

**STUDY AND CLASSIFICATION OF CHAGASIC PATIENTS USING
MULTIVARIATE STATISTICAL TECHNIQUES****ESTUDIO Y CLASIFICACIÓN DE PACIENTES CHAGÁSICOS UTILIZANDO
TÉCNICAS ESTADÍSTICAS MULTIVARIANTES****MSc. Luis Enrique Mendoza*, PhD. Delfina M. Padilla ****

* Postgrado de Ingeniería Biomédica, **Universidad de los Andes**, Facultad de Ingeniería, Mérida 5101, Venezuela, E-mail: luenmecaib@gmail.com.

** Departamento de Cálculo, **Universidad de los Andes**, Facultad de Ingeniería, Mérida 5101, Venezuela. E-mail: pdelfina@ula.ve.

Abstract: The study of cardiac arrhythmia from electrocardiogram signals (ECG), has been an extensive field research in biomedical engineering, with the pathologies associated with Chagas disease one of the highest scientific interest in recent years. This article presents results of recent studies to extract information from a subspace of features smaller than the original using multivariate statistical techniques, and thus properly classified and quick healthy patients, Chagas patients and mild chronic Chagas patients. Other techniques used in the study are: the wavelet transform and dynamic programming.

Resumen: El estudio de arritmias cardíacas a partir de señales electrocardiográficas (ECG), ha sido un campo de extensa investigación en la ingeniería biomédica, siendo las patologías relacionadas con la enfermedad del chagas una de las de mayor interés científico en los últimos años. Este artículo presenta resultados de recientes estudios realizados para extraer información de un subespacio de características de menor dimensión que el original utilizando técnicas estadísticas multivariantes, y así clasificar de forma correcta y rápida pacientes sanos, pacientes chagas leve y pacientes chagas crónico. Otras técnicas utilizadas en el estudio son: la transformada wavelet y programación dinámica.

Keywords: ECG signals, Chagas disease, multivariate techniques, classification.

1. INTRODUCCIÓN

En la fisiología humana, el análisis del sistema cardíaco ha sido de gran importancia durante muchas décadas, esto se debe a que el sistema cardíaco tiene gran repercusión del funcionamiento fisiológico total del ser humano. El estudio del sistema cardiaco con base en señales ECG, han proporcionado resultados los cuales muestran que se pueden detectar arritmias cardíacas tales como: taquicardias ventriculares y mal de chagas, entre otros (Rasoul *et al.*, 2007); (Mark *et al.*, 1998);

(Bonneyoy *et al.*, 2000). El estudio, caracterización y clasificación de arritmias cardíacas, específicamente las encontradas en el complejo QRS, el cual se encuentra presente en la señal ECG, tienen gran importancia en el mundo de la medicina, ya que fisiológicamente, el complejo QRS, representan la despolarización ventricular del ciclo cardiaco (Mithilesh *et al.*, 2007); (Florim *et al.*, 2007). El mal de chagas es un enfermedad típica en América, generando una alta tasa de mortalidad en la última década, debido al tardío diagnóstico médico; el mal de chagas se puede

detectar específicamente en los complejos QRS o en el segmento ST de la señal ECG (Gan-Xin *et al.*, 1999), Figura 1. Diferentes trabajos han desarrollado técnicas para la detección de arritmias cardíacas, específicamente para el estudio y clasificación de pacientes con mal de chagas (Marta *et al.*, 2006); (Tomás *et al.*, 2006), una desventaja de estos métodos desarrollados es la sensibilidad que las señales tienen al ruido, esto es causado por la poca tasa de muestreo que poseen los dispositivos de captura de la señal ECG (Caryn *et al.*, 2007).

La enfermedad del chagas es causada por un parásito llamado: *Trypanosoma cruzi*, usualmente se transmite por: picadura del insecto, transfusiones sanguíneas, trasplante de órganos y raramente por ingestión de comida o bebidas contaminada (Caryn *et al.*, 2007). La transmisión del parásito ocurre en el continente Americano donde aproximadamente entre 8-10 millones de personas poseen esta mal.

En países como Colombia y Venezuela, más del 40% de la población esta expuesta a padecer este mal, por estos motivos, este trabajo se basa en el estudio de pacientes chagásicos, con el fin de ayudar al médico a diagnosticar esta enfermedad en fases tempranas, y lograr reducir la tasa de mortalidad que aquejan a los países Americanos.

