

Extracción de características de ECG basadas en transformaciones no lineales y wavelets

Extracting ECG signal characteristics based on non-linear transformations and wavelets

Victoria Eugenia Montes R,¹ Gustavo A. Guarín,² Germán Castellanos Domínguez⁵

RESUMEN

Se presentan diferentes métodos de extracción de características en señales ECG normales y en presencia de eventos relacionados con cardiopatía isquémica, basados en mediciones de diagnóstico, la transformada *wavelet* y el análisis no lineal de componentes principales. Con el fin de determinar las características que contribuyen de mejor manera con el modelo, se aplican dos técnicas de selección efectiva de características empleando métodos estadísticos multivariados y univariados. La evaluación de las técnicas de extracción propuestas se realiza mediante análisis discriminante lineal y máquinas de soporte vectorial, comparando el error en la clasificación de diferentes estados de funcionalidad cardíaca. Como resultado del análisis discriminante lineal se obtiene que las características más efectivas se consiguen empleando el análisis no lineal de componentes principales sobre un latido. En este caso, el error obtenido de clasificación es de hasta el 0.22%, contra 6.78% en el caso de las *wavelets*, y 24.22% en el caso de las mediciones de diagnóstico. Con las máquinas de soporte vectorial se obtiene que las características más discriminantes se obtienen empleando *wavelets* aplicadas al latido con una precisión de clasificación hasta del 0.1%, contra 0.12% en el caso del análisis no lineal de componentes principales y 5.11% en el caso de las mediciones de diagnóstico.

Palabras clave: ECG, cardiopatía isquémica, extracción de características, *wavelets*, PCA, KPCA.

ABSTRACT

Different extraction methods were compared regarding the characteristics of normal ECG signals and those emitted in the presence of events related to ischemic cardiopathy based on diagnosis measurements, wavelet transformation and nonlinear analysis of main components. Methods were developed for automatic recognition between normal and ischemic ECG signals. Two effective feature selection techniques were proposed; one used multivariate statistical methods and the second univariate ones. Linear discriminatory evaluation and vector support machines were used for evaluating the proposed feature extraction techniques, comparing error when classifying different states of cardiac functionality. Nonlinear PCA offered slightly better performance compared to wavelet representation but was much better compared to diagnosis measurement. There was up to 0.22% error compared to 6.78% in the case of wavelets and 24.22% in the case of diagnostic measurements. Support vector machines increased the performance for all analysed feature extraction methods; more discriminating characteristics were obtained when using wavelets applied to heartbeat having up to 0.1% classification precision compared to 0.12% in the case of nonlinear analysis of main components and 5.11% in the case of diagnostic measurements.

Key words: ECG, ischemic heart disease, feature extraction, wavelets, nonlinear transformations, PCA, KPCA.

Recibido: abril 17 de 2005

Aceptado: septiembre 16 de 2005

1 Ingenieria electrónica. Grupo de Control y Procesamiento Digital de Señales, Universidad Nacional de Colombia Sede Manizales, e-mail: victimontes@hotmail.com

2 Ingeniero electrónico. Estudiante de Maestría en Ingeniería y de Computadores de la Universidad de los Andes. Grupo de Control y Procesamiento Digital de Señales, Universidad Nacional de Colombia Sede Manizales, e-mail: gustavoguarin@hotmail.com

3 Ph.D. en Ingeniería del Nauchno-Isseledovatelskiy Institut, MUTSI, Rusia. Grupo de Control y Procesamiento Digital de Señales, Universidad Nacional de Colombia Sede Manizales, e-mail: gcastell@telesat.com.co

Introducción

El empleo de métodos de control y prevención de las enfermedades cardiovasculares implica, entre otros, el desarrollo de herramientas más efectivas en el diagnóstico de la funcionalidad cardíaca, en particular el análisis de la actividad eléctrica del corazón a través de los registros ECG, los cuales están relacionados con la cuantificación directa de sus diferentes parámetros morfológicos (ritmo, estimación de intervalos y amplitudes de eventos, etc.) con ayuda de reglas relativamente simples. Sin embargo, las limitaciones de los métodos de análisis son severas, particularmente cuando las señales ECG deben ser evaluadas en relación con factores de influencia internos (sincronismo del mismo fenómeno de ECG en sus diferentes derivaciones) o externos (estado funcional de otros órganos, hábitos del paciente, factores congénitos, etc.). El ECG de cada estado de actividad cardíaco-eléctrica (normal o patológico) algunas veces es extremadamente difícil de describir mediante ponderaciones exactas de las formas de sus picos, ondulaciones bruscas u otros patrones anormales que el especialista detecta a simple vista. Lo anterior plantea el problema de la cuantificación (extracción de características) y la clasificación de las señales ECG que envuelve elementos de reconocimiento de patrones, suponiendo de antemano que pueden ser estimadas las características principales de actividad cardíaco-eléctrica (normal o patológica).

La extracción de características es la etapa encargada de obtener la información clínica relevante de un registro ECG y antecede a la selección efectiva de características, la cual consiste en reducir la dimensionalidad de las características iniciales que alimentan el clasificador, manteniendo un nivel discriminante que permita el reconocimiento de las diferentes clases de estado funcional cardíaco. Entre las técnicas de extracción de características más conocidas se encuentran las basadas en mediciones de diagnóstico (Kunzmann et al., 2002), (GholamHosseini y Nazeran, 1999), (Silipo, 1999), que han sido empleadas principalmente en la detección de arritmias debido a la facilidad que ofrecen en el hallazgo de variaciones presentes en los intervalos que componen la señal ECG. En el campo de la representación de señales, la aproximación mediante funciones (Olszewski, 2001) permite realizar la aproximación a tramos de la señal mediante funciones ya definidas (constante, recta, triangular, etc.), mientras la expansión de Hermite (Haraldsson et al., 2004) consigue describir la señal ECG en forma de funciones de descomposición ortogonal. Mediante el análisis de componentes principales las características pueden ser extraídas por medio de transformaciones (lineales o no lineales) del conjunto original de características (Fujimura y Kiyasu, 2001), (Jager, 2002), (Stamkopoulos et al., 1998); sin embargo, las técnicas basadas en transformaciones lineales ignoran las correlaciones de más de segundo orden en los datos, por lo que es más apropiado asumir que las características ocul-

tas son funciones no lineales de las variables a analizar. Las técnicas descritas anteriormente no tienen en cuenta el carácter no estacionario de las señales ECG, por lo cual presentan alta sensibilidad en su estimación; otras técnicas como las distribuciones tiempo-frecuencia (Rosado et al., 2002), (Vila, 1996) y la transformada wavelet (Hughes et al., 2004), (Reyna y Jané, 2001), (Lemire et al., 2000) permiten la localización conjunta de eventos en tiempo-frecuencia, siendo adecuadas para el análisis de señales no estacionarias.

