

Ramírez López, Leonardo

Modelo de clasificación dinámico de arritmias cardiacas mediante aprendizaje de maquina con interfaz
a usuario

Ciencia e Ingeniería Neogranadina, vol. 16, núm. 2, agosto-diciembre, 2006, pp. 86-95
Universidad Militar Nueva Granada
Bogotá, Colombia

Disponible en: <http://www.redalyc.org/articulo.oa?id=91116209>

MODELO DE CLASIFICACIÓN DINÁMICO DE ARRITMIAS CARDIACAS MEDIANTE APRENDIZAJE DE MAQUINA CON INTERFAZ A USUARIO

MODEL OF DYNAMIC CLASSIFICATION OF ARRHYTHMIAS CARDIAC BY MEANS OF LEARNING OF MACHINE WITH INTERFACE TO USER

*Leonardo Ramírez López*¹

Fecha de Recepción: 14 de Octubre de 2006

Fecha de Aprobación: 17 de Noviembre de 2006

RESUMEN: *Este trabajo presenta el desarrollo de un modelo de clasificación de arritmias cardiacas, basado en la duración de las ondas de registros electrocardiográficos, utilizando técnicas de clasificación supervisada. Los clasificadores desarrollados se basaron en técnicas de aprendizaje de máquina, Se implementaron metaclasificadores para los diferentes clasificadores. El desempeño de los diferentes clasificadores se probó aplicando criterios objetivos de comparación, y se obtuvieron resultados prometedores. Para presentar los resultados al usuario final, se diseñó una interfaz para ser usada por el especialista como ayuda a la clasificación de arritmias.*

PALABRAS CLAVES: *Arritmias, holter, aprendizaje de maquina, clasificador débil, metaclasificador*

ABSTRACT: *This work presents the development of a classification model of heart arrhythmias, based on the duration ECG wave records, using supervised classification techniques and their corresponding meta-classifiers. The classifiers were tested and compared obtaining promising results. For the presentation to the specialist an interface was designed as a support for arrhythmia classification.*

KEYWORDS: *Arrhythmias, holter, learning of machine, weak, metaclasificador sort key*

¹ *Ing. Leonardo Ramírez López, M.Sc., docente asistente de la UMNG, director del grupo TIGUM, Miembro del capítulo profesional de Procesamiento de Señales IEEE-Colombia, contacto 6343200 ext. 386 – correo leonardo.ramirez@umng.edu.co*

I. INTRODUCCIÓN

Las enfermedades cardiovasculares han aumentado significativamente en la última década¹, ocupando el segundo lugar en crecimiento después del cáncer que sigue ocupando el primer lugar de mortalidad a nivel mundial. Este crecimiento ha dado lugar a un significativo aumento de estudios científicos que analizan las señales electrocardiográficas en morfología, amplitud y/o duración desarrollando técnicas automáticas de clasificación de arritmias cardiacas que soporten objetivamente el diagnóstico del especialista. Los registros electrocardiográficos obtenidos de equipos Holter son analizados e interpretados por programas propietarios, que luego el médico especialista interpreta para emitir el diagnóstico al paciente, siendo algunas veces demorado y agotador debido al creciente número de registros. Sin embargo, la aplicación de nuevas técnicas de clasificación basadas en el aprendizaje de máquina aportan investigaciones sobre métodos de clasificación supervisada y su evolución en los meta-clasificadores, los cuales mejoran el poder predictivo de los sistemas de clasificadores convencionales. Algunos autores como Breiman (1996)², Bahuer y Kohavi (1999)³, entre otros, han desarrollado combinaciones de clasificadores en un solo clasificador basado en el remuestreo del conjunto de datos.

El aprendizaje de máquina (ML, por su sigla en inglés Machine Learning) utiliza técnicas que ayudan a inducir el conocimiento basado en ejemplos o reglas. Los métodos de aprendizaje de máquina son útiles en casos donde las soluciones algorítmicas no están disponibles, falta de modelos formales o el conocimiento sobre el dominio de la aplicación se define pobremente. El ML proporciona métodos, técnicas y herramientas que ayudan a resolver el diagnóstico y pronóstico de problemas en una variedad de dominios médicos [7].

