



Universidad de los Andes
Facultad de Ciencias
Departamento de Física
Laboratorio de Física Aplicada

Búsqueda de agrupaciones en data proveniente de electrocardiogramas (ECG), mediante el Análisis de Componentes Principales (PCA) y el uso de Redes Neuronales.

Trabajo especial de grado.

Br. Abrahan David Quintero Terán

Tutor: Prof. Juan Villegas

Jurados:
Prof. Marcos Rodríguez
Prof. Jhon Ferreira

Mérida – Venezuela
2024

Resumen

Este es el resumen de tu tesis en español.

Índice general

I	Primer Seminario	1
1	El problema	3
1.1.	Justificación falta por corregir y leer	3
1.2.	Antecedentes	4
1.3.	Objetivos	6
2	Marco Teórico	7
2.1.	Antecedentes	7
2.2.	Conceptos básicos	7
2.3.	Conceptos específicos	8
II	Segundo seminario	9
3	Marco experimental	10
4	Resultados	11
III	Conclusiones	12
5	Conclusiones	13

Parte I

Primer Seminario

Introducción

Capítulo 1

El problema

1.1. Justificación falta por corregir y leer

Según la organización mundial de la salud, las enfermedades cardiovasculares son las principal causa de muerte en hombres y mujeres en el mundo, con alrededor de 17,9 millones de personas que mueren al año a causa de estas, es por esto que surge la necesidad de desarrollar herramientas que ayuden a reducir los errores provenientes de la interpretación manual de los ECG, como lo son la variabilidad entre los observadores y métodos manuales de identificación de anomalías que son subjetivos y pueden llevar a ocasionar diagnósticos imprecisos y en consecuencia, tratamientos inadecuados. Además, la gran cantidad de datos generados por los ECGs y la presencia de ruidos e interferencias en las señales complican aún más su análisis, por lo que la identificación de patrones específicos y la clasificación precisa en grupos de pacientes siguen siendo desafíos importantes ya que los métodos tradicionales de análisis de ECGs a menudo no son lo suficientemente robustos para manejar estas complejidades de manera eficiente. Es por esto que la debida identificación de diversas enfermedades cardiovasculares, puede tener un impacto positivo en el proceso de diagnostico de enfermedades y facilitar un tratamiento oportuno. Una opción para la detección de arritmias, por ejemplo, es el análisis de las señales de electrocardiograma [4], tradicionalmente, este se divide en dos pasos: extracción de características y clasificación de patrones. Por lo general, los médicos emplean características heurísticas diseñadas manualmente o utilizan arquitecturas de aprendizaje de características superficiales para la interpretación de los electrocardiogramas [5], sin embargo el problema radica en la posibilidad de no encontrar la mayoría de las características apropiadas que darán una alta exactitud en el diagnostico a partir de las señales.

1.2. Antecedentes

El electrocardiograma (ECG) es una técnica no invasiva que permite registrar y medir las señales eléctricas generadas por el corazón. Consiste en la captación de la variación temporal del potencial bioeléctrico durante cada ciclo cardíaco, utilizando electrodos colocados en la superficie cutánea del paciente [1]. El análisis del electrocardiograma proporciona datos sobre el sistema cardiovascular, en particular el corazón, lo que permite detectar diversas enfermedades que puedes afectar su funcionamiento adecuado. Estas enfermedades incluyen arritmias cardíacas, insuficiencia cardíaca y ataques al corazón [2]. En este sentido, según la Organización Mundial de la Salud [3], 17.9M de personas al año, siendo este el 32 % de las muertes anuales en el mundo, siendo la principal causa de muerte a nivel mundial. Es por esto que la debida identificación de diversas enfermedades cardiovasculares, puede tener un impacto positivo en el proceso de diagnostico de enfermedades y facilitar un tratamiento oportuno. Una opción para la detección de arritmias, por ejemplo, es el análisis de las señales de electrocardiograma [4], tradicionalmente, este se divide en dos pasos: extracción de características y clasificación de patrones. Por lo general, los médicos emplean características heurísticas diseñadas manualmente o utilizan arquitecturas de aprendizaje de características superficiales para la interpretación de los electrocardiogramas [5], sin embargo el problema radica en la posibilidad de no encontrar la mayoría de las características apropiadas que darán una alta exactitud en el diagnostico a partir de las señales. Desde finales del siglo pasado ha habido un creciente interés en los métodos computacionales aplicados a la salud, por ejemplo Pan y Tompkins (1985) [6] diseñaron un algoritmo para detección de los complejos QRS en tiempo real y con bajo costo computacional, este algoritmo sentó las bases de la implementación de algoritmos de detección y análisis para las señales de electrocardiogramas. En 1996, Laguna *et al.* [7] presentaron un sistema de estimación del modelo de Hermite adaptativo (AHMES) para la estimación en línea latido a latido de las características que describen el complejo QRS con el modelo de Hermite. Gómez Herrero *et al.* [8], presentaron un algoritmo conocido como "Matching Pursuit" que ofrece la capacidad de descomponer cualquier señal en un combinación lineal de formas de onda extraídas de un diccionario redundante de funciones llamado Gabor. Este algoritmo se ha reconocido como una herramienta eficaz para realizar transformaciones adaptativas de tiempo-frecuencia en señales de ECG, lo que permite obtener características relevantes en el dominio tiempo-frecuencia. Siguiendo esta línea de algoritmos que realizan transformaciones adaptativas en el dominio tiempo-frecuencia, se encuentran Martínez *et al.* [9] quienes proponen un delineador de ECG basado en las transformadas Wavelet

