

Multiclass Classification of Cardiac Rhythms on Short Single Lead ECG Recordings using Bidirectional Long Short-Term Memory Networks

Miguel Altuve, *Senior Member, IEEE*, and Fabio Hernández

Abstract—The recognition of cardiac rhythms is a topic of great relevance, particularly using short single-lead ECG recordings, due to its potential to early detect cardiovascular diseases and take actions quickly to preserve people’s well-being. Among cardiac arrhythmias, atrial fibrillation is the most common sustained cardiac arrhythmia, with significant mortality and morbidity rates. Several approaches have been conceived to identify cardiac rhythms, from the comparison of heart rate with adaptive and fixed thresholds to the application of deep and machine learning techniques. In this work, the classification of three cardiac rhythms (normal, atrial fibrillation, and other arrhythmias), as well as the identification of noise recordings, were performed using bidirectional LSTM networks that exploit the ECG signal (a representation of the cardiac electric activity) and time series containing information about the auricular and ventricular activities. A Monte Carlo 10-fold cross-validation of 10 iterations was performed to assure the generalization of the classifiers and the replicability of the results. An average accuracy of 77.97% was obtained to recognize the four classes but increase up to 85.95% when noise recordings were left out of the classification process. Moreover, micro $F1$ scores of 89.96%, 79.23%, and 79.77% were obtained for normal rhythm, atrial fibrillation, and other arrhythmias, respectively. The imbalance of classes and the characteristic patterns of normal rhythm and atrial fibrillation were the main factors associated with these performances.

Index Terms—Cardiac Rhythm, Atrial Fibrillation, Electrocardiography, LSTM, Classification, Deep Learning.

I. INTRODUCCIÓN

De acuerdo con los datos mostrados por el Observatorio Nacional de Salud, en Colombia, al igual que en el resto del mundo, las enfermedades cardiovasculares son la principal causa de muerte [1], produciendo un enorme gasto a la nación de 6,4 billones de pesos cada año (aproximadamente 1700 millones de dólares), de los cuales 3,84 billones (aproximadamente 1000 millones de dólares) es el costo del sistema de salud, equivalente al 6,55% del gasto total en salud del país.

La sucesión regular de las contracciones de los músculos del corazón produce el ritmo cardiaco. En condiciones normales, estas contracciones son controladas por el nódulo sinoauricular [2]. Una arritmia cardiaca ocurre cuando el ritmo cardiaco se altera y esta puede ser, de acuerdo con su origen, auricular, de unión o ventricular, de acuerdo con su frecuencia, taquicardia (mayor que 100 latidos por minuto en adultos) o

bradicardia (menor que 60 latidos por minuto en adultos), de acuerdo con su mecanismo, por automatismo, de reentrada o por actividad desencadenada, de acuerdo con su conducción, auriculoventricular, normal, de retardo o de bloqueo, y de acuerdo con su duración, no sostenida (menos de 30 s) o sostenida (30 s o más) [3]. La fibrilación auricular (FA), la arritmia cardíaca sostenida más común (afecta a casi el 2% de la población) y con significativas tasas de mortalidad y morbilidad [4], es un trastorno caracterizado por una actividad eléctrica auricular descoordinada, con el consecuente deterioro de la función mecánica auricular, en el que los ventrículos, al no recibir los impulsos eléctricos de forma regular, se contraen de forma acelerada, descontrolada e irregular [2].

El electrocardiograma (ECG), una herramienta de exploración de la actividad eléctrica cardíaca, es la herramienta preferida para el diagnóstico de arritmias cardíacas, entre ellas la FA [5]. En un ECG normal, las ondas P representan la despolarización auricular, no obstante, ondas de fibrilación (ondas f) de amplitud y duración variables reemplazan a las ondas P durante la FA. Por su condición episódica, el desconocimiento de los mecanismos exactos que la generan y su similitud con otros tipos de arritmias cardíacas, la detección y caracterización de la FA puede ser complicada y desafiante, tanto para el médico especialista como para algoritmos computacionales, y conducir a diagnósticos erróneos (falsos negativos). Además, la alta variabilidad de la amplitud y duración de las ondas f hacen difícil reconocer un patrón de FA en registros ECG de corta duración, por tal motivo, los registros de Holter son usados tradicionalmente para obtener suficientes datos y realizar el diagnóstico de FA.

Los métodos de detección de la FA pueden basarse en el análisis de la actividad auricular (ausencia/presencia de ondas P), en el análisis de la actividad ventricular o en el análisis conjunto de las actividades auricular y ventricular [6]. Por ejemplo, la reducción de la actividad ventricular se ha logrado con una red neuronal de estado de eco [7], mientras que el modelado estadístico se ha empleado para caracterizar las ondas P durante el ritmo sinusal normal, identificar la ausencia de ondas P y detectar la FA [8]. Las gráficas de Poincarè de intervalos RR (longitud del ciclo cardíaco) se han empleado para tener en cuenta la actividad ventricular [9], [10], mientras que el análisis combinado de las actividades auricular y ventricular se ha llevado a cabo utilizando el modelado de procesos de Markov para analizar la irregularidad del intervalo RR y la localización y morfología de la onda P [11] o utilizando redes neuronales artificiales (RNA) con características

M. Altuve, Facultad de Ingeniería Eléctrica y Electrónica, Universidad Pontificia Bolivariana, Bucaramanga, Colombia. E-mail: miguel.altuve@ieec.org.

A. Hernández, Facultad de Ingeniería Eléctrica y Electrónica, Universidad Pontificia Bolivariana, Bucaramanga, Colombia.

ondículas extraídas de los registros de ECG [12]. El desafío PhysioNet/Computing in Cardiology Challenge 2017 [13], [6] se propuso recientemente para clasificar registros de un canal de ECG de corta duración de acuerdo con el ritmo cardiaco: ritmo sinusal normal (RN), FA, otros ritmos anormales (OR) o demasiado ruidoso para identificar el ritmo (SR) [6]. Varios enfoques de clasificación multiclase, basados tanto en aprendizaje de máquina como en aprendizaje profundo, han sido propuestos para reconocer estos ritmos con desempeños de clasificación prometedores [14], [15], [16], [17]. Usando RNA que explotan características extraídas de la descomposición en ondículas de la señal ECG obtuvimos un $F1$ macro de 64% como desempeño de clasificación de estos cuatro tipos de ritmos [18].

