

Algoritmos genéticos aplicados a la optimización de características en la clasificación de arritmias cardiacas utilizando los clasificadores KNN y naive Bayes

Christian Padilla-Navarro¹, Sheila González-Reyna³,
Gabriel Aguilera-González¹, Martín Ortega-Yepez¹, Sandra Bombela-Jiménez¹,
Manuel Rangel-Huerta¹, Carlos Lino-Ramírez²

¹ Universidad Politécnica de Juventino Rosas, Gto., México

² Instituto Tecnológico de León, León, Gto., México

³ Instituto Tecnológico Superior de Irapuato, Irapuato, Gto., México

{jpadilla_ptc,direccion_ite,
martin.ortega,sandra.bombela,manuel.rangel}@upjr.edu.mx
sheila.gonzalez@itesi.edu.mx, carlos.lino@itleon.edu.mx

Resumen. Muchas muertes en el mundo suceden a consecuencia de enfermedades cardiovasculares. El método propuesto combina metaheurísticas —Algoritmos Genéticos (AG)—, y los clasificadores KNN y Naive Bayes. Las pruebas se realizaron a través de una base de datos del Massachusetts Institute of Technology-Beth Israel Hospital (MIT-BIH) [1]. Las metaheurísticas se implementan para mejorar el rendimiento de los clasificadores. Los resultados experimentales demuestran que se logra hasta un 94 % de precisión en la clasificación.

Palabras clave: Electrocardiograma, arritmia, AG, KNN, naive Bayes.

Genetic Algorithms Applied to the Optimization of Characteristics in the Classification of Cardiac Arrhythmias Using the KNN and Naive Bayes Classifiers

Abstract. Many deaths in the world happen as a result of cardiovascular diseases. The proposed method combines metaheuristic – Genetic Algorithms (AG) – and the KNN and Naive Bayes classifiers. The tests were performed through a database of the Massachusetts Institute of Technology-Beth Israel Hospital (MIT-BIH) [1]. Metaheuristics are implemented to improve the performance of classifiers. Experimental results show that up to 94 % accuracy is achieved in the classification.

Keywords: Electrocardiogram, arrhythmia, GA, KNN, naive Bayes.

1. Introducción

Las arritmias son alteraciones de la frecuencia cardíaca o del ritmo cardíaco [2]. Hay en existencia dos grupos para las arritmias, el primero incluye las taquicardias y fibraciones ventriculares, las cuales provocan una grave amenaza para la vida del paciente, se requiere terapia inmediata con desfibrilador. El segundo grupo está compuesto por arritmias que no ponen en peligro la vida pero también requiere terapia [3].

Las Enfermedades Cardiovasculares (ECV) son la principal causa de muerte en todo el mundo. Cada año mueren más personas por ECV que por cualquier otra causa. Se calcula que en 2012 murieron por esta causa 17.5 millones de personas, lo cual representa un 31 % de todas las muertes registradas en el mundo [4]. Para las personas con ECV o con alto riesgo cardiovascular (debido a la presencia de uno o más factores de riesgo, como la hipertensión arterial, la diabetes, la hiperlipidemia o alguna ECV ya confirmada), son fundamentales la **detección precoz y el tratamiento temprano**, por medio de servicios de orientación o la administración de fármacos, según corresponda [4].

Ante una situación de emergencia lo más importante a considerar es el tiempo que se demora en atender dicha circunstancia [5]. Un paso importante hacia la identificación de una arritmia es la clasificación de los latidos del corazón, debido a la gran cantidad de información para analizar, los especialistas y aparatos tienen dificultades para detectar las irregularidades [6] y por lo tanto, cualquier sistema que ayude a realizar este proceso sería de gran apoyo en la detección de anomalías en las señales, siendo el objetivo principal de esta investigación. El problema a resolver es asegurar y determinar que la clasificación de los electrocardiogramas sea precisa, haciendo un sistema eficaz que en caso de salir alguna anomalía en el ECG notifique al personal encargado del paciente [3].

2. Estado del arte

Un electrocardiograma (ECG) es un procedimiento de diagnóstico con el que se obtiene un registro de la actividad eléctrica del corazón. Es la técnica más usada para el estudio electrofisiológico del corazón, debido a que es un método no invasivo y permite registrar la actividad eléctrica del corazón desde la superficie del cuerpo humano [7].