II. INFORMACIÓN DEL PROBLEMA A RESOLVER

Durante varias décadas el estudio de la electrocardiografía se ha convertido en una herramienta de diagnóstico de desórdenes cardíacos, debido a su bajo costo y que usa técnicas no invasivas. En la década de los cincuenta, se desarrolló un equipo portátil y ambulatorio llamado Holter para grabar continuamente registros de la actividad cardiaca de un paciente entre 24 y 48 horas. Estos registros son analizados e interpretados por médicos especialistas que revisan minuto a minuto los cambios registrados, con el ánimo de descubrir anomalías o patologías en el paciente, lo cual resulta tortuoso por la cantidad de pacientes. Adicionalmente a esto, en Colombia para el mes de Julio del 2005, la clínica Shaio registró entre 12 a 20 Holter diarios, obligando a los médicos especialistas a invertir entre veinte a treinta minutos en el análisis, esto con la ayuda de programas propietarios (propios del fabricante) para emitir su diagnóstico al paciente. Las investigaciones desarrolladas en aprendizaje supervisado, abren la posibilidad de plantear nuevos modelos de clasificación supervisada de arritmias cardiacas, que puedan brindar un correcto y oportuno diagnóstico objetivo, que sirva de apoyo al médico especialista.

III. TEORÍA

A. Corazón

Órgano muscular hueco, situado entre los pulmones y sobre el diafragma. Su función principal es operar como una bomba para impulsar la sangre por los vasos de todo el cuerpo. Potenciales de acción del músculo cardíaco: el potencial de membrana en reposo del músculo cardíaco es aproximadamente de 85 a 95 mV y de 90 a 100 mV en las fibras conductoras especializadas. Ciclo cardíaco: es el periodo que transcurre entre el inicio de un latido hasta el inicio

1 Reporte Organización Mundial de la Salud, Mayo de 2005, p. 17-18

2 Creador del método Bagging (Bootstrap Aggregating)

3 Creadores del método Boosting y su aplicación Adaboosting

del siguiente, cada ciclo comienza con la generación espontánea de un potencial de acción en el nodo sinusal.

B. Electrocardiograma

El ECG registra de la actividad eléctrica del corazón medida entre dos puntos de la superficie del cuerpo, el cual puede suministrar información sobre el funcionamiento del corazón (ritmo cardíaco y cavidades del corazón). El ECG es el estudio más eficaz para diagnosticar una arritmia, ya que puede interpretarse según distintos criterios: eje, ritmo, frecuencia, formación del estímulo y conducción. Es importante determinar si la actividad del corazón es rítmica o arrítmica y si el marcapasos del ritmo es el nódulo sinusal funcionando correctamente.

C. Ondas, segmentos e intervalos

Los siguientes valores de duración de cada onda, intervalo y segmento se toman como referencia pacientes normales con parámetro de 72 ppm (pulsaciones por minuto) [14].

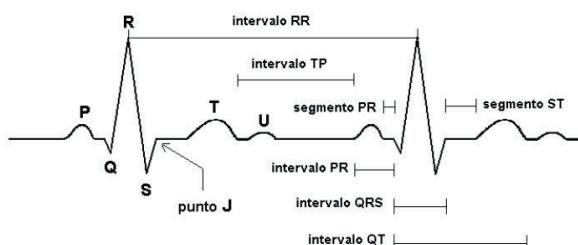


Figura 1. Caracterización del ECG

La figura 1 muestra las diferentes ondas, segmentos, intervalos y complejos del ECG.

Onda P: Es el resultado de la despolarización auricular. Su duración, en condiciones normales, debe ser menor de 0.12 segundos.

Intervalo P-R: Su duración se mide desde el inicio de la onda P hasta el inicio del complejo QRS. En condiciones normales, el intervalo P-R de un adulto se encuentra entre 0.12 y 0.20 segundos en un adulto,

en un niño entre 0.1 y 0.18 segundos.

Segmento P-R: Se encuentra entre el final de la onda P y el inicio del QRS.

Complejo QRS: Está compuesto por las ondas Q, R y S. Normalmente tiene una duración que varía entre 0.06 y 0.11 segundos.

Onda Q: Las ondas Q normales tienen una duración entre 0.04 y 0.08 segundos.

Onda R: Es la primera deflexión positiva del QRS, sin importar si existe onda Q.

Onda S: Es cualquier deflexión negativa que siga a la onda R.

Punto J: Es el sitio de unión entre el complejo QRS y el segmento ST.

Onda T: Presenta gran variabilidad en su morfología, polaridad y duración.

Segmento ST: Representa el tiempo que dura la despolarización ventricular.

Intervalo QT: Se mide entre el inicio del complejo QRS y la terminación de la onda T. QT debe ser menor de 0.42 segundos.

