

# Desarrollo de un Clasificador de Latidos de ECG con Capacidad de Generalización

M Llamedo Soria<sup>1,2,3</sup>, JP Martínez Cortés<sup>1,3</sup>

<sup>1</sup>Grupo de Tecnologías de las Comunicaciones, Instituto de Investigación en Ingeniería de Aragón, Universidad de Zaragoza, {llamedom, jpmart}@unizar.es

<sup>2</sup>Universidad Tecnológica Nacional, Buenos Aires, Argentina

<sup>3</sup>CIBER de Bioingeniería, Biomateriales y Nanomedicina (CIBER-BBN)

## Resumen

*En este trabajo estudiamos el desempeño de modelos de características seleccionados mediante un algoritmo de búsqueda flojante, cuyo objetivo fue la capacidad de generalización. Las características utilizadas se relacionan con el ritmo cardíaco y la morfología de cada latido. Los experimentos de clasificación se realizaron en bases de datos públicas, disponibles en Physionet, siguiendo las recomendaciones AAMI-EC57. El mejor modelo encontrado consta de 8 características y fue entrenado y evaluado en conjuntos de datos completamente disjuntos. Los resultados obtenidos fueron: exactitud global de 93%; para la clase de latidos normales, sensibilidad (S) 95%, valor predictivo positivo ( $P^+$ ) 98%; para la clase de latidos supraventriculares S 77%,  $P^+$  39%; para los latidos ventriculares S 81%,  $P^+$  87%. Este modelo de clasificación contiene menos características y su rendimiento es mejor que otros métodos actuales.*

## 1. Introducción

Las enfermedades cardiovasculares son en la actualidad la principal causa aislada de muerte en los países desarrollados de acuerdo con sus agencias de salud. El análisis de la señal electrocardiográfica (ECG) provee una técnica de bajo coste y no invasiva para el estudio de la función cardíaca. Uno de los análisis importantes a realizar sobre el ECG es la clasificación de latidos, importante en el estudio de arritmias.

Diversos algoritmos para la clasificación de latidos fueron desarrollados en la última década [1–4], pero pocos trabajos adoptaron metodologías similares [2, 3], dificultando así la comparación de los resultados. Para aliviar este inconveniente, la *Association for the Advancement of Medical Instrumentation* (AAMI) hizo una serie de recomendaciones respecto a la presentación de resultados y a la definición de clases que ha ido ganando aceptación hasta la actualidad [5]. La recomendación AAMI establece 5 clases, que son los latidos Normales y de bloqueo de ramas (N), los ectópicos Ventriculares (V) o Supraventriculares (S), los latidos de Fusión (F) y aquellos dudosos o que no pueden clasificarse (Q). La mayoría de los traba-

jos revisados abordó la clasificación entre las clases N y V, pero muy pocos han desarrollado clasificadores multiclase [2, 3]. Respecto a la división de los datos para establecer los conjuntos de entrenamiento y evaluación, algunos autores dividieron latido por latido. Esta estrategia, conlleva el inconveniente de incluir latidos del mismo sujeto tanto en el entrenamiento como en la evaluación del clasificador [4]. Este enfoque produce un sesgo optimista en la estimación del desempeño, siendo más adecuada la división orientada a sujetos [2].

El objetivo de este trabajo es desarrollar y evaluar un clasificador de latidos que incluya las características más relevantes, de manera de obtener el mejor desempeño en distintas bases de datos. Este algoritmo será completamente automático, cumplirá con las recomendaciones AAMI, y estará basado en un clasificador simple, con características robustas y con un claro significado fisiológico. El clasificador desarrollado será comparado con el mejor clasificador multiclase publicado [2].

