

# ECG Feature Extraction for Automatic Classification of Ischemic Events

G. Chávez, C. Rose, M. Serna, and O. Rodríguez

**Abstract**—A fundamental part of the analysis of electrocardiographic signals (ECG) lies in the transformation of the data or samples of the signal, with the aim of improving the effectiveness of the processing and the subsequent interpretation of the results obtained from it. Thus, a certain number of features are extracted, which are expected to have high discriminatory properties. In this way, the next step of the analysis must be the construction of a vector of features, which must be a set of descriptors that completely transmit the essence of the signal being studied. These features can focus on different aspects of the signal, for example, durations of intervals or segments, amplitudes of different wave peaks, deviations with respect to the isoelectric line, and areas under curves, among others. In this paper we present a method that includes the detection of temporal events of the ECG signal, such as ST segment, T wave, QRS complex and QT interval, several features are extracted using these events. The features are validated with the K-means algorithm and the feature vector is used as input in the ischemia classifier.

**Index Terms**—Feature extraction, Electrocardiogram (ECG), Ischemia, K-means clustering.

## I. INTRODUCCIÓN

DE acuerdo a estadísticas recopiladas por la Organización Mundial de la Salud (OMS), en el año 2015 el 54% de las muertes a nivel mundial ocurrieron a causa de cardiopatías isquémicas y accidentes cerebrovasculares. Lo anterior convierte a las enfermedades cardiovasculares en la principal causa de muerte a nivel mundial [1]. Tal como sucede en el escenario mundial, en México las enfermedades del corazón también están entre las principales causas de muerte. De acuerdo a datos del Instituto Nacional de Estadística y Geografía (INEGI), en el año 2016 se registraron un total de 136,342 muertes a consecuencia de enfermedades del corazón en el país; equivalentes al 23.66% del total de defunciones

Se agradece el apoyo al Consejo Nacional de Ciencia y Tecnología por la beca #446217 otorgada al primer autor. Asimismo, el apoyo del Tecnológico Nacional de México al proyecto de investigación clave 5753.16-P.

G. Chávez es estudiante de la División de Estudios de Posgrado e Investigación, Tecnológico Nacional de México / Instituto Tecnológico de Hermosillo, Hermosillo, México (e-mail: gilbertochavezlopez@gmail.com).

C. E. Rose es parte de la División de Estudios de Posgrado e Investigación, Tecnológico Nacional de México / Instituto Tecnológico de Hermosillo, Hermosillo, México (e-mail: crose@ith.mx).

M.T. Serna es parte de la División de Estudios de Posgrado e Investigación, Tecnológico Nacional de México / Instituto Tecnológico de Hermosillo, Hermosillo, México (e-mail: tserna@ith.mx).

O.M. Rodríguez es parte de la División de Estudios de Posgrado e Investigación, Tecnológico Nacional de México / Instituto Tecnológico de Hermosillo, Hermosillo, México (e-mail: omrodriguez@ieec.org).

ocurridas ese año; de estas muertes, 97,743 corresponden a enfermedades isquémicas del corazón [2].

Si bien las estadísticas por sí solas son alarmantes, existe otro factor crítico que lleva a poner particular atención en estos padecimientos: las muertes prematuras. Éstas son aquellas que ocurren en personas cuya edad aún se encuentra por debajo de la esperanza de vida. A nivel mundial se define el rango de edad para muerte prematura entre los 30 y los 69 años, aunque algunas fuentes lo extienden hasta los 75 años en países altamente desarrollados. Según datos de la Organización Panamericana de la Salud, un 46% de las muertes ocurridas a consecuencia de enfermedades cardiovasculares se pueden clasificar como prematuras [3].

El electrocardiograma (ECG) registra los patrones de activación eléctrica del corazón. En una gráfica de ECG típica se aprecia una línea guía, llamada línea isoelectrica o base, cuyas oscilaciones o deflexiones representan los latidos cardiacos. En conjunto estas oscilaciones forman ondas, segmentos e intervalos, comúnmente llamados ondas del ECG. En la Fig. 1 se muestran las principales morfologías que forman un latido cardiaco en el ECG [4][5][6].

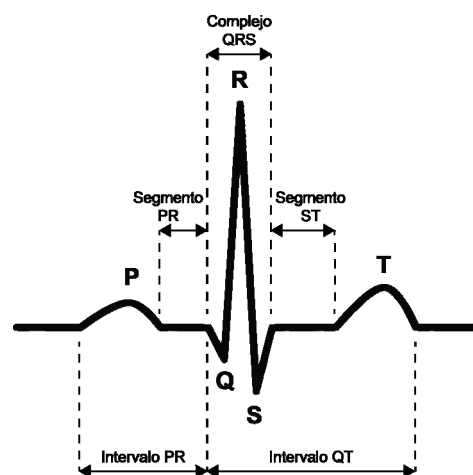


Fig. 1. Principales ondas, segmentos e intervalos del ECG.

El origen de este trabajo se enmarca en un proyecto cuyo objetivo es el desarrollo de un sistema inteligente de monitoreo para pacientes ambulatorios con antecedentes de anomalías cardiacas, de modo que sea posible detectar y predecir la ocurrencia, o reincidencia, de dichas anomalías.

En este artículo particularmente se describe el método para determinar las características correspondientes a eventos isquémicos, de tal manera, que se presenta el análisis de la señal de ECG para la selección de características a extraer.

Permitiendo caracterizar la señal, la cual es la entrada de un algoritmo clasificador para detectar eventos de anomalías isquémicas. En la sección 2 se presentan trabajos relacionados con la problemática en cuestión. La sección 3 presenta las características seleccionadas y la metodología propuesta para su extracción. La sección 4 presenta los resultados y discusión de un experimento exploratorio sobre la capacidad discriminativa de las características seleccionadas, asimismo, se presenta un clasificador binario. En la quinta y última sección se muestran las conclusiones.