(TW), para así poder detectar los inicios, picos y finales de las ondas P y T como también las ondas individuales del complejo QRS incluyendo el inicio y fin del complejo, todo esto para luego determinar los diversos intervalos y segmentos dentro del ECG, este método propuesto es tan robusto que no se ve afectado por los diversos ruidos que pueden existir dentro del ECG como lo es por ejemplo, el desplazamiento de la línea base, para remover este último, Sharma y Sharma en 2015 [10] usan Descomposición Vibracional de Hilbert para descomponer la señal original del ECG en una serie de funciones modales intrínsecas para luego remover el primer término, que corresponde a la componente de mayor energía y así eliminar el desplazamiento de la línea base del ECG. Otros métodos relevantes que implican la extracción de características a partir de dominios transformados, tales como la Transformada de Coseno Discreta (DCT), la transformada Wavelet Continua (CWT) y la Transformada Wavelet Discreta (DWT). Khorrami y Moavenian (2010) [11] utilizaron la CWT con dos técnicas de transformación de datos DWT y DCT, con el fin de mejorar la capacidad de dos clasificadores de patrones en la clasificación de arritmias ECG. Song *et al.* (2005) [12] extrajeron diecisiete características de entrada originales de señales pre-procesadas mediante TW, utilizando el análisis discriminante lineal (LDA). El rendimiento del clasificador SVM (Máquina de Soporte Vectorial) con características reducidas por LDA mostró ser mayor que con el análisis de componentes principales (PCA) e incluso con características originales, sin embargo, esta técnica requirió de un mayor costo computacional. Yu y Chen (2007) utilizaron la transformación wavelet y la red neuronal probabilística (PNN), para descomponer las señales de latido de ECG en diferentes sub-bandas utilizando la transformada wavelet discreta. Posteriormente, seleccionaron tres conjuntos de características estadísticas de las señales compuestas para caracterizar las señales de ECG, así como la potencia de AC y el intervalo RR instantáneo de la señal original. Ye, Coimbra y Kumar (2010) propusieron un enfoque de combinación de características morfológicas y dinámicas refiriéndose a la TW y el análisis de componentes independientes (ICA) aplicándose por separado a cada latido del corazón para extraer los coeficientes correspondientes como características morfológicas. Además se concatenaron la información del intervalo RR y estos dos tipos diferentes de características, y se utilizó SVM para la clasificación. Rojas, Medina y Dugarte diseñan un sistema de adquisición multicanal y analizan la señal electrocardiográfica de alta resolución, en el que utilizan Máquinas de Soporte Vectorial de mínimos cuadrados (LSSVM) para determinar el inicio del complejo QRS y el final de la onda T, entrenadas en base a atributos extraídos de la señal preprocesada y de señales obtenidas mediante descomposiciones con Wavelets. Estas técnicas permiten estimar el intervalo QT así como el intervalo QT cor-

regido (QTc). También Zhang *et al.* (2024) proponen un modelo de Redes Neuronales Convolucionales (CNN) para clasificar insuficiencias cardíacas por clases, según la Asociación del Corazón de Nueva York, a partir de imágenes electrocardiográficas en el que consiguen el mejor resultado segmentando las imágenes en fragmentos de 12 segundos. Así también Astudillo *et al.* prueban cinco arquitecturas de CNN diferentes para clasificación de arritmias cardíacas. Estas dos ultimas investigaciones consiguen predicciones superiores al 98.98 % en el mejor de los escenarios.