La detección de FA de una manera rápida, automática y precisa, en registros de un canal ECG de corta duración es una necesidad urgente y un tema de investigación activa, particularmente usando dispositivos vestibles, como Checkme™ (Viatom Technology, Shenzhen, China) [19], teléfonos y relojes inteligentes, donde sus capacidades para almacenar, procesar y transmitir información los hace idóneos para advertir a los servicios de salud de manera oportuna y en tiempo real sobre la presencia de ritmos cardíacos irregulares. En ese sentido, en este trabajo, por medio de registros de un canal de ECG de corta duración, adquiridos usando un dispositivo económico y fácil de utilizar, se clasifican de manera automática ritmos cardíacos. En un primer experimento de clasificación multiclase, desarrollamos un clasificador para reconocer cuatro ritmos (RN, FA, OR, y SR), y, en un segundo experimento de clasificación, desarrollamos un clasificador para reconocer tres ritmos (RN, FA, y OR). La clasificación es llevada a cabo usando redes de gran memoria de corto plazo (LSTM, por sus siglas en inglés *Long Short-Term Memory*) bidireccionales. Este es un tipo de red neuronal artificial recurrente particularmente adaptada al procesamiento de series de tiempo, ya que tienen celdas de memoria para almacenar la información a lo largo del tiempo y aprender dependencias a largo plazo [20], [21]. Además, se empleó un enfoque basado en redes bidireccionales LSTM (Bi-LSTM) [22] para considerar la información futura y pasada a largo plazo de las señales ECG en la tarea de clasificación.

La detección automática de ritmos cardíacos, y particularmente la FA, en registros de ECG de corta duración sería una herramienta muy valiosa en los sistemas de apoyo a la decisión clínica ya que podría contribuir a que se realicen tratamientos a tiempo y a que se prevengan eventos fatales o el aumento de enfermedades cerebrovasculares crónicas. Además, reconocer automáticamente ritmos cardíacos podría ayudar a la detección temprana de arritmias cardíacas, lo que impactaría favorablemente en la vida de los pacientes, ya que podrían obtener un tratamiento oportuno que mejoren su calidad de vida y eviten que su salud se deteriore, disminuyendo así la necesidad de cirugías u hospitalizaciones, lo que a su vez reduciría los gastos en salud, en especial, en los países en vías de desarrollo, como Colombia.

II. MATERIALES Y MÉTODOS

A. Conjunto de Datos

Los registros de una derivación de ECG de corta duración utilizados en este trabajo fueron tomados de la página web de PhysioNet/Computing in Cardiology Challenge del año 2017¹. Estos registros fueron adquiridos usando dispositivos KardiaMobile² (AliveCor®, California, USA). Los datos fueron almacenados a una frecuencia de muestreo de 300 muestras por segundos y 16 bits de resolución, con un ancho de banda de 0.5 a 40 Hz y ± 5 mV [6]. La base de datos consta de 8528 registros de ECG divididos en cuatro clases: 5076 RN, 758 FA, 2415 OR, y 279 SR. En la tabla I se muestra un resumen estadístico de la duración de los registros ECG por cada clase, y en la figura 1 se observan ejemplos de señales ECG con estos ritmos.

Debido a que las redes LSTM procesan series de tiempo con la misma longitud, se realizó un histograma para conocer la distribución de la duración de los registros de ECG, tal como se muestra en la figura 2.

Del histograma se puede observar que la mayoría de los registros tienen 9000 muestras de longitud (30 s). Se seleccionaron entonces los registros ECG cuya duración fuera mayor o igual que 9000 muestras como entradas a la red LSTM. Aquellos registros con más de 9000 muestras se dividieron en registros de 9000 muestras consecutivas (sin solapamiento). Este procedimiento es resumido en el pseudocódigo mostrado en el algoritmo 1.

Este algoritmo toma un registro de ECG (RECG) de longitud L y obtiene(n) nuevo(s) registro(s) (NRECG) de 9000 muestras de longitud fija.;

Obtener L de (RECG);

si $L \geq 9000$ muestras y $L < 18000$ muestras

entonces

Obtener nuevo registro NRECG[0:8999] = RECG[0:8999];

de lo contrario

si $L \geq 18000$ muestras **entonces**

Obtener dos nuevos registros;

NRECG1[0:8999]=RECG[0:8999];

NRECG2[9000:17999]=RECG[9000:17999] ;

de lo contrario

Eliminar registro ;

fin

fin

Algoritmo 1: Pseudocódigo para unificar la longitud de los registros de ECG.

El nuevo conjunto de datos está compuesto por 8408 registros de ECG de 9000 muestras de duración, divididos como se detalla en la tabla II. Este es el conjunto de datos utilizado por el los modelos de clasificación.

¹<https://physionet.org/content/challenge-2017/1.0.0/>

²<https://www.alivecor.com/kardiamobile>

TABLE I
DURACIÓN DE LOS REGISTROS DE ECG POR CADA CLASE DE LA BASE DE DATOS.

Ritmo	# registros	Duración (s)			
		Media ± desv. est.	Mínimo	Mediana	Máximo
RN	5076	32,11 ± 9,97	9,00	30	60,95
FA	758	32,34 ± 12,31	9,98	30	60,20
OR	2415	34,29 ± 11,76	9,12	30	60,86
SR	279	24,38 ± 10,40	9,60	30	60,00

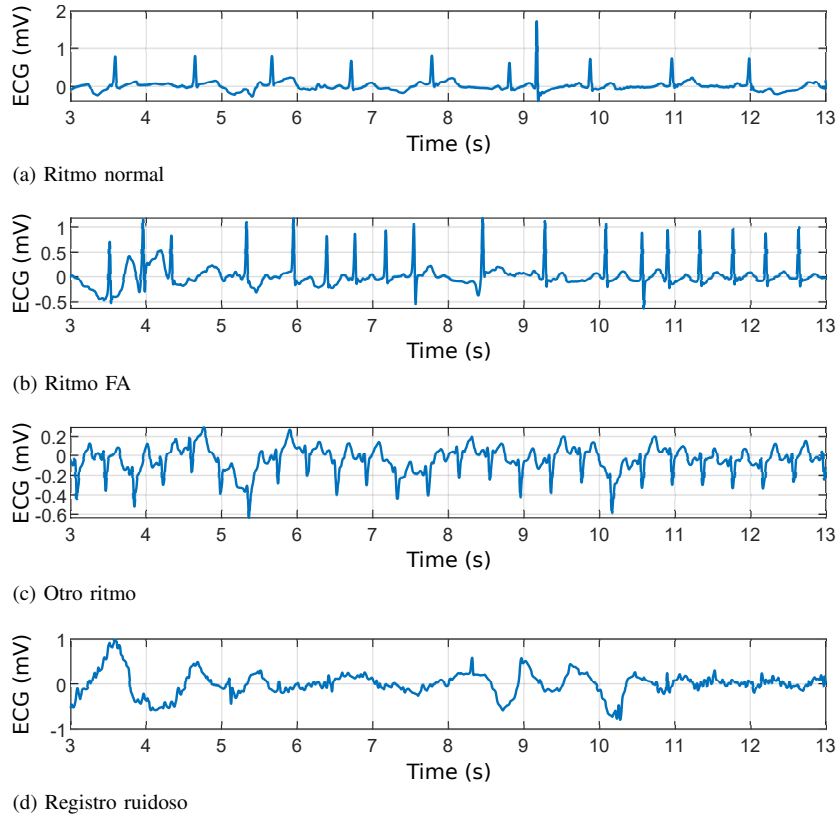


Fig. 1. Ejemplos de señales ECG tomadas de la base de datos.