En el reconocimiento de señales ECG, dependiendo del tipo de patrones que se tenga, se pueden emplear tanto los métodos estadísticos, como los contemporáneos, basados en inteligencia artificial. En los primeros, que básicamente incluyen técnicas clásicas de discriminación y bayesianas, la selección de características se lleva a cabo mediante la metodología secuencial, que exige suficiente experiencia y conocimiento del experto que procesa la información. En los segundos, que incluyen redes neuronales, máquinas de soporte vectorial, etc., se aprovecha tanto la capacidad de interpolación, sobre todo en los casos en donde no existe el suficiente número de datos de entrenamiento, como la habilidad para realizar discriminantes no lineales en un posible espacio multidimensional, lo que permite la formación de regiones no conjuntas para cada una de las clases. En este caso, no se necesita información a priori de la distribución estadística o de la independencia de los vectores de características.

En este trabajo se presenta un método de identificación de eventos relacionados con cardiopatía isquémica sobre registros ECG, el cual consta del análisis de señales mediante diferentes técnicas de extracción de características, además de una etapa de selección efectiva con el objetivo de determinar el conjunto que brinda mayor discriminación entre las clases. La evaluación de la separabilidad de las características se hace mediante el rendimiento de clasificadores (precisión de clasificación) basados en análisis discriminante lineal y máquinas de soporte vectorial. El artículo está organizado como sigue: en la segunda sección se muestran los principios básicos relacionados con las técnicas de extracción y selección efectiva empleadas. En la sección III se presenta la metodología empleada en cada una de las etapas del proceso desde la lectura de los registros de la base de datos hasta su clasificación, mientras en la sección IV se presentan los resultados obtenidos, y finalmente, en la sección V, se mencionan las conclusiones del trabajo.

Marco teórico

En esta sección se presentan los principios teóricos básicos relacionados con las técnicas de extracción de características y selección efectiva que se emplean en este trabajo.

Técnicas de extracción de características

El objetivo de la etapa de extracción de características es transformar el segmento de la señal que se va a analizar, de tal manera que se obtenga en un número reducido de coeficientes la información clínica relevante. De esta forma es posible representar la señal en un espacio cuya métrica minimice la distancia entre patrones de una misma clase y maximice la distancia entre patrones de distinta clase (Aguirre et al., 2002).

1. Características de diagnóstico: corresponden a mediciones realizadas sobre diferentes puntos de la señal ECG, los cuales provienen de la experiencia médica.

2. Características transformada wavelet (WT): la WT descompone la señal en sus diferentes componentes espectrales, de tal manera que cada una de estas tenga una resolución de acuerdo con su escala. La función $\Psi(t)$ de variable real t se conoce como función wavelet madre y debe oscilar en el tiempo, además de estar bien localizada en el dominio temporal. El parámetro de escala a queda asociado a un estiramiento o encogimiento de la función madre. El parámetro de traslación b permite la localización temporal de la distribución de energía. A partir de la función madre $\Psi(t)$, se generan las funciones wavelet $\Psi_{a,b}(t)$ mediante operaciones conjuntas de cambio de escala y traslación, en la forma,

$$\Psi_{a,b}(t) = \frac{1}{\sqrt{|a|}} \Psi\left(\frac{t-b}{a}\right)$$

Para el caso de la transformada wavelet discreta, los parámetros de dilatación a y traslación b toman solamente valores discretos. La dilatación de la wavelet madre, se relaciona con potencias enteras de una escala de referencia a_0 , normalmente mayor que 1, así $a = a_0^j$. Si se seleccionan escalas y posiciones basadas en potencias de 2 ($a_0 = 2$, llamadas escalas y posiciones *diádicas*), el análisis será mucho más eficiente e igual de preciso que el análisis continuo. En este caso, la señal $f(t)$ se representa como una serie de aproximaciones (baja frecuencia) y detalles (alta frecuencia) en diferentes resoluciones. En cada etapa, un par de filtros (h, g) son aplicados a la señal de entrada para producir una señal de aproximación y una de detalle, respectivamente. La señal de detalle representa la información perdida, desde una resolución alta hasta una más baja. La representación wavelet es, entonces, el conjunto de coeficientes de detalle en todas las resoluciones y los coeficientes de aproximación en la resolución más baja.

3. Análisis no lineal de componentes principales: la técnica de PCA no lineal usa funciones *kernel* para transformar, de una manera no lineal, un conjunto de datos en un nuevo conjunto de variables no correlacionadas que representa la mayor parte de la información (Schölkopf et al., 1996). El procedimiento se puede describir de la siguiente manera:

- Cálculo de la matriz K mediante la expresión, $K_{ij} = (\Phi(x_i) \cdot \Phi(x_j))$.

- Determinación de los autovectores para la matriz K y normalización de estos en el espacio de características F .

- Cálculo de las proyecciones de un punto de prueba en los autovectores con el fin de extraer los componentes principales mediante

$$(\Psi^k \cdot \Phi(x)) = \sum_{i=1}^M \alpha_i^k (\Phi(x_i) \cdot \Phi(x)) \quad (1)$$

- Eliminación de los componentes principales que contribuyen con menos de un determinado porcentaje de la variación total de los datos.