Se ha descubierto que se pueden detectar muchos de los padecimientos del corazón (arritmia, fibrilación auricular, auriculoventricular (AV) disfunciones, y la enfermedad arterial coronaria, etc.) en base a las anomalías en las señales de ECG [8].

Estas señales son clasificadas en base a sus características para determinar a que clase de arritmia pertenece.

Nait-Hamound M. y Moussaoui A. [9] presentaron dos métodos para la clasificación de arritmias multiclase aplicando Análisis de Componentes Principales (PCA), las Máquinas de Soporte de Vectores Difusos (FSVM, por sus siglas en inglés), y la Closterización Desbalanceada (UC, por sus siglas en inglés).

Thanapatay et al. [10] propusieron un nuevo método de clasificación de ECG usando PCA y SVM.

Los clasificadores tienen diferentes aplicaciones dentro de las señales de ECG, su función principal es separar a través de sus clases los distintos tipos de arritmias.

Nait-Hamoud et al. [11], muestran dos nuevos métodos para la clasificación de ECG para discriminar cinco tipos de ritmos cardíacos combinando Análisis de Componentes Principales (PCA) y Modified Fuzzy One-Against-One (MFOAO).

Molina F. y Benítez D. [12] describen las características más importantes de las señales de ECG mediante el filtrado con la transformada de Wavelet para detectar los puntos significativos en las señales y clasificarlas.

Gutiérrez et al. [13] desarrollaron e implementaron un filtro digital recursivo de señales de ECG para detectar complejos QRS basándose en Wavelet de Haar usando la base de datos de ECG del Massachusetts Institute of Technology-Beth Israel Hospital (MIT-BIT).

Fira et al. [14] investigaron los resultados de la clasificación de la comprensión de señales de ECG basándose en diferentes tipos de matrices de proyección. Se clasifican las señales comprimidas usando KNN, se analiza respecto a las matrices de proyección y a los resultados obtenidos.

Mientras tanto en Martínez [15] se propone un modelo de clasificación de señales de ECG usando sistemas inteligentes para la detección de anomalías en el corazón usando Algoritmos Genéticos y Redes Neuronales (RN).

Algunos de los padecimientos requieren de un monitoreo constante de la actividad eléctrica del corazón para detectar alteraciones periódicas en las señales. Sannino et al. [16] muestran un sistema para monitorear aceleraciones cardíacas anormales en tiempo real, se basa en un método que analiza las señales por medio de un algoritmo basado en umbral para reducir el ruido.

Elena et al. presenta [17] un algoritmo eficiente para la compresión de ECG y monitoreo en tiempo real, actualizándose con cada nueva señal de entrada por medio de un umbral de ruido óptimo.

Hoffman et al. [25] presentan un algoritmo de procesamiento de ECG basado en Multiple Model Adaptive Estimator (MMAE) para un sistema de vigilancia fisiológica, utilizando veinte señales de la base de datos de ECG del MIT.

En arritmias cardíacas la cantidad de datos que se procesa para la detección suele ser muy grande, en consecuencia aumenta el tiempo de computación, así como los requerimientos de memoria y la cantidad de características usadas. En diferentes investigaciones se busca reducir el tiempo de ejecución y la cantidad de recursos computacionales, incrementar la efectividad y hacerla más precisa con menor cantidad de datos.

Kim en la investigación [18] propone Cuatro Niveles Vectoriales (Quad Level Vector-QLV) para el flujo de compresión y flujo de clasificación para mejorar el rendimiento con baja complejidad computacional. En el algoritmo de clasificación se emplean la segmentación de los latidos del corazón y los métodos de detección de R-peak.

Bilgin et al. en la investigación [19] aplican JPEG2000 para comprimir las señales de ECG usando la base de datos de MIT-BIH como ejemplo.

Cuesta-Frau et al. [20], muestran la comparación de distintos métodos que son usados para extraer la información principal de señales de ECG, también se busca reducir el costo computacional usando la base de datos del MIT.

Kallas et al. [21] mostraron la combinación de la clasificación de arritmias multiclase usando SVM con la extracción de características usando PCA en las señales de ECG.