## 2. Metodología

### 2.1. Bases de datos de ECG

En este trabajo hemos utilizado la base de datos (BD) MIT-BIH Arrhythmia (MIT-BIH-AR), tanto para entrenamiento como para evaluación del desempeño. Adicionalmente hemos utilizado la BD MIT-BIH Supraventricular Arrhythmia (MIT-BIH-SUP) para propósitos de validación. Ambas bases de datos se pueden obtener de manera libre en Physionet [6]. En MIT-BIH-AR se adoptó la misma división de entrenamiento ( $DS1$ ) y evaluación ( $DS2$ ) utilizada en [2] para facilitar la comparación de resultados. La clase Q ha sido descartada en este trabajo debido a que se encuentra marginalmente representada en ambas bases de datos. Una limitación similar ocurre con la clase F, que está pobremente representada en ambas bases de datos, razón por la cual se planteó un etiquetado alternativo al AAMI (denominado AAMI2 en este trabajo). El mismo consiste en considerar a las clases de fusión (entre un latido normal y ventricular) y ventricular como la misma clase ventricular extendida ( $V'$ ). La división de los datos y la presencia de

Conj. de datos	MIT-BIH-AR Arrhythmia					#Reg.
	prop.	N	S	V	F	
DS1	entr.	45673	929	3755	412	22
DS2	eval.	44053	1833	3202	388	22
MIT-BIH-SUP	val.	161902	12083	9897	193	78

Conj. de datos	MIT-BIH-AR registros
DS1	101, 106, 108, 109, 112, 114, 115, 116, 118, 119, 122, 124, 201, 203, 205, 207, 208, 209, 215, 220, 223, 230
DS2	100, 103, 105, 111, 113, 117, 121, 123, 200, 202, 210, 212, 213, 214, 219, 221, 222, 228, 231, 232, 233, 234

**Tabla 1.** Esquema de la división de las BD MIT-BIH-AR y MIT-BIH-SUP. Los registros con latidos marcapaseados fueron excluidos. Las clases de latidos son Normales (N), Supraventriculares (S), Ventriculares (V) y de Fusión (F). También se observa en la otra tabla la presencia de registros en los conjuntos de entrenamiento (DS1) y evaluación (DS2).

clases en cada BD está resumida en la tabla 1.

## 2.2. Procesado de señales

La frecuencia de muestreo de la MIT-BIH-SUP fue primero convertida a 360 Hz para igualarla a la MIT-BIH-AR. El remuestreo se realizó con un filtro pasa-bajos FIR de décimo orden. Todos los registros han sido primero pre-procesados para la eliminación de ruidos según se describe en [2]. Algunas de las características extraídas se calculan de la transformada discreta wavelet (DWT) del ECG. Como función prototipo usamos la derivada de una función de suavizado (spline cuadrática), de esta manera se obtiene la derivada suavizada del ECG para cada escala analizada en la DWT. Por este motivo, la DWT contiene en forma de máximos y cruces por cero la información importante del ECG. Referimos a [7] para los detalles en la implementación de la DWT para delineación de ECG. Siguiendo las conclusiones de [7], el análisis de la DWT permite un análisis robusto a las interferencias típicas presentes en registros de ECG, por lo tanto las características extraídas de la DWT podrían heredar dichas propiedades.

## 2.3. Clasificación de latidos: clasificadores y características

Bajo la asunción de características independientes y normalmente distribuidas, utilizando el criterio de máximo *a posteriori* (MAP) obtenemos las conocidas funciones de clasificación cuadráticas. La función discriminante cuadrática para un latido representado por un vector de características  $\mathbf{x}$ , de la clase  $i$ -ésima puede escribirse

$$g_i(\mathbf{x}) = -\frac{1}{2}\mathbf{x}^T\boldsymbol{\Sigma}_i^{-1}\mathbf{x} + \boldsymbol{\mu}_i^T\boldsymbol{\Sigma}_i^{-1}\mathbf{x} - \frac{1}{2}\boldsymbol{\mu}_i^T\boldsymbol{\Sigma}_i^{-1}\boldsymbol{\mu}_i - \frac{1}{2}\log(|\boldsymbol{\Sigma}_i|) + \log(P(\omega_i)). \quad (1)$$