D. Ritmo cardíaco

Los efectos de la acción sostenida, organizada en forma de espacios de tiempo sometidos a un incremento gradual, dependen de un importante hecho biológico. El ritmo cardíaco “ideal”, se sitúa entre el 70 y el 85 % de su máximo ritmo cardíaco. El funcionamiento del corazón y la fisiología varían en cada individuo, por lo que el ritmo cardíaco ideal de cada uno está comprendido dentro de una cierta “zona de objetivo” de ritmos ideales.

E. Arritmias

Arritmia es un ritmo anormal del corazón, que puede hacer que éste bombee de forma menos eficaz. Se considera como arritmia cualquier ritmo cardíaco que no es el sinusal normal debido a alteraciones en la formación y/o en la conducción de impulsos. Según su origen se dividen en: supraventriculares (nodo sinusal, aurículas y unión AV) y ventriculares; según su presentación pueden ser paroxísticas y permanentes.

Las arritmias también se definen según la velocidad de los latidos en bradicardia que es un pulso muy lento, es decir, una frecuencia cardiaca inferior a los 60 latidos por minuto y taquicardia que es un pulso muy rápido, es decir, una frecuencia cardiaca superior a los 100 latidos por minuto [22].

IV. DESARROLLO DEL MODELO

En este aparte se presenta el diseño y el desarrollo del modelo de clasificación de arritmias cardíacas. En particular la capacidad de los algoritmos de clasificación basados en técnicas de aprendizaje de máquina.

La figura 2 especifica el modelo propuesto mostrando los pasos que siguen los datos desde la toma del registro ECG, donde primero se realiza el preprocesamiento, en segundo lugar la extracción de características, luego se aplican las técnicas de clasificación supervisada y finalmente se realiza la entrega de la respuesta al médico especialista.



Figura 2. Modelo desarrollado

A. Datos

Se toma la base de datos de arritmias MIT-BIH, estos registros fueron digitalizados a 2.77 ms (360 muestras por segundo) por un canal de resolución de 11 bits (2048 pasos digitales). Actualizados en Abril de 2005.

B. Preprocesamiento

Para procesar los datos obtenidos de los registros electrocardiográficos se realiza primero dos tipos de preprocesamiento: primero una normalización (0,1)

y un filtraje de la señal para obtener la ubicación del complejo QRS. Se normalizan los datos mediante la siguiente ecuación:

$$\text{Valor normal} = (\text{Veval} - \text{Vmin}) / (\text{Vmax} - \text{Vmin})$$

Siendo:

Valor normal: El Valor Normalizado

Veval: El valor antes de ser normalizado

Vmin: El valor más bajo comprendido para la variable a la cual corresponde el valor

Vmax: El valor más alto de los valores pertenecientes a la columna de la variable que contiene el valor a evaluar

C. Filtro

A la señal electrocardiográfica, se le realiza proceso de filtraje con el objetivo de extraer la ubicación del complejo QRS dentro del registro, basado en el algoritmo que emplea el Cygwin (interfase de WFDB para visualizar la señal y extraer información de los registros de ECG) llamado algoritmo de Pam-Tompkins [10].

D. Caracterización del ECG

En la tabla 1 se muestran los datos obtenidos de un registro ECG de paciente normal a 72 ppm.

E. Aprendizaje de máquina

El aprendizaje de máquina (ML) consiste en desarrollar métodos para aprender de la experiencia, es decir, que ayuden a inducir conocimiento a partir de ejemplos o reglas. Formalmente, dado un conjunto de datos de entrenamiento, se trata de seleccionar la hipótesis más probable que genera dichos datos. Los métodos de aprendizaje de máquina son útiles en casos donde las soluciones algorítmicas no están disponibles, falta de modelos formales o el conocimiento sobre el dominio de la aplicación se define pobremente [21].

En el contexto médico ML proporciona métodos, técnicas y herramientas que ayudan a resolver problemas de diagnóstico (subjetividad) y de pronóstico

(tendencias de comportamiento clínico) en una variedad de situaciones clínicas, apoyando al médico en situaciones asociadas a eliminar los problemas de asociación con la fatiga humana y habituación, dando una rápida identificación de anomalías y habilitando el diagnóstico en tiempo real.