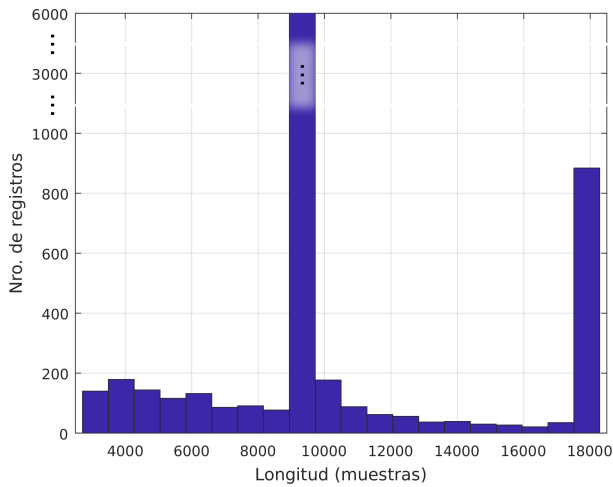


Fig. 2. Histograma de la longitud de los registros ECG de la base de datos.

TABLE II
CONJUNTO DE DATOS DE REGISTROS ECG DE 30 s DE DURACIÓN.

Ritmo	# registros
RN	4962
FA	744
OR	2552
SR	150

B. Secuencias de Entradas al Clasificador

Llamemos $x[n]$ a la serie de tiempo (secuencia o señal en tiempo discreto) correspondiente a la versión muestreada de la señal ECG en tiempo continuo $x(t)$, esto es [23]:

$$x[n] = x(nT) \tag{1}$$

donde T es el período de muestreo ($T = 1/300 = 3.33$ ms en este caso).

Para cada secuencia ECG del nuevo conjunto de datos (tabla II), se crearon tres secuencias de duración 30 s: *i*) secuencia pico onda P, $x_P[n]$, *ii*) secuencia pico onda R, $x_R[n]$, y *iii*) secuencia pico onda S, $x_S[n]$. La secuencia $x_P[n]$ (ver figura 3b) tiene todos sus valores en cero excepto cuando ocurre el pico de la onda P, P_{peak} , en donde el valor corresponde a ese pico en la secuencia ECG, $x[n]$ (ver figura 3a), esto es:

$$x_P[n] = [0, 0, \dots, x[P_{peak}], \dots, 0, 0]. \quad (2)$$

De manera similar, la secuencia $x_R[n]$ (ver figura 3c) tiene todos sus valores en cero excepto en cuando ocurre el pico de la onda R en el ECG:

$$x_R[n] = [0, 0, \dots, x[R_{peak}], \dots, 0, 0]. \quad (3)$$

Finalmente, la secuencia $x_S[n]$ (ver figura 3d) tiene todos sus valores en cero excepto en cuando ocurre el pico de la onda S en el ECG:

$$x_S[n] = [0, 0, \dots, x[S_{peak}], \dots, 0, 0]. \quad (4)$$

Los instantes de tiempo donde ocurren P_{peak} , R_{peak} y S_{peak} fueron obtenidos automáticamente usando un algoritmo de detección de límites de ondas en señales ECG, a través de la función `ecgpuwave` de la WFDB Toolbox for MATLAB and Octave [13], [24].

Las secuencias $x[n]$, $x_P[n]$, $x_R[n]$, y $x_S[n]$ fueron estandarizadas utilizando el método de escalado descrito en la ecuación 5, con el fin de tener secuencias con amplitudes entre 0 y 1.

$$\hat{x}[n] = \frac{x[n] - \min(x[n])}{\max(x[n]) - \min(x[n])}. \quad (5)$$

Finalmente, para cada registro de ECG del conjunto de datos, la matriz de entrada a los clasificadores basados en redes Bi-LSTM se puede representar como:

$$\mathbf{Z} = \begin{bmatrix} \hat{x}[n] \\ \hat{x}_P[n] \\ \hat{x}_R[n] \\ \hat{x}_S[n] \end{bmatrix} \quad (6)$$

Esta matriz es de tamaño 4×9000 .

C. Clasificador Multiclase Basado en Bi-LSTM

En este trabajo se utilizaron redes LSTM para procesar las secuencias ya que este tipo de redes toma en cuenta la información temporal en las celdas de memoria, donde, al almacenar información a lo largo del tiempo, facilitan el aprendizaje de las dinámicas de la señal. Además, dado que se cuenta con toda la señal en el tiempo, se utilizaron redes LSTM bidireccionales (Bi-LSTM) para tomar en cuenta la información del pasado y del futuro.

Dos clasificadores multiclases fueron desarrollados, el primero para identificar las clases FA, RN, OR y SR, y el segundo para reconocer las clases FA, RN y OR. El segundo clasificador no considera las señales ruidosas, tal como lo recomiendan los organizadores del desafío [6], debido a su

poca cantidad de registro y minimizar así el desequilibrio de clases.

Los clasificadores propuestos están compuestos por dos capas Bi-LSTM seguidos de una capa RNA de clasificación. Ambos clasificadores tienen la misma arquitectura y se diferencian únicamente en la última capa. La entrada de la primera capa es un arreglo de 4×9000 que pasa a través de 50 neuronas para obtener una secuencia de 50 muestras de longitud como salida. Esta secuencia de salida corresponde a la entrada de la segunda capa, la cual consta de 25 neuronas, y cuya salida es un arreglo de 4 valores para el primer clasificador y de 3 valores para el segundo clasificador. Finalmente, la última capa es una RNA, de 4 neuronas para el primer clasificador y de 3 neuronas para el segundo clasificador, que organiza la salida de la capa anterior para obtener un vector de clasificación donde cada posición del vector está asociada a una clase: [RN FA OR SR] para el primer clasificador y [RN FA OR] para el segundo clasificador. La figura 4 muestra el diagrama de estos clasificadores.

Los siguientes parámetros de configuración fueron utilizados para ambos clasificadores: tamaño del lote igual 100, 500 periodos de cambio, regularización de abandono de 0,2, función de activación Softmax para las RNA, y función sigmoidea y tangente hiperbólica para la celda de memoria de las redes Bi-LSTM, Adam como algoritmo de optimización, entropía cruzada categórica como función de pérdida. Además, se utilizó la opción *class weight* de Keras durante el entrenamiento para asignar pesos a cada clase, con el fin de ponderar la función de pérdida durante el entrenamiento. Esto es útil para que el modelo preste más atención a las clases con menos registros y así minimizar un poco el impacto del desequilibrio de clases.