## II. TRABAJOS RELACIONADOS

Se realizó una revisión bibliográfica con el fin de identificar las principales características utilizadas en la literatura para detectar eventos de isquemia. Los resultados más relevantes de dicha revisión se resumen en la Tabla I.

TABLA I  
PRINCIPALES CARACTERÍSTICAS UTILIZADAS EN LA LITERATURA PARA CARACTERIZAR SEÑALES DE ECG EN EL CONTEXTO DE DETECCIÓN DE ISQUEMIA

Descripción	Morfología de Interés	Usado en
Desviación del segmento ST	Segmento ST	[7][8][9][10][11][12][13][14][15]
Pendiente del segmento ST	Segmento ST	[8][12][14][16][17]
Forma de onda del intervalo ST-T	Segmento ST	[12][15]
Área del segmento ST	Segmento ST	[10][13][14][16]
Duración de Jpoint a pico T	Segmento ST	[13]
Función de energía isoelectrica del segmento ST	Segmento ST	[18]
Amplitud o nivel de la onda T	Onda T	[9][14][17][19]
Duración de la onda T	Onda T	[9][16][17][19]
Área de la onda T	Onda T	[17]

Algunas de las características mostradas en la Tabla I se evalúan respecto al tiempo, tales como la duración de intervalos, segmentos y ondas. En algunos casos dichas características se evalúan en conjunto con alguna otra en forma de ratios o proporciones. Generalmente las características de duraciones de morfologías se evalúan en conjunto con la duración del intervalo RR. El cual se refiere a la duración entre dos picos consecutivos de la onda R, para evitar perturbaciones en el análisis producto de cambios en la frecuencia cardiaca.

La Tabla I permite observar cuáles son las características más utilizadas en otros trabajos, siendo las principales la desviación y pendiente del segmento ST. También en el segmento ST se evalúan la forma de onda y el área bajo la curva. Asimismo, la onda T aparece como una morfología de interés en lo referente a su nivel, duración, área bajo la curva y características morfológicas. El complejo QRS es otra morfología sobre la que se calculan distintas características, en aspectos como duración, área bajo la curva y pendientes.

El objetivo de determinar las características a través del método propuesto en este artículo es obtener las características más relevantes, esto es, que nos permitan usarlas como entrada a un algoritmo de clasificación para la detección de isquemia. Entre los trabajos encontrados en la literatura se encuentra el

de Bhoi et al. [20], en el cual se exponen varios puntos muy relevantes, dado que desarrollan un modelo que busca atacar dos problemas a la vez, centrado tanto en la detección de arritmias como en la detección de isquemia. Comparan el desempeño de dos clasificadores, basados en árboles de decisión y una máquina de soporte vectorial, así mismo muestran una implementación de agrupamiento usando *K*-means.

Afsar et al. [21] presentan un método para la detección de episodios de desviación del segmento ST usando un conjunto de redes neuronales. Se detectan episodios de desviación del segmento ST usando un clasificador compuesto por un conjunto de redes neuronales con retropropagación. Asimismo, Hadjem et al. [22] predicen 7 anomalías del segmento ST y la onda T relacionadas con isquemia miocárdica. Se combinan técnicas de potenciación de árboles de decisión y sub-muestreo aleatorio para mejorar el desempeño y resolver el problema de desbalance de clases.

Por otra parte, Jeong et al. [15] proponen un algoritmo para la detección de cambios en el patrón del segmento ST. Tal como puntualizan los autores, resulta de especial importancia detectar cambios transitorios en el ECG ocasionados por episodios de isquemia miocárdica, los cuales en parte se reflejan electrocardiográficamente en el segmento ST. El algoritmo analiza la señal de ECG, detecta los cambios en el nivel del segmento ST y clasifica la forma del mismo.

La revisión bibliográfica de la que emanan estas observaciones permite tener un panorama más amplio sobre las características utilizadas para caracterizar señales electrocardiográficas en aplicaciones enfocadas a la detección de isquemia.

## III. MATERIALES Y MÉTODOS

### A. Características Seleccionadas para Evaluación y Análisis

Se determinó evaluar las características mostradas en la Tabla II. En subsecciones posteriores se describen los algoritmos utilizados para calcular algunas de las características seleccionadas.

TABLA II  
CARACTERÍSTICAS SELECCIONADAS PARA LA DETECCIÓN DE ISQUEMIA

Etiqueta	Descripción	Morfología de Interés
STdesv	Desviación del segmento ST con respecto a la línea isoelectrica	Segmento ST
STpend	Pendiente del segmento ST	Segmento ST
STform	Forma de onda del segmento ST	Segmento ST
Tamp	Amplitud o nivel de la onda T	Onda T
Tmorf	Morfología de la onda T	Onda T
QRSTrat	Ratio de duración QRS vs ST-T (%)	Intervalo QT
QRSup	Pendiente de subida del complejo QRS	Complejo QRS
QRSdw	Pendiente de bajada del complejo QRS	Complejo QRS
QTdur	Duración intervalo QT	Intervalo QT

### B. Conjunto de Datos para Experimentación

Se cuenta con la base de datos del intervalo ST-T de la Sociedad Europea de Cardiología (ESC), comúnmente llamada *European ST-T Database* o EDB, la cual está

disponible en el portal Physionet [23][24][25].

