4	Ver Capítulo 5.3: Implementación y desarrollo de la interfaz grafica
Objetivo 5	Ver Capítulo 5.4: Prueba de la herramienta

Fuente: Autores del Proyecto

6.2 RECOMENDACIONES Y TRABAJO FUTURO

Se recomienda para próximas adaptaciones el uso de ventanas adaptativas, ya que todas las personas no producen pulso en la misma distancia, lo cual hace que si en una ventana queda la mitad de un pulso y en otra el otro, este pulso no sea tenido en cuenta.

Realizar la conexión del software del Biopac con la herramienta de manera directa, para que el análisis de los datos se lleve a cabo al mismo tiempo que la adquisición de los mismos.

Realizar la implementación de secciones de usuario para que la aplicación pueda no solo ser usada por usuarios expertos, sino con carácter educativo.

7. PRODUCCIÓN INTELECTUAL

En el transcurso del desarrollo de este proyecto se escribió un artículo que fue aceptado en el Congreso Nacional de Diseño e Ingenierías Fisicomecánicas de la UNIVERSIDAD INDUSTRIAL DE SANTANDER; titulado Detección y caracterización de señales electrocardiográficas en flujo de datos en la modalidad de Ponencia Oral.

Autores: Lidis Mayerly Rondón Suárez, Germán Eduardo Córdoba Carreño, Nelson Enrique León Martínez, Lidis Mayerly Rondón Suárez.

Ciudad y fecha: Bucaramanga, Colombia 14 al 16 de noviembre 2012.

BIBLIOGRAFIA

- [1] MICHELINE KAMBER, JiaweiHan, Data Mining Concepts and Techniques. Tercera Edición. Edición Digital. 32 p.
- [2] ARINDAM BANERJEE, Varun Chandola, and Vipin Kumar. 3.7. Sensor Networks. Anomaly Detection: A Survey, University of Minnesota, Publicado en 2009. 21 p.
- [3] INTRAMED, artículo El infarto es la primera causa de muerte en Latinoamérica Fecha de consulta, enero de 2012. Disponible en internet: http://www.intramed.net/buscar_resultado.asp
- [4] FUNDACIÓN COLOMBIANA DEL CORAZÓN. Fecha de consulta: enero de 2012. Disponible en internet: <http://www.corazonesresponsables.org/novedades12>.
- [5] CLÍNICA CARDIOVASCULAR. Fecha de consulta: Junio de 2012. Disponible en internet: <http://www.fcv.org/site/index.php>.
- [6] FERNADEZ, N., Suarez, C. Web Service para implementación del estándar SCP-ECG orientado al geoposionamiento. Trabajo de grado para optar por el título de Ingeniero de Sistemas, Bucaramanga. Universidad Industrial de Santander. 2011.
- [7] FRANCISCO RUIZ, Macario Polo, Mario Piattini. Utilización de investigación-acción en la definición de un entorno para la gestión del proceso de mantenimiento del software. Universidad de Castilla-La Mancha. Servicios de publicaciones, Madrid 2003.

- [8] JIM HIGSMITH. Reitering lifecycle dinosaurs: Using adaptive software development to meet the challenges of a high-speed, high-change environment. Software Testing and Quality Engineering (SQET) magazine, 2000.
- [9] MICHELINE KAMBER, JiaweiHan, Data Mining Concepts and Techniques. Tercera Edición. Edición Digital.
- [10] AHMAD KHOUREICHKA, DimitriPetritis, Automated localization and classification of abnormal beats in electrocardiograms using parsimonious wavelet analysis. Association for Computing Machinery. Octubre 2011. Francia. p.1-4.
- [11] ARINDAM BANERJEE, Varun Chandola, and Vipin Kumar. Anomaly Detection: A Survey, University of Minnesota, Publicado en 2009. 15 p.
- [12] CARMONA R., Juan. Carmona R., Luis. Navarro J., Manuel. Electrocardiograma. Malaga, 2008.
- [13] CRUZ DEL VALLE W. Lectura comprensión e interpretación del Electrocardiograma. División de educación continua módulo instruccional. EDIC College, Puerto Rico, 2011.
- [14] ARINDAM BANERJEE, Varun Chandola, and Vipin Kumar. Anomaly Detection: A Survey, University of Minnesota, Publicado en 2009. 9 p.
- [15] MINERÍA DE DATOS, Fecha de consulta: enero de 2012. Disponible en internet: <http://mineriadedatos.blogspot.es/>.
- [16] NINA, MISHRA. Sudipto, Guha. Rajeev, Motwani. Liadan O'Callaghan. Clustering Data Stream, Fecha de consulta, noviembre de 2012.