D. Métricas de Evaluación de la Clasificación Multiclase

Se utilizó una matriz de confusión para visualizar las discrepancias entre las clases predichas y las clases verdaderas, como se muestra en la tabla III, donde las filas corresponden a la clase verdadera (clase de referencia) y las columnas corresponden a la clase predicha (clase de salida) [21]. La diagonal corresponde a las clasificaciones correctas hechas por el clasificador. A partir de la matriz de confusión, la exactitud (*Acc*), la precisión (*Pre*) y la exhaustividad (*Rec*) se calcularon como se muestra en la tabla III. Además, se determinó el valor *F1* definido como el promedio ponderado entre la precisión y la exhaustividad:

$$F1 = \frac{2 \times Pre \times Rec}{Pre + Rec} \quad (7)$$

Los valores de las métricas *Pre*, *Rec* y valor *F1* determinados para cada clase corresponden a métricas micro. El promedio de las métricas micro corresponde a una métrica macro. Por ejemplo, para el caso de las cuatro clases RN, FA, OR y SR, el valor *F1* macro corresponde al promedio de los valores *F1* micro, uno por cada clase, esto es:

$$F1 = \frac{F1_{RN} + F1_{FA} + F1_{OR} + F1_{SR}}{4}. \quad (8)$$

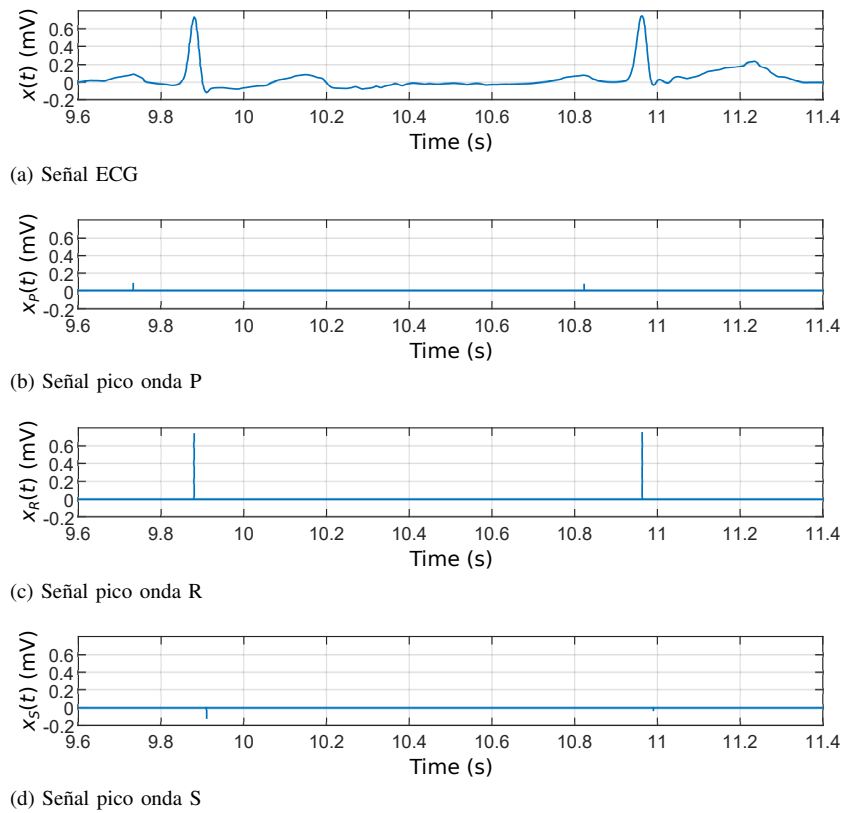


Fig. 3. Ejemplo de extractos de las señales en tiempo continuo correspondiente a: ECG (a), posición del pico de las ondas P (b), posición del pico de las ondas R (c), y posición del pico de las ondas S (d).

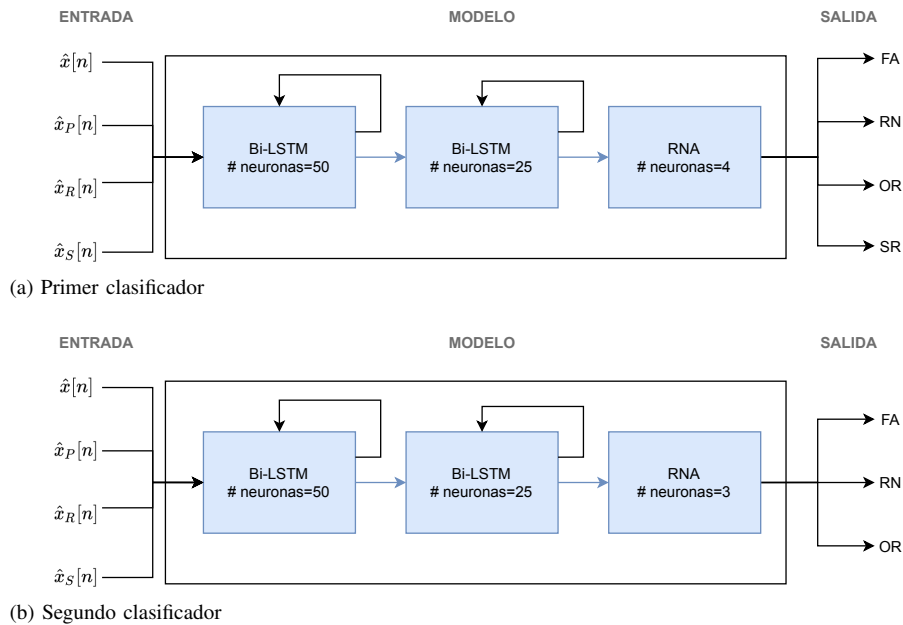


Fig. 4. Estructura de los clasificadores multiclase propuestos.

Para evaluar el desempeño, y garantizar la repetibilidad y reproducibilidad de los resultados, se llevaron a cabo dos validaciones cruzadas anidadas. Se realizó una validación cruzada por Monte Carlo (MCCV, por sus siglas en inglés, *Monte Carlo cross-validation*) de 10 iteraciones, en la que cada iteración se dividió el conjunto de datos de forma aleatoria

en 80% para llevar a cabo el entrenamiento y el 20% restante para ejecutar la prueba. En cada iteración de la MCCV se llevó a cabo una validación cruzada de 10 iteraciones (*10-fold cross-validation*).

TABLE III
MATRIZ DE CONFUSIÓN. LAS ECUACIONES DE PRECISIÓN (Pre), EXHAUSTIVIDAD (Rec) Y EXACTITUD (Acc) SE MUESTRAN COMO REFERENCIA.

Referencia	Predicción				Rec
	$Clase_1$	$Clase_2$...	$Clase_n$	
$Clase_1$	a_{11}	a_{12}	...	a_{1n}	$\frac{a_{11}}{\sum_{j=1}^n a_{1j}}$
$Clase_2$	a_{21}	a_{22}	...	a_{2n}	$\frac{a_{22}}{\sum_{j=1}^n a_{2j}}$
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
$Clase_n$	a_{n1}	a_{n2}	...	a_{nn}	$\frac{a_{nn}}{\sum_{j=1}^n a_{nj}}$
Pre	$\frac{a_{11}}{\sum_{i=1}^n a_{i1}}$	$\frac{a_{22}}{\sum_{i=1}^n a_{i2}}$...	$\frac{a_{nn}}{\sum_{i=1}^n a_{in}}$	$Acc = \frac{\sum_{i=1}^n a_{ii}}{\sum_{i=1}^n \sum_{j=1}^n a_{ij}}$

TABLE IV
EJEMPLO DE UNA MATRIZ DE CONFUSIÓN EN UN SUBCONJUNTO DE PRUEBA DE UNA ITERACIÓN DE LA MCCV.