La EDB consta de 90 grabaciones de ECG de dos horas de duración con una frecuencia de muestreo de 250Hz, tomadas de 79 pacientes con sospecha o diagnóstico de isquemia miocárdica. Además, cada registro incluye anotaciones de referencia realizadas y contrastadas por 2 cardiólogos expertos. Estas anotaciones indican eventos de cambio en el intervalo ST característicos de isquemia, específicamente elevaciones y depresiones en el segmento ST y la onda T [23][24].

**C. Cálculo de la Desviación del Segmento ST**

La primera característica evaluada es la desviación del segmento ST, esta se refiere al nivel de este segmento con respecto a la línea isoeletrica de la señal.

Diversas fuentes especifican distintos valores de referencia para la línea isoeletrica, siendo los más comunes el nivel del segmento PQ [8][11][15], la región más plana del segmento PR [7][26][27], o el segmento TP [26][27]. Se seleccionó el segmento PQ como referencia para la línea isoeletrica dado que al explorar ejemplos del conjunto de datos se observó que este presenta menos variaciones comparado con las otras opciones. En la Fig. 2 se muestra el algoritmo general utilizado para calcular la desviación del segmento ST.

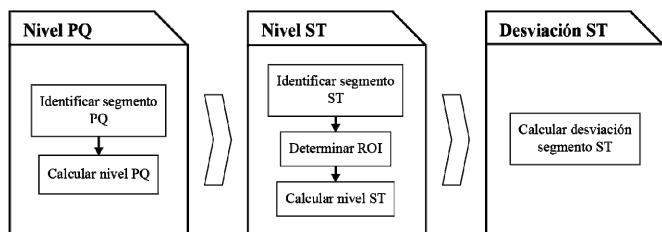


Fig. 2. Algoritmo general para el cálculo de desviación del segmento ST.

El primer paso es identificar el segmento PQ, para lo cual se calcula la media aritmética de la señal entre el final de la onda P y el inicio de la onda Q, y se toma dicho valor como el nivel isoeletrico de la señal.

Posteriormente se determina el nivel del segmento ST. La mayoría de los casos en la literatura toman como nivel del segmento ST un punto 80 ms después del J-point, que marca el inicio de dicho segmento, para casos de frecuencia cardiaca normal, o 60 ms en casos donde la frecuencia cardiaca supera los 90 latidos por minuto (lpm) [9][11][14][15].

La duración normal del segmento ST se estima en 120 ms [17]. Con esto en mente se decidió tomar como referencia para el nivel un punto ubicado a dos terceras partes de la duración del segmento ST, el cual denominamos punto ST. Para evitar perturbaciones, producto de variaciones en la señal, se decidió tomar el valor de una región de interés (ROI) en lugar de una lectura puntual. La ROI se define como 20 ms alrededor del punto ST. Se calcula la media aritmética de la señal en la ROI y se toma como nivel del segmento ST. Este procedimiento se ilustra en la Fig. 3.

Una vez obtenidos los niveles de los segmentos ST y PQ se comparan dichos valores para calcular la desviación del

segmento ST con respecto a la línea isoeletrica.

Adicionalmente, para uniformizar las mediciones entre latidos, se evalúa el valor de la desviación del segmento ST con respecto al valor pico de la onda R del latido respectivo.

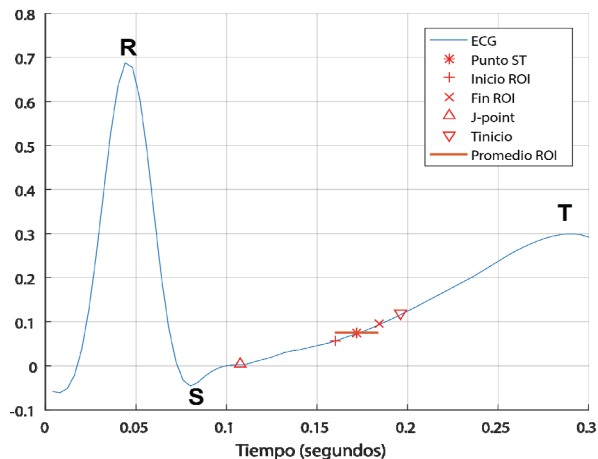


Fig. 3. Determinación de nivel del segmento ST.

**D. Cálculo de la Pendiente del Segmento ST**

Para calcular la pendiente del segmento ST en el punto ST se utiliza un método de aproximación polinomial para encontrar un segmento de recta. Dicha recta es, desde el punto de vista de mínimos cuadrados, la mejor aproximación a la señal en la ROI descrita anteriormente. Se divide la pendiente de dicha recta entre un factor  $\Delta t$ , correspondiente al incremento temporal de una muestra de la señal, y se toma este valor como la pendiente del segmento ST.

En la Fig. 4 se muestra un ejemplo de la recta aproximada para el cálculo de la pendiente del segmento ST.

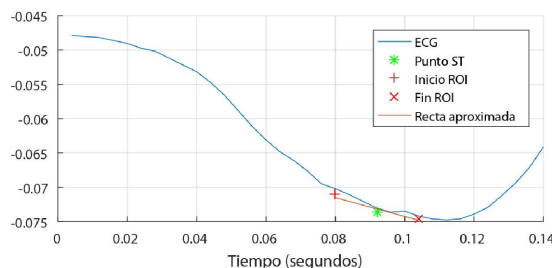


Fig. 4. Aproximación polinomial de una recta para el cálculo de la pendiente del segmento ST.

**E. Determinación de la Forma de Onda del Segmento ST**

Entre las características evaluadas también se tiene la forma del segmento ST, aunque propiamente se evalúa una sección que abarca el segmento ST y la onda T. Se considera la forma de la curva que se extiende desde el J-point hasta el pico de la onda T. De manera similar a lo mostrado en [12][15], se establecen 6 tipos de forma para el segmento ST, los cuales se muestran en la Fig. 5.