La regla de clasificación asigna  $\mathbf{x}$  a la clase  $i$  que tenga máxima probabilidad *a posteriori*  $g_i(\mathbf{x})$ . Siendo  $\boldsymbol{\mu}_i$ ,  $\boldsymbol{\Sigma}_i$  y  $P(\omega_i)$  el vector medio, la matriz de covarianza y la probabilidad *a priori* de la  $i$ -ésima clase. Los valores de  $\boldsymbol{\mu}_i$  y  $\boldsymbol{\Sigma}_i$  se calculan como la media y la matriz de covarianza muestral, mientras que las probabilidades *a priori* se consideraron iguales para todas las clases. En el caso que la matriz de covarianza  $\boldsymbol{\Sigma}$  sea la misma para todas las clases ( $\boldsymbol{\Sigma}_i = \boldsymbol{\Sigma}_j = \boldsymbol{\Sigma}$ ,  $\forall i \neq j$ ), el clasificador discriminante cuadrático (QDC) se convierte en lineal en  $\mathbf{x}$  dando lugar al LDC. En este caso  $\boldsymbol{\Sigma}$  puede estimarse como la matriz de covarianza muestral pesada

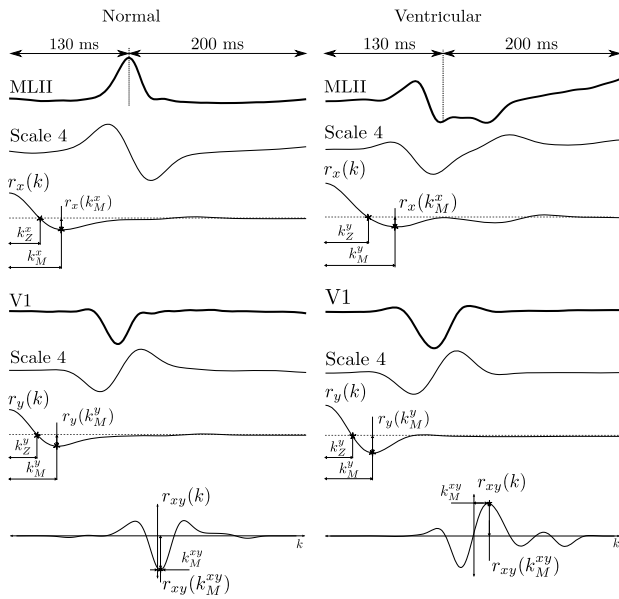
$$\boldsymbol{\Sigma} = \frac{\sum_{i=1}^C w_i \sum_{m=1}^{M_i} (\mathbf{x}_i(m) - \boldsymbol{\mu}_i) \cdot (\mathbf{x}_i(m) - \boldsymbol{\mu}_i)^T}{\sum_{i=1}^C w_i \cdot M_i}. \quad (2)$$

La posibilidad de asignar más relevancia a algunas clases es de mucha importancia en esta aplicación dado que la clase normal por lo general está al menos un orden de magnitud más representada que el resto. En el resto del trabajo referiremos al clasificador LDC cuando  $\mathbf{w}_i = \mathbf{w}_j$ ,  $\forall i \neq j$ , en cualquier otra situación lo denominaremos clasificador lineal compensado (LDC-C).

Siguiendo las conclusiones de trabajos anteriores [1,2], estudiamos la utilidad de características de ritmo y morfología. Como características de ritmo estudiamos aquellas obtenidas de la secuencia de intervalos  $RR$ , como  $RR[i-1]$ ,  $RR[i]$  y  $RR[i+1]$  para describir la evolución local del ritmo cardíaco. Para estudiar la variación local del ritmo, se definió la característica  $RR_V[i] = \sum_{j=-1}^1 |dRR[i-j]|$  (siendo  $dRR[i] = RR[i] - RR[i-1]$ ). También incluimos estimaciones del ritmo local y global mediante el intervalo medio en los últimos 1, 5, 10 y 20 minutos ( $RR_P$  siendo  $P \in \{1, 5, 10, 20\}$  el tiempo de promediado).