La matriz de productos punto K puede ser calculada seleccionando un kernel $k(x,y)$ con el fin de evitar cualquier cálculo en el espacio de características con dimensión alta (Schölkopf et al., 1996).

Técnicas de selección efectiva de características

La selección efectiva de características identifica y selecciona el mejor subconjunto de características del conjunto de entrada con respecto a la identificación automatizada de estados funcionales. En este sentido, un número limitado de características simplifica la representación tanto del patrón como del clasificador, lo que resulta en un clasificador más rápido y que usa menos memoria. Por otro lado, una reducción exagerada en el número de características podría llevar a una pérdida en el poder discriminante, empobreciendo la precisión del sistema de reconocimiento (Jain et al., 2000).

1. Pruebas de independencia estadística y PCA:

a. Pruebas de hipótesis: esta prueba permite comparar las clases desde el punto de vista de los promedios de cada una de las componentes del espacio de características. Se analizan las siguientes hipótesis:

H_0 : No existe diferencia significativa en la media de cada característica para discriminar las dos clases.

H_1 : Existe una diferencia significativa del promedio en cada característica.

Para probar las anteriores hipótesis se emplea el método *t-student*. Sea $\xi_i = [x_1, x_2, \dots, x_{N_i}]$ el vector correspondiente a las mediciones del conjunto de características, con media μ y varianza σ^2 , ambas desconocidas. A partir de las N_i observaciones por clase, se estiman los valores de $\bar{\mu}$ y $\bar{\sigma}^2$. Así, un intervalo de confianza bilateral al 100 $(1-\alpha)\%$ para la media verdadera es:

$$(\bar{\mu}_1 - \bar{\mu}_2) - t_{(\alpha/2, N_1+N_2-2)} \bar{\sigma}_{\bar{\mu}_1-\bar{\mu}_2} \leq \mu_1 - \mu_2 \leq (\bar{\mu}_1 - \bar{\mu}_2) + t_{(\alpha/2, N_1+N_2-2)} \bar{\sigma}_{\bar{\mu}_1-\bar{\mu}_2} \quad (2)$$

con

$$\bar{\sigma}_{\bar{\mu}_1 - \bar{\mu}_2}^2 = \frac{N_1 \bar{\sigma}_1^2 + N_2 \bar{\sigma}_2^2}{N_1 + N_2 - 2} \left(\frac{1}{N_1} + \frac{1}{N_2} \right) \quad (3)$$

donde $t_{(\alpha/2, N_1 + N_2 - 2)}$ representa el punto porcentual de la distribución t con $N_1 + N_2 - 2$ grados de libertad. Si el intervalo de confianza dado por la ecuación (2) contiene el valor 0, entonces no se rechaza la hipótesis nula (H_0), en caso contrario se acepta la hipótesis alternativa H_1 . Así, si la hipótesis nula es rechazada, se asume que existe diferencia entre la media de cada clase.

b. Análisis de correlación por rangos de Spearman: como medida de asociación entre las características, el análisis no paramétrico de correlación por rangos es utilizado para observar su mutua dependencia. Una de estas medidas de asociación es el coeficiente de rango de Spearman (Doltsinis et al., 1999). Los valores de los coeficientes de correlación por rangos están entre -1 y 1. Un valor cercano a cero indica que no existe una asociación entre las variables. El coeficiente de correlación de Spearman r_s es definido como el coeficiente de correlación lineal entre los rangos R_i de ζ_i y los rangos S_i de χ_i con $\zeta_i, \chi_i \in F$ es

$$\mathbf{H} = n \sum_{i=1}^k (\bar{y}_i - \bar{y}_{..})(\bar{y}_i - \bar{y}_{..})' = \sum_{i=1}^k \frac{1}{n} \mathbf{y}_i \mathbf{y}_i' - \frac{1}{kn} \mathbf{y}_{..} \mathbf{y}_{..}' \quad (4)$$

c. Análisis de componentes principales (PCA): la técnica de PCA concentra la información de la señal en el mínimo número de parámetros, empleada con el propósito de reducir la dimensión del espacio de características, además de detectar dependencia o independencia en dicho espacio.

Combinación de técnicas estadísticas:

a. MANOVA: su objetivo es comparar los vectores media de un conjunto de k muestras para encontrar diferencias significativas. La hipótesis por lo tanto es

$H_0: \mu_1 = \mu_2 = \dots = \mu_k$ vs. H_1 : al menos dos μ 's son diferentes.

La igualdad de los vectores media implica que las k medias son iguales para cada variable; esto es, $\mu_{1r} = \mu_{2r} = \dots = \mu_{kr}$ para $r = 1, 2, \dots, p$ (Rencher, 1992). Se tienen las matrices "entre" e "intra" muestras (H y E):

$$\mathbf{H} = n \sum_{i=1}^k (\bar{y}_i - \bar{y}_{..})(\bar{y}_i - \bar{y}_{..})' = \sum_{i=1}^k \frac{1}{n} \mathbf{y}_i \mathbf{y}_i' - \frac{1}{kn} \mathbf{y}_{..} \mathbf{y}_{..}' \quad (5) \quad \text{y}$$

$$\mathbf{E} = \sum_{i=1}^k \sum_{j=1}^n (\mathbf{y}_{ij} - \bar{y}_i)(\mathbf{y}_{ij} - \bar{y}_i)' = \sum_{ij} \mathbf{y}_{ij} \mathbf{y}_{ij}' - \sum_i \frac{1}{n} \mathbf{y}_i \mathbf{y}_i' \quad (6)$$

La proporción de probabilidad de $H_0: \mu_1 = \mu_2 = \dots = \mu_k$ está dada por

$$\Lambda = \frac{|\mathbf{E}|}{|\mathbf{E} + \mathbf{H}|} \quad (7)$$

la cual es conocida como el de Wilks. Se rechaza H_0 si $\Lambda \leq \Lambda_{\alpha, p, v_H, v_E}$. El rechazo es para valores pequeños de Λ . Los parámetros en la distribución Λ de Wilks son: $p =$ número de variables (dimensión), $v_H =$ grados de libertad para la hipótesis y $v_E =$ grados de libertad para el error.