- [17] KHOUREICHKA, AHMAD. Petritis, Dimitri. Automated localization and classification of abnormal beats in electrocardiograms using parsimonious wavelet analysis, 2011.
- [18] MARK, RG. Moody, GB, The impact of the MIT-BIH Arrhythmia Database. IEEE Eng in Med and Biol Junio, 2001. Disponible en internet: <http://www.physionet.org/physiobank/database/mitdb/>
- [19] KHOUREICHKA, AHMAD. Petritis, Dimitri. Automated localization and classification of abnormal beats in electrocardiograms using parsimonious wavelet analysis, 2009. 4 p.
- [20] BOGDAN HAGIU, Ovidiu Apostu, Sever Pașca, Wireless ECG Monitoring and Alarm System Using ZigBee; Print ISBN: 978-1-4577-0507-6, Julio de 2011 IEEE.
- [21] FONOLLOSA, JOSÉ ADRIÁN. Mariño Acebal, José B. Moreno Bilbao, Asunción. Vallverdú Bayés, Francesc Rodríguez. Tratamiento digital de la señal una introducción experimental. Departamento de Teoría de la Señal y Comunicaciones. Barcelona, 2004.
- [22] SHENG, Y., "The Transforms and Applications Handbook". CRC Press , 1996.
- [23] DAUBECHIES. "Ten Lectures on Wavelets". The Society for Industrial and Applied Mathematics, 1992.
- [24] GARCIA M., Mancilla A., Montes F. Optimización de la Transformada Wavelet Discreta (DWT). Universidad Complutense de Madrid, España 2005.

- [25] KHALIL I., Sufi F., Mahmood A. A clustering based system for instant detection of cardiac abnormalities from. RMIT University, School of Computer Science and Information Technology, Australia 2010.
- [26] NIEBERL S, Kail E., Khor S., Ambulatory Wireless Internet Electrocardiography: New Concepts & Maths. University of Technology and Economics. Hungary 2005.
- [27] CELENTANO L., Hu F., Jiang M., Xiao Y. Robust medical ad hoc sensor networks (MASN) with wavelet-based ECG data mining. Rochester Institute of Technology, Rochester, NY, United States, 2007.
- [28] CHAO-HUI LEE, Jin-Shang Wu, Lee-Cheng Chen, Vincent S. Tseng, Yu-Chia. Development of a Vital Sign Data Mining System for Chronic Patient Monitoring. Institute for Information Industry, Taiwan 2008.
- [29] CHEN X., Leng M., Li L. Variable Length Methods for Detecting Anomaly Patterns in Time Series. Shangrao Normal College and Engineering, Lanzhou University. China, 2008.
- [30] COSTAS P.A., Themis P., Markos G. Methodology for the automated creation of fuzzy expert systems for ischaemic and arrhythmic beat classification based on a set of rules obtained by a decision tree. Departamento de Cardiología, Escuela de Medicina, University of Ioannina, 2007.
- [31] ALI NIZAM, Mehmet Korürek. Clustering MIT-BIH arrhythmias with Ant Colony Optimization using time domain and PCA compressed wavelet coefficients. Istanbul Technical University, Istanbul, Turkey 2010.
- [32] MICHELINE KAMBER, Jiawei Han, Data Mining Concepts and Techniques. Segunda Edición. Edición Digital. 2006

[33] MAIMON, O., Rokach, L. Data Mining and Knowledge Discovery Handbook. Segunda Edición, Editorial Springer, New York

[34] MOLINA J.M, García J. Técnicas de análisis de datos, aplicaciones prácticas utilizando Microsoft, Excel y Weka. Universidad Carlos III, Madrid 2004.