Tabla 1. Valores de ECG paciente normal

PARAMETROS DE LAS ONDAS, SEGMENTOS E INTERVALOS			
	Amplitud	Duración	Morfología
Onda	P 0.25 mv	0.08 - 0.10 s	Simétrica de rama ascendente
	Q -0.1 mv	20 ms o 1/4 de R	Simétrica de descendente a ascendente
	R 5 mv	60 ms	Simétrica de ascendente a descendente.
	S -0.08	20 ms	Simétrica de descendente a ascendente.
	T 2.2 mv	360 ms	Asimétrica de rama ascendente
Segmento	PR Isoeléctrico < 0.1 mv	0.02 - 0.12s	Isoeléctrica
	ST Isoeléctrico < 0.1 mv	0.05 – 0.15s	Isoeléctrica
Intervalo	PR -	0.12 a 0.20s	Isoeléctrica
	QT -	0.35 – 0.44s (45% del ciclo cardiaco)	Isoeléctrica
	RR -	0.6 – 1s	Isoeléctrica
Complejo QRS	-	0.04 y < 0.10	Asimétrica en zigzag con vértices en cada amplitud de onda

f. Técnicas de ML utilizadas

Para desarrollar un sistema basado en aprendizaje de máquina es necesario contar con un conjunto de datos obtenidos a partir de la actividad cardiaca de un grupo de pacientes. Para cada registro se deben extraer las características temporales seleccionadas anteriormente, las cuales se presentan al sistema para la identificación y clasificación de las arritmias.

En este trabajo se implementaron diferentes técnicas de aprendizaje de máquina, las cuales se listan a continuación:

Bayes: Clasificador basado en la teoría de la probabilidad, usa frecuencias para calcular probabilidades

condicionales para calcular predicciones sobre nuevos casos. Específicamente se utilizaron dos clasificadores de este tipo:

Naïve Bayes: (Minsky, 1961) es un método de clasificación probabilístico. Se utiliza una distribución normal para clasificar una nueva instancia de un conjunto D dentro de un conjunto finito C de clases pre-determinadas. Así se tiene que:

$$P(C|W) = \frac{P(W|C) P(C)}{P(W)}$$

Donde $P(C)$ es la probabilidad a priori de la clase y $P(W|C)$ es la probabilidad condicional la palabra W dada la clase C .

Bayes Net: (Pearl 1988) son métodos basados en “score” y búsqueda, en los cuales la medida de bondad (verosimilitud, entropía, porcentaje de bien clasificados) de una estructura particular es definida, y ahora se lleva a cabo un procedimiento de búsqueda a lo largo del espacio de todas posibles estructuras de redes bayesianas.

Funciones: Clasificador que utiliza el análisis matemático para su distinción. En este principio operan: Logistic: (Hosmer y Lemeshow 1989) es un método proveniente de la estadística cuyo objetivo es obtener una relación funcional entre una transformación de una variable cualitativa llamada logit y p variables predictoras que pueden ser cuantitativas o cualitativas.

$$g(x) = \ln \frac{P(Y=1)}{P(Y=0)} = \beta_0 + \beta_1 X_1 + \dots + \beta_p X_p$$

La determinación de los coeficientes es realizada a través del método de máxima verosimilitud. Estos coeficientes son interpretados en términos de “radios”, y la selección de variables puede realizarse mediante tres métodos: hacia adelante, hacia atrás o escalonado por pasos.

Perceptrón multicapa (MLP): con algoritmo de aprendizaje backpropagation

SMO: (Sequential Minimal Optimization) algoritmo de entrenamiento para el clasificador de máquinas con vectores de soporte usando polinomios. Una máquina de soporte vectorial (SVM) es un clasificador de una sola clase.

RBF network: Las funciones de base radial (Romhead y Lowe, 1988) es una función cuya característica principal es que su respuesta disminuye (o aumenta) monótonamente con la distancia a un punto fijo llamado centroide. Una función de base radial es Gausiana, de forma generalizada:

$$e_1(X) = e_2(X) = \dots = e_K(X).$$

K-vecino más cercano KNN (Fix y Hodges, 1951). En este método se estima la función de densidad de donde provienen un conjunto de datos, aplicado a discriminación se utiliza para estimar la función de densidad $f(x/C_j)$ de las predictoras x por cada clase. Es el método más básico de discriminación no paramétrico, ya que no se hace ninguna suposición distribucional.

Combinación de clasificadores: El concepto de METACLASIFICADORES ha sido propuesto como una nueva dirección para mejorar el rendimiento de los clasificadores individuales. Esta técnica de combinar clasificadores consiste en construir un clasificador a partir de una sucesión de conjuntos de entrenamiento \mathcal{E}_k , cada uno con N observaciones diferentes. Este proceso se lleva a cabo usando el método de votación.