El Λ de Wilks compara la matriz E de sumas de cuadrados y productos "intra" con la matriz $E + H$ del "total" de sumas y productos. Mediante el uso de determinantes, el test Λ se reduce a un escalar. Así, la información multivariada en E y H acerca de la separación de los vectores media $\bar{y}_1, \bar{y}_2, \dots, \bar{y}_k$ es canalizada hacia una escala simple, en la cual se decide si la separación de los vectores media es significativa (Rencher, 1992).

b. ANOVA: en este caso se realiza la comparación de las medias de las muestras $\bar{y}_i, i = 1, \dots, k$ para ver si son lo suficientemente diferentes como para creer que las medias de la población difieren. La hipótesis puede expresarse como $H_0: \mu_1 = \mu_2 = \dots = \mu_k$. Si la hipótesis es cierta, todas las y_{ij} pertenecen a la misma población $N(\mu, \sigma^2)$, y se pueden obtener dos estimaciones de σ^2 , una basada en las varianzas de las muestras $s_1^2, s_2^2, \dots, s_k^2$ y la otra basada en las medias de las muestras $\bar{y}_1, \bar{y}_2, \dots, \bar{y}_k$. El estimador "intra" muestras combinado de σ^2 es

$$s_e^2 = \frac{1}{k} \sum_{i=1}^k s_i^2 = \frac{\sum_{ij} (y_{ij} - \bar{y}_i)^2}{k(n-1)} = \frac{\sum_{ij} y_{ij}^2 - \sum_i y_i^2 / n}{k(n-1)} \quad (8)$$

El segundo estimador de σ^2 (bajo H_0) está basado en la varianza de las medias de las muestras,

$$ns_y^2 = \frac{n \sum_i (\bar{y}_i - \bar{y}_{..})^2}{k-1} = \frac{\sum_i y_i^2 / n - y_{..}^2 / kn}{k-1} \quad (9)$$

donde $y_{..} = \sum_i y_i = \sum_{ij} y_{ij}$ es el total del conjunto. Si H_0 es falsa, $E(ns_y^2) = \sigma^2 + n \sum_i \alpha_i^2 / (k-1)$, y ns_y^2 tenderá a reflejar una dispersión mayor en $\bar{y}_1, \bar{y}_2, \dots, \bar{y}_k$. Dado que ns_y^2 y s_e^2 son independientes y ambos estiman a σ^2 , su relación forma un F -estadístico:

$$F = \frac{ns_y^2}{s_e^2} = \frac{\left(\sum_i y_i^2 / n - y_{..}^2 / kn \right) / (k-1)}{\left(\sum_{ij} y_{ij}^2 - \sum_i y_i^2 / n \right) / [k(n-1)]} = \frac{SSH / (k-1)}{SSE / [k(n-1)]} = \frac{MSH}{MSE} \quad (10)$$

donde $SSH = \sum_i y_i^2 / n - y_{..}^2 / kn$ y $SSE = \sum_{ij} y_{ij}^2 - \sum_i y_i^2 / n$ son la suma de los cuadrados "entre" muestras (debido a las medias) y la suma de los cuadrados "intra" muestras, respectivamente, y MSH y MSE son los correspondientes cuadrados de la media de las muestras. El F -estadístico está distribuido como $F_{k-1, k(n-1)}$ cuando H_0 es cierta. Se rechaza H_0 si $F > F_{\alpha}$ (Rencher, 1992).

Desarrollo experimental

En esta sección se presenta la metodología empleada en cada una de las etapas del proceso, desde la lectura de los registros de la base de datos hasta el reconocimiento de los patrones extraídos de la señal ECG.

Lectura de los registros de la base de datos

Se utilizó la base de datos ST-T Europea, la cual consta de 90 registros de ECG que contienen episodios del complejo ST-T anotados latido a latido por cardiólogos. Cada registro de esta base de datos se compone de un archivo de cabecera (.hea), archivos de anotaciones (.atr y .xws) y un archivo de la señal (.dat). Los archivos binarios de la señal y de las anotaciones fueron convertidos a formato de texto mediante los programas *rdsamp* y *rdann* disponibles en Physionet. Se analizaron 19 registros que contenían la derivación V4. El archivo de anotaciones de cada uno de los registros seleccionados fue revisado para buscar las secciones de señal en las que se presenta alguno de los episodios del complejo ST-T propios de la cardiopatía isquémica, tales como: Supradesnivel del segmento ST: ST+, infradesnivel del segmento ST: ST-, elevación de la onda T: T+ ó T++, inversión de la onda T: T- o T- -.

Dado que el sistema fue entrenado en el reconocimiento de las clases normal e isquémica, también se buscaron las secciones de señal anotadas como normales (N) en los mismos registros.

Preprocesamiento

En el preprocesamiento de señales ECG se hace necesario reducir ciertas características que aparecen con frecuencia en la señal debidas a efectos como la interferencia de la red (60 Hz), ruido térmico de los aparatos, movimiento del paciente y respiración, entre otros, que contaminan el contenido de la información y disminuyen la efectividad en las tareas de clasificación y reconocimiento orientadas al diagnóstico automatizado de diversas patologías (Ramírez, 2004 ; Orozco, 2003). Así mismo, para analizar la señal ECG se deben ubicar los puntos fiduciales de esta que permiten descomponer la señal en cada una de sus ondas constitutivas.

1. Reducción de perturbaciones: Para reducir perturbaciones se empleó la WT. El proceso de reducción de perturbaciones se describe a continuación (Ramírez, 2004):

- Escogencia de una *wavelet*, un nivel y cálculo de la descomposición wavelet de la señal s en el nivel J .
- Aplicación de un umbral a los coeficientes de detalle, del nivel 1 al J .
- Reconstrucción de la señal, basándose en los coeficientes de aproximación y los coeficientes modificados de detalle.

