

# Proposal of Asymmetric Multi-classifier of Arrhythmias

Luis Alejandro Mora

Laboratorio de Instrumentación, Control y Automatización

Universidad Nacional Experimental del Táchira

San Cristóbal, Venezuela

lmora@unet.edu.ve

Jhon Edgar Amaya

Laboratorio de Computación de Alto Rendimiento

Universidad Nacional Experimental del Táchira

San Cristóbal, Venezuela

jedgar@unet.edu.ve

**Abstract**—This paper presents a new methodology for the development of multi-classifiers SVM with One-Against-One (OAO), which allows each node to use different features or attributes to differentiate each pair of classes, called asymmetric OAO-SVM. We evaluated this method by developing a classification system to identify four types of arrhythmias (Atrial Fibrillation, Atrial Flutter, Supraventricular Tachyarrhythmia and Ventricular Tachycardia) and Normal ECG, using non-linear characteristics such as Shannon entropy and Lempel-Ziv complexity. This method presents a positive prediction of 90.72% which represent an improve with respect a typical multi-classifier OAO-SVM.

**Keywords**—Cardiac Arrhythms, Shannon entropy, Lempel-Ziv complexity, Support vector machines.

## I. INTRODUCCIÓN

En los procesos de clasificación e identificación patrones, es necesario extraer de las señales u objetos a clasificar, ciertos parámetros característicos o atributos que permitan diferenciar entre las diversas clases existentes. En la literatura existen diferentes propuestas de algoritmos para realizar el proceso de clasificación, como por ejemplo, árboles de decisión, redes bayesianas, máquinas de soporte vectorial, entre otros [32]. Las máquinas de soporte vectorial (SVM) son una herramienta muy utilizada en este tipo de procesos, debido a su robustez y la eficiencia presentada en la solución de problemas en diferentes dominios [4, 16].

Se ha mostrado que un solo esquema de clasificación generalmente no presenta buenos resultados en problemas complejos, por lo que se han propuesto los denominados multclasificadores. Se pueden definir los multclasificadores con un conjunto de clasificadores diferentes que realizan predicciones que se obtienen como resultado la combinación de cada una de ellos. En general, cuando existen 3 o más clases el problema de multclasificación se divide en  $k$  problemas binarios o nodos. La construcción de estos nodos depende de la técnica desarrollada, sin embargo, como regla general todos los nodos se entrenan y prueban con todas las características o atributos extraídos de los objetos a identificar [1–6]. El problema fundamental radica en que no todas las características aportan suficiente información para diferenciar entre cada par de clases de cada nodo, es decir, cada nodo puede requerir de solo un subconjunto de características necesarias para realizar

la diferenciación entre un par de clases, descartando así el uso de los demás atributos.

Dado que la detección de arritmias cardíacas se considera un problema importante y complejo, se desarrolló un multclasificador de arritmias cardíacas basado en SVM con la técnica llamada *uno-contra-uno* (OAO por las siglas en inglés de *One-Against-One*) donde cada nodo es construido usando las características de forma asimétrica para diferenciar entre señales ECG normales, fibrilaciones auriculares (AFIB), *flutters* auriculares (AFL), taquiarritmias supraventriculares (SVTA) y taquicardias ventriculares (VT), a partir de los registros encontrados en la base de datos del MIT-BIH [7]. Para ello se extrajeron de esta base de datos 2124 segmentos, con una longitud de 720 muestras cada uno, los cuales fueron preprocesados utilizando los filtros pasa-bajo y pasa-alto diseñados por Apaclicia [8]. Se realiza además, el análisis experimental correspondiente para determinar la efectividad de la propuesta de multclasificación.

El artículo está organizado como sigue. En la Sección II se muestran los métodos utilizados para la extracción de características del ECG, así como los métodos de codificación respectivos. En la Sección III muestra el proceso de clasificación mediante SVM y la estrategia OAO. La Sección IV se muestra la propuesta de la multclasificación asimétrica basada en SVM. Los resultados se muestran en la Sección V. Finalmente se presentan las conclusiones y el trabajo futuro.

## II. CARACTERÍSTICAS

Para el desarrollo del sistema de clasificación, se extrajeron las características no lineales a partir de cada segmento, para ser más específicos: la entropía de Shannon y la complejidad de Lempel-Ziv. Para entender mejor como se calculan, consideremos que se tiene un segmento  $S$  de longitud  $n$ , el cual está formado por palabras del alfabeto  $A$  y de longitud  $\alpha=|A|$ .

### A. Entropía Shannon

También conocida como la entropía de información. La entropía Shannon (ESH) es una medida matemática de la falta de información asociada a una distribución de probabilidad de un conjunto de datos [9]. Fue definida en el año 1948 por C.E. Shannon [10] como la sumatoria de una función basada en la

probabilidad de ocurrencia de un dato, como se observa en (1), siendo  $p_i$  la probabilidad de que al extraer de forma aleatoria un elemento del segmento  $S$  sea igual a la  $i$ -ésima palabra del alfabeto  $A$ .