ANEXOS

A. MANUAL DE USUARIO DIAGECG

1. DESCRIPCIÓN

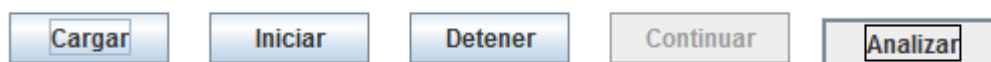
La aplicación DiagECG tiene como objetivo la detección de anomalías en una señal electrocardiográfica a partir del análisis de una cantidad de datos determinada por el usuario por vez, para posibilitar al profesional de la salud una visualización de cada pulso producido por el paciente y centrar su atención en los puntos problema de la señal.

2. ESTRUCTURA DE LA HERRAMIENTA

La herramienta costa de cinco botones en la parte central los cuales le dan acceso al usuario a:

1. Cargar: Permite cargar la señal a partir de un archivo de texto plano formato txt.
2. Iniciar: Inicia el análisis de dicha señal, con un tamaño dado en el cuadro de Cantidad de Datos
3. Detener: Detiene el análisis de la señal, para poder exportar la tabla de datos que se lleva hasta el momento, o generar gráfica de estadística.
4. Continuar: Continúa con la ejecución del programa en el punto en el cual se detuvo.
5. Analizar: Muestra una ventana en la cual se pueden realizar ajustes manualmente por el usuario según sea el caso.

Figura A1. Botones de la aplicación DiagECG



Fuente: Autores del Proyecto

También costa de dos botones inferiores los cuales le dan acceso al usuario a:

6. Exportar: Exporta el contenido de la tabla a un archivo de Excel.
7. Limpiar: Reinicia todos los valores para comenzar nuevamente con otro análisis.

Figura A2. Botones de Limpiar y Exportar

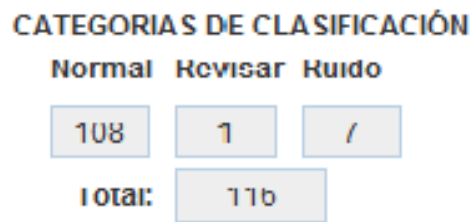


Fuente: Autores del Proyecto

En la parte inferior derecha se encuentra una sección de dato modificable y resultados, acá encontramos:

8. Cantidad de datos: Lugar para colocar la cantidad de datos a analizar por vez.
9. Frecuencia Cardíaca: Cuadro de texto que muestra la frecuencia cardíaca del paciente calculada por distancia entre RR
10. Clases de Clasificación: en esta sección se encuentra las tres clases en las cuales se divide cada pulso, las cuales cada una es un contador que aumenta cuando se produce la condición.
 - Normal: Cuando los valores de amplitud de todas las señales están entre el rango de valores normales
 - Revisión: Cuando los valores de alguna de las señales
 - Ruido: Cuando por tamaño de ventana alguna(s) ondas no son detectadas por pulso
 - Total: Totaliza los tres cuadros anteriores, mostrándonos la cantidad de pulsos normales, que requieren revisión y los no aplica.

Figura A3. Categorías de clasificación

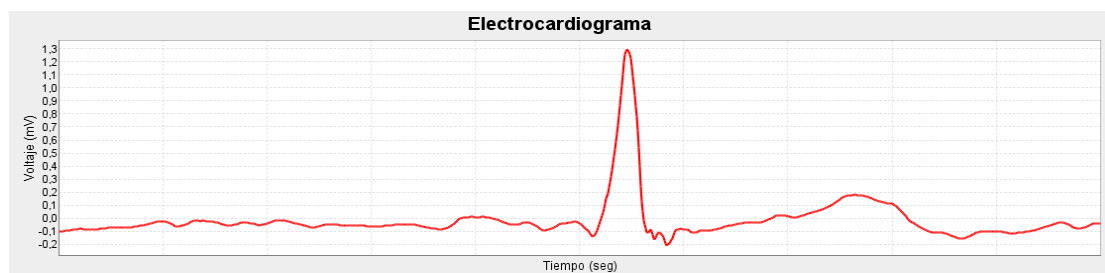


Fuente: Autores del Proyecto

11. Estadísticas: Despliega una ventana con un diagrama de torta 3D con los datos de las clases de clasificación.

12. Vista de la Cantidad de Datos Graficada

Figura A4. Vista de los datos cargados y su respectiva grafica



Fuente: Autores del Proyecto

3. EMPEZAR A TRABAJAR CON DIAGECG

Al ingresar a la herramienta, encontrara en la parte central de la misma el botón de carga el cual despliega un cuadro de diálogo para buscar la señal en formato .txt en su computador.