Para determinar el tipo de forma del segmento ST se desarrolló un algoritmo que produce los resultados mostrados en la Fig. 6. Primeramente se aísla el segmento de interés en la señal. Posteriormente se realiza un ajuste polinomial de curva

sobre estos datos para tener una curva con menos variaciones y representativa de la forma del segmento ST. Dicha curva es dividida, de forma automática, en 3 secciones, las cuales son: (1) del J-point a la mitad del segmento ST, (2) de la mitad del segmento ST al inicio de la onda T, y (3) del inicio de la onda T al pico de la misma.

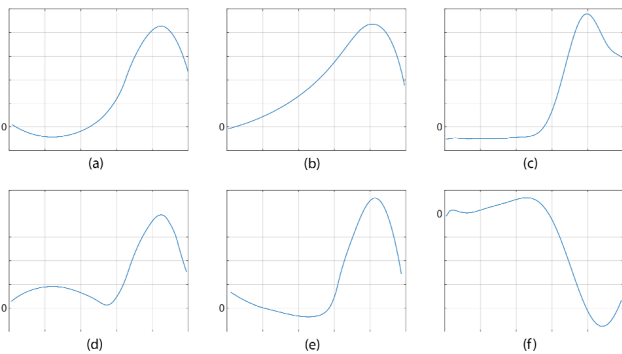


Fig. 5. Tipos de forma del segmento ST-T; (a) cóncavo, (b) pendiente ascendente, (c) horizontal, (d) convexo, (e) pendiente descendente y (f) onda T invertida.

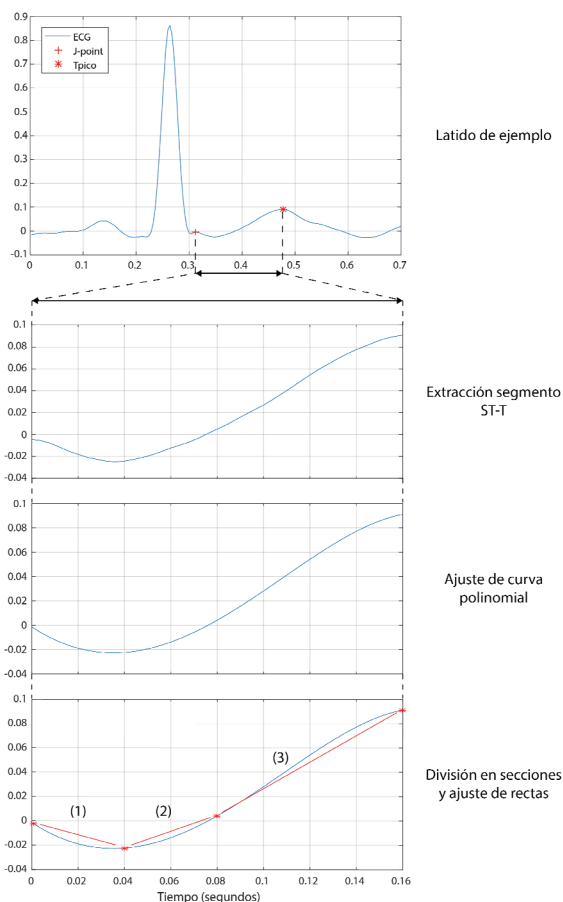


Fig. 6. Algoritmo para la determinación del tipo de forma del segmento ST-T.

Para cada sección se aproxima una recta y con base a la pendiente de la misma, y apoyándonos en un umbral, se infiere cómo es la pendiente que predomina en la sección respectiva del segmento ST, ya sea plana (0), pendiente

negativa (-1) o pendiente positiva (+1).

Con la pendiente de cada una de las tres secciones del segmento ST, y mediante el conjunto de reglas mostrado en la Tabla III, se determina a cuál de los tipos de forma establecidos corresponde cada latido.

TABLA III  
CONJUNTO DE REGLAS PARA DETERMINAR LA FORMA DE ONDA DEL SEGMENTO ST.

Tendencia de la pendiente			Tipo de forma
Segmento 3	Segmento 1	Segmento 2	
-1	-	-	Onda T invertida
+1	-1	-1	Pendiente descendente
+1	-1	0	Pendiente ascendente
+1	-1	+1	Cóncavo
+1	0	-1	Pendiente descendente
+1	0	0	Horizontal
+1	0	+1	Pendiente ascendente
+1	+1	-1	Convexo
+1	+1	0	Pendiente ascendente
+1	+1	+1	Pendiente ascendente

### F. Cálculo de la Duración del Intervalo QT

Distintos trabajos reportan evidencia de que durante eventos de isquemia se presenta un acortamiento en la duración del intervalo QT [13][28]. Lo que sugiere que esta característica puede aportar en el objetivo de conseguir una mejor diferenciación entre latidos normales y latidos isquémicos. En la Fig. 7 se muestra el intervalo QT en un latido de ejemplo de la EDB.

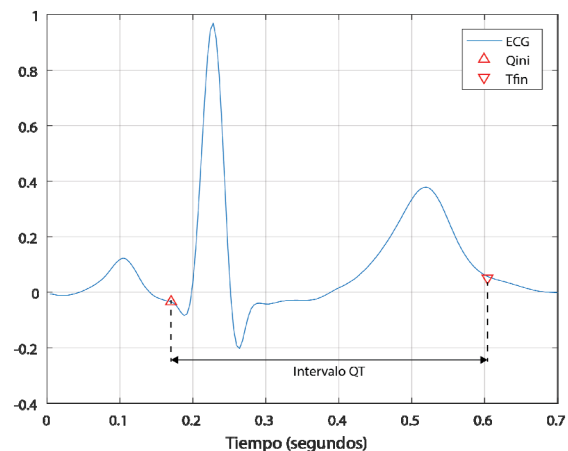


Fig. 7. Intervalo QT en un latido de ejemplo de la EDB.