Principio de voto por la mayoría: el problema es el producir una nueva etiqueta $E(X) = j$ a partir de una

serie de etiquetas $e_k(X) = j_k$, $k=1, \dots, K$, donde la siguiente ecuación no tiene que ser necesariamente verdadera:

$$e_1(X) = e_2(X) = \dots = e_K(X).$$

Esto es, pueden existir conflictos entre las decisiones de los K clasificadores. Para ello se toma el principio de voto por la mayoría. En él si la mayoría de los K clasificadores votan por la misma etiqueta, entonces el clasificador combinado toma esta etiqueta como resultado final, donde la mayoría significa que más de la mitad de los clasificadores votan por la misma etiqueta. La regla más general de voto por la mayoría es la siguiente:

$$E(X) = \begin{cases} j, & \text{si } T_E(X = j) = \max_{i \in \mathcal{E}} T_E(X = i) > \frac{K}{2} \\ M+1, & \text{en otro caso.} \end{cases}$$

En la metodología de la combinación de clasificadores, existen dos métodos generales tales como: Bagging [19] y Boosting [20].

Bagging (Bootstrap Aggregating) (Breiman 1996) el algoritmo de Bagging genera clasificadores de varias muestras de los ejemplo, esto funciona especialmente para algoritmos de aprendizaje inestables (cambian mucho sus estructuras al cambiar un poco los ejemplos). Una muestra de ejemplos bootstrap se genera al muestrear uniformemente m instancias del conjunto de entrenamiento con reemplazo. Se generan T muestras, B_1, B_2, \dots, B_T y se construye un clasificador C_i para cada muestra. Con estos, se construye un clasificador final C^* de todos los C_1 a C_T cuya salida es la salida mayoritaria de los clasificadores. Bagging construye N clasificadores de una muestra de ejemplos tomando aleatoriamente cada uno de ellos, donde posiblemente faltan ejemplos o existen repetidos, para clasificar con todos los clasificadores y tomar la mayoría por votación.

Boosting (Freund-Schapire 1996) su variante más

usada AdaBoost (Adapting Boosting) generan igual un conjunto de clasificadores, sin embargo, AdaBoost los genera secuencialmente (Bagging los puede generar en paralelo). A todos los ejemplos, les asigna inicialmente un peso igual ($1/m$). Cada vez que se genera un clasificador, se cambian los pesos de los nuevos ejemplos usados para el siguiente clasificador. La idea es obligar al nuevo clasificador a minimizar el error esperado. Para esto se les asigna más peso a los ejemplos mal clasificados y menos a los bien clasificados.

V. IMPLEMENTACIÓN

Para medir el desempeño del modelo de clasificación es necesario conocer la clase a la que pertenece un ciclo cardíaco, para ello se utilizaron los registros 233,217,214,209,228 y 200 de MIT-BIH. Las arritmias a clasificar son PVC, APC, Asistolia, CAP y TV.

Para cada ciclo detectado como arrítmico, se toma el pico R como referencia, con tres ciclos previos y tres ciclos posteriores, obteniendo siete ciclos por cada anomalía. Para las pruebas de desempeño de las técnicas de clasificación se usaron 105 datos representativos de cada arritmia.

Las pruebas de los clasificadores se realizaron con librería WEKA, que contiene una colección de algoritmos de aprendizaje de máquina. Para cada clasificador esta librería provee su modelo, la clasificación de cada dato, además el desempeño de cada clasificador mediante la matriz de confusión. Los clasificadores se probaron usando validación cruzada, para un número de particiones de cinco y diez ($n=5$ y $n=10$). Este método divide los datos en n partes y por cada parte se construye el clasificador con las $n-1$ partes restantes, el cual se prueba con ésta.

Para comparar los diferentes métodos de aprendizaje de máquina se registran: el número de instancias

clasificadas correctamente, el número de instancias clasificadas incorrectamente, el porcentaje de clasificación y la estadística Kappa, que es un coeficiente de medida de acuerdo y mide la concordancia entre dos jueces cuando ambos están valorando el mismo objeto. La diferencia entre la proporción de casos observada en la que los jueces concuerdan y la proporción esperada por azar se divide por la máxima diferencia posible entre las proporciones observadas y esperadas, dados los totales marginales. Si el resultado es igual a 1 indica un acuerdo perfecto, igual a cero indica que el acuerdo no es mejor que el que se obtendrá por azar.