Como características morfológicas consideramos primeramente la anchura del complejo QRS. A partir del loop del vectocardiograma (VCG) construido con las dos derivaciones disponibles, calculamos el ángulo y módulo del máximo vector hallado en el complejo QRS. Otras características morfológicas fueron calculadas de la cuarta escala de la DWT, debido a la buena proyección que tiene el complejo QRS en esta escala (comprende desde 12.25 a 22.5 Hz). A partir de esta escala, se calculó la secuencia de autocorrelación para cada derivación ( $r_x(k)$  y  $r_y(k)$ ) y la correlación cruzada entre ellas ( $r_{xy}(k)$ ), en una ventana que comienza 130 ms antes y termina 200 ms después, del punto fiducial. Luego para  $r_{xy}(k)$  se extrajeron dos características, el máximo absoluto y su posición, mientras que para  $r_x$  y  $r_y$  el primer cruce por cero, y la posición y amplitud del primer máximo de módulo, como se ilustra en la Figura 1. Estas características morfológicas tienen la particularidad que estarán sincronizadas en tiempo aún en el caso que el punto fiducial no esté localizado con precisión.

El conjunto completo consta de 39 características relacionadas con el ritmo y la morfología del complejo QRS. Es sabido que aquellos modelos con muchas características tienden a sobre-ajustarse a los ejemplos presentados durante el entrenamiento, perdiendo capacidad de generalización. Por este motivo, utilizaremos un algoritmo secuencial de búsqueda de características flotante (SFFS) [8] para encontrar el modelo de características más pequeño y con el mejor desempeño.



**Figura 1.** Características calculadas de las secuencias de autocorrelación de escala 4 de la DWT del ECG, para un latido normal y uno ventricular. Se muestran las señales de autocorrelación para ambas derivaciones ( $r_x$  y  $r_y$ ), como también la correlación cruzada ( $r_{xy}$ ) en la parte inferior. Las características calculadas de cada señal se indican con un asterisco.

## 2.4. Esquema experimental

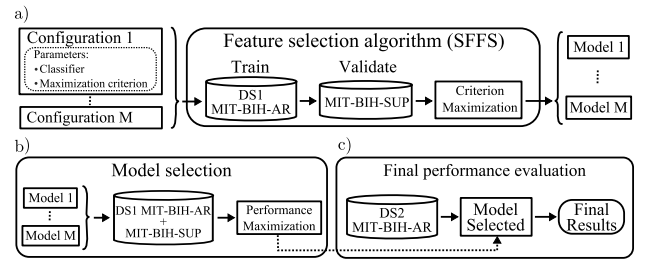
En este trabajo nos interesamos en encontrar un modelo reducido y del mejor rendimiento posible en las BD utilizadas. Podemos describir el experimento en tres pasos:

1) El primer paso consistió en encontrar el mejor modelo entre el conjunto total de características utilizando como entrenamiento  $DS1$  de MIT-BIH-AR y como validación MIT-BIH-SUP, como se muestra en la Figura 2a. En cada iteración del algoritmo SFFS, el modelo se entrena en  $DS1$  y su desempeño se evalúa en MIT-BIH-SUP, asegurando la capacidad de generalización del modelo seleccionado. Se estudiaron varias configuraciones de búsqueda, variando tanto el clasificador utilizado (LDC, LDC-C y QDC), como el criterio de optimización. Los criterios utilizados han sido los promedios tanto del valor predictivo positivo como de la sensibilidad de cada clase ( $J_{P+}$  y  $J_S$ ). La compensación utilizada para el caso del LDC-C ha sido  $w_N = 1$ ,  $w_S = 10$  y  $w_V = 10$ .

2) El segundo paso será la elección del mejor modelo, entre aquellos seleccionado por el SFFS para cada configuración de búsqueda en el paso anterior. Para ello, se evaluaron dichos modelos en la unión de los conjuntos de entrenamiento y validación por medio de una validación cruzada con  $k = 10$  registros, como puede verse en la Figura 2b.

3) Finalmente el modelo seleccionado en el paso anterior se entrena en  $DS1$  y se lo evalúa en  $DS2$  para la comparación final con [2], como se muestra en la Figura 2c.