Referencia	Predicción				$Rec(\%)$
	RN	FA	OR	SR	
RN	820	11	123	3	85,68
FA	32	112	21	1	67,46
OR	177	9	349	2	64,99
SR	5	4	3	9	42,85
$Pre(\%)$	79,30	82,35	70,36	60,00	76,74

E. Análisis Estadístico

La prueba estadística de Friedman, seguida de la prueba de diferencia honestamente significativa de Tukey, se emplearon para encontrar diferencias significativas de los valores $F1$ micro entre las diferentes clases, para los dos clasificadores. Específicamente, en cada corrida de la MCCV, se realizaron estas pruebas estadísticas para encontrar las diferencias de los valores de $F1$ entre las cuatro clases, para el caso del primer clasificador, y entre las tres clases, para el caso del segundo clasificador. Un valor de $p < 0.05$ se consideró estadísticamente significativo. Además, barras de error, que representan la mediana $\pm 95\%$ del intervalo de confianza de los valores $F1$ micro en cada iteración de la MCCV, fueron graficadas para inspeccionar visualmente los resultados.

III. RESULTADOS

La tabla IV muestra un ejemplo de clasificación para un subconjunto de prueba en una iteración de la MCCV. En esta tabla podemos observar que 820 registros de ECG fueron clasificados como RN ($F1_{RN} = 82,37\%$), 112 como FA ($F1_{FA} = 74,17\%$), 349 como OR ($F1_{OR} = 67,57\%$) y 9 como SR ($F1_{SR} = 50,00\%$). Para este ejemplo, el valor $F1$ macro es de 68,52%.

La tabla V muestra los resultados del desempeño del primer clasificador, para cada iteración de la MCCV. Se puede observar que RN fue la clase más fácil de detectar (en promedio $F1_{RN} = 83,54\%$) mientras que las señales ruidosas fue la clase más difícil de detectar (en promedio $F1_{SR} = 55,83\%$). El valor $F1$ macro promedio para las 10 iteraciones en el conjunto de prueba fue de 70,71%. Del análisis estadístico se

pudo observar una diferencia estadísticamente significativa de los valores $F1$ de RN con respecto a los valores $F1$ de las otras tres clases. También existe una diferencia estadísticamente significativa de los valores $F1$ de FA con respecto a los valores $F1$ de OR (excepto en la tercera iteración) y a los valores $F1$ de SR. Finalmente, existe una diferencia estadísticamente significativa de los valores $F1$ de OR con respecto a los valores $F1$ de SR.

La figura 5 muestra las barras de error de los valores $F1$ micro del primer clasificador. En esta figura podemos observar la gran variabilidad del desempeño para reconocer la clase SR en las diferentes iteraciones de la MCCV; las clases RN y FA son identificadas con desempeños similares en las diferentes iteraciones de la MCCV.

La tabla VI muestra los resultados del desempeño del segundo clasificador, para cada iteración de la MCCV. Se puede observar nuevamente que RN fue la clase más fácil de detectar (en promedio $F1_{RN} = 89,96\%$) mientras que el desempeño de clasificación de las otras dos clases fue muy similar (en promedio $F1_{FA} = 79,23\%$ y $F1_{OR} = 79,77\%$). El valor $F1$ macro promedio para las 10 iteraciones en el conjunto de prueba fue de 82,99%. Del análisis estadístico se pudo observar una diferencia estadísticamente significativa de los valores $F1$ de RN con respecto a los valores $F1$ de las otras dos clases. También existe una diferencia estadísticamente significativa de los valores $F1$ de FA con respecto a los valores $F1$ de OR (excepto en las iteraciones 1, 4 y 5).

La figura 6 muestra las barras de error de los valores $F1$ micro del segundo clasificador. En esta figura podemos observar una mayor variabilidad del desempeño para reconocer la clase FA en las diferentes iteraciones de la MCCV.

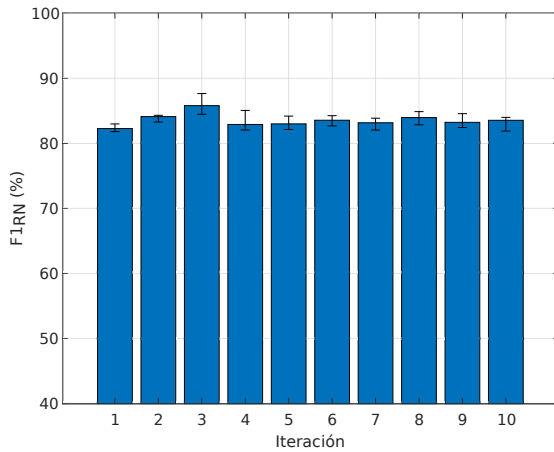
IV. DISCUSIÓN

En ambos clasificadores, el ritmo normal fue el ritmo más fácil de reconocer, con desempeños superiores a 80%. La cantidad de registros de este tipo y los patrones característicos del ritmo cardiaco normal, extraídos de las series de tiempo por ambos modelos, fueron factores relevantes en la obtención de este resultado, ya que los modelos pudieron extraer una mayor cantidad de información útil de las secuencias con más ejemplos para aprender. Lo contrario ocurrió con los registros ruidosos, en donde, dado que solo se contaron con 150 registros de este tipo y la inexistencia de un patrón característico de registros ruidosos, la red tuvo inconvenientes

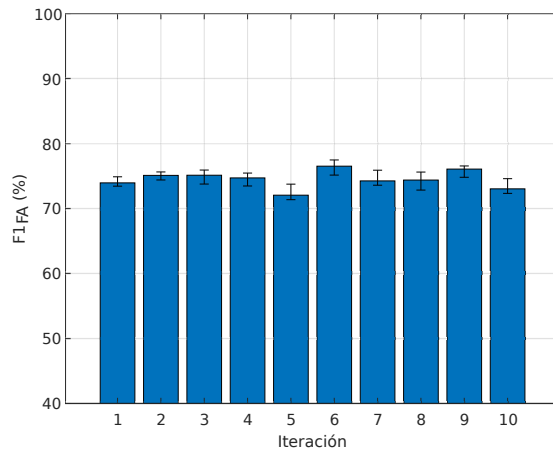
TABLE V
DESEMPEÑO DEL PRIMER CLASIFICADOR EN CADA ITERACIÓN DE LA MCCV.