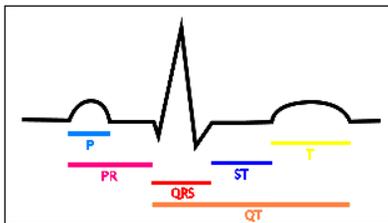


Fig. 1. Señal ECG

2. MÉTODOS

La base de datos fue tomada del grupo de ingeniería biomédica de la universidad de los Andes GIBULA, la base de datos contiene 10 sujetos sanos, 8 sujetos con chagas leve y 10 sujetos con chagas crónico, los datos fueron adquiridos a una frecuencia de muestreo de 1Khz.

2.1. Detección complejo QRS

La transformada wavelet (TW), es una técnica multiresolución, que permite descomponer una señal o una imagen en diferentes niveles en un plano tiempo-escala (Zhang, 2005). En este trabajo se usa la wavelet madre *symle 7*, con un nivel de

descomposición 5, y se eligió el nivel de detalle 2, para encontrar la duración del complejo QRS, ya que representa de forma adecuada el inicio y fin de las componentes frecuenciales de la onda QRS (Laciar *et al.*, 2000).

Matemáticamente la TW se expresa (Sydney *et al.*, 1998):

$$c(\mathbf{t}, s) = \frac{1}{\sqrt{|s|}} \int_{-\infty}^{\infty} f(t) * \mathbf{y} \left[\frac{t - \mathbf{t}}{s} \right] dt \quad (1)$$

$\Psi[\mathbf{t}]$: Wavelet madre

$f(t)$: Señal original

\mathbf{t} : Coeficientes traslación,

s : Coeficientes de escala.

2.2. Alineamiento de secuencias

El Alineamiento de secuencias se realiza como una aplicación de la programación dinámica, en la cual se permiten identificar correspondencias exactas o aproximadas entre subsecuencias de caracteres (Weichun *et al.*, 2005), en este trabajo se implementó el algoritmo de lineamiento global, con el objetivo de alinear los complejos QRS de los diferentes pacientes. Las ondas QRS se alinearon utilizando un punto fiduciario, en este trabajo el punto es identificado como el pico máximo de la onda R.

Matemáticamente el alineamiento global se expresa (Weichun *et al.*, 2005):

Condiciones iniciales

$$M(i, 0) = -d * i \quad 0 \leq i \leq n \quad (2)$$

$$M(0, j) = -d * j \quad 0 \leq j \leq m$$

n, m : Tamaños secuencias 1 y 2 respectivamente.

d : Penalización por inserción de espacios.

Fórmula recursiva

$$M(i, j) = \begin{cases} M(i-1, j-1) + score(x_i, y_j) \\ M(i, j-1) + d \\ M(i-1, j) + d \end{cases} \quad (3)$$

Para $1 \leq i \leq n$ y $1 \leq j \leq m$

$score(x_i, y_j)$: Puntaje de alinear elemento i de la secuencia x con el elemento j de la secuencia y .

2.3. Análisis de componentes principales (ACP) y Clasificadores Multivariantes

Cuando el investigador dispone habitualmente de muchas variables medidas en una colección de individuos y pretende estudiarlas conjuntamente, suele acudir al análisis estadístico de datos (Lebart *et al.*, 1982); (Diday *et al.*, 1982). Al observar muchas variables sobre un muestra es presumible que una parte de la información recogida pueda ser redundante, en cuyo caso los **métodos multivariantes de reducción de la dimensionalidad** tratan de eliminarla. Por otro lado, los individuos pueden presentar ciertas características comunes en sus respuestas, que permitan intentar **su clasificación en grupos** de cierta homogeneidad (José et al., 2004).

En este trabajo se utilizaron técnicas estadísticas como ACP, con el objetivo de disminuir dimensionalidad de los datos y obtener un plano de control donde se ubicaron los pacientes sanos, pacientes chagas leve y pacientes chagas crónico. El centroide y discriminante canónico, se utilizaron para clasificar un nuevo paciente que ingresa al plano de control.

Modelo matemático ACP y clasificadores (José *et al.*, 2004); (César 2004).

$$X = \{x_{ij}\} \quad (4)$$

X matriz original $n \times p$, $i = 1 \dots n$ son las observaciones y $j = 1 \dots p$ y son las variables.