Todos los experimentos descritos en este trabajo tienen como objetivo clasificar automáticamente las tres clases de AAMI2 (N, S, V'), dado que la clase F de AAMI está po-



**Figura 2.** La figura muestra un esquema de los experimentos realizados en este trabajo. En el panel a) se resume la búsqueda de características, indicando el conjunto de entrenamiento y validación, como también los parámetros para guiar la búsqueda. En el panel b) se muestra cómo se obtiene el modelo final entre los obtenidos en a). Finalmente en c) el mejor modelo es evaluado en el conjunto de evaluación.

brevemente representada en las BD que utilizamos.

## 3. Resultados

Los resultados de los experimentos descritos en la sección anterior se presentan en las Tablas 2 y 3. En la tabla 3, la evaluación del desempeño balanceada significa que las filas de la matriz de confusión fueron escaladas de manera tal que todas sumen lo mismo, es decir cada clase tenga igual representación. El mejor modelo encontrado consta de 8 características, usando un clasificador LDC-C; dichas características son  $\ln(RR[i])$ ,  $\ln(RR[i + 1])$ ,  $\ln(RR_1)$ ,  $\ln(RR_{20})$ ,  $k_Z^x$ ,  $k_Z^y$ ,  $k_M^x$  y  $k_M^y$ .

## 4. Discusión y conclusiones

El mejor modelo encontrado consta de 8 características, todas ellas mediciones de intervalos temporales. Esto puede explicarse debido a que los registros utilizados en los experimentos no siempre contienen las mismas derivaciones, y por lo tanto la capacidad de clasificación de aquellas características basadas en amplitudes se ve seriamente disminuida. Las primeras 4 características del modelo están claramente relacionadas a la evolución del ritmo cardíaco, mientras que las otras cuatro pueden interpretarse como mediciones alternativas de la anchura del complejo QRS, y por lo tanto morfológicas. Como resultado, el modelo obtenido tiene la ventaja evidente de un menor tamaño, lo que redundante tanto en un ahorro computacional como en una mejor estimación de los parámetros del modelo durante el entrenamiento. Como ventaja adicional, este modelo depende exclusivamente de la detección de cada latido, haciendo este clasificador especialmente útil en aquellos casos donde la delineación de las ondas del ECG no puede realizarse de manera confiable. En este trabajo hemos prestado especial atención a la capacidad de generalización del modelo, para ello incluimos en el desarrollo la MIT-BIH-SUP [9], disponible en Physionet [6] de manera libre. La limitación de no contar con una clase de fusión correctamente representada en ambas BD, fue resuelta por medio del etiquetado alternativo AAMI2. Este reetiquetado puede justificarse dado que la clase fusión incluye los latidos ventriculares que suceden al mismo tiempo que los normales, por lo que asumir que se trata de latidos ventriculares

Configuración de búsqueda			Evaluación del modelo								
Clasificador	Crit. Opt.	# Caract.	Normal		Suprav.		Ventr.		Total		
			S	P <sup>+</sup>	S	P <sup>+</sup>	S	P <sup>+</sup>	A	S	P <sup>+</sup>
<b>LDC-C</b>	$J_{P^+}$	<b>8</b>	<b>93</b>	<b>98</b>	<b>78</b>	<b>40</b>	<b>68</b>	<b>70</b>	<b>91</b>	<b>80</b>	<b>70</b>
QDC	$J_{P^+}$	7	80	98	7	12	89	22	77	59	44
LDC	$J_S$	10	92	98	74	37	70	67	89	78	67
QDC	$J_S$	9	87	98	43	32	80	33	84	70	55
de Chazal et al. [2]		48	87	98	57	30	63	36	84	69	55

**Tabla 2.** Resumen de los modelos que tuvieron mejor desempeño separando las 3 clases AAMI2 durante la búsqueda SFFS, según la Figura 2b. El mejor modelo ha sido resaltado en negrita y se utilizará para la evaluación final. Los resultados están en porcentajes.

de Chazal et al. [2]					Este trabajo				
Referencia	Algoritmo				Referencia	Algoritmo			
	n	s	v'	Total		n	s	v'	Total
N	40718	1863	1677	44258	N	41950	2002	236	44188
S	307	1361	169	1837	S	216	1422	197	1835
V'	235	845	2529	3609	V'	473	222	2911	3606
Total	41260	4069	4375	49704	Total	42639	3646	3344	49629