Iteración	$F1_{RN}$ (%)	$F1_{FA}$ (%)	$F1_{OR}$ (%)	$F1_{SR}$ (%)	$F1$ macro (%)	Acc (%)
1	82,37 ^{a,b,c}	74,17 ^{d,e}	67,57 ^f	50,00	68,53	76,74
2	83,77 ^{a,b,c}	75,00 ^{d,e}	70,54 ^f	53,33	70,66	78,47
3	86,03 ^{a,b,c}	74,85 ^e	73,74 ^f	55,81	72,61	80,84
4	83,53 ^{a,b,c}	73,47 ^{d,e}	77,63 ^f	63,41	72,02	77,63
5	83,13 ^{a,b,c}	72,56 ^{d,e}	67,40 ^f	57,14	70,06	77,16
6	83,44 ^{a,b,c}	76,32 ^{d,e}	67,84 ^f	48,89	69,12	77,57
7	82,93 ^{a,b,c}	74,75 ^{d,e}	69,25 ^f	61,54	72,12	77,75
8	83,84 ^{a,b,c}	74,23 ^{d,e}	69,90 ^f	53,06	70,26	78,26
9	83,47 ^{a,b,c}	75,69 ^{d,e}	68,84 ^f	62,50	72,63	78,11
10	82,91 ^{a,b,c}	73,47 ^{d,e}	67,46 ^f	52,63	69,12	77,10
Promedio	83,54	74,45	69,02	55,83	70,71	77,97

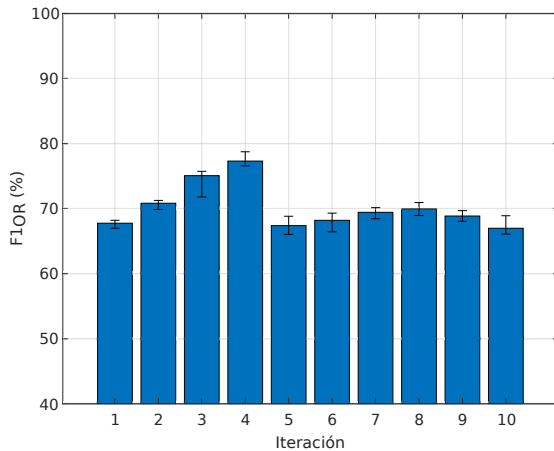
- ^a Diferencia estadísticamente significativa entre RN y FA.
- ^b Diferencia estadísticamente significativa entre RN y OR.
- ^c Diferencia estadísticamente significativa entre RN y SR.
- ^d Diferencia estadísticamente significativa entre FA y OR.
- ^e Diferencia estadísticamente significativa entre FA y SR.
- ^f Diferencia estadísticamente significativa entre OR y SR.



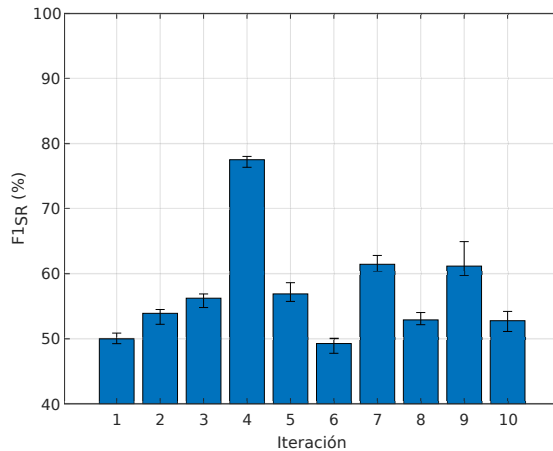
(a) F1 de RN



(b) F1 de FA



(c) F1 de OR



(d) F1 de SR

Fig. 5. Barras de error de los valores $F1$ micro del primer clasificador, en cada iteración de la MCCV.

para identificarlos correctamente ya que no tuvo información suficiente para lograr aprender dinámicas relevantes de pocos

ejemplos. No obstante, es importante resaltar que la FA, a pesar de también contar con pocos registros (menos de la mitad

TABLE VI
DESEMPEÑO DEL SEGUNDO CLASIFICADOR EN CADA ITERACIÓN DE LA MCCV.

Iteración	$F1_{RN}$ (%)	$F1_{FA}$ (%)	$F1_{OR}$ (%)	$F1$ macro (%)	Acc (%)
1	91,01 ^{a,b}	82,05	81,64	84,90	87,44
2	89,82 ^{a,b}	77,18 ^c	80,00	82,33	85,61
3	90,99 ^{a,b}	85,16 ^c	80,69	85,61	87,29
4	88,92 ^{a,b}	78,67	78,47	82,02	84,91
5	90,04 ^{a,b}	79,71	79,91	83,22	86,19
6	89,98 ^{a,b}	80,00 ^c	78,91	82,97	85,80
7	88,96 ^{a,b}	76,03 ^c	79,29	81,43	84,86
8	89,40 ^{a,b}	77,13 ^c	79,12	81,89	85,17
9	90,64 ^{a,b}	80,60 ^c	79,77	83,67	86,58
10	89,86 ^{a,b}	75,79 ^c	79,86	81,84	85,64
Promedio	89,96	79,23	79,77	82,99	85,95

^a Diferencia estadísticamente significativa entre RN y FA

^b Diferencia estadísticamente significativa entre RN y OR

^c Diferencia estadísticamente significativa entre FA y OR.

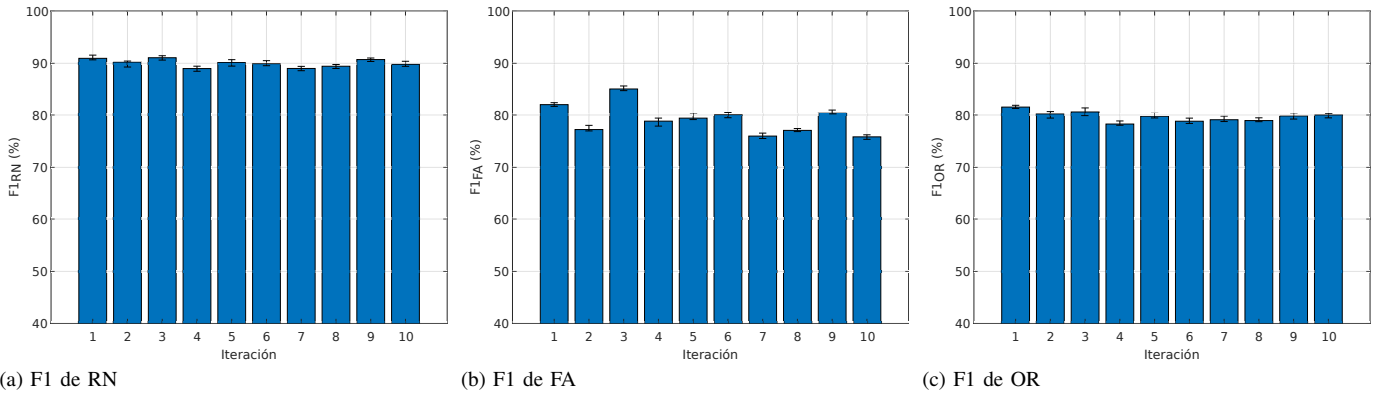


Fig. 6. Barras de error de los valores $F1$ micro del segundo clasificador, en cada iteración de la MCCV.

de los otros ritmos), tiene patrones característicos que ambos clasificadores pudieron aprovechar para extraer de las series de tiempo y lograr su clasificación correctamente; incluso, el primer clasificador fue capaz de clasificar mejor la FA que los otros tipos de ritmos cardiacos (74,45% vs 69,02% de $F1$ micro promedio), mientras que el desempeño en el segundo clasificador fue similar para estas dos clases (79,23% vs 79,77% de $F1$ micro promedio). Finalmente, la eliminación de los registros ruidosos del proceso de clasificación, lo que redujo un poco el desequilibrio de las clases, permitió que el segundo clasificador pudiera reconocer mejor los tres ritmos y mejorar significativamente el desempeño de la clasificación (82,99% vs 70,71% de $F1$ macro promedio, y 85,95% vs 77,97% de exactitud promedio).