Mientras tanto Abdeel-Badeeh M et al. [8] aplicó el aprendizaje automático de máquinas modernas en el diagnóstico de ECG.

Soman T., Bobbie P. [22] utilizaron sistemas de aprendizaje automático, Oner, J48 y Naive Bayes para clasificar el conjunto de datos de arritmias.

Gao et al. [23] describen un sistema para detectar arritmias cardiacas basandose en una red neuronal artificial bayesiana (ANN). Su desempeño en esta tarea se compara con otros clasificadores como Naive Bayes, árboles de decisión, regresión logística y redes RBF.

En Daamouche [24] se logró optimizar el nivel de clasificación usando Máquinas de Soporte Vectorial (Support Vector Machines-SVM, Optimización por Enjambre de Partículas (PSO) y se usaron Wavelets para reducir el ruido.

3. Metaheurísticas

Las metaheurísticas pueden concebirse como estrategias generales de diseño de procedimientos heurísticos (capacidad de un sistema para realizar de forma inmediata innovaciones positivas) para la resolución de problemas con un alto rendimiento o al diseño de alguno de los tipos fundamentales de procedimientos heurísticos de solución de un problema de optimización [26].

3.1. Algoritmos genéticos

Los Algoritmos Genéticos son métodos adaptativos, generalmente usados en problemas de búsqueda y optimización de parámetros, basados en la reproducción sexual y en el principio supervivencia del más apto.

Más formalmente, y siguiendo la definición dada por Goldberg, los Algoritmos Genéticos son algoritmos de búsqueda basados en la mecánica de selección natural y de la genética natural. Combinan la supervivencia del más apto entre estructuras de secuencias con un intercambio de información estructurado, aunque aleatorizado, para constituir así un algoritmo de búsqueda que tenga algo de las genialidades de las búsquedas humanas [27].

Para alcanzar la solución a un problema, se parte de un conjunto inicial de individuos, llamado población, generado de manera aleatoria. Cada uno de estos individuos representa una posible solución al problema. Estos individuos evolucionarán tomando como base los esquemas propuestos por Darwin [28] sobre la selección natural, y se adaptarán en mayor medida tras el paso de cada generación a la solución requerida.

3.2. Orígenes

El desarrollo de los Algoritmos Genéticos se debe en gran medida a John Holland, investigador de la Universidad de Michigan. A finales de la década de los 60 desarrolló una técnica que imitaba en su funcionamiento a la selección natural. Aunque originalmente esta técnica recibió el nombre de planes reproductivos, a raíz de la publicación en 1975 de su libro *Adaptation in Natural and Artificial Systems* [29] se conoce principalmente con el nombre de Algoritmos Genéticos.

La Programación Evolutiva surge principalmente a raíz del trabajo *Artificial Intelligence Through Simulated Evolution* de Fogel, Owens y Walsh, publicado en 1966 [30]. En este caso los individuos, conocidos aquí como organismos, son máquinas de estado finito. Los organismos que mejor resuelven alguna de las funciones objetivo obtienen la oportunidad de reproducirse. Antes de producirse los cruces para generar la descendencia se realiza una mutación sobre los padres.

3.3. Bases biológicas

En la naturaleza, los individuos de una población compiten constantemente con otros por recursos tales como comida, agua y refugio. Los individuos que tienen más éxito en la lucha por los recursos tienen mayores probabilidades de sobrevivir y generalmente una descendencia mayor. Al contrario, los individuos peor adaptados tienen un menor número de descendientes, o incluso ninguno. Esto implica que los genes de los individuos mejor adaptados se propagarán a un número cada vez mayor de individuos de las sucesivas generaciones.

La combinación de características buenas de diferentes ancestros puede originar en ocasiones que la descendencia esté incluso mejor adaptada al medio que los padres. De esta manera, las especies evolucionan adaptándose más y más al medio a medida que transcurren las generaciones [31].

3.4. Operadores genéticos

Selección La selección se realiza por medio del método de Vasconcelos. Para aplicar este método necesitamos ordenar aptitud de todos los individuos, ascendente o descendientemente, y tomar el individuo más apto y el menos apto.