La familia de wavelets Daubechies ofrece las mejores características para la filtración de la señal (Ramírez, 2004). Los umbrales fueron obtenidos mediante métodos estadísticos. Las perturbaciones removidas fueron: interferencia de la red (Daubechies 3 con nivel 3), fluctuación de línea de base causada por la respiración (Daubechies 4 con nivel 8), ruido electromiográfico (Daubechies 3 con nivel 4).

2. Detección de complejos QRS: las señales de la base de datos ST-T Europea vienen muestreadas a 250 Hz y debido a que todos los algoritmos posteriores de proceso funcionan a una frecuencia nominal de 500 Hz, la señal debe ser muestreada nuevamente. Este procedimiento se realiza mediante la duplicación de cada uno de los puntos de la señal, obteniendo así una señal con el doble de longitud de la señal original. El método utilizado para determinar la ubicación del QRS es el cruce por cero de la WT. Esto es, porque el módulo máximo y los cruces por cero de la WT corresponden a bordes finos de la señal (Morales, 2003). El complejo QRS produce dos módulos máximos con signos opuestos con un cruce por cero entre ellos, por lo tanto este es determinado aplicando reglas de decisión (umbrales) a la WT de la señal ECG. La mayor parte de energía del complejo QRS se encuentra entre las escalas 2^3 y 2^4 , siendo más grande en 2^4 . La energía decrece si la escala es más grande que 2^4 . La energía de artefactos debidos al movimiento y al error de línea base se incrementa para escalas mayores a 2^5 . Por lo tanto, se decidió usar las escalas 2^1 a 2^4 para la WT. El complejo QRS se detectó así (Ladino, 2004):

- Se establece un umbral para escoger el módulo máximo.
- Se toman las vecindades de la escala 2^3 y se detecta el módulo máximo correspondiente a esta escala; lo mismo para las escalas 2^2 y 2^1 , esto para reducir el efecto del ruido de alta frecuencia, que se presenta más en las escalas bajas.
- Se encuentra el cruce por cero entre los módulos máximos de la escala 2^1 (positivo y negativo), que equivale al pico de la onda R.
- En el cálculo del inicio del QRS se toma una ventana de 100 ms a la izquierda del pico R para hallar un punto máximo a partir del módulo máximo negativo encontrado.

Para el final del QRS se toma una ventana de 60 ms a la derecha del pico R, para encontrar un punto mínimo, después de hallado ese punto se establece un umbral de sobrepaso para determinar el final del par módulo máximo que es el 25% del valor mínimo, en el instante que se sobrepase dicho umbral se establece el final del QRS.

3. Segmentación: debido a que desplazamientos del eje eléctrico del complejo QRS pueden ocasionar cambios no isquémicos en el segmento ST (Jager, 2002), se analizaron dos conjuntos de observaciones, uno de latidos y otro de complejos ST-T.

a. Segmentación de latidos: la derivación del conjunto de latidos se hizo de la siguiente manera:

- El inicio del latido se estableció con una ventana de 360 ms a la izquierda del pico *R*.
- El final del latido se estableció con una ventana de 352 ms a la derecha del pico *R*.

b. Segmentación de complejos ST-T: La derivación del conjunto de complejos ST-T se realizó así:

- El inicio del complejo ST-T se estableció en el punto final del QRS.
- El final del complejo ST-T se estableció con una ventana de 280 ms a la derecha del final del QRS.

Ambos conjuntos fueron normalizados (rango unitario y remoción de media). Finalmente, se construyó una matriz con 1800 observaciones (900 normales y 900 patológicas).

Extracción de características

1. Extracción de características mediante mediciones de diagnóstico: se consideraron diferentes mediciones en puntos específicos de la señal ECG, los cuales se usan tradicionalmente en el diagnóstico clínico (ver Tabla 1). El conjunto de mediciones construido se basó en el trabajo realizado por (Silipo, 1999). El cálculo de cada uno de estos parámetros se realizó a partir de los valores *picor* (posición del pico *R*), *iqrs* (inicio del QRS), *fqrs* (final del QRS), *it* (inicio de la onda *T*), *ft* (fin de la onda *T*) e *ip* (inicio de la onda *P*) obtenidos por el detector de QRS. Las mediciones se calcularon de la siguiente manera:

- Ancho del QRS: Intervalo que va desde el inicio hasta el fin del QRS: $qrs_w = fqrs - iqrs$

Tabla 1. Conjunto de mediciones caracterizando cada latido

Nombre	Descripción
<i>qrs_w</i>	Ancho del QRS
<i>pamp</i>	Amplitud positiva del QRS
<i>namp</i>	Amplitud negativa del QRS
<i>pqrs</i>	Área positiva del QRS
<i>nqrs</i>	Área negativa del QRS
<i>areat</i>	Área de la onda <i>T</i>
<i>ivr</i>	Repolarización ventricular invertida
<i>lst</i>	Nivel del segmento <i>ST</i>
<i>stsl</i>	Pendiente del segmento <i>ST</i>
<i>pr</i>	Intervalo PR
<i>ta</i>	Amplitud de la onda <i>T</i>
<i>tp</i>	Posición de la onda <i>T</i> respecto al pico <i>R</i>
<i>Qtd</i>	Intervalo QT

- Amplitud positiva del QRS: primero se halla el parámetro *bl* (línea de base) como un punto de la señal que se encuentra 80 ms antes de picor. La amplitud positiva del QRS se define como la amplitud del pico *R* menos el valor de *bl*.

- Amplitud negativa del QRS: se define como el valor mínimo que hay en el trozo de señal que va desde *iqrs* hasta *fqrs*, restando el valor de *bl*.

- Área positiva del QRS: se define como la sumatoria de todos los puntos positivos de la señal que están entre *iqrs* y *fqrs*, restando el valor de *bl*.

- Área negativa del QRS: se define como la sumatoria de todos los puntos negativos de la señal que están entre *iqrs* y *fqrs*, restando el valor de *bl*.