$$H = -\sum_{i=1}^{\alpha} p_i \log_2(p_i) \quad (1)$$

La entropía escero cuando  $S$  es una señal constante, es decir, no hay información, y es máxima cuando es una señal aleatoria, es decir, posee la mayor información posible. Para efectos de comparación entre secuencias que pudieran tener diferentes alfabetos se hace necesaria la normalización de la entropía  $H$ . La normalización se logra dividiendo  $H$  entre el máximo valor posible que puede tomar la entropía, cuyo valor está determinado por el logaritmo de la longitud del alfabeto que compone el segmento  $S$ . Por lo tanto, la entropía de información normalizada se define como,

$$Hn = \frac{-\sum_{i=1}^{\alpha} p_i \log_2(p_i)}{\log_2(\alpha)} \quad (2)$$

### B. Complejidad de Lempel-Ziv

Existen varias técnicas para medir la complejidad de una secuencia, una de ellas es la complejidad de Lempel-Ziv (CLZ), la cual es muy utilizada en la criptografía y compresión de datos [11]. En esta medición de complejidad—descrita como  $c(S)$ —, se explora la secuencia  $S$ , leyendo sus elementos de izquierda a derecha y se agrega una nueva palabra a una memoria cada vez que se encuentra una sub-secuencia de dígitos consecutivos que no ha sido encontrada previamente. Se utiliza como parámetros para medir la mencionada complejidad: el tamaño del alfabeto y el número de palabras encontradas en  $S$ , es decir, la complejidad  $c(S)$  está determinada por el menor número posible de sub-secuencias necesarias para construir  $S$  [12]. Para el cálculo de la CLZ, se utilizó el algoritmo presentado por Borowska et al. [13]. De igual manera que con la entropía, el valor de la complejidad depende del alfabeto y de la longitud de la secuencia, por lo que para comparar secuencias con diferentes alfabetos y longitudes, muchos autores normalizan la CLZ, basados en el teorema presentado en [12], el cual indica que la complejidad de una secuencia  $c(S)$  es menor a :

$$c(S) < \frac{n}{(1-\varepsilon_n) \log_{\alpha}(n)} \quad (3)$$

donde

$$\varepsilon_n = 2^{\frac{1+\log_{\alpha}(\log_{\alpha}(an))}{\log_{\alpha}(n)}} \quad (4)$$

El termino derecho en (3) llega a su máximo valor  $b(S)$  cuando  $\varepsilon_n \rightarrow 0$ , lo cual ocurre cuando  $n \rightarrow \infty$ .

$$b(S) = \frac{n}{\log_{\alpha}(n)} \quad (5)$$

Por lo tanto, la complejidad normalizada  $C(S)$  se suele calcular mediante la expresión que se muestra en (6).

$$C(S) = \frac{c(S)}{b(S)} \quad (6)$$

### C. Métodos de codificación

Como se puede observar las características son dependientes de la representación, i.e., de la longitud del alfabeto asociado a esta característica. Sin embargo, las señales electrocardiográficas son señales análogas adquiridas con una resolución de  $n$  bits, normalmente entre 10 y 12 bits, por lo cual el tamaño del alfabeto de estas señales puede llegar hasta  $2^n$  elementos, implicando una mayor complejidad en el cálculo de las características. Por lo tanto, se hace necesario realizar una codificación previa de la señal con la finalidad de simplificar el cálculo. En la investigación se utilizaron los métodos de la pendiente y del valor umbral para generar señales codificadas con alfabetos binarios y ternarios. En el caso del método del valor umbral, se utilizó el valor medio de la señal más un factor de desviación  $E$ , para establecer los límites de la codificación.

### D. Evaluación de las distribuciones

Como se menciona en [14], se hace necesario evaluar cual de los métodos de codificación descritos anteriormente genera la mayor separación, para facilitar la clasificación de los datos, para ello se utilizó la función  $\lambda$ , la cual describimos a continuación. Sea  $D$  un conjunto de datos distribuidos en  $m$  clases, de forma tal que  $D = \{C_1, C_2, \dots, C_m\}$ , donde cada clase  $C_i$  contiene  $\eta_i$  elementos  $C_i = \{e_{i1}, e_{i2}, \dots, e_{i\eta_i}\}$ , y existe un centroide  $\Phi_i$  para cada clase.

*Lemma 1:* Se puede evaluar la distribución de los elementos del conjunto  $D$  a partir de la relación de la distancia existente entre cada elemento  $e_{ij}$  hasta el centroide  $\Phi$  de su propia clase, respecto a su distancia hasta los centroides de las demás clases. Considerando un centroide como el centro espacial de la distribución de los elementos de una clase, el centroide de la clase  $C_i$  estará determinado por

$$\Phi_i = \frac{\sum_{j=1}^{\eta_i} e_{ij}}{\eta_i} \quad (7)$$

*Definición 1:* Sea  $l_i$  la distancia existente entre un elemento  $e_{ij}$  y el centroide  $\Phi_i$  de su propia clase, y sea  $l_h$  la distancia entre dicho elemento y el centroide  $\Phi_h$ , para  $h = 1, 2, \dots, m$  con  $h \neq i$ . Se entiende por la función de distribución  $\lambda_i$  de una clase como el número de elementos de la clase cuya distancia a su centroide sea mayor o igual a la distancia respecto a algún centroide de otra clase. Es decir,

$$\lambda_i = \sum_{j=1}^{\eta_i} \psi(e_{ij}) \therefore \mathbb{E}\psi(e_{ij}) = \begin{cases} 1, l_i \geq l_h, \forall h \neq i \\ 0, \text{caso contrario} \end{cases} \quad (8)$$

La función de distribución  $\lambda_i$  se anula, cuando todos los elementos de la clase  $C_i$  están agrupados cerca de su centroide, y es máxima cuando existe una superposición o solapamiento total entre 2 o más clases. El máximo valor que puede tomar la función de distribución de una clase es igual al número de elementos que conforman dicha clase, por lo tanto la distribución normalizada de una clase está dada por (9).

$$\lambda_{in} = \frac{\lambda_i}{\eta_i} \quad (9)$$

*Definición 2:* Se entiende como la distribución de un conjunto de datos  $\lambda_D$  como la sumatoria de las distribuciones de cada clase que lo conforman, como se observa en (10).