Cruza Se realiza la cruce a partir de dos puntos aleatorios (P1 y P2). De 0 a la posición del primer corte se obtenían los datos del Padre (el mejor individuo), del primer al segundo corte se obtenían los datos de la Madre (el peor individuo), del segundo corte y hasta terminar se agregaban los datos del Padre como se muestra en la Figura 1.

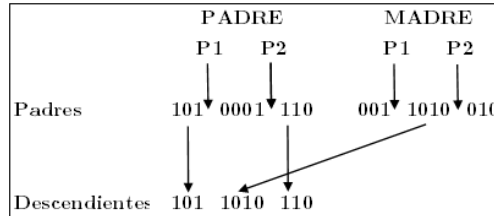


Fig. 1. Operador de cruce basado en un punto de corte.

Muta La muta de un individuo provoca que alguno de sus genes, generalmente uno sólo, varíe su valor de forma aleatoria. La muta se muestra en la Figura 2.

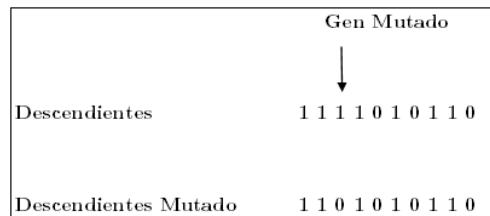


Fig. 2. Operador de mutación.

Elitismo Se clonan los mejores individuos de un porcentaje representativo de la población. Es decir, en caso de tener una población de 100 individuos, y un elitismo del 0.20, pasará a los 20 mejores individuos en espacios aleatorios.

3.5. Algoritmo genético simple

Los Algoritmos Genéticos trabajan sobre una población de individuos. Cada uno de ellos representa una posible solución al problema que se desea resolver. Todo individuo tiene asociado un valor de acuerdo a la aptitud con respecto al problema de la solución que representa (en la naturaleza el equivalente sería una medida de la eficiencia del individuo en la lucha por los recursos). El funcionamiento genérico de un Algoritmo Genético puede apreciarse en el siguiente pseudocódigo:

Algoritmo 1 Algoritmo genético simple.

```
1: Inicializar una población aleatoria
2: mientras No se cumpla el criterio de terminación hacer
3:   Crear población temporal vacía
4:   mientras población temporal no llena hacer
5:     Seleccionar padres
6:     Cruzar padres con probabilidad Pc
7:     si Se ha producido el cruce entonces
8:       Mutar uno de los descendientes con probabilidad Pm
9:       Evaluar descendientes
10:      Añadir descendientes a la población temporal
11:     si no
12:       Añadir padres a la población temporal
13:     fin si
14:   fin mientras
15:   Aumentar contador generaciones
16:   Establecer como nueva población actual la población temporal
17: fin mientras
```

3.6. Algoritmo genético con elitismo

El funcionamiento de un Algoritmo Genético con la inclusión de elitismo puede apreciarse en el siguiente pseudocódigo:

Algoritmo 2 Algoritmo genético con elitismo.

```
1: Inicializar una población aleatoria
2: mientras No se cumpla el criterio de terminación hacer
3:   Crear población temporal vacía
4:   mientras población temporal no llena hacer
5:     Realizar proceso de elitismo, según el porcentaje indicado
6:     Seleccionar padres
7:     Cruzar padres con probabilidad Pc
8:     si Se ha producido el cruce entonces
9:       Mutar uno de los descendientes con probabilidad Pm
10:      Evaluar descendientes
11:      Añadir descendientes a la población temporal
12:     si no
13:       Añadir padres a la población temporal
14:     fin si
15:     Agregar intensificador
16:   fin mientras
17:   Aumentar contador generaciones
18:   Establecer como nueva población actual la población temporal
19: fin mientras
```

4. Datos de prueba

El conjunto de datos experimentales proviene de la bases de datos del sitio Physionet. Los registros fueron obtenidos del conjunto MIT/BIH arrhythmia. Este contiene 48 registros de 30 minutos de duración [1]. Todos sus los registros fueron etiquetados.