Dado que el intervalo QT depende directamente de la frecuencia cardiaca es recomendable ajustarlo en base a esta para ser interpretado. Es así que se utiliza el intervalo QT corregido (QTc), el cual se calcula mediante la fórmula de Bazett, la cual se muestra en (1).

$$QTc = QT / (RR)^{1/2} \tag{1}$$

### G. Clasificador de Latidos Normales o Anormales Isquémicos

En esta sección se presenta un clasificador binario que permite detectar latidos normales y latidos anormales

isquémicos. Esta anomalía es obtenida en base a las características previamente definidas. Se considera esta clasificación como un primer paso en la detección de eventos isquémicos que se ilustra en la Fig. 8.

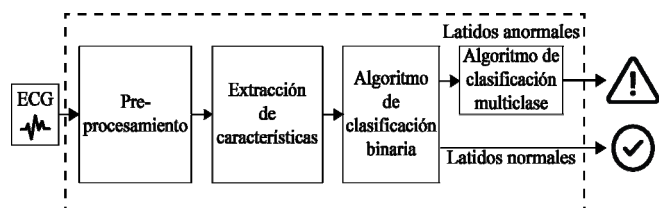


Fig. 8. Diagrama general del algoritmo propuesto para la detección de eventos isquémicos.

El clasificador binario se basa en una Máquina de Soporte Vectorial (SVM, acrónimo en inglés para *Support Vector Machine*). La idea básica de su funcionamiento es que dado un conjunto de entrenamiento, la SVM construye un hiperplano como una frontera de decisión que maximiza el margen de separación entre las observaciones positivas y negativas.

Se determinó usar un algoritmo basado en aprendizaje supervisado. Esto permite generar un modelo respaldado por un mapeo previo de los tipos de latidos que se busca identificar. Así como evaluar el desempeño del modelo para la identificación de dichos tipos de latidos.

A partir de la EDB se obtienen los archivos de señales de ECG digitalizadas y anotaciones. Los archivos de señales son procesados para extraer las características de la señal, sobre las que se realiza un proceso de remoción de valores atípicos u *outliers*. Posteriormente, se crea el vector de características para cada latido con la etiqueta correspondiente al tipo identificado mediante el archivo de anotaciones. El 70% de los datos se destinan a un conjunto de entrenamiento, el cual se usa para entrenar la SVM, y el 30% restante a un conjunto de validación, con el cual se evalúa el desempeño posterior al entrenamiento.

Para el entrenamiento de la SVM, se introdujeron los datos mediante una matriz de características y un vector de etiquetas de clase. Adicionalmente, se especificaron parámetros como la función kernel a utilizar, la cual se determinó que fuera de tipo Gaussiano, así como la estandarización de los datos de la matriz de características de entrada.

#### IV. RESULTADOS Y DISCUSIÓN

##### A. Validación de las Características Seleccionadas

Con el objetivo de validar la utilidad de las características seleccionadas en base a su capacidad discriminatoria, se realizaron experimentos sobre un caso particular del conjunto de datos. Se utilizó el archivo e0104 de la EDB, el cual de antemano se sabía que presentaba distintos eventos tanto de elevación del segmento ST como elevaciones y depresiones de la onda T.

Se tomaron dos intervalos de la señal donde se identificaron, mediante las anotaciones de la base de datos, un caso de control y un caso de elevación del segmento ST. Se obtuvieron un total de 223 latidos para llevar a cabo el

experimento, 113 del intervalo de control y 110 del intervalo con el evento de elevación del segmento ST.

Sobre los datos seleccionados se aplicaron los distintos algoritmos para el cálculo de características, con lo que se construyó el vector de características de cada latido, agrupando estos es una matriz conteniendo los datos respectivos para cada una de las 223 observaciones.

En la Fig. 9 se muestran los latidos utilizados, se grafican en tres dimensiones usando como ejes la desviación del segmento ST, el valor pico de la onda T y la pendiente de la onda T en el cruce por cero, adicionalmente, también se grafican en dos dimensiones utilizando como ejes la desviación del segmento ST y el valor pico de la onda T, con la intención de apreciar de una mejor manera los latidos correspondientes al evento de elevación del segmento ST.

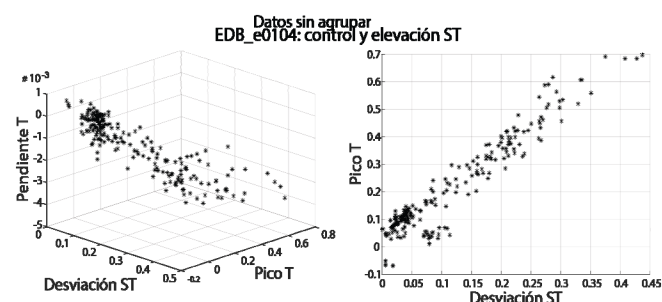


Fig. 9. Datos sin agrupar para dos intervalos del archivo EDB\_e0104.

Una vez construido el vector de características se utilizó el algoritmo de agrupamiento  $K$ -means, una técnica de aprendizaje no supervisado, con un número de clústeres  $k=2$  para identificar dos grupos y asignar cada latido a uno de ellos [29][30].

El objetivo de esto fue validar que las características seleccionadas fueran de utilidad para separar dos posibles grupos de latidos: normales y anormales. Se utilizó una técnica de aprendizaje no supervisado para mostrar que  $n$  número de características seleccionadas permiten a un algoritmo computacional identificar y separar los datos en dos grupos.