VI. RESULTADOS

Se realizaron un total de 42 experimentos, en los cuales el clasificador débil Bayes net, obtiene el 100% de clasificación correcta. Los clasificadores MLP y KNN obtienen el 99.04% de clasificación correcta (104 de 105).

Prueba 1: Clasificadores débiles con crossvalidation para n=5

Salida	CLASIFICADOR						
	Bayes net	Naïve Bayes	Logistic	MLP	RBF	SMO	IBk
Instancias clasificadas correctamente	105	102	102	104	101	103	104
Instancias clasificadas incorrectamente	0	3	3	1	4	2	1
Porcentaje de clasificación	100%	87.14%	97.14%	99.04%	96.19%	98.09%	99.04%
Estadística Kappa	1	0.9556	0.9556	0.9851	0.941	0.9703	0.9851

Prueba 2: Clasificadores débiles con crossvalidation para n=10

Salida	CLASIFICADOR						
	Bayes net	Naïve Bayes	Logistic	MLP	RBFnet	SMO	IBk
Instancias clasificadas correctamente	105	102	101	104	101	103	104
Instancias clasificadas incorrectamente	0	3	4	1	4	2	1
Porcentaje de clasificación	100%	87.14%	96.19%	99.04%	96.19%	98.09%	99.04%
Estadística Kappa	1	0.9556	0.9405	0.9851	0.941	0.9703	0.9851

Prueba 3: Metaclasificador Bagging con crossvalidation para n=5

Salida	CLASIFICADOR META-BAGGING cross validation 5						
	Bayes Net	Naïve Bayes	Logistic	MLP	RBFnet	SMO	IBk
Instancias clasificadas correctamente	105	102	102	104	101	104	104
Instancias clasificadas incorrectamente	0	3	3	1	4	1	1
Porcentaje de clasificación	100%	87.14%	97.14%	99.04%	96.19%	99.04%	99.04%
Estadística Kappa	1	0.9556	0.9556	0.9851	0.941	0.9851	0.9851

Prueba 4: Metaclásificador Bagging con cross n=10

Salida	CLASIFICADOR META-BAGGING cross validation 10						
	Bayes Net	Naive Bayes	Logistic	MLP	RBFnet	SMO	IBk
Instancias clasificadas correctamente	105	102	103	104	101	103	103
Instancias clasificadas incorrectamente	0	3	2	1	4	2	2
Porcentaje de clasificación	100%	87.14%	88.09%	99.04%	96.19%	88.09%	98.09%
Estadística Kappa	1	0.9556	0.9702	0.9851	0.941	0.9703	0.9702

Prueba 5: Metaclásificador AdaBoost con crossvalidation para n=5

Salida	CLASIFICADOR META-ADABOOST cross validation 5						
	Bayes Net	Naive Bayes	Logistic	MLP	RBFnet	SMO	IBk
Instancias clasificadas correctamente	105	102	102	104	101	104	104
Instancias clasificadas incorrectamente	0	3	3	1	4	1	1
Porcentaje de clasificación	100%	87.14%	97.14%	99.04%	96.19%	99.04%	99.04%
Estadística Kappa	1	0.9556	0.9556	0.9851	0.941	0.9851	0.9851

Prueba 6: Metaclásificador AdaBoost con crossvalidation para n=10

Salida	CLASIFICADOR META ADABOOST cross validation 10						
	Bayes net	Naive Bayes	Logistic	MLP	RBF-net	SMO	IBk
Instancias clasificadas correctamente	105	102	101	104	101	103	104
Instancias clasificadas incorrectamente	0	3	4	1	4	2	1
Porcentaje de clasificación	100%	97.14%	88.19%	99.04%	96.19%	88.09%	99.04%
Estadística Kappa	1	0.9556	0.9405	0.9851	0.941	0.9702	0.9851

La tendencia del error de clasificación se marca en un ciclo de PVC. El clasificador mas bajo fue RBF con el 96.19 de acierto.

La validación cruzada con n=5 y n=10 demostró que si se disminuye el numero de subconjuntos los clasificadores decrecen en desempeño, el número óptimo fue de 10.

Al utilizar los metaclásificadores Bagging y Adaboost el clasificador débil SMO aumento su desempeño a 99.04% y KNN llego al 100% de clasificación correcta.

Interfaz de usuario

Para hacer amigable la interacción con el usuario final, se diseño una pantalla con campos abiertos al ingreso de nuevo conocimiento, generado por el especialista y almacenado en el sistema; para luego, ser comparado con una nueva entrada de registro de ECG. Así, el la herramienta aumentará progresivamente el tipo de arritmias a clasificar. Esto de muestra en la figura 3.