El estándar de la Association for the Advancement of Medical Instrumentation (AAMI) [32] recomienda las siguientes consideraciones para etiquetar los tipos de arritmias existentes: ritmo normal (etiquetado como N), latido ectópico supraventricular (S), latido ectópico ventricular (V), latido de fusión (F), y latidos desconocidos (Q). Cualquier registro debería estar representado en alguno de estos tipos. Una descripción completa de todos los registros se incluyen en la Tabla 1, mostrando la equivalencia entre el AAMI y el MIT/BIH.

Extracción de características Tomamos como referencia el trabajo de [33] para realizar la extracción de características usadas en el conjunto de datos del MIT. El vector de características de entrada $x_j = \{x_{1j}, x_{2j}, \dots, x_{pj}\}$, con $p = 100$ está compuesto por:

$$\begin{aligned} x_{1j} &= l_j - l_{j-1} \text{ (intervalo RR)}, \\ x_{2j} &= l_{j-1} - l_{j-2} \text{ (pre-intervalo RR)}, \\ x_{3j} &= l_{j+1} - l_j \text{ (post-intervalo RR)}, \\ x_{4j} &= x_{1j} - x_{2j}, \\ x_{5j} &= x_{3j} - x_{1j}, \\ x_{6j} &= \left(\frac{x_{3j}}{x_{1j}}\right)^2 + \left(\frac{x_{2j}}{x_{1j}}\right)^2 - \left(\frac{1}{3} \sum_{k=1}^3 x_{kj}^2 \log(x_{kj}^2)\right). \end{aligned}$$

Representación PQRST En x_{7j} se cuantifica la disimilitud entre el actual complejo QRS, y la linealidad acumulada de los últimos 10 complejos QRS [34] por medio de un enfoque de distorsión de tiempo dinámico (DTW).

$$\begin{aligned} x_{8j} &= \left| \frac{\max\{QRS_j[t]\}}{\min\{QRS_j[t]\}} \right|, \\ x_{9j} &= \sum_{k=0}^{L_j} QRS_j[t]^2. \end{aligned}$$

Coefficientes de Hermite De x_{10j} a x_{19j} correspondiente a los primeros 10 polinomios de Hermite, usando el software de MATLAB.

Wavelet Daubechies Se utilizó el 4to. nivel de los coeficientes (A-Aproximación, D-Detalle) de la descomposición Wavelet Daubechies, (Db4)

$$\begin{aligned} \text{De } x_{20j} \text{ a } x_{25j} &\Leftarrow A_4, \\ \text{De } x_{26j} \text{ a } x_{31j} &\Leftarrow D_4, \\ \text{De } x_{32j} \text{ a } x_{42j} &\Leftarrow D_3, \\ \text{De } x_{43j} \text{ a } x_{58j} &\Leftarrow D_2, \end{aligned}$$

De x_{59j} a $x_{90j} \leftarrow D_1$,
De x_{91j} a $x_{95j} = \text{var}\{A_4, D_4, D_3, D_2, D_1\}$,
De x_{95j} a $x_{100j} = \text{max}\{A_4, D_4, D_3, D_2, D_1\}$.

5. Definición de la función objetivo

La función objetivo se definió utilizando diversos clasificadores, se ingresaron los vectores característicos hasta obtener el porcentaje de clasificación mayor (siendo 100 el máximo) y el menor número de características del vector.

5.1. Naive Bayes

Se decidió utilizar la función del clasificador Naive Bayes que viene contenida en el software *Weka*, de esta forma tendríamos datos reales y probados.

5.2. KNN

Se aplicó la función del clasificador KNN que viene contenida en el software *Weka*, fue utilizada para 1 y 3 vecinos.

6. Implementación de las metaheurísticas

6.1. Algoritmos genéticos

Los operadores genéticos fueron utilizados de la siguiente manera:

Selección. La selección se realizó a través del método Vasconcelos. Para aplicar este método necesitamos ordenar el fitness de todos los individuos, ascendente o descendientemente, y tomar el mejor y el peor individuo.

Cruza. Realizamos la cruce a partir de dos puntos aleatorios. Se toma de la cadena del Padre (mejor individuo) desde la posición 0 hasta el primer punto aleatorio, desde el primer punto aleatorio hasta el segundo punto aleatorio se toma de la cadena de datos de la Madre (peor individuo), y finalmente desde el segundo punto hasta terminar la cadena se obtiene del Padre.