Inicialmente se alimentó el algoritmo  $K$ -means con solo tres características: la desviación del segmento ST, el nivel pico de la onda T y la pendiente de la onda T. Con dichas características se obtuvieron los clústeres y asignaciones que se muestran en la Fig. 10.

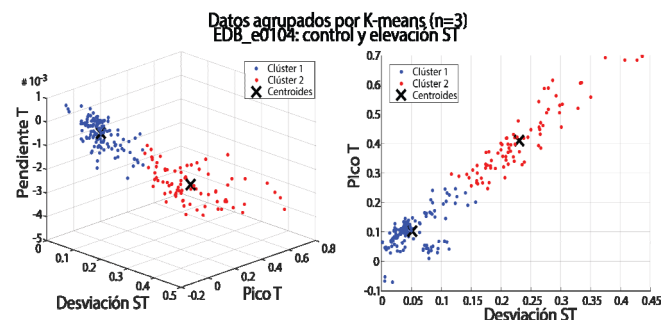


Fig. 10. Latidos del archivo EDB\_e0104 agrupados en dos clústeres mediante  $K$ -means.

Posteriormente se evaluó qué tan adecuadamente separados se encuentran los clústeres determinados por el algoritmo, y por consiguiente los datos asignados a cada clúster. Para esto se calculó para cada observación el coeficiente silueta, el cual, denotado por el término  $S$ , es una medida para validación de la consistencia en clústeres de datos. Evalúa la cohesión y separación de los datos; toma un valor entre  $-1$  y  $+1$ , donde un valor positivo alto indica que el elemento es muy similar a los demás elementos de su clúster y poco similar a los elementos de otros clústeres [30].

De esta forma, el coeficiente silueta de un conjunto de datos está dado por el valor promedio del coeficiente para todos los elementos que forman el conjunto. Por lo que si la mayoría de los elementos tienen valores positivos altos se interpreta que el agrupamiento por clústeres es adecuado.

En la Fig. 11 se muestran los resultados obtenidos del coeficiente silueta para el caso de  $K$ -means con un número de características  $n=3$ . Se puede apreciar que, a pesar de que existen algunos elementos con valores negativos en el clúster 2, la mayoría de los elementos en ambos clústeres presenta valores mayores a 0.8. En este caso se obtuvo un valor de coeficiente silueta promedio para el conjunto de datos de  $S=0.8296$ .

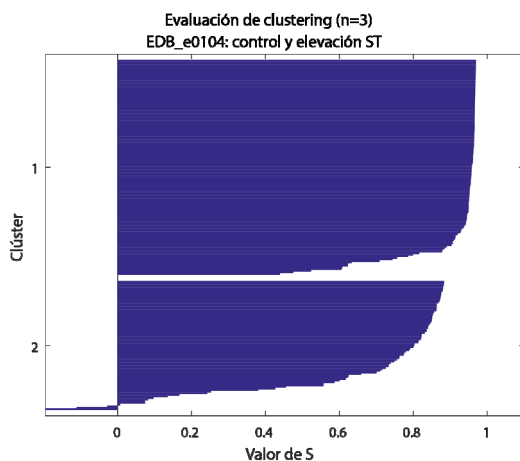


Fig. 11. Representación gráfica del coeficiente silueta para el agrupamiento en 2 clústeres mediante 3 características.

Posteriormente se probó el agrupamiento en 2 clústeres mediante el algoritmo  $K$ -means utilizando el resto de las características calculadas, agregando una característica a la vez al conjunto de datos proporcionado como entrada al algoritmo. Los resultados obtenidos para el coeficiente silueta del conjunto de datos para los casos de  $n = [3,11]$  características se resumen en la Tabla IV.

En la Tabla IV se puede apreciar que a medida que se agregan características, el coeficiente silueta disminuye en la mayoría de los casos, sin embargo, es importante recordar que este coeficiente no dice cuál de los grupos de características es el mejor, solo proporciona información sobre qué tan separados están los agrupamientos de datos, en este caso, los posibles latidos normales y anormales. En la Tabla IV se aprecia que el valor del coeficiente silueta se encuentra entre 0.6066 y 0.8296, lo cual es cercano a  $+1$ , que indica mayor

separación de los grupos de datos, lo cual asegura una buena clasificación como se discute en la siguiente sección.

TABLA IV  
COEFICIENTE SILUETA OBTENIDO CON DISTINTAS CANTIDADES DE CARACTERÍSTICAS UTILIZADAS EN EL ALGORITMO  $K$ -MEANS

n	Características Utilizadas	Coefficiente Silueta
3	STdesv, Tamp, Tpend	0.8296
4	STdesv, Tamp, Tpend, STpend	0.8296
5	STdesv, Tamp, Tpend, STpend, QRStrat	0.8009
6	STdesv, Tamp, Tpend, STpend, QRStrat, Tratdur	0.4892
7	STdesv, Tamp, Tpend, STpend, QRStrat, Tratdur, QRSup	0.7743
8	STdesv, Tamp, Tpend, STpend, QRStrat, Tratdur, QRSup, Tratarea	0.7649
9	STdesv, Tamp, Tpend, STpend, QRStrat, Tratdur, QRSup, Tratarea, QRSdw	0.6190
10	STdesv, Tamp, Tpend, STpend, QRStrat, Tratdur, QRSup, Tratarea, QRSdw, STform	0.6067
11	STdesv, Tamp, Tpend, STpend, QRStrat, Tratdur, QRSup, Tratarea, QRSdw, STform, QTdur	0.6066

### B. Validación del Clasificador Binario

Para el entrenamiento y validación del clasificador binario se construyó un conjunto de datos de 72,080 latidos, de los cuales 50,454 latidos corresponden al conjunto de entrenamiento. Los datos usados para validación del clasificador binario corresponden a un total de 21,626 latidos, de los cuales 50% son latidos normales y el 50% restante son latidos anormales o isquémicos.