$$\lambda_D = \sum_{i=1}^m \lambda_i \quad (10)$$

**Definición 3:** Se entiende como la distribución normalizada de un sistema  $\lambda_{DN}$  como la relación entre la distribución del sistema y el número total de elementos que lo conforman.

$$\lambda_{DN} = \frac{\lambda_D}{N} \therefore N = \sum_{i=1}^m \eta_i \quad (11)$$

**Definición 4:** Se entiende como la distribución promedio de un conjunto de datos  $\lambda_{DP}$  como la relación entre la sumatoria de las distribuciones normalizadas de las clases que conforman el sistema y el número total clases.

$$\lambda_{DP} = \frac{\sum_{i=1}^m \lambda_{in}}{m} \quad (12)$$

### III. MÉTODO OAA-SVM

Como herramienta para la clasificación de los datos se utilizó máquinas de soporte vectorial (SVM), las cuales son un tipo específico de algoritmos basados en la teoría del aprendizaje estadístico, presentados por Boser, Guyon&Vapnik [15]. Las SVM buscan maximizar el margen entre los patrones de entrenamiento y los límites de decisión, siendo una de las técnicas de clasificación más utilizada en los últimos años, debido a su versatilidad y excelentes resultados obtenidos en diferentes tipos de aplicaciones [16]. Esta técnica representa los datos en un espacio vectorial, donde cada clase de datos se agrupan en regiones separables del espacio de representación, con lo cual se trata de buscar un hiperplano que separe a cada clase y que maximice la distancia entre los datos y el propio hiperplano, lo cual se denomina clasificador de máximo margen, donde el hiperplano está definido por (13)

$$f(x) = \omega \cdot x + b \quad (13)$$

La optimización de (13) consistiría en obtener los valores de  $\omega$  y de  $b$  que maximicen los márgenes entre el hiperplano y los datos.

$$\min_{\omega, b} \frac{1}{2} \|\omega\|^2 + \gamma \sum_{i=1}^n \xi_i \quad (14)$$

para  $y_i \cdot (\omega \cdot x_i + b) \geq 1 - \xi_i$ , donde  $\xi_i \geq 0, \forall i$ , donde  $\xi_i$  es la distancia entre el dato y el hiperplano,  $\gamma$  es conocida como la constante de penalización de error y  $n$  el número de datos a clasificar.

Para el proceso de clasificación se utilizó la técnica *uno-contra-uno* (OAO)[17-18]. Este método de multi-clasificación divide el problema de clasificar  $m$  clases en  $k$  clasificadores binarios o nodos, donde cada nodo está diseñado para diferenciar solo entre 2 clases o pares, generando un voto solo para aquellas clases para las cuales fue entrenado. En la Figura 1 se puede observar la estructura típica de los clasificadores OAO, donde cada dato procesado, representado por el vector de atributos  $X$ , es pasado por todos los nodos y al final la función de pertenencia lo etiqueta con la clase que haya recibido más votos.

En el caso de existir un empate en el número de votos recibidos por 2 o más clases, se puede etiquetar el dato como desconocido, lo cual es conveniente si en la fase de pruebas

existen datos que no pertenecen a ninguna de las clases con las que ha sido entrenado el clasificador. Sin embargo, en caso de que las pruebas solo se utilicen datos de las clases con las que fue entrenado el multi-clasificador, se puede utilizar una variante de la función de pertenencia, de acuerdo a lo indicado por Abe [17]. Es decir, si consideramos las funciones de decisión primarias, también conocidas como predicciones numéricas parciales,

$$z_i = w_i^T x + b_i \quad (16)$$

donde  $w_i$  es el vector de pesos  $l$ -dimensional del nodo  $i$  y  $b_i$  es el termino *bias*.

En caso de existir un empate en las votaciones, la asignación de la etiqueta se realiza considerando como función de pertenencia la sumatoria de las predicciones numéricas parciales  $z_i$  de cada clase involucrada, de esta forma la identificación se realiza de acuerdo la clase, dentro de las que se encuentran empatadas, cuya función de pertenencia sea mayor.

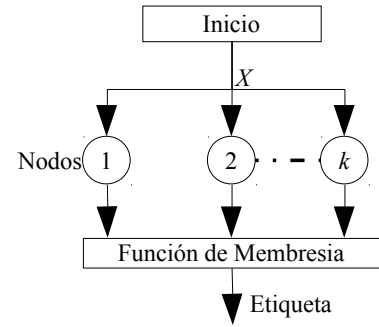


Figure 1. Estructura de un clasificador OAO típica.

### IV. MÉTODO OAO-SVM ASIMÉTRICO

Antes de describir el método propuesto, es importante definir el concepto de simetría asociado a la multclasificación, como se muestra a continuación.

#### A. Simetría

Se considera que un multi-clasificador es simétrico si cada nodo utiliza exactamente los mismos atributos para diferenciar entre el par de clases para el que fue entrenado. En la Figura 2.a se observa un gráfico de red para un multi-clasificador OAO desarrollado mediante las técnicas tradicionales. Este clasificador estaría diseñado para diferenciar entre 4 diferentes clases, utilizando 6 nodos, con un vector de atributos  $X$  de 12 dimensiones y cada nodo utiliza todos atributos. Se observa la perfecta simetría de los nodos en la gráfico.

Por lo tanto, se considera que un multi-clasificador es asimétrico si los nodos que lo conforman utilizan diferentes combinaciones de atributos. En la Figura 2.b se observa un multi-clasificador OAO asimétrico, que al igual al clasificador anterior está diseñado para identificar 4 clases de datos, sin embargo, cada nodo utiliza diferentes combinaciones de los atributos que se extraen en  $X$ , por lo que el gráfico de red es asimétrico.