Muta. Se muta un porcentaje de la población, se toma un dato de la cadena de manera aleatoria y se cambia su valor de 0 a 1 o de 1 a 0.

Elitismo. Se toma el mejor individuo y se clona un porcentaje de veces.

Algoritmo genético simple para la clasificación de ECG Para realizar el proceso del Algoritmo Genético se utilizó el Algoritmo 3.5.

Algoritmo genético con elitismo para la clasificación de ECG Para realizar el proceso del Algoritmo Genético con elitismo se implementó el Algoritmo 3.6.

7. Pruebas y resultados

Las primeras pruebas se realizaron con los clasificadores 1NN, 3NN y Naive Bayes, sin metaheurísticas. Consideramos todas las características para tener un parámetro inicial y poder comparar. Después, probamos con los mismos clasificadores, pero utilizando los algoritmos genéticos y algoritmos genéticos con elitismo.

Se puede ver la Media del Porcentaje de Clasificación obtenido (ver la Tabla 2) y la Media del Número de Características empleadas en la clasificación (ver la Tabla 3).

8. Conclusiones

Los resultados fueron satisfactorios al implementar las metaheurísticas como optimización en la clasificación de arritmias cardiacas en señales de ECG. Como consecuencia de la reducción de características permite eliminar información innecesaria, reduciendo el ruido y mejorando el desempeño de los clasificadores.

Aplicando los clasificadores KNN y Naive Bayes y las metaheurísticas Algoritmos Genéticos y Algoritmos Genéticos con Elitismo a la Base de Datos MIT-BIH como auxiliares en la clasificación de arritmias cardiacas, el porcentaje de clasificación aumentó hasta en un 79.76%, en el caso de las características se logró reducir la cantidad hasta en un 67%. La reducción de la información innecesaria, quita el ruido en la clasificación, y hace que el porcentaje de la clasificación sea más alto y por tanto la clasificación sea mucho más óptima.

Referencias

1. Goldberger, A.L., Amaral, L.A.N., Glass, L., Hausdorff, J.M., Ivanov, P.Ch., Mark, R.G., Mietus, J.E., Moody, G.B., Peng, C.-K., Stanley, H.E.: PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals. *Circulation* 101(23):e215-e220 (June 2000)
2. MedlinePlus Biblioteca Nacional de Medicina de los EE.UU. (2014)
3. De Chazal, P., O'Dwyer, M., Reilly, R. B.: Automatic classification of heartbeats using ECG morphology and heartbeat interval features. *IEEE Transactions on Biomedical Engineering* 51, pp. 1196–1206 (2004)
4. Organización Mundial de la Salud: Enfermedades cardiovasculares (2015)
5. Fuentes, C. R., Ordoñez, H. F. J.: Capa de sensado de un nodo para monitoreo de signos vitales. *Ingenium* (2010)
6. Médica Sur: Mexicanos desarrollan dispositivo portátil que detecta arritmias cardiacas en tiempo real. Instituto Carlos Slim de la Salud (2016)
7. Arévalo, A. M.: Electrocardiografía ECG DALCAME. Grupo de Investigación Biomédica (2005)
8. Abdeel-Badeeh, M., Revett, K., et al.: Machine Learning in Electrocardiogram Diagnosis. In: International Multiconference on Computer Science and Information Technology, pp. 429–433 (2009)