Estos resultados son comparables con resultados reportados por otros autores con enfoques similares. Por ejemplo, Datta y colaboradores [14], utilizando una serie de clasificadores binarios en cascada de múltiples capas, obtuvieron $F1$ macro de 83% para clasificar tres clases, Zihlmann y colaboradores [15], mediante la combinación de capas convolucionales para la extracción de características con capas de LSTM para la agregación temporal de características, obtuvieron un $F1$ macro 79,2% para clasificar las cuatro clases, y 82% para clasificar tres clases respectivamente, Zabihi y colaboradores [16], a

partir de la extracción manual de 491 características de los dominios del tiempo, de la frecuencia y del tiempo-frecuencia, y de la reconstrucción del espacio de fase de las señales ECG, la selección de 150 de estas características, y la clasificación usando bosques aleatorios, obtuvieron un desempeño de 83% para clasificar las tres clases. No obstante, consideramos que el enfoque de clasificación propuesto en este trabajo tiene ventajas en los siguientes aspectos: *i*) utilizan series de tiempo en sus entradas, con información relevante respecto a la actividad auricular y ventricular, ya que estas indican los momentos donde ocurren las ondas P, R y S, *ii*) al ser desarrollado bajo plataformas de programación como MATLAB y Google Colab, este trabajo es fácilmente escalable y replicable sin necesidad de tener demasiados recursos de hardware, y *iii*) al no necesitar grandes recursos de hardware, este algoritmo podría ser implementado en sistemas embebidos y en teléfonos celulares inteligentes.

Las principales contribuciones de este trabajo pueden resumirse en el uso de redes Bi-LSTM para explotar, automáticamente y sin necesidad de pasar por un proceso laborioso de extracción manual de características, la información contenida en series de tiempo que reflejan las actividades auriculares y ventriculares, así como el recorrido del potencial de acción en el miocardio (i.e. el ECG). Esta propuesta de

representación de las entradas a la red LSTM permitió reducir el costo computacional al momento del entrenamiento, algo muy valioso en aplicaciones de aprendizaje profundo. No obstante, algunas dificultades y limitaciones en este trabajo deben resaltarse, tales como el alto desequilibrio entre las clases, en donde las clases RN, FA, OR y SR correspondieron al 61%, 8%, 28% y 3% del total de la base de datos, respectivamente. Aunque se implementaron métodos para solventar este problema, los entrenamientos se vieron afectados por este desequilibrio, por lo que, en trabajos futuros, sería deseable contar con una base de datos más balanceada para facilitar el aprendizaje de todas las clases. También es importante resaltar que los registros ECG utilizados son de un solo canal y de corta duración, sería interesante utilizar estos modelos de aprendizaje en señales de varios canales, de más larga duración y con otros tipos de arritmias [25]. Finalmente, dado que el aprendizaje profundo es un campo recientemente explorado y con una evolución muy rápida, no existe una regla sobre las técnicas y enfoques más apropiados para resolver un problema en particular; en este trabajo, y con el ánimo de aportar soluciones que impacten en el mejoramiento de la calidad de vida de las personas, hemos utilizados redes Bi-LSTM para extraer la información, tanto pasada como futura, de series de tiempo sin necesidad de realizar una etapa de extracción/selección de características, haciendo la propuesta simple y generalizable (menos dependiente del problema).

V. CONCLUSIÓN

El reconocimiento de ritmos cardiacos en registros de una derivación de ECG de corta duración, llevado a cabo con un enfoque de redes profundas, específicamente redes de gran memoria de corto plazo bidireccionales, que explotan la información contenida en series de tiempo que reflejan la actividad eléctrica del corazón, y en particular, las actividades auriculares y ventriculares, arrojó desempeños de clasificación alrededor de 70% cuando registros ruidosos son incluidos en el proceso de aprendizaje, pero supera el 80% cuando los registros ruidosos son eliminados del proceso de clasificación. La inexistencia de un patrón característico y el bajo número de ejemplos de los registros ruidosos fueron los principales factores para que el clasificador cometiera errores de asignación, tanto falsos positivos como falsos negativos. Además, la presencia de un patrón característico tanto en el ritmo cardiaco normal como durante la fibrilación auricular facilitó el reconocimiento de estos ritmos por parte del clasificador, a pesar de que este último tipo de ritmo contaba con un menor número de ejemplos. Si bien se agruparon en una sola clase a los otros ritmos cardiacos (bradicardia, taquicardia, etc.), los clasificadores fueron capaces de reconocer estos ritmos con tan variadas características.

REFERENCES

[1] Observatorio Nacional de Salud, "Enfermedad cardiovascular: principal causa de muerte en Colombia," https://www.ins.gov.co/Direcciones/ONS/Boletines/boletin_web_ONS/boletin1.html, Diciembre 2013, (Accedido el 03/05/2020).