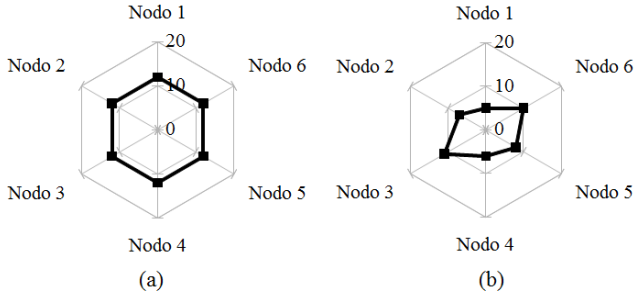


Figure 2. Estructura espacial de un clasificador OAO-SVM. En a) se muestra un clasificador simétrico. En b) se observa un clasificador asimétrico.

### B. Estructura

En la Figura 3 se observa la estructura del clasificador OAO asimétrico propuesto, donde un sistema de selección de características se encarga de entregarle a cada nodo solo las características necesarias para diferenciar entre el par de clases correspondiente. Este sistema se basa en una matriz de selección de características, donde a cada nodo le corresponden solo las características marcadas con un “1” en la fila correspondiente, descartando aquellas marcadas con un “0”. La Tabla I muestra un ejemplo de la matriz de selección de características para el clasificador OAO asimétrico que se observa en la Figura 2.b.

TABLA II: EJEMPLO DE MATRIZ DE SELECCIÓN DE CARACTERÍSTICAS

Nodos	Características											
	1	2	3	4	5	6	7	8	9	10	11	12
Nodo 1	1	1	0	1	0	0	1	0	0	1	0	0
Nodo 2	0	0	1	0	1	1	0	1	1	1	0	1
Nodo 3	1	1	1	1	1	1	1	1	1	0	1	1
Nodo 4	1	0	0	0	1	1	0	1	0	0	1	1
Nodo 5	1	1	1	1	0	1	1	0	0	1	0	1
Nodo 6	1	1	1	1	1	0	1	1	1	1	1	0

### C. Selección de las características

La correcta construcción de la matriz de características es de fundamental importancia para el funcionamiento del multi-clasificador asimétrico, ya que una selección errónea de las características correspondientes a cada nodo generaría una disminución en la eficiencia del multi-clasificador. Se proponen tres formas de realizar la selección de las características: por basamentos teóricos, de forma empírica o utilizando la función  $\lambda_{DP}$  para evaluar las distribuciones generadas por cada combinación de características.

En el caso de la técnica de basamentos teóricos, el investigador o desarrollador del multi-clasificador puede seleccionar las características para diferenciar cada par de clases, a partir de sus conocimientos sobre los datos a procesar, opiniones de expertos y/o consulta de referencias bibliográficas.

Utilizado la forma empírica, la selección de las características se realiza probando cada una de las combinaciones para seleccionar aquellas que maximicen el rendimiento de cada nodo.

Si consideramos que  $\lambda_{DP}$ , es una medida relativa del número de elementos que están más propensos a ser

clasificados de forma errónea, entonces la selección de las características se puede realizar a partir de aquellas combinaciones que menor  $\lambda_{DP}$  generen.

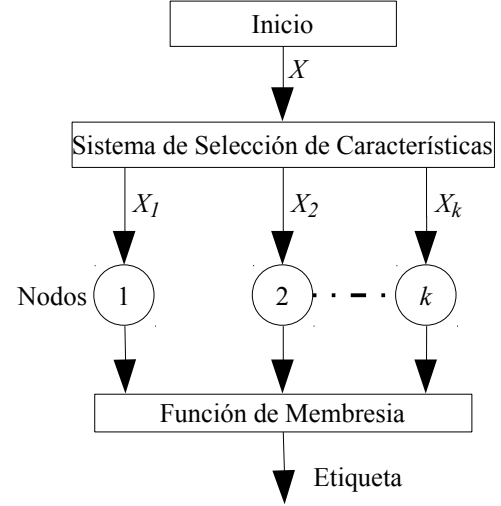


Figure 3. Estructura de un clasificador OAO Asimétrico

## V. RESULTADOS

Se extrajeron 2124 segmentos de base de datos del MIT-BIH, los cuales se distribuyen de la siguiente forma: 800 segmentos de ECG Normales, 325 segmentos de ECG con AFIB, 324 segmentos de ECG con AFL, 325 segmentos de ECG con SVTA y 350 segmentos de ECG con VT. Es importante destacar para los procesos de entrenamiento del multi-clasificador se seleccionaron de forma aleatoria el 30% de las muestras de cada clase, dejando el restante 70% para la fase de pruebas.

Para la extracción de las características se aplicaron cuatro métodos de codificación a las señales electrocardiográficas, estas codificaciones fueron: pendiente ternaria (PT), valor umbral binario (MB) con  $E=1/10$  y  $E=1/20$  y valor umbral ternario (MT) con  $E=1/10$ . La selección de estos métodos se basó en la separación entre clases para cada par de arritmias, según la función  $\lambda_{DP}$  de acuerdo a lo indicado en [14]. Extrayendo la entropía y complejidad de las secuencias generadas por cada uno de estos métodos se construye un vector de características para cada dato de acuerdo a la Tabla III.