Modo de evaluación	Clasificador	# Caract.	Normal		Suprav.		Ventr.		Total		
			S	P <sup>+</sup>	S	P <sup>+</sup>	S	P <sup>+</sup>	A	S	P <sup>+</sup>
Desbalanceado	Este trabajo	8	95	98	77	39	81	87	93	84	75
	de Chazal et al. [2]	48	92	99	74	33	70	58	90	79	63
Balanceado	Este trabajo	8	95	79	77	88	81	88	84	84	85
	de Chazal et al. [2]	48	92	80	74	73	70	84	79	79	79

**Tabla 3.** Comparación del desempeño del modelo sugerido en este trabajo y [2] separando las 3 clases AAMI2 en DS2 de MIT-BIH-AR. Ambos modelos fueron entrenados en DS1 de MIT-BIH-AR. Arriba se muestran las dos matrices de confusión obtenidas por ambos modelos, y debajo se resume el desempeño obtenido. El desempeño está en porcentajes para ambos tipos de evaluación.

puede ser una alternativa razonable.

De los resultados obtenidos de la búsqueda de modelos (Tabla 2), puede verse que varios modelos superan el desempeño en los conjuntos de entrenamiento y validación al clasificador de [2]. El modelo seleccionado en este trabajo, corrobora su desempeño como puede verse en la Tabla 3 y evidencia su capacidad de generalización ya que el conjunto de evaluación se ha sido utilizado exclusivamente para tal fin. Cabe destacar que el desempeño de ambos clasificadores comparados en este trabajo durante la fase de desarrollo (Tabla 3), es inferior para todas las clases que el obtenido en la fase de evaluación final (Tabla 3). Este fenómeno ya fue reportado en [2], sugiriendo que el desempeño real del clasificador debe corroborarse en otras BD. A pesar de esta limitación, el grado de generalización de este modelo debería ser superior al de los trabajos revisados, que solamente consideraron para el desarrollo la BD MIT-BIH-AR. Finalmente, los resultados presentados en este trabajo constituyen una mejora respecto al tamaño del modelo y al desempeño obtenido.

## Agradecimientos

Este trabajo fue financiado por los proyectos TEC2010-21703-C03-02 de CICYT y GTC T-30 de la DGA. El CIBER de Bioingeniería, Biomateriales y Nanomedicina es una iniciativa de ISCIII.

## Referencias

[1] Hu YH, Palreddy S, Tompkins W. A patient-adaptable ecg beat classifier using mixture of experts approach. IEEE Transactions on Biomedical Engineering 1997; 44:891–899.

[2] de Chazal P, O'Dwyer M, Reilly RB. Automatic classification of heartbeats using ecg morphology and heartbeat interval features. IEEE Transactions on Biomedical Engineering 2004;51:1196–1206.

[3] Park K, Cho B, Lee D, Song S, Lee J. Hierarchical support vector machine. In Computers in Cardiology 2008, volume 35. IEEE Computer Society Press, 2008; 229–232.

[4] Ince T, Kiranyaz S, Gabbouj M. A generic and robust system for automated patient-specific classification of ecg signals. IEEE Transactions on Biomedical Engineering 2009;56:1415–1426.

[5] Testing and reporting performance results of cardiac rhythm and st-segment measurement algorithms. American National Standard, ANSI/AAMI/ISO EC57, 1998–(R)2008.

[6] Goldberger AL, et al. PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. Circulation 2000; 101(23):e215–e220.

[7] Martínez JP, Almeida R, Olmos S, Rocha A, Laguna P. A wavelet-based ecg delineator: Evaluation on standard databases. IEEE Transactions on Biomedical Engineering 2004;51:570–581.

[8] Pudil P, Novovicova J, Kittler J. Floating search methods in feature selection. Pattern Recognition Letters 1994;15(11):1119–1125.

[9] Mark R, Moody G, Greenwald S. Mit-bih supraventricular arrhythmia database. <http://www.physionet.org/physiobank/database/svdb/>, 1990.