Se determinó utilizar la totalidad de las características de la señal de ECG calculadas para conformar el vector de características de entrada del clasificador.

Se introdujeron los datos para validación al modelo y se obtuvieron las etiquetas de clase predichas para cada uno de los latidos presentes en el vector de características de entrada. Se compararon las etiquetas de clase predichas con las etiquetas reales y se creó la matriz de confusión mostrada en la Fig. 12. La etiqueta de clase "0" corresponde a latidos normales, mientras que la etiqueta "1" indica latidos anormales o isquémicos.

		Clase predicha		
		0	1	
Clase real	0	10114 46.77%	699 3.23%	93.54% 6.45%
	1	445 2.06%	10368 47.94%	95.88% 4.12%
		95.79% 4.21%	93.68% 6.32%	94.71% 5.29%

Fig. 12. Matriz de confusión para la validación del modelo de clasificación.

## V. CONCLUSIONES

Mediante los resultados obtenidos al utilizar el algoritmo de agrupamiento  $K$ -means, podemos inferir que las características calculadas poseen propiedades discriminativas que permiten una separación aceptable de los datos. Especialmente si tenemos en consideración el hecho de que esta es una técnica de aprendizaje no supervisado.

Es importante resaltar que no se considera el coeficiente silueta como una medida para determinar cuál es el mejor conjunto de características a utilizar, sino como una medida exploratoria para evaluar si las características seleccionadas permiten obtener una separación de los datos, especialmente dado que no se usa en conjunto con un método de validación que permita evaluar si los clústeres determinados efectivamente separan los latidos en normales y anormales.

Adicionalmente, los resultados obtenidos en la validación del clasificador binario implementado, nos permiten considerar que la técnica de clasificación basada en aprendizaje supervisado consigue obtener un porcentaje de aciertos elevado al clasificar latidos para diferenciarlos entre normales y anormales isquémicos.

Con esta primera clasificación binaria se está en posibilidades de al menos generar una alerta en un sistema de monitoreo de pacientes con anomalías cardíacas. Esto da pie a proponer, como trabajo futuro, un clasificador multiclase que permitirá detectar eventos específicos de isquemia miocárdica y así proporcionar mayor información.