- [2] C. T. January, L. S. Wann, J. S. Alpert, H. Calkins, J. E. Cigarroa, J. C. Cleveland, J. B. Conti, P. T. Ellinor, M. D. Ezekowitz, M. E. Field *et al.*, "2014 aha/acc/hrs guideline for the management of patients with atrial fibrillation: executive summary: a report of the american college of cardiology/american heart association task force on practice guidelines and the heart rhythm society," *Journal of the American College of Cardiology*, vol. 64, no. 21, pp. 2246–2280, 2014.
- [3] L. Gaztañaga, F. E. Marchlinski, and B. P. Betensky, "Mecanismos de las arritmias cardiacas," *Revista española de cardiología*, vol. 65, no. 2, pp. 174–185, 2012.
- [4] A. J. Camm, P. Kirchhof, G. Y. Lip, U. Schotten, I. Savelieva, S. Ernst, I. C. Van Gelder, N. Al-Attar, G. Hindricks, B. Prendergast, H. Heidbuchel, O. Alfieri, A. Angelini, D. Atar, P. Colonna, R. De Caterina, J. De Sutter, A. Goette, B. Gorenek, M. Haldal, S. H. Hohloser, P. Kolh, J.-Y. Le Heuzey, P. Ponikowski, F. H. Rutten, E. C. for Practice Guidelines (CPG), A. Vahanian, A. Auricchio, J. Bax, C. Ceconi, V. Dean, G. Filippatos, C. Funck-Brentano, R. Hobbs, P. Kearney, T. McDonagh, B. A. Popescu, Z. Reiner, U. Sechtem, P. A. Sirmes, M. Tendera, P. E. Vardas, P. Widimsky, D. Reviewers, P. E. Vardas, V. Agladze, E. Aliot, T. Balabanski, C. Blomstrom-Lundqvist, A. Capucci, H. Crijns, B. Dahlöf, T. Folliguet, M. Glikson, M. Goethals, D. C. Gulba, S. Y. Ho, R. J. M. Klautz, S. Kose, J. McMurray, P. Perrone Filardi, P. Raatikainen, M. J. Salvador, M. J. Schalij, A. Shpektor, J. Sousa, J. Stepinska, H. Uetoea, J. L. Zamorano, and I. Zupan, "Guidelines for the management of atrial fibrillation: The Task Force for the Management of Atrial Fibrillation of the European Society of Cardiology (ESC)," *European Heart Journal*, vol. 31, no. 19, pp. 2369–2429, 10 2010. [Online]. Available: <https://doi.org/10.1093/eurheartj/ehq278>
- [5] A. I. Hernández, J. Dumont, M. Altuve, A. Beuchée, and G. Carrault, "Evolutionary optimization of ECG feature extraction methods: Applications to the monitoring of adult myocardial ischemia and neonatal apnea bradycardia events," in *ECG Signal Processing, Classification and Interpretation: A Comprehensive Framework of Computational Intelligence*. Springer London, 2012, pp. 237–273.
- [6] G. D. Clifford, C. Liu, B. Moody, H. L. Li-wei, I. Silva, Q. Li, A. Johnson, and R. G. Mark, "Af classification from a short single lead eeg recording: the physionet/computing in cardiology challenge 2017," in *2017 Computing in Cardiology (CinC)*. IEEE, 2017, pp. 1–4.
- [7] A. Petrenas, V. Marozas, L. Sornmo, and A. Lukosevicius, "An echo state neural network for QRST cancellation during atrial fibrillation," *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 10, pp. 2950–2957, 2012.
- [8] S. Ladavich and B. Ghoraani, "Rate-independent detection of atrial fibrillation by statistical modeling of atrial activity," *Biomedical Signal Processing and Control*, vol. 18, pp. 274–281, 2015.
- [9] J. Park, S. Lee, and M. Jeon, "Atrial fibrillation detection by heart rate variability in poincare plot," *BioMedical Engineering OnLine*, vol. 8, no. 1, p. 38, 2009.
- [10] S. Sarkar, D. Ritscher, and R. Mehra, "A detector for a chronic implantable atrial tachyarrhythmia monitor," *IEEE Transactions on Biomedical Engineering*, vol. 55, no. 3, pp. 1219–1224, 2008.
- [11] S. Babaeizadeh, R. E. Gregg, E. D. Helfenbein, J. M. Lindauer, and S. H. Zhou, "Improvements in atrial fibrillation detection for real-time monitoring," *Journal of Electrocardiology*, vol. 42, no. 6, pp. 522–526, 2009.
- [12] S. Kara and M. Okandan, "Atrial fibrillation classification with artificial neural networks," *Pattern Recognition*, vol. 40, no. 11, pp. 2967–2973, 2007.
- [13] A. L. Goldberger, L. A. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley, "Physiobank, physiotoolkit, and physionet: components of a new research resource for complex physiologic signals," *circulation*, vol. 101, no. 23, pp. e215–e220, 2003.
- [14] S. Datta, C. Puri, A. Mukherjee, R. Banerjee, A. D. Choudhury, R. Singh, A. Ukil, S. Bandyopadhyay, A. Pal, and S. Khandelwal, "Identifying normal, af and other abnormal eeg rhythms using a cascaded binary classifier," in *2017 Computing in Cardiology (CinC)*. IEEE, 2017, pp. 1–4.
- [15] M. Zihlmann, D. Perekrestenko, and M. Tschannen, "Convolutional recurrent neural networks for electrocardiogram classification," in *2017 Computing in Cardiology (CinC)*. IEEE, 2017, pp. 1–4.
- [16] M. Zabihi, A. B. Rad, A. K. Katsaggelos, S. Kiranyaz, S. Narkilahti, and M. Gabbouj, "Detection of atrial fibrillation in eeg hand-held devices using a random forest classifier," in *2017 Computing in Cardiology (CinC)*. IEEE, 2017, pp. 1–4.

- [17] P. Warrick and M. N. Homsí, "Cardiac arrhythmia detection from ecg combining convolutional and long short-term memory networks," in *2017 Computing in Cardiology (CinC)*. IEEE, 2017, pp. 1–4.
- [18] F. Hernández, D. Méndez, L. Amado, and M. Altuve, "Atrial fibrillation detection in short single lead ecg recordings using wavelet transform and artificial neural networks," in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2018, pp. 5982–5985.
- [19] M. Weenk, H. van Goor, M. van Acht, L. J. Engelen, T. H. van de Belt, and S. J. Bredie, "A smart all-in-one device to measure vital signs in admitted patients," *PloS one*, vol. 13, no. 2, p. e0190138, 2018.
- [20] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [21] M. Altuve, P. Lizarazo, and J. Villamizar, "Human activity recognition using improved complete ensemble emd with adaptive noise and long short-term memory neural networks," *Biocybernetics and Biomedical Engineering*, vol. 40, no. 3, pp. 901 – 909, 2020.
- [22] M. Schuster and K. K. Paliwal, "Bidirectional recurrent neural networks," *IEEE transactions on Signal Processing*, vol. 45, no. 11, pp. 2673–2681, 1997.
- [23] A. V. Oppenheim and R. W. Schaffer, *Discrete-time signal processing*. Pearson Education, 2014.
- [24] I. Silva and G. B. Moody, "An open-source toolbox for analysing and processing physionet databases in matlab and octave," *Journal of open research software*, vol. 2, no. 1, 2014.
- [25] G. B. Moody and R. G. Mark, "The impact of the mit-bih arrhythmia database," *IEEE Engineering in Medicine and Biology Magazine*, vol. 20, no. 3, pp. 45–50, 2001.



Miguel Altuve (IEEE S'01-M'12-SM'17) Doctor en Procesamiento de Señales y Telecomunicaciones (2011), Universidad de Rennes 1, Rennes, Francia, Magíster en Ingeniería Electrónica (2006), Universidad Simón Bolívar, Caracas, Venezuela, Ingeniero Electrónico (2002), Universidad Nacional Experimental Politécnica de las Fuerzas Armadas, Maracay, Venezuela. Consultor Externo (2020), Universidad Internacional de Valencia, España, Profesor Asociado (2015-2019), Facultad de Ingeniería Eléctrica y Electrónica, Universidad Pontificia Bolivariana seccional Bucaramanga, Colombia, Profesor Asistente (2005-2014), Departamento de Tecnología Industrial, Universidad Simón Bolívar, Caracas, Venezuela. Sus intereses de investigación incluyen el procesamiento digital de señales y el aprendizaje automático en bioingeniería.



Fabio Hernández Ingeniero Electrónico (2016) y Magíster en Ingeniería Electrónica (2020), Universidad Pontificia Bolivariana seccional Bucaramanga, Colombia. Ingeniero de Investigación y Desarrollo, Stericlinic SAS, Bucaramanga, Colombia, (2016, 2018), Profesor Cátedra, Universidad Pontificia Bolivariana, Bucaramanga, Colombia, (2017-2018), Ingeniero de Investigación y Desarrollo (remoto), Golden Security Services, Hallandale Beach, EEUU (2018-actual). Sus intereses de investigación incluyen procesamiento digital de señales, aprendizaje automático, Internet de las cosas y desarrollo de productos basados en sistemas embebidos para uso industrial.