9. Nait-Hamoud, M.: Two Novel Methods for Multiclass ECG Arrhythmias Classification Based on PCA. In: Fuzzy Support Vector Machine and Unbalanced Clustering IEEE, pp. 140–145 (2010)
10. Thanapatay, D., Suwansaroj, C.: ECG Beat Classification Method for ECG Printout with Principle Components Analysis and Support Vector Machines. In: International Conference on Electronics and Information Engineering (ICEIC), 1, pp. 72–75 (2010)
11. Nait-Hamoud, M. C., Moussaoui, A.: Two novel methods for multiclass ECG arrhythmias classification based on PCA, fuzzy support vector machine and unbalanced clustering. In: Machine and Web Intelligence (ICMWI), 2010 International Conference on, pp. 140–145 (2010)
12. Molina, F. et al.: Desarrollo e Implementación de un Algoritmo Para la Caracterización de los Límites de Forma de Onda de Un Electrocardiograma (ECG) Utilizando Ondillas (Wavelets). En: XIX Jornadas en Ingeniería Eléctrica y Electrónica, 19, pp. 95–105 (2005)
13. Gutiérrez, A., Lara, M.: Evaluación de un Detector de Complejo QRS Basado en la Wavelet de Haar, Usando las Bases de Datos MIT-BIH de Arritmias y Europea del Segmento ST y de la Onda T. *Computación y Sistemas*, 8, pp. 293–302 (2005)
14. Fira, M., Goras, L., et al.: On the Projection Matrices Influence in the Classification of Compressed Sensed ECG Signals. *International Journal of Advanced Computer Science and Applications* (2012)
15. Martínez, A., Rojas, I., et al.: Utilización de Sistemas Inteligentes para la Detección de Problemas de Corazón Mediante ECG. (2005)
16. Sannino, G., Pietro, G.: An Advanced Mobile System for Indoor Patients Monitoring In: Proc. 2nd International Conference on Networking and Information Technology (ICNIT) (2011)
17. Elena, M., Quero, J., Borrego, I.: An optimal technique for ECG noise reduction in real time applications. *Computers in Cardiology*, pp. 225–228 (2006)
18. Kim, H., Yazicioglu, R. F., Merken, P., Van Hoof, C., Yoo, H.-J.: ECG signal compression and classification algorithm with quad level vector for ECG holter system. *IEEE Transactions on Information Technology in Biomedicine*, IEEE, 14, pp. 93–100 (2010)
19. Bilgin, A., Marcellin, M. W., Altbach, M. I.: Compression of electrocardiogram signals using JPEG2000. *IEEE Transactions on Consumer Electronics*, 49, pp. 833–840 (2003)
20. Cuesta-Frau D., et al.: Feature Extraction Methods Applied to the Clustering of Electrocardiographic Signals. A Comparative Study. In: 16th International Conference on Pattern Recognition, pp. 961–964 (2002)
21. Kallas, M., Francis, C., et al.: Multi-Class SVM Classification Combined with Kernel PCA Feature Extraction of ECG Signals. In: International Conference on Telecommunications (2012)
22. Soman, T., et al.: Classification of Arrhythmia Using Machine Learning Techniques. In: ICOSSE (2005)
23. Gao, D., Madden M., et al.: Bayesian ANN Classifier for ECG Arrhythmia Diagnostic System: A Comparison Study. In: IJCNN (2005)
24. Daamouche, A., Hamami, L., et al.: A Wavelet Optimization Approach for ECG Signal Classification. *Biomedical Signal Processing and Control*, 7, pp. 342–349 (2012)
25. Hoffman, G. S., Miller, M., Kabrisky, M., Maybeck, P., Raquet, J.: Novel electrocardiogram segmentation algorithm using a multiple model adaptive estimator.

- In: Decision and Control, Proceedings of the 41st IEEE Conference on, 3, pp. 2524–2529 (2002)
26. Vega, J. M. M., Batista, M. B. M., Pérez, J. A. M.: Metaheurísticas: Una visión global. *Inteligencia artificial: Revista Iberoamericana de Inteligencia Artificial, Asociación Española para la Inteligencia Artificial (AEPIA)*, 7, pp. 7–28 (2002)
 27. Goldberg, D.E.: *Genetic Algorithms in Search, Optimization and Machine Learning*. Addison-Wesley Longman Publishing Co. (1989)
 28. Darwin, Ch.: *On the Origin of Species by Means of Natural Selection*. John Murray (1859)
 29. Holland, J.: *Adaptation in Natural and Artificial Systems*. (1975)
 30. Fogel, L., Owens, A., et al.: *Artificial Intelligence through Simulated Evolution*. (1966)
 31. Beasley, D., Bull, D. R., et al.: An Overview of Genetic Algorithms: Part 1, Fundamentals University computing. pp. 58–69 (1993)
 32. De Chazal, P., O’Dwyer, M., et al.: Automatic Classification of Heartbeats Using ECG Morphology and Heartbeat Interval Features. *IEEE Transactions on Biomedical Engineering* (2004)
 33. Rodríguez-Sotelo, J., Peluffo, D., et al.: Unsupervised Feature Relevance Analysis Applied to Improved ECG Heartbeat Clustering. (2012)
 34. Sornmo, L. P.: *Bioelectrical Signal Processing In Cardiac And Neurological Applications*. Academic Press Series in Biomedical Engineering (2005)

Tabla 1. Conjunto de registros de la base de datos MIT/BIH utilizados en los experimentos.