Los pasos que debe seguir son los siguientes:

1. Dar clic en el botón de Cargar y seleccionar el archivo .txt

2. Ajustar el valor de Cantidad de Datos (por defecto se encuentra en 1000 datos).

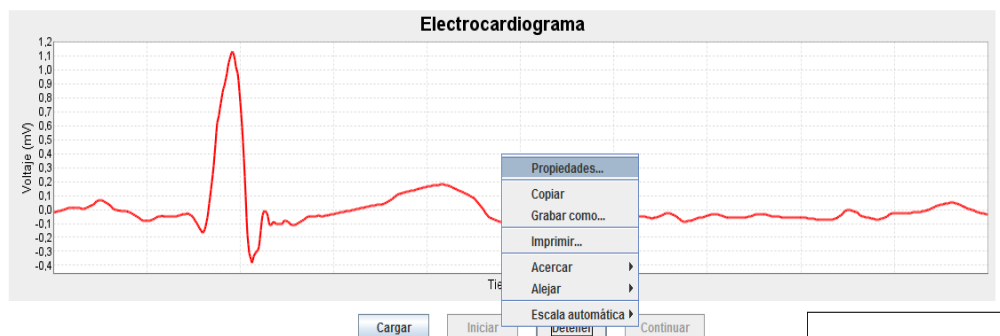
Figura A5. Datos ajustables



Fuente: Autores del Proyecto

3. Dar clic en el botón Iniciar, el cual comenzara a arrojar los resultados del análisis en la tabla que se encuentra en la parte inferior del programa.
4. En cualquier momento de la ejecución usted puede:
 - 4.1. Ampliar, Guardar la imagen de la señalque está apareciendo, estas dos opciones las encontrara dando clic derecho sobre la imagen de la señal ECG

Figura A6. Guardar la imagen de la señal



Fuente: Autores del Proyecto

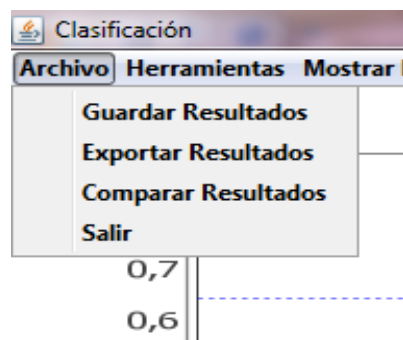
4.2. Detener el análisis y exportar la tabla de datos o ver las estadísticas de cómo están consolidado los pulsos.

4.3 Analizar muestra una ventana en la cual se pueden hacer modificaciones dependiendo los casos que considere el usuario.

A continuación se describe cada una de los menús que aparecen en esta ventana.

El menú Archivo tiene las siguientes opciones: Guardar Resultados, con la cual se guardan los resultados de la señal una vez analizados por el usuario; la opción Exportar Resultados envía los datos a una ubicación del disco en formato .xlsx. La opción Comparar resultados sirve para comparar los resultados que se obtienen por parte del especialista y los que arroja la aplicación. La opción salir cierra esta ventana.

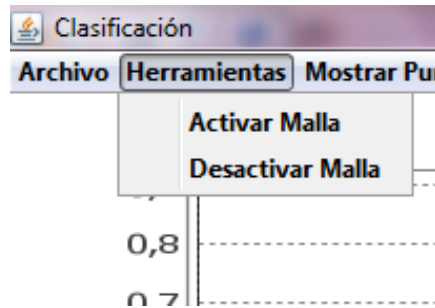
Figura A7. Menú desplegable archivo



Fuente: Autores del Proyecto

El menú herramientas cuenta con dos opciones : Activar Malla y Desactivar Malla, las cuales se dejan a escogencia del usuario. En la Figura 37 se muestra graficamente este menú.