La construcción de los clasificadores se realizó utilizando el *Toolbox LS-SVMLab* desarrollado por Suykens et al. [19], el cual contiene los algoritmos y funciones para la entonación, entrenamiento y simulación de multi-clasificadores con los kernels: lineal (Lin), polinomial (Poly) y función de base radial (RBF). Para evaluar la eficiencia de los diferentes clasificadores y realizar una comparación entre ellos, se utilizó como parámetros: la sensibilidad (SE), la especificidad (SP), la predicción de positivos (PP) y el porcentaje elementos clasificados correctamente, es decir, el rendimiento del clasificador (RE).

Inicialmente los multi-clasificadores OAO fueron entrenados y luego probados 100 veces, seleccionando al azar en cada ocasión las muestras para el proceso de entrenamiento

según la TablaIV, dejando el resto de las muestras para el proceso de pruebas. En el caso del multi-clasificador OAO simétrico se obtuvo una predicción de positivos global del clasificador de  $90.32\% \pm 0.71$  con un máximo de  $91.99\%$ . Sin embargo, de acuerdo a los datos recabados en la Tabla Vse puede observar la sensibilidad y la predicción de positivos para la fibrilación auricular y la taquicardia ventricular siguen siendo bajas.

TABLA VI: CONSTRUCCIÓN DEL VECTOR DE CARACTERÍSTICAS A PARTIR DE LOS PARÁMETROS EXTRAÍDOS

Método de Codificación	PT		MB E=1/10		MB E=1/20		MT E=1/10	
Parámetro Extraído	ESH	CLZ	ESH	CLZ	ESH	CLZ	ESH	CLZ
Característica #	1	2	3	4	5	6	7	8

TABLA VII: RESULTADOS OBTENIDOS USANDO UN MULTI-CLASIFICADOR OAO CON TODAS LAS CARACTERÍSTICAS

Arritmias	SE	SP	PP
Normal	99.03% $\pm 0.57$	98.12% $\pm 0.57$	96.95% $\pm 0.87$
AFIB	74.21% $\pm 3.80$	96.66% $\pm 0.61$	79.43% $\pm 2.60$
AFL	94.10% $\pm 1.80$	98.76% $\pm 0.34$	93.13% $\pm 1.72$
SVTA	89.40% $\pm 2.33$	98.43% $\pm 0.42$	91.05% $\pm 2.10$
VT	82.71% $\pm 2.82$	96.438% $\pm 0.76$	81.51% $\pm 3.00$
TOTAL	90.32% $\pm 0.71$	97.64% $\pm 0.17$	90.32% $\pm 0.71$

En el caso de los multi-clasificadores OAO asimétricos, se construyeron dos modelos. El primero a partir de la generación de la matriz de características mediante el método empírico y el segundo a partir de la función  $\lambda_{DP}$ . No se construyó una matriz a partir de bases teóricas debido a que la bibliográfica consultada [20–23], las arritmias están caracterizadas por parámetros temporales y morfológicos, y no se encontró documentación referente a las condiciones para identificar arritmias a partir de la ESH y la CLZ.

En la Tabla VIII se observa la matriz de características obtenidas para el primer clasificador asimétrico, donde el rendimiento de cada nodo es mayor 87%. Obteniendo una predicción de positivos global promedio del clasificador de  $90.73\% \pm 0.66$  con un máximo de  $92.32\%$ . Si bien el rendimiento global del este clasificador es similar al simétrico, se observa un aumento en la sensibilidad para las AFIB de 4.84% y la predicción de positivos de la VT en 1.23%, haciendo de este un clasificador más equilibrado de acuerdo los datos presentados en la Tabla IX.

El segundo multi-clasificador asimétrico la matriz de características se construyó a partir de la función  $\lambda_{DP}$ . En la Figura 4 se observa como a medida que disminuye la función  $\lambda_{DP}$  el rendimiento promedio del nodo tiende a aumentar, por lo que se puede interpretar que la combinación de características que generan un  $\lambda_{DP}$  bajo presentan en la mayoría de los casos un rendimiento mayor a aquellas combinaciones con un  $\lambda_{DP}$  alto.

La matriz de características construida a partir de la combinación de atributos con  $\lambda_{DP}$ , se observa en la Tabla X. Se obtuvo una predicción de positivos global promedio de

$83.44\% \pm 0.69$ , como se observa en la Tabla XI, con un máximo de  $85.25\%$ .

TABLA VIII: MATRIZ DE CARACTERÍSTICAS Y RENDIMIENTO DE CADA NODO, OBTENIDOS DE FORMA EMPÍRICA

Nodos	Características								RE(%)
	1	2	3	4	5	6	7	8	
Normal vs AFIB	1	1	1	0	1	0	1	0	98.32% $\pm 0.30$
Normal vs AFL	1	1	1	1	1	1	0	1	99.63% $\pm 0.21$
Normal vs SVTA	1	1	1	0	0	1	1	0	99.65% $\pm 0.28$
Normal vs VT	1	1	1	0	1	0	1	1	98.43% $\pm 0.32$
AFIB vs AFL	1	1	1	1	1	1	0	1	97.21% $\pm 0.77$
AFIB vs SVTA	1	1	1	0	1	1	0	0	92.27% $\pm 1.17$
AFIB vs VT	1	0	1	1	1	0	0	1	87.48% $\pm 1.18$
AFL vs SVTA	1	1	1	1	0	1	1	1	98.63% $\pm 0.43$
AFL vs VT	0	1	1	0	0	1	0	1	96.82% $\pm 0.97$
SVTA vs VT	1	0	1	0	1	1	1	0	97.27% $\pm 0.74$

TABLA IX: RESULTADOS DEL MULTI-CLASIFICADOR ASIMÉTRICO UTILIZANDO CARACTERÍSTICAS SELECCIONADAS DE FORMA EMPÍRICA

Arritmia	SE	SP	PP
Normal	98.89% $\pm 0.66$	98.30% $\pm 0.61$	97.23% $\pm 0.95$
AFIB	77.60% $\pm 3.50$	96.40% $\pm 0.67$	79.05% $\pm 2.69$
AFL	93.83% $\pm 1.88$	98.88% $\pm 0.36$	93.76% $\pm 1.81$
SVTA	90.04% $\pm 3.18$	98.58% $\pm 0.39$	91.89% $\pm 1.89$
VT	81.98% $\pm 2.94$	96.69% $\pm 0.80$	82.74% $\pm 3.17$
TOTAL	90.72% $\pm 0.66$	97.73% $\pm 0.16$	90.72% $\pm 0.66$

TABLA X: MATRIZ DE CARACTERÍSTICAS Y RENDIMIENTO DE CADA NODO, OBTENIDOS MEDIANTE  $\lambda_{DP}$

Nodos	Características								RE(%)
	1	2	3	4	5	6	7	8	
Normal vs AFIB	0	1	0	0	0	1	1	1	96.18% $\pm 0.64$
Normal vs AFL	1	0	0	1	0	1	0	1	99.12% $\pm 0.22$
Normal vs SVTA	1	1	0	0	0	1	1	1	99.25% $\pm 0.30$
Normal vs VT	1	1	0	0	0	0	1	0	96.52% $\pm 0.37$
AFIB vs AFL	1	0	1	0	0	1	1	1	96.03% $\pm 1.01$
AFIB vs SVTA	0	1	0	0	0	1	0	1	82.60% $\pm 1.18$
AFIB vs VT	1	0	0	0	1	0	0	0	79.55% $\pm 1.66$
AFL vs SVTA	1	1	1	1	1	1	0	1	98.00% $\pm 0.81$
AFL vs VT	1	0	0	0	1	1	0	1	90.78% $\pm 1.30$
SVTA vs VT	1	0	0	0	0	0	1	1	84.50% $\pm 1.36$

Comparando la matriz de características encontrada con función  $\lambda_{DP}$  y la matriz construida mediante el método empírico, podemos observar que se generan distribuciones asimétricas muy diferentes entre sí, tal como se muestra en la Figura 5.

En la Tabla XII se puede observar el costo computacional promedio para los procesos de entrenamiento y pruebas de las de los diferentes métodos de multclasificación, destacando una reducción del tiempo de ejecución por muestra del

multiclasificador en la fase de prueba/muestra de 2.09% para el primer multiclasificador asimétrico y del 3.13% para el segundo con respecto al método simétrico.

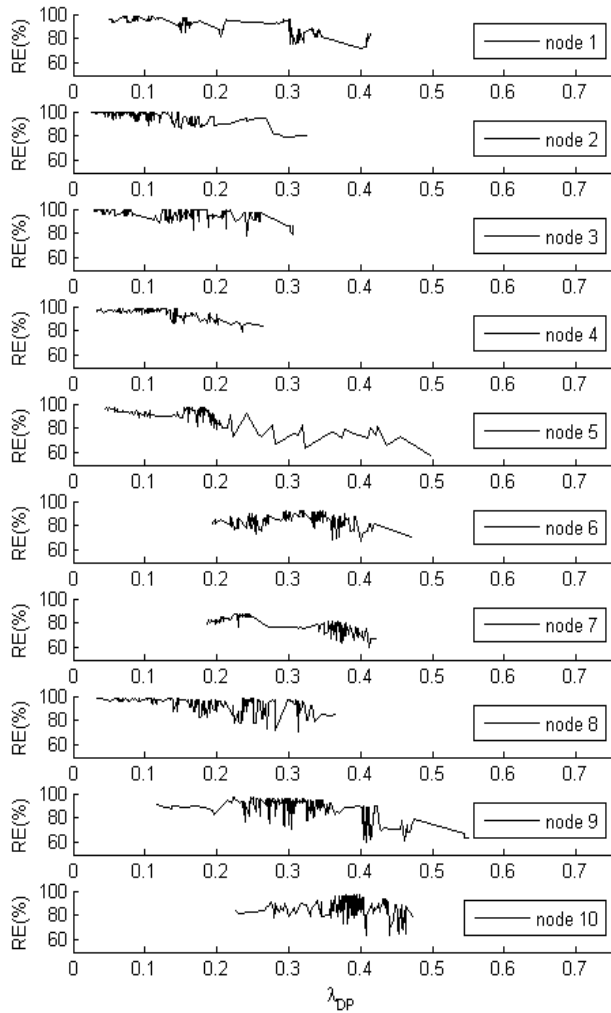


Figure 4. Comportamiento del rendimiento de cada nodo respecto a la distribución generada en cada combinación de características

TABLA XI: RESULTADOS DEL MULTICLASIFICADOR ASIMÉTRICO UTILIZANDO CARACTERÍSTICAS SELECCIONADAS MEDIANTE  $\lambda_{DP}$

Arritmia	SE	SP	PP
Normal	97.20% $\pm$ 1.31	96.51% $\pm$ 0.88	94.34% $\pm$ 1.28
AFIB	66.55% $\pm$ 3.90	92.87% $\pm$ 0.99	61.51% $\pm$ 2.82
AFL	90.00% $\pm$ 2.72	98.22% $\pm$ 0.47	90.01% $\pm$ 2.20
SVTA	74.72% $\pm$ 3.27	97.32% $\pm$ 0.63	82.87% $\pm$ 2.92
VT	70.12% $\pm$ 3.76	95.42% $\pm$ 0.95	74.34% $\pm$ 3.22
TOTAL	83.44% $\pm$ 0.69	96.02% $\pm$ 0.16	83.44% $\pm$ 0.69

Otro punto a destacar es la duración del proceso de construcción de la matriz de características. En caso del método empírico se realizó el proceso donde entrenó y evaluó el rendimiento de cada nodo en 30 ocasiones con cada combinación posible de características, tardando 78.09 horas en encontrar las mejores combinaciones; muy superior a los 1.13 minutos que se tardó evaluando la función  $\lambda_{DP}$  para todas combinaciones de características en cada nodo.

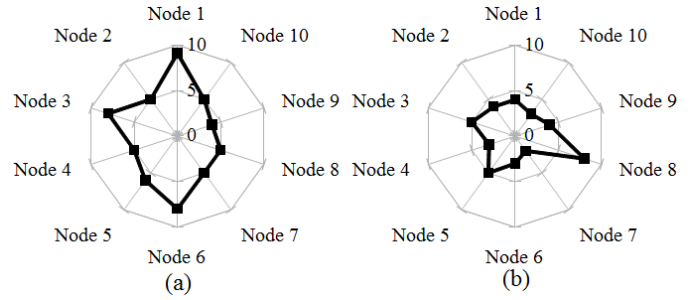


Figure 5. Simetría de los multi-clasificadores asimétricos construidos. a) Usando el método empírico. b) Usando la función  $\lambda_{DP}$

TABLA XII: RENDIMIENTO COMPUTACIONAL DE LOS CLASIFICADORES OAO

Caso	Tiempo de duración (seg.)		
	Entrenamiento	Prueba	Prueba/muestra*
OAO Simétrico	31.15 $\pm$ 0.83	9.96 $\pm$ 0.18	6.71 $\pm$ 0.12
OAO Asimétrico Empírico	31.06 $\pm$ 0.59	9.76 $\pm$ 0.19	6.57 $\pm$ 0.13
OAO Asimétrico $\lambda_{DP}$	31.55 $\pm$ 0.62	9.66 $\pm$ 0.18	6.50 $\pm$ 0.12

\* Tiempo en mili-segundos

## VI. CONCLUSIONES

Mediante la selección de las características en cada nodo, es posible desarrollar un multiclasificador de arritmias asimétrico que presente buenos resultados en comparación con [24-31], con un aumento en la predicción de positivos y disminuyendo el tiempo de ejecución por muestra en comparación con un multiclasificador simétrico.

Para el desarrollo de un multiclasificador asimétrico la construcción de la matriz de características es muy importante para mejorar el rendimiento del sistema de clasificación, siendo el método empírico el que mejores resultados presentó, sin embargo, el tiempo que se tarda en determinar la mejor combinación de características es muy superior al tiempo que se tarda usando  $\lambda_{DP}$ . En trabajos futuros se recomienda el uso de algoritmos heurísticos para una búsqueda eficiente de las combinaciones que formaran parte de la matriz de características, así como también, el desarrollo de otras técnicas de multi-clasificación basadas en SVM utilizando la filosofía de clasificadores asimétricos, tales como “One-Againts-All” (OAA), “Error Correcting Output Code” ECOC, “Minimum Output Coding” (MOC) y/o “DirectedAcyclicGraph SVM” (DAGSVM).

## RECONOCIMIENTOS

Nuestro más sincero agradecimiento al Decanato de Investigación de la Universidad Nacional Experimental del Táchira por el apoyo prestado durante el desarrollo de esta investigación bajo el código de proyecto 01-20-2011.

## REFERENCIAS

- [1] D. Abdelhamid, B. Chaouki and T.-A. Abdelmalik. "A fast multi-class SVM learning method for huge databases". *International Journal of Computer Science Issues*, vol. 8 (5), 2011. 544–550.
- [2] R. Wang, Sa. Kwong and D.-G. Chen. "A new method for multi-class support vector machines by training least number of classifiers". In *International Conference on Machine Learning and Cybernetics, ICMMLC*, Guilin, China, 2011, vol. 2, pp. 648 – 653.
- [3] B. Tang and D. Mazzoni. "Multiclass Reduced-Set Support Vector Machines". In *23rd International Conference on Machine Learning*, Pittsburgh, PA, USA, 2006, pp. 921–928.
- [4] N. Verma, A. Roy and A. Salour. "An optimized fault diagnosis method for reciprocating air compressors based on SVM". In *IEEE International Conference on System Engineering and Technology, ICSET*, Shah Alam, Malaysia, 2011, pp. 65 – 69.
- [5] G.-B. Huang, H. Zhou, X. Ding and R. Zhang. "Extreme Learning Machine for Regression and Multiclass Classification". *IEEE Transactions on Systems, Man, and Cybernetics—PART B: Cybernetics*, vol. 42, n<sup>o</sup>. 2, pp. 513–529, 2012.
- [6] M. Kallas, C. Francis, L. Kanaan, D. Merhed, P. Honeine and H. Amoud. "Multi-class SVM classification combined with kernel PCA feature extraction of ECG signals". In *19th International Conference on Telecommunications, ICT' 2012*, Jounieh, Lebanon, 2012, pp. 1–5.
- [7] MIT-BIH. "MIT-BIH Arrhythmia Database". *PhysioBank Archive Index*, 2009. [Online]. Available: <http://www.physionet.org/physiobank/database/mitdb/>.
- [8] L. Apacía. "Módulo para el procesamiento digital de señales electrocardiográficas en FPGA". Tesis de Grado en Maestría, Universidad Nacional Experimental del Táchira, San Cristóbal, Venezuela, 2009.
- [9] A. Nepomuceno, F. Salguero, y J. Quesada. "Información: tratamiento y representación". Sevilla, España: Universidad de Sevilla, 2001.
- [10] C. Shannon. "A Mathematical Theory of Communication". *The Bell System Technical Journal*, vol. 27, n<sup>o</sup>. 3, pp. 379–423, 623–656, 1948.
- [11] F. Göloğlu y A. Doğanaksoy. "On Lempel-Ziv Complexity of Sequences", in *SETA 2006*, Beijing, China, 2006, pp. 180–189.
- [12] A. Lempel and J. Ziv. "On the Complexity of Finite Sequences". *IEEE Transactions on Information Theory*, vol. 22, n<sup>o</sup>. 1, pp. 75–81, 1976.
- [13] M. Borowska, E. Oczerelko, A. Mazurek, A. Kitlas and P. Kuc. "Application of the Lempel-Ziv complexity measure to the analysis of biosignals and medical images". *Annales Academiae Medicae Bialostocensis*, vol. 50, n<sup>o</sup>. Supplement 2, pp. 29–30, 2005.
- [14] L. Mora and J. Amaya. "Efecto del Método de Codificación sobre la Distribución de Arritmias Cardíacas". In *III Congreso Internacional de Ingeniería Mecatrónica*, Bucaramanga, Colombia, 2011, vol. 2.
- [15] B. Boser, I. Guyon and V. Vapnik. "A Training Algorithm for Optimal Margin Classifiers". In *Fifth Annual Workshop on Computational Learning Theory*, Pittsburgh, Pennsylvania, United States, 1992, pp. 144–152.
- [16] J. Shawe and N. Cristianini. "An introduction to support vector machines and other kernel-based learning methods". Cambridge University Press, 2000.
- [17] S. Abe. "Support Vector Machines for Pattern Classification". 1<sup>o</sup> Edición. Springer, 2005.
- [18] M. Athimethphat and B. Lerteerawong. "Binary classification tree for multiclass classification with observation-based clustering". In *ECTI-CON*, Phetchaburi, Thailand, 2012, pp. 1–4.
- [19] J. Suykens, K. De Brabanter, P. Karsmakers, F. Ojeda, C. Alzate, J. De Brabanter, K. Pelckmans, B. De Moor and J. Vandewalle. "LS-SVMlab Toolbox". Leuven, Belgium: Katholieke Universiteit Leuven, 2010.
- [20] S. Mahmoodabadi, A. Ahmadian, M. Abolhassani, J. Alirezai and P. Babyn. "ECG Arrhythmia Detection Using Fuzzy Classifiers". In *Annual Meeting of the North American Fuzzy Information Processing Society, NAFIPS'07*. San Diego, California, 2007, pp. 48–53.
- [21] S. Kara and M. Okandan. "Atrial fibrillation classification with artificial neural networks". *Pattern Recognition*, vol. 40, n<sup>o</sup>. 11, pp. 2967 – 2973, 2007.
- [22] A. Dallali, A. Kachouri and M. Samet. "Classification of Cardiac Arrhythmia Using WT, HRV, and Fuzzy C-Means Clustering". *Signal Processing: An International Journal*, vol. 5, n<sup>o</sup>. 3, pp. 101–109, 2011.
- [23] J. Narisi, M. Naghibzadeh, S. Yazdi and B. Naghibzadeh. "ECG Arrhythmia Classification with Support Vector Machines and Genetic Algorithm". In *EMS'09*, Athens, 2009, pp. 187 – 192.
- [24] P. Bhardwaj, R. Choudhary and R. Dayama. "Analysis and Classification of Cardiac Arrhythmia using ECG Signals". *International Journal of Computer Applications*, vol. 38, n<sup>o</sup>. 1, pp. 37–40, 2012.
- [25] S. Graja and J. M. Boucher. "SVM Classification of patients prone to atrial fibrillation". In *WISP 2005*, Faro, Portugal, 2005, pp. 370–374.
- [26] V. Ilankumaran and S. ThamaraiSelvi. "Ventricular Arrhythmias Detection using Wavelet Decomposition". *International Journal of Computer Applications*, vol. 20, n<sup>o</sup>. 1, pp. 11–18, 2011.
- [27] P. Sanaz and A. Ahmad. "Automatic Atrial Fibrillation Detection Using Autoregressive Modeling". In *International Conference on Biomedical Engineering and Technology IPCBEE*, Singapore, Malaysia, 2011, vol. 11, pp. 105–108.
- [28] J. C. Fook, F. Qiang and I. Cosic. "A computationally light-weight real-time classification method to identify different ECG signals". In *ISBB'2011*, Suzhou, China, 2011, pp. 287 – 290.
- [29] J. Lee, B. Reyes, D. Memanus, O. Maitas and H. Chon. "Atrial Fibrillation Detection using an iPhone 4S". *IEEE Transactions on Biomedical Engineering*, vol. PP, n<sup>o</sup>. 99, p. 1, 2012.
- [30] R. Alcaraz, F. Sandberg, L. Sörnmo and J. Rieta. "Classification of Paroxysmal and Persistent Atrial Fibrillation in Ambulatory ECG Recordings". *IEEE Transactions on Biomedical Engineering*, vol. 58, n<sup>o</sup>. 5, pp. 1441 – 1449, 2011.
- [31] Q. Xiao, C. W. Jian and G. D. Fei. "ECG signal classification based on BPNN". In *International Conference on Electric Information and Control Engineering*, Wuhan, China, 2011, pp. 62 – 1364.
- [32] I. Witten & E. Frank. "Data Mining: Practical Machine Learning Tools and Techniques". Morgan Kaufmann Publisher. Second Edition. 2005.