El propósito es hallar un subespacio de dimensión m , $m \leq p$

$$Z_1 = (z_1, z_2, \dots, z_n)^T = x_1 a_1 \quad (5)$$

Finalmente, $(S - II)a = 0$ a, I, S son vector propio, valor propio y covarianza respectivamente.

$$d_1 = \frac{1}{l} \sum_{k=1}^l x_k \quad d_2 = \frac{1}{l} \sum_{k=1}^l y_k \quad (6)$$

x_k y y_k son las componentes en x y y respectivamente de las secuencias.

$$C = (d_1, d_2) \quad \text{Centroide} \quad (7)$$

$$d = \sqrt{\sum_i (a_i - b_i)^2} \quad \text{Distancia euclidiana} \quad (8)$$

a_i , Secuencia a, y b_i Secuencia b.

3. RESULTADOS

La figura 2, muestra el nivel de descomposición utilizado para detectar la ubicación del complejo QRS y la duración del mismo. La tabla 1 ilustra la relación comparativa de la duración del complejo QRS encontrada en el dominio del tiempo y la detección en el dominio frecuencial. El algoritmo se ensayó en 100 complejos QRS, con error de 9%.

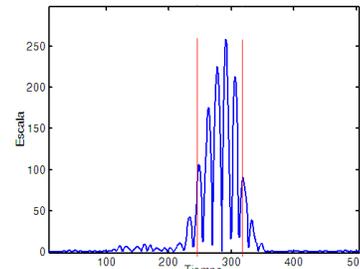
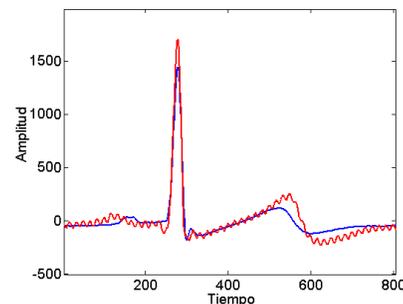


Fig. 2. Ubicación complejo QRS

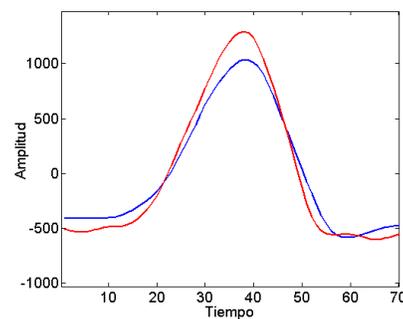
Tabla 1. Comparación duración ondas QRS

	Complejo QRS			
	1(ms)	2(ms)	3(ms)	Error
Detección	118±3	120±4	120±3	9%
Tiempo	117±5	120±4	120±2	
Frecuencia (TW)				

En la figura 3, a y b, se muestra el resultado al aplicar el algoritmo de alineamiento entre dos porciones de señal ECG, el alineamiento toma como punto clave el pico máximo de la onda R.



(a)



(b)

Fig. 3. Alineamiento de señales ECG

La figura 4, presenta el plano que se diseñó con el uso de la técnica ACP, este plano se utilizó para clasificar los diferentes pacientes: chagas leve, chagas crónico y normales. En este plano existen 3 regiones, la primera presenta sujetos normales (color amarillo), la segunda región representa sujetos chagas leve (color violeta) y la tercera región representa sujetos chagas crónico (color verde), la clasificación se desarrolló teniendo en cuenta la posición de un nuevo paciente y los resultados de los métodos de clasificación usados en este trabajo. La tabla 2, muestra algunos resultados totales que se han obtenido.

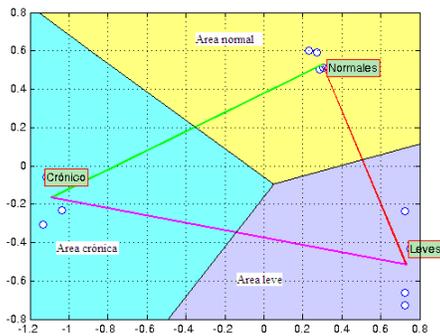


Fig. 4. Plano de clasificación

Tabla 2. Resultados de clasificación

	# pacientes	% clasificación	% error
Pacientes sanos	15	93.33	6.67
Pacientes leves	10	80	20
Pacientes crónicos	13	92.3	7.7
Total resultados	38	89.47	10.52