- Área de la onda *T*: se define como la sumatoria de todos los puntos de la señal que se encuentran desde *it* hasta *ft*, restando el valor de *bl*.

- Repolarización ventricular invertida (IVR): $IVR = \frac{pqrs + nqrs}{areat}$

- Nivel del segmento *ST*: este valor se tomó 60 ms después de *fqrs*.

- Pendiente del segmento *ST*: se define como una variación en amplitud dividida entre una variación en tiempo. La variación en amplitud se toma entre los puntos *picor + 180 ms* y *picor + 220 ms*; en tiempo se toman 40 ms de variación.

- Intervalo PR: este intervalo va desde el inicio de la onda *P* hasta el comienzo del QRS: $pr = iqrs - ip$

- Amplitud de la onda *T*: se define como el punto máximo del trozo de señal comprendido entre *it* y *ft* menos el valor de *bl*.

- Posición de la onda *T* respecto al pico *R*: se define como el número de muestras que existen entre el pico *R* y la amplitud de la onda *T*.

- Intervalo QT: este intervalo va desde el comienzo del QRS hasta el final de la onda *T*: $qtd = ft - iqrs$

El espacio de características queda representado por una matriz de $N_s \times n$, donde: N_s es el número de latidos, n es el número de mediciones de diagnóstico tomadas para cada latido, en este caso $n = 13$ y la última columna contiene las etiquetas para cada una de las dos clases: 1: señal normal, -1: señal isquémica.

2. Extracción de características mediante la transformada wavelet: existe una abundante variedad de funciones wavelet madre y el problema fundamental a resolver, es decidir cuál wavelet producirá los mejores resultados para una aplicación particular. La forma más común de selección de la función madre es el análisis

extensivo de un conjunto amplio de funciones y la escogencia de la función que dé el mejor rendimiento. Para extraer características con la WT se emplearon dos esquemas diferentes:

a. Esquema A: descomposición wavelet multinivel usando diferentes combinaciones de wavelets y niveles de descomposición empleando los coeficientes de descomposición de aproximación y detalle para formar un vector de características que represente adecuadamente la señal (Orozco, 2003). Se realizó un análisis para diferentes wavelets: ortogonales (*Daubechies-dbN* y *Symlets-symN*) y pares de wavelets bi-ortogonales (*biorNr.Nd* y *rbioNr.Nd*) aplicadas a los conjuntos tanto de latidos como de complejos ST-T. Los coeficientes de aproximación de los niveles de descomposición 3 y 4 se tomaron como componentes del vector de características. El espacio de características queda representado por una matriz de $N_s \times D+1$, donde: N_s es el número de patrones, D es la dimensión del vector de características para cada combinación, siendo diferente para cada wavelet madre, y la última columna contiene las etiquetas para cada una de las dos clases: 1: señal normal, -1: señal isquémica.

b. Esquema B: descomposición wavelet multinivel usando diferentes combinaciones de wavelets, empleando los coeficientes de aproximación máximos de distintos niveles de descomposición para formar un vector de características independiente de la longitud en muestras de cada señal (Ojeda, 2003). Este esquema se ha empleado para el caso en el que las muestras del conjunto inicial tengan tamaños diferentes, ya que la forma en que se seleccionan los coeficientes es independiente de la longitud en muestras de la señal. Del mismo modo que para el Esquema A, se realizó un análisis para diferentes wavelets: ortogonales y pares de wavelets bi-ortogonales aplicadas a los conjuntos tanto de latidos como de complejos ST-T. Se hizo una descomposición wavelet hasta el nivel 6 y en cada escala de aproximación se seleccionaron p coeficientes máximos, los cuales en conjunto formaron el vector de características. El parámetro p , correspondiente al número de coeficientes seleccionados por escala, se fijó primero en 3 y luego en 4. Para el conjunto de los complejos ST-T, sólo se empleó $p = 3$. El espacio de características queda representado por una matriz de $N_s \times D+1$, donde: N_s es el número de patrones, D es la dimensión del vector de características para cada combinación, definido como $D = 6 \times p$, y la última columna contiene las etiquetas para cada una de las dos clases: 1: señal normal, -1: señal isquémica.

3. Extracción de características mediante PCA no lineal: para el cálculo de la matriz K se empleó un conjunto conformado por 100 patrones, con el fin de obtener un número máximo de 100 autovectores. El tipo de kernel utilizado fue RBF, descrito por la siguiente expresión:

$$k(x, y) = \exp\left(-\frac{\|x-y\|^2}{2\sigma^2}\right) \quad (11)$$

donde el parámetro σ fue aumentado heurísticamente desde 0.1 hasta 1 en intervalos de 0.1. El conjunto de prueba para hacer las proyecciones estuvo constituido por 1800 patrones. Se escogieron los criterios del 90, 95 y 99% de variación de la señal para determinar el número de componentes principales que generaron la nueva representación. El espacio de características queda representado por una matriz de $N_s \times CP+1$, donde: N_s es el número de patrones, CP es la cantidad de componentes principales obtenidos para cada combinación entre σ y el porcentaje de representación, y la última columna contiene las etiquetas para cada una de las dos clases: 1: señal normal, -1: señal isquémica.

Selección efectiva de características

1. Pruebas de independencia estadística y PCA: con el propósito de observar y evaluar las características que discriminan adecuadamente las dos clases de señales ECG, se realizaron:

a. Pruebas de hipótesis: se analizan las siguientes hipótesis:

H_0 : No existe una diferencia significativa en la media de cada característica para discriminar las clases. En consecuencia, la resta de las medias dadas por cada una de las características de estas clases es cero.

H_1 : Existe una diferencia significativa en la media de cada característica para discriminar las clases. Por lo tanto, la resta de las medias dadas por cada una de las características de estas clases es diferente de cero.