## REFERENCIAS

- [1] Organización Mundial de la Salud, "Las 10 principales causas de defunción," March 2018 [Online]. Available: <http://www.who.int/mediacentre/factsheets/fs310/es/>. [Accessed: May 14, 2018]
- [2] Instituto Nacional de Estadística y Geografía, "Mortalidad," [Online]. Available: <http://www.beta.inegi.org.mx/proyectos/registros/vitales/mortalidad/>. [Accessed: May 14, 2018]
- [3] Organización Panamericana de la Salud, "Prevención de las enfermedades cardiovasculares: Directrices para la evaluación y el manejo del riesgo cardiovascular," 2007 [Online]. Available: [http://www.paho.org/hq/index.php?option=com\\_docman&task=doc\\_view&gid=13815&Itemid=270&lang=en](http://www.paho.org/hq/index.php?option=com_docman&task=doc_view&gid=13815&Itemid=270&lang=en). [Accessed: May 14, 2018]
- [4] A. López Farre and C. Macaya Miguel, *Libro de la salud cardiovascular del Hospital Clínico San Carlos y de la Fundación BBVA*, 1ra ed. España: Ed. Fundación BBVA, 2009.
- [5] J. E. Hall, *Tratado de fisiología médica*, 13ra ed. Italia: Elsevier España, 2016.
- [6] V. Fuster, R. A. Walsh, and R. A. Harrington, *Hurst's The Heart*, 13ra ed. New York: McGraw-Hill Medical, 2011.
- [7] R. Magrans, P. Gomis, and P. Caminal, "Myocardial ischemia event detection based on support vector machine model using QRS and ST segment features." in *2016 Computing in Cardiology Conference (CinC)*, Vancouver, BC, 2016, pp. 405-408.
- [8] S. Thulasi Prasad and S. Varadarajan, "Classification of ST-segments in ECG using ICA and Triangle method" in *2016 International Conference on Communication and Signal Processing (ICCS)*, Melmaruvathur, 2016, pp. 2119-2123.
- [9] M. Hadjem, F. Naït-Abdesselam, and A. Khokhar, "ST-segment and T-wave anomalies prediction in an ECG data using RUSBoost" in *2016 IEEE 18th International Conference on e-Health Networking, Applications and Services (Healthcom)*, Munich, 2016, pp. 1-6.
- [10] R. Twerenbold, R. Abächerli, T. Reichlin, S. Osswald, and C. Müller, "Early diagnosis of acute myocardial infarction by ST-segment deviation score" in *2012 Computing in Cardiology*, Krakow, 2012, pp. 657-660.
- [11] J. Presedo et al., "Detection of significant variations of the ST segment for the determination of Ischemic episodes," in *1992 14th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Paris, France, 1992, pp. 540-541.
- [12] G. Y. Jeong and K. H. Yu, "Development of ambulatory ECG monitoring device with ST shape classification" in *2007 International Conference on Control, Automation and Systems*, Seoul, 2007, pp. 1591-1595.
- [13] L. Maršánová et al., "ECG features and methods for automatic classification of ventricular premature and ischemic heartbeats: A comprehensive experimental study" *Nature Scientific Reports*, 2017. doi: 10.1038/s41598-017-10942-6.
- [14] Y. L. Tseng, K. S. Lin, and F. S. Jaw, "Comparison of Support-Vector Machine and Sparse Representation Using a Modified Rule-Based Method for Automated Myocardial Ischemia Detection" *Computational and Mathematical Methods in Medicine*, pp. 1-8, 2016.
- [15] G. Y. Jeong and K. H. Yu, "Design of Ambulatory ECG Monitoring System to detect ST pattern change" in *SICE-ICASE International Joint Conference 2006*, Busan, 2006, pp. 5873-5877.
- [16] C.A. Bustamante, S.I. Duque, A. Orozco-Duque, and J. Bustamante, "ECG delineation and ischemic ST-segment detection based in wavelet transform and support vector machines" in *2013 Pan American Health Care Exchanges (PAHCE)*, Medellín, 2013, pp. 1-7.
- [17] A. Gacek and W. Pedrycz, *ECG Signal Processing, Classification and Interpretation: A Comprehensive Framework of Computational Intelligence*. London: Springer, 2012.
- [18] A. Kumar and M. Singh, "Ischemia detection using Isoelectric Energy Function", *Computers in Biology and Medicine*, vol. 68, pp. 76-83, 2016.
- [19] F. Simon, P. Laguna, and E. Pueyo, "Study of the dynamic relationship between T wave morphology and heart rate during ischemia" in *2007 Computers in Cardiology*, Durham, NC, 2007, pp. 297-300.
- [20] A. K. Bhoi, K. S. Sherpa, and B. Khandelwal, "Arrhythmia and ischemia classification and clustering using QRS-ST-T (QT) analysis of electrocardiogram", *Cluster Computing*, vol. 21, pp. 1033-1044, 2018.
- [21] F. A. Afsar and M. Arif, "Detection of ST Segment Deviation Episodes in the ECG using KLT with an Ensemble Neural Classifier" in *2007 International Conference on Emerging Technologies*, Islamabad, 2007, pp. 11-16.
- [22] M. Hadjem, F. Naït-Abdesselam, and A. Khokhar, "ST-segment and T-wave anomalies prediction in an ECG data using RUSBoost" in *2016 IEEE 18th International Conference on e-Health Networking, Applications and Services (Healthcom)*, Munich, 2016, pp. 1-6.
- [23] PhysioNet, "European ST-T Database," Oct 2016 [Online]. Available: <https://physionet.org/physiobank/database/edb/>. [Accessed: May 14, 2018]
- [24] A. Taddei et al., "The European ST-T database: standard for evaluating systems for the analysis of ST-T changes in ambulatory electrocardiography" *European Heart Journal*, vol. 13, no. 9, pp. 1164-1172, Sep 1992.
- [25] A. L. Goldberger et al., "PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals" *Circulation*, vol. 101, no. 23, pp. e215-e220, June 2000.
- [26] E. B. Hanna and D. L. Glancy, "ST-segment depression and T-wave inversion: Classification, differential diagnosis, and caveats" *Cleveland Clinic Journal of Medicine*, vol. 78, no. 6, pp. 404-414, June 2011.
- [27] P. M. Rautaharju, B. Surawicz, and L. S. Gettes, "AHA/ACCF/HRS Recommendations for the Standardization and Interpretation of the Electrocardiogram: Part IV: The ST Segment, T and U Waves, and the QT Interval" *Journal of the American College of Cardiology*, vol. 53, no. 11, pp. 982-991, 2009.
- [28] R. Zayas Molina, R. E. Díaz Garriga, and M. Dorantes Sánchez, "Dispersión del intervalo QT: Un predictor de arritmias ventriculares malignas" *Revista Cubana de Cardiología y Cirugía Cardiovascular*, vol. 14, no. 2, pp. 116-123, 2000.
- [29] J. T. Palma Méndez and R. Marín Morales, *Inteligencia Artificial: Métodos, técnicas y aplicaciones*, 1ra ed. España: McGraw-Hill, 2008.
- [30] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*, 3ra ed. Estados Unidos de América: Elsevier, 2012.



**Gilberto Chávez López** es Ingeniero en Mecatrónica por la Universidad de Sonora (2013). Maestro en Ciencias de la Computación por el Tecnológico Nacional de México / Instituto Tecnológico de Hermosillo (2018).



**César Enrique Rose Gómez** es Ingeniero Industrial en Electrónica (1978) y Maestro en Ciencias en Ingeniería Electrónica (1990) por el Instituto Tecnológico de Chihuahua, es Maestro en Ciencias de la Computación (1996) por New Mexico State University en USA. Actualmente es profesor investigador en el Tecnológico Nacional de México / Instituto

Tecnológico de Hermosillo.



**María Trinidad Serna Encinas** es Licenciada en Informática (1992) por el Instituto Tecnológico de Hermosillo, Maestra en Ciencias en Ciencias Computacionales (2000) por el Instituto Tecnológico de Nogales. Es Doctora en Informática (2005) por la Universidad Joseph Fourier de Grenoble, Francia. Actualmente es profesora investigadora en

el Tecnológico Nacional de México / Instituto Tecnológico de Hermosillo.



**Oscar Mario Rodríguez Elías** es Ingeniero en Sistemas Computacionales (1999) por el Instituto Tecnológico de Hermosillo, Maestro en Ciencias de la Computación (2003) y Doctor en Ciencias de la Computación (2007) por el Centro de Investigación Científica y de Educación Superior de Ensenada (CICESE) en México. Actualmente es

profesor investigador en el Tecnológico Nacional de México / Instituto Tecnológico de Hermosillo.