En este sentido, Pan y Tompkins obtuvieron un 99.3 % de complejos QRS detectados haciendo uso de la base de datos MIT-BIH Arritmia (MITDB). Song *et al.* (2005) identificaron seis tipos diferentes de arritmias obteniendo una exactitud del 98.94 %. Ye, Coimbra y Kumar (2010) reconocieron 15 clases de latidos con una exactitud del 99.66 % en un grupo de prueba de 85,945 muestras.

Martínez *et al.* usando MITDB encuentran un 99.8 % de complejos QRS detectados, un resultado superior al obtenido por Pan y Tompkins. Así también Sharma y Sharma compara su método usando criterios de correlación y de relación señal/ruido donde concluyen que la técnica propuesta se desempeña mejor en la mayoría de los casos.

1.3. Objetivos

Objetivo General

Analizar los datos provenientes de electrocardiogramas mediante la aplicación de análisis de componentes principales (PCA) y redes neuronales, con el objetivo de identificar agrupaciones en dicha data.

Objetivos específicos

1. Analizar y catalogar los datos de los electrocardiogramas para mejorar la calidad de los datos y la precisión de los modelos predictivos.
2. Desarrollar un marco metodológico solido que combine el análisis de componentes principales (PCA) y redes neuronales para la identificación de grupos de datos.
3. Evaluar el rendimiento de los modelos propuestos utilizando los datasets de Physionet.

Capítulo 2

Marco Teórico

2.1. Antecedentes

2.2. Conceptos básicos

Introducción al Electrocardiograma (ECG)

La humanidad siempre, en su continuo deseo de aprender y entender más, ha querido desentrañar los secretos del cuerpo humano entendiendo su funcionamiento interno. Al principio con técnicas invasivas acordes a la tecnología disponible a la época, pero evolucionando continuamente, creando así exámenes cada vez menos invasivos, con el fin de mejorar el diagnóstico, siendo mas preciso y oportuno. Entre esos exámenes se destaca el electrocardiograma (ECG) el cual es una representación visual de la actividad eléctrica del corazón en función del tiempo, que se obtiene desde la superficie corporal, con un electrocardiógrafo, este es el instrumento principal de la electrofisiología cardíaca y tiene una función relevante en el cribado y diagnóstico de las enfermedades cardiovasculares, alteraciones metabólicas y demás utilidades.

El corazón consta de cuatro partes, dos aurículas y dos ventrículos, el ECG registra los impulsos eléctricos que estimulan estas partes y producen su contracción. Las células cardíacas en reposo se encuentran cargadas o polarizadas; pero la estimulación eléctrica las "despolariza", y se contraen.

Como se observa en la figura 2.1 el ECG consta de varias ondas representativas de cada etapa de un latido cardíaco, estas son:

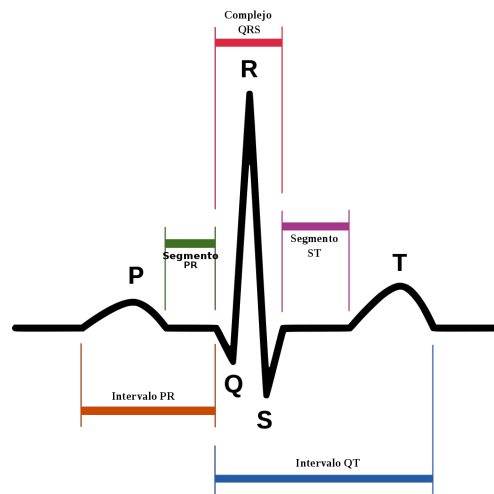


Figura 2.1: ECG del corazón con ritmo sinusal normal

- Onda P:

2.3. Conceptos específicos

Parte II

Segundo seminario

Capítulo 3

Marco experimental

...

Capítulo 4

Resultados

...

Parte III

Conclusiones

Capítulo 5

Conclusiones

...