En el método *t-student* se emplea un intervalo de confianza del 95%. Este análisis lleva a la selección de aquellas características que rechazan la hipótesis nula, es decir, son discriminantes entre clases (Suárez et al., 2004).

b. Análisis de correlación por rangos de Spearman: tomando como base las características seleccionadas en el proceso anterior y normalizadas a media cero y varianza unitaria, se calcula la correlación entre sus datos, esto con el fin de garantizar la menor cantidad de información redundante. Finalmente, se produce la matriz de correlación, en la que se establecen los valores permitidos de relación entre los diferentes pares de variables características. En general, entre dos características que presentan una alta correlación se elimina aquella variable cuya suma de correlaciones con las demás características sea mayor (Suárez et al., 2004). El rechazo de las características se da para valores de correlación superiores a 0.5.

c. PCA: con el fin de seleccionar el número de componentes principales (autovalores) que van a generar la nueva representación, se escogió el criterio del 90% de varianza acumulada.

2. Combinación de técnicas estadísticas: esta metodología conlleva la aplicación subsecuente de análisis multivariado de varianza (MANOVA), PCA, análisis de varianza (ANOVA) y análisis de correlación.

a. MANOVA: se usa con el fin de extraer del hiperespacio inicial de características un subespacio que sea altamente discriminante como grupo. Para determinar si un conjunto de características es verdaderamente significativo, se utiliza el test estadístico de Wilks a través de la evaluación del coeficiente Λ_a de Wilks transformado a valores que puedan ser evaluados sobre la distribución F (Λ a F -estadístico) (Rencher, 1992). Para encontrar el conjunto de características idóneas se efectúa el siguiente algoritmo de árbol de decisión:

- Cálculo de los F -estadísticos para cada característica.
- Ordenamiento descendente de los F -estadísticos.
- Se toma la característica correspondiente al mayor valor de los F -estadísticos y se asocia con el F -estadístico que tenga el segundo valor en magnitud.
- Se evalúa el conjunto inicial de dos características mediante el test de Wilks y se determina el valor de probabilidad acumulada del nuevo F -estadístico. Si dicho valor es mayor al del conjunto anterior, se asume que el nuevo conjunto es más discriminante y por tanto dicha característica es seleccionada; en caso contrario, se descarta la característica agregada. Una característica también puede ser descartada si su valor de Λ es una indeterminación, ya que esto implica redundancia de información.
- Los pasos anteriores se ejecutan de manera secuencial y progresiva agregando o descartando características según el criterio de la probabilidad acumulada.

b. PCA: la intención de PCA es obtener un nuevo espacio de características en el cual estas posean máxima varianza y mínima correlación. La transformación dada por PCA permite que el conjunto de características seleccionado por MANOVA tenga independencia estadística lineal y de esta forma se pueda aplicar el análisis univariado (ANOVA) para obtener una nueva reducción espacial.

c. ANOVA: mediante ANOVA se pretende hallar las proyecciones PCA más discriminantes, aprovechando la propiedad de independencia estadística entre las componentes. ANOVA realiza la estimación de los F -estadísticos que permiten determinar el grado de separabilidad entre las características.

d. Análisis de correlación: una vez se tiene la proyección de PCA más discriminante según ANOVA, se busca la correlación de las características que devuelve MANOVA con las proyecciones, con el fin de determinar las características que tienen mayor contribución en el espacio reducido. Con base en los errores de entrenamiento y validación de los clasificadores, se busca el número de características que verdaderamente son significativas.

Clasificación y reconocimiento de señales ECG

Para evaluar la utilidad del clasificador se debe medir el porcentaje de observaciones que fueron clasificadas correctamente. Esto genera una estimación de la probabilidad de casos correctamente clasificados. El método empleado para estimar esta probabilidad fue la prueba de muestras independientes (ITS). Si el conjunto de muestras es grande, se puede dividir en un conjunto de entrenamiento y otro de validación. Se usa el conjunto de entrenamiento para construir el clasificador y se clasifican las observaciones del conjunto de validación usando la regla de clasificación. La proporción de observaciones correctamente clasificadas es el porcentaje de clasificación estimado. Como el clasificador no ha visto los patrones en el conjunto de validación, el porcentaje de clasificación estimado no está sesgado. Los pasos para evaluar el clasificador usando este método son:

- Separar aleatoriamente la muestra en dos conjuntos de tamaño n_{TEST} y n_{TRAIN} , donde $n_{TEST} + n_{TRAIN} = N_s$.
- Construir el clasificador usando el conjunto de entrenamiento.
- Presentar cada patrón del conjunto de validación al clasificador y obtener una etiqueta de clase para él. Dado que se conoce la clase correcta de estas observaciones, se pueden contar los patrones correctamente clasificados N_{CC} .
- El porcentaje en que las observaciones son correctamente clasificadas es

$$P(CC) = 100 * \frac{N_{CC}}{n_{TEST}} \quad (12)$$

La evaluación de este método puede tener una alta varianza, que puede depender en gran medida de los datos que finalmente quedan tanto para el conjunto de entrenamiento como para el de validación. Para los experimentos realizados se utilizaron dos conjuntos de entrenamiento diferentes con 1.800 observaciones (balanceadas, esto es, 900 patológicas), para construir clasificadores binarios que discriminaron entre las clases patológica y normal.

Resultados

Evaluación de los métodos de extracción de características

Cuando se empleó LDA con las mediciones de diagnóstico se obtuvo un error del 24.22%. Con la WT, para el Esquema A el mínimo error de clasificación fue 9.78% (*Symlet 8* en el nivel 3); para el Esquema B disminuye el error de clasificación hasta el 6.78% (*Daubechies 2* con cuatro coeficientes máximos). Con PCA no lineal se obtuvieron errores hasta del 0.22%. Los mejores resultados de LDA se muestran en las Tablas 2, 3 y 4.

Al clasificar con SVM⁴ las mediciones de diagnóstico, se obtuvo un error del 5.11%. El Esquema A de la WT tuvo un mejor desempeño en el nivel 3 de descomposición con las madres biortogonales y Daubechies, y en el Esquema B los mejores resultados se pueden observar en la Tabla 5. En las características de PCA no lineal, el parámetro en que hubo mayor discriminación fue 0.4 para los tres porcentajes de representación.