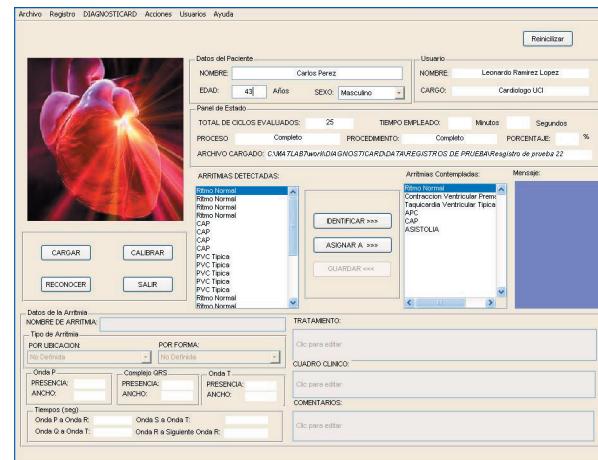


Figura 3. Interfaz de usuario

VII. CONCLUSIONES

La flexibilidad en el diseño del modelo permite emplear varios clasificadores hace posible utilizar metaclásificadores a partir de la combinación de clasificadores básicos o débiles aumentando la precisión en la clasificación, los cuales son convenientes de usar si el error de cada clasificador débil es menor al 50%, para el modelo planteado el error no supera el 6% en el peor de los casos (redes de función de base radial). Al aumentar el número de características extraídas del registro ECG, aumenta el buen desempeño del clasificador, pero hace en varias ocasiones que valores repetitivos (presencia o ausencia de señal) representen un tiempo mayor en el procesamiento. Para el diseño del modelo propuesto el tiempo de ejecución máximo es de menos de un segundo, lo que es aceptable, ya que se ha propuesto mejorar la especificidad en la clasificación y utilizar algoritmos que realicen un ordenamiento previo de todos los atributos, reduciendo el tiempo de procesamiento.

Al dividir el conjunto de datos en submuestras, para aplicar la técnica de validación cruzada lo que demostró un error promedio de clasificación nulo para Naive Bayes y máximo para RBFnet. La validación cruzada produce intervalos de confianza muy an-

chos para el error del clasificador, por ello se utilizó la combinación de clasificadores Bagging y Boosting, para corregir estimar el sesgo del error aparente y luego corregirlo combinando clasificadores débiles que permiten que errores no correlacionados de clasificadores individuales pueden eliminarse aplicando el principio de voto por mayoría.

Los metaclásificadores mejoraron la precisión en la clasificación porque: (a) Al seleccionar los datos de entrenamiento, al ser al azar, pueden no proporcionar información suficiente al peor clasificador (existiendo varias hipótesis de donde escoger). (b) el espacio de búsqueda puede no contener la función objeto por oscurecer decisiones tomadas ya que al iniciar una nueva clasificación puede tomar un nuevo grupo de entrenamiento y otro nuevo de prueba.

Todos los modelos consultados soportan el análisis y diagnóstico del especialista, ya que definen variables que inciden en el comportamiento cardíaco, pero ninguno toma diez variables diferentes de entrada al ML, demostrando su alto nivel de especificidad y selectividad de los datos extraídos de un registro de ECG.