AAMI	N					S				V		F		Q	
Código MIT	N	L	R	e	j	A	a	J	S	V	E	F	f	P	Q
Clase	1	2	3	34	11	8	4	7	9	5	10	6	38	12	13
	0000	1010	0111	1011	0101	0010	1000	0110	1100	0001	1101	1001	0100	0011	1110
100	2237					33				1					
101	1858					3									2
102	99									4		56	2026		
103	2080					2									
104	163									2		666	1378		18
105	2524									41					5
106	1505									520					
107										59			2076		
108	1738				1	4				16		2			
109		2490								38		2			
111		2121								1					
112	2535					2									
113	1787						6								
114	1818					10		2		43		4			
115	1951														
116	2300					1				109					
117	1532					1									
118			2164			96				16					
119	1541									444					
121	1859					1				1					
122	2474														
123	1513									3					
124		1529		5	2		29			47		5			
200	1742					30				825		2			
201	1623			10	30	97	1			198		2			
202	2059				36	19				19		1			
203	2527					2				444		1			4
205	2569				3					71		11			
207		1457	85		106					105	105				
208	1585							2		992		372			2
209	2619				383					1					
210	2421						22			194	1	10			
212	922		1824												
213	2639				25	3				220		362			
214		2001								256		1			2
215	3194				2					164		1			
217	244									162			260	1540	
219	2080				7					64		1			
220	1952				94										
221	2029									396					
222	2060			212	208		1								
223	2027		16		72	1				473		14			
228	1686				3					362					
230	2253									1					
231	314		1252		1					2					
232		396		1	1381										
233	2229				7					830		11			
234	2698						50			3					
TOTAL	74986	8069	7250	16	229	2543	150	83	2	7127	106	802	982	7020	33

La primera fila corresponde a las etiquetas utilizadas de acuerdo al estándar AAMI.
 La segunda fila corresponde a las etiquetas utilizadas en la base de datos MIT/BIH.
 La tercera fila corresponde al código numérico de las etiquetas anteriores.
 La primera columna es el nombre de los registros, mientras que los otros contienen el número de latidos del corazón de cada tipo.

Tabla 2. Media del porcentaje de clasificación.

Clasificador	Población	Media de clasificación	Capas	Llamadas a función
Linea de base				
1NN	-	94.11 %	2	-
3NN	-	94.71 %	2	-
Naive Bayes	-	9.72 %	2	-
Naive Bayes	-	9.93 %	5	-
Naive Bayes	-	9.93 %	10	-
Algoritmos Genéticos				
1NN	5	93.38 %	2	20
3NN	5	93.78 %	2	20
Naive Bayes	5	14.42 %	2	20
Naive Bayes	5	15.59 %	5	20
Naive Bayes	10	15.75 %	10	50
Algoritmos Genéticos con Elitismo				
1NN	5	93.81 %	2	20
3NN	5	94.32 %	2	20
Naive Bayes	5	16.01 %	2	20
Naive Bayes	5	17.46 %	5	20
Naive Bayes	10	17.85 %	10	50

Tabla 3. Media de las características en la clasificación (la línea base implica 100 características en todos los casos).

Clasificador	Población	Media de las características	Capas	Llamadas a función
Algoritmos Genéticos				
1NN	5	49	2	20
3NN	5	52	2	20
Naive Bayes	5	41	2	20
Naive Bayes	5	34	5	20
Naive Bayes	10	45	10	50
Algoritmos Genéticos con Elitismo				
1NN	5	46	2	20
3NN	5	51	2	20
Naive Bayes	5	40	2	20
Naive Bayes	5	33	5	20
Naive Bayes	10	43	10	50