Selección efectiva de características

Los métodos de selección efectiva de características se aplicaron a un conjunto conformado por:

- Mediciones de diagnóstico, de dimensión (1800 x 13).
- Mejores resultados de la WT: Esquema A aplicado al latido con las madres *Daubechies 2, 6 y Symlet 8*.
- Conjunto de latidos originales, de dimensión (1800 x 180).

Las características obtenidas en la etapa de selección efectiva fueron nuevamente evaluadas mediante LDA y SVM, con el fin de analizar la incidencia que tiene esta etapa sobre la exactitud en la clasificación. Para las pruebas de independencia estadística y PCA los resultados se muestran en la Tabla 6, en las que dimensión original hace referencia a las dimensiones del vector antes de aplicarle PCA. Mediante el método de combinación de técnicas estadísticas se obtuvieron los resultados mostrados en la Tabla 7.

Tabla 2. Mejores resultados para el esquema A de WT con LDA.

Latido				Complejo STT			
Madre	Dimensiones	Error (%)	Nivel	Madre	Dimensiones	Error (%)	Nivel
sym8	35	9.78	3	sym8	21	21.78	3
db6	32	10.67	3	db8	21	21.22	3
bior26	33	10.33	3	bior55	18	21.11	3
rbio26	33	10.11	3	rbio26	16	21.78	4

Tabla 3. Mejores resultados para el esquema B de WT con LDA.

Madre	Latido				Complejo STT	
	Coeficientes	Error (%)	Coeficientes	Error (%)	Coeficientes	Error (%)
db2	4	6.78	3	10.33	3	14.33
sym2	4	6.78	3	10.33	3	14.33
bior15	4	9.00	3	10.22	3	19.78
rbio13	4	11.22	3	12.00	3	21.44

Tabla 4. Mejores resultados de PCA no lineal LDA.

Porcentaje	σ	Latido		Complejo STT	
		CP	Error (%)	CP	Error (%)
kpca99	1	62	0.22	80	6.11
kpca95	1	42	0.22	60	7.56
kpca90	0.5	54	0.22	73	10.56
kpca90	1	32	0.33	49	8.56

Tabla 5. Mejores resultados para el esquema B de WT con SVM.

Latido				Complejo STT			
Madre	Coeficientes	Error (%)	SV	Madre	Coeficientes	Error (%)	SV
sym9	4	2.22	56	sym3	3	11.56	105
db3	4	2.89	68	db3	3	11.56	105
bior44	4	3.56	58	bior55	4	12.00	142
rbio55	4	3.56	71	rbio22	4	11.11	108

Tabla 6. Resultados de clasificación para las características seleccionadas mediante las pruebas de independencia estadística y PCA.

Madre	CP	Dimensión orig.	Nivel	LDA	SVM
				Error (%)	Error (%)
db2	3	8	4	1.44	0.10
sym8	3	8	3	8.00	1.44
db6	7	18	3	3.00	0.00

Tabla 7. Resultados de clasificación para las características seleccionadas mediante las técnicas estadísticas.

Madre	Nivel	Dimensiones	LDA	SVM
			Error (%)	Error (%)
db2	4	8	0.11	2.00
sym8	3	20	0.11	0.00
db6	3	14	0.00	0.00

Conclusiones

Existe ambigüedad respecto a la técnica más efectiva para la extracción de características, ya que no hay coincidencia entre los resultados arrojados por ambos clasificadores. Esto sugiere la necesidad de encontrar un método objetivo e independiente del clasificador que permita evaluar la calidad de las características extraídas. Sin embargo, si se tiene en cuenta la complejidad computacional, además de la robustez del algoritmo ante la presencia de ruido, las *wavelets* son mejores debido a que se basan en técnicas de filtrado, por lo tanto las componentes de alta frecuencia son suprimidas en el momento de aplicar la transformada, mientras que las técnicas no lineales implementadas actúan directamente sobre la información sin suprimir ningún tipo de componente frecuencial, por lo que se ven afectadas en presencia de ruido.

En general los resultados obtenidos mediante LDA y SVM indican que al aplicar la extracción de características a un latido completo

hay un mejor desempeño en la identificación de la cardiopatía isquémica. Sin embargo, el análisis basado en mediciones de diagnóstico resulta ser muy dependiente del desempeño de la etapa de segmentación, es por esto que se sugiere la implementación de un segmentador robusto ante condiciones de ruido y de anomalías.

4 Para SVM, SV son los vectores de soporte y CP los componentes principales

Bibliografía

- Aguirre, P.; Cardelina, J. y Loeff, N. "Sistema de detección, clasificación e identificación en línea de complejos QRS". Tesis doctoral, Instituto de Ingeniería Eléctrica, Facultad de Ingeniería Universidad de la República, Montevideo, Uruguay, 2002.
- Burges, C. "A tutorial on support vector machines for pattern recognition". Knowledge Discovery and Data Mining, Vol. 2, p. 22, 1998. [Online]. Disponible: <http://www.kernel-machines.org/papers/Burges98.ps.gz>
- Doltsinis, I., Rau, F. y Werner, M. "Analysis of random systems". Stochastic analysis of multivariate systems in computational mechanics and engineering, 1^o ed., International Center for Numerical Methods in Engineering, 1999, pp. 9-159.
- Fujimura, S. y Kiyasu, S. "Application of feature extraction scheme to the discrimination of electrocardiogram (ECG)". TIEE Japan, Vol. 121-A, N^o 8, 2001, pp. 725-730.
- Gholamhosseini, H. y Nazeran, H. "Efficient features for ann-based ECG classifiers". School of Engineering, the Flinders University of South Australia, 1999.
- Haraldsson, H., Edenbrandt, L. y Ohlsson, M. "Detecting acute myocardial infarction in the 12-lead ECG using hermite expansions and neural networks". Artif Intell Med, Vol. 32, N^o 2, 2004, pp. 127-136.
- Hughes, N., Tarassenko, L. y