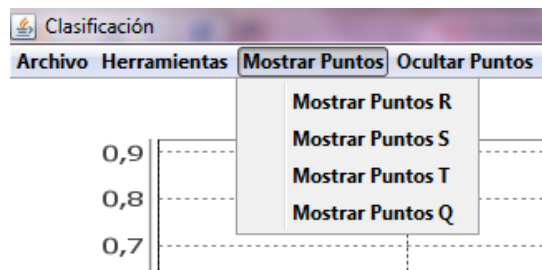
Figura A8. Menú Activar y desactivar malla



Fuente: Autores del Proyecto

El menú Mostrar puntos sirve para mostrar los puntos del ECG que el usuario desee que se vean.

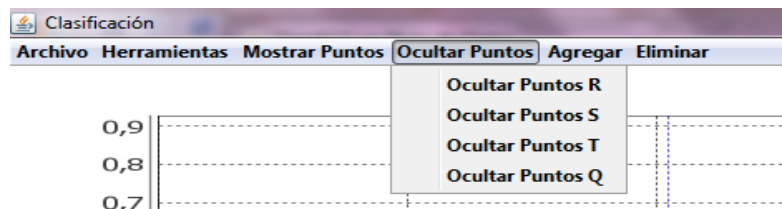
Figura A9. Menú Mostrar Puntos



Fuente: Autores del Proyecto

En el menú ocultar puntos se puede escoger cual punto el usuario no desea ver.

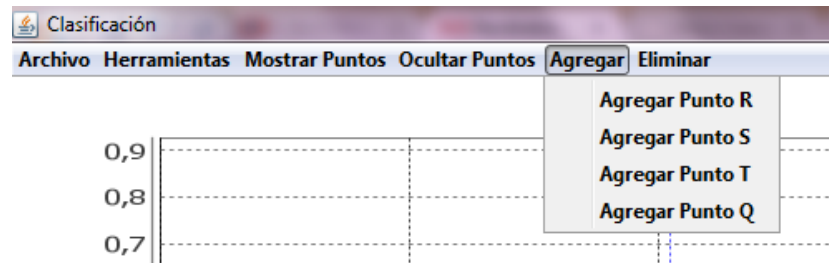
Figura A10. Menú ocultar Puntos



Fuente: Autores del Proyecto

En el menú agregar se puede agregar los puntos que por alguna razón no identifica la aplicación en la ubicación en la que el usuario desee.

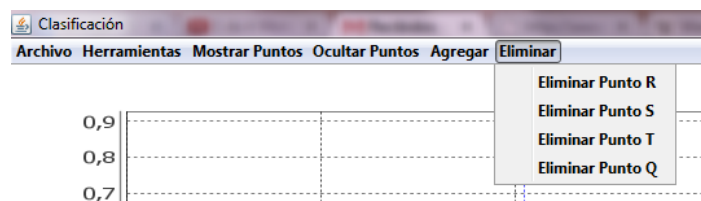
Figura A11. Menú Agregar



Fuente: Autores del Proyecto

En el menú Eliminar el usuario puede eliminar los puntos que desee si considera que no están correctamente ubicados.

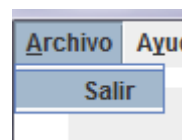
Figura A12. Menú Eliminar



Fuente: Autores del Proyecto

5. Para finalizar puede dirigirse a al menú superior Archivo, Salir.

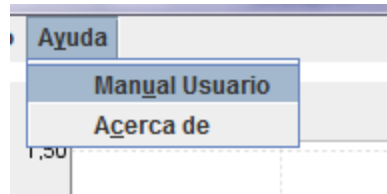
Figura A13. Menú Archivo



Fuente: Autores del Proyecto

6. En ayuda encontrara una descripción de los autores.

Figura A14. Menú ayuda



Fuente: Autores del Proyecto

Esperemos que esta herramienta sea de su entero agrado.

4. DATOS DE CONTACTO

Cualquier inquietud se puede comunicar a los correos:
german.cordoba@correo.uis.edu.co, lidis.rondon@correo.uis.edu.co