4. DISCUSIÓN Y CONCLUSIONES

Con el uso de la TW se pueden realizar análisis más profundos y más exigentes en el plano tiempo-frecuencia que proporciona esta técnica. El estudio e implementación de la TW, fue de gran ayuda para disminuir el tiempo de ejecución del trabajo total, ya que se pudo ubicar ciertas zonas de interés como lo fue la ubicación del inicio y del fin del complejo QRS. La programación dinámica proporciona características muy importantes si se logra utilizar de forma correcta o si se tiene claro que tipo de aplicación se debe utilizar para el desarrollo del proceso a trabajar. El alineamiento de secuencias es muy provechoso para comparar

información de cualquier tipo y conseguir el mayor grado de similitud entre los datos presentes, este grado de similitud se da ya que se obtienen secuencias de los datos originales de forma que sus correlaciones sean más altas. Los métodos multivariantes de datos son muy importantes para lograr clasificar ciertas clases de objetos (sujetos en nuestro caso), el método de clasificación dependerá de las características de los datos, y en ciertas ocasiones se da con la ayuda de las propiedades estadísticas de los datos a trabajar.

Los métodos multivariantes lineales no son muy adecuados para clasificar datos que provengan de sistemas fisiológicos como por ejemplo datos que muestran el comportamiento del sistema cardíaco, por esto se recomienda utilizar métodos de clasificación no lineales tales como métodos bayesianos, máquinas de vectores soporte y kernel-ACP, que están siendo estudiados actualmente, con el fin de mejorar los resultados obtenidos hasta la fecha.

AGRADECIMIENTOS

Al grupo de ingeniería biomédica de la Universidad de los Andes Mérida-Venezuela GIBULA.

REFERENCIAS

- Rasoul, M.J. de Boer, H. Suryapranata, Hoorntje, A. Gosselink, F. Zijlstra, J.P. Ottervanger, J.H.E. Dambrink, A.W.J. van 't Hof. (2007). Circumflex artery-related acute myocardial infarction: limited ECG abnormalities but poor outcome. Netherlands.
- Mark, A. Marinella, M. (1998). Electrocardiographic Manifestations and Differential Diagnosis of Acute Pericarditis.
- Bonnefoy, E. Godon, P. Kirkorian, G. Fatemi, M. Chevalier, P. and Touboul, P. (2000). Serum cardiac troponin I and ST-segment elevation in patients with acute pericarditis. Lyon, France.
- Mithilesh, K. Bilal, K. Sony, J. Awaneesh, K. Jo, M. (2007). Significance of a Fragmented QRS Complex Versus a Q Wave in Patients With Coronary Artery Disease.
- Florim, C. Richard, K. Paul, E. (2007). ECG changes following cardioversion and defibrillation. Department of Cardiology, Kantonsspital Luzern, Switzerland.

- Gan-Xin, Y. Charles, A. (1999). Cellular Basis for the Brugada Syndrome and Other Mechanisms of Arrhythmogenesis Associated With ST-Segment Elevation.
- Marta, V. Richard, R. and Ricardo, E. (2006). Use of an Immuno-chromatographic Dipstick Test for Rapid Detection of *Trypanosoma cruzi* in Sera from Animal Reservoir Hosts.
- Tomás, F. Jorge, A. María, C. Alonso, P. Luis, A. Nilda, G. Adriana, N. Adelina, R. and Horacio, A. (2006). Early detection of left ventricular diastolic dysfunction in Chagas' disease.
- Caryn, B. Susan, P. Barbara, L. Anis, R. Jose, A. Marin, N. James, H. Harry, A. Carlos, M. Louis, V. Robert, H. Pedro, A. Roberto, S. Anne C. (2007). Evaluation and Treatment of Chagas Disease in the United States.
- Zhang, D. (2005). Wavelet Approach for ECG Baseline Wander Correction and Noise Reduction.
- Laciar, E. Jané, R. and Brooks, D.H. (200). Beat-to-Beat Analysis of QRS Duration in Time and Wavelet Domains.
- Sidney, B. Ramesh, G. Haitao, G. (1998). Introduction to Wavelets and Wavelet Transforms A primer Texas, 200-532.
- Weichun, H. David, M. and Leping L. (2005). Accurate anchoring alignment of divergent sequences. USA.
- Diday, E. Lemaire, J. Pouget, J. and Testu, F. (1982). Elements d'analyse de données, Dunod. France. 116-375.
- Lebart, L. Morineau, A. Fénelon, and J-P. (1982). Traitement des données statistiques, Dunod. France. 273-373.
- José, H. José, R. and César F. (2004). Introducción a la minería de datos. PEARSON Printice-Hill. España. 166-252.
- César, P. (2004). Técnicas de análisis multivariante de datos aplicados con SPSS. PEARSON, Prentice-Hall. 121-153.