REFERENCIAS

- [1] MINAMI, K. NAKAJIMA, H. and TOYOSHIMA, T. "Real-time discrimination of ventricular tachyarrhythmia with Fourier-transform neural network," IEEE Trans. Biomed. Eng., vol. 46, pp. 179–185, Feb. 1999.
- [2] AFONSO, V. X. and TOMPKINS, W. J. "Detecting ventricular fibrillation, selecting the appropriate time-frequency analysis tool for the application," IEEE Eng. Med. Biol. Mag., pp. 152–159, Mar. 1995.
- [3] KHADRA, L. AL-FAHOUM, A. S. and AL-NASHASH, H. "Detection of lifethreatening cardiac arrhythmias using thewavelet transformation," Med.Biolo. Eng. Comput., vol. 35, pp. 626–632, Nov. 1997.
- [4] AL-FAHOUM A. S. and HOWITT, I. "Combined wavelet transformation and radial basis neural networks for classifying life-threatening cardiac arrhythmias," Med. Biology. Eng. Comput., vol. 37, pp. 566–573, Sept. 1999.
- [5] THAKOR, N. V. and ZHU, Y. "Applications of adaptive filtering to ECG analysis: Noise cancellation and arrhythmia detection," IEEE Trans.Biomed. Eng., vol. 38, pp. 785–794, Aug. 1991.
- [6] ITURRALDE TORRES, P. "Arritmias cardíacas," Ed McGrawHill Interamericana. Segunda edición p. 25.
- [7] MAGOULAS, G. y PRENTZA, A. Machine learning in medical applications, lecture notes in computer science. 1999.
- [8] AGUIRRE, P. CARDELINO, J. y LOEFF, N. CARDIDENT, Instituto de Ingeniería eléctrica de la Universidad de la república de Montevideo. Abril 2001.
- [9] PALLAS ARENY, R. director tesis doctoral, Análisis de la variabilidad de intervalos de tiempo del ECG. Departamento de Ingeniería Electrónica. Universidad Politécnica de Cataluña (UPC), España. (2001).
- [10] PAN, J. TOMPKINS, W. A real time QRS detection algorithm, IEEE transactions on biomedical engineering, vol. BME-32, No. 3, Marzo 1985
- [11] FRITSCHY, D. RISK, M. Detección de latidos cardíacos ectópicos con redes neuronales. U. College London. 2000.
- [12] VÉLEZ, A. H. ROJAS M. W. BORRERO, R. J. RESTREPO M. J. Equipo Médico del Centro Cardiovascular Colombiano, Clínica Santa María FUNDAMENTOS DE MEDICINA. CARDIOLOGIA. Sexta edición, 2002.
- [13] Grupo Red Heart, Colombia. 2005.

- [14] GOLDBERGER, E. Derivaciones unipolares en electrocardiografía y vectorgrafía incluyendo las derivaciones estándar, las arritmias cardíacas y los principios de cetoecardiografía. P. 16 – 51.
- [15] MITCHELL, T. Machine learning. Mc Graw Hill, 1997.
- [16] HILERA, J. y MARTÍNEZ, V. Redes neuronales artificiales, fundamentos, modelos y aplicaciones. Ed. RaMa 1995.
- [17] HAU, D. COIERA, E. Learning qualitative models of dynamic system. Machine learning. 1997.
- [18] BAUER E. and KOHAVI. An empirical comparison of voting classification algorithms. Machine Learning, Boston. 1999.
- [19] BREIMAN L. BAGGING PREDICTOR. Machine Learning, Technical report. U. of. California. 1994.
- [20] SCHAPIRA, R. E. and SINGER, Y. Improved boosting algorithms. Machine Learning, 297-336. 1999.
- [21] LAVRAE Relational data mining applications. 1998
- [22] ISS y ASCOFAME. Guía práctica clínica basadas en evidencia. 2004
- [23] VIÑUELA, I. P. y GALVAN, I. M. Redes neuronales: Un enfoque práctico. Ed. Pearson Educación, 2004. p. 261
- [24] ARAGON TORRES, A. SASADO, S. Uso de las redes neuronales para optimizar simuladores. Ed. Burgos, 2002. p. 39
- [25] Fuente: <http://mit-bih/physionet>. Registros de electrocardiología, Instituto Tecnológico de Massachusetts. 2004.
- [26] Fuente: www.physionet.org/physiotools/wfdb.shtml 2005.
- [27] Fuente: http://www.texasheartinstitute.org/arcat_sp del hospital St. Luke's Episcopal, Instituto del cardiólogía de Texas, 2005
- [28] Fuente: <http://www.cardiologiahmm.com/hmm/index> Unidad de Cardiología, Hospital General Universitario Morales Messenguer, Abril 2005.
- [29] Fuente: <http://www.cardioglobal.com.ar/sociedades>, Directorio de la Cardiología mundial, Junio 2005.
- [30] Fuente: <http://www.hemodinamiadelsur.com.ar>, HDS Hemodinámica del sur, Junio 2005.
- [31] Fuente: <http://www.portalcardio.com/>, una excelente guía de recursos cardíacos, Diciembre 2003
- [32] Fuente: <http://www.mmhs.com/clinical/adult/spanish/cardiac/arrhythm.htm> Sistema de Salud Martín Memorial, Marzo 2005.
- [33] Fuente: <http://philips.com/ecg>. 2004
- [34] Fuente: <http://www.weka.com/manual> Completo manual de clasificadores mediante la librería weka, de la Universidad de Waikato. 2005.
- [35] Fuente: http://www.umm.edu/esp_imagepages/8772. Universidad de Maryland, Centro Médico, ADAM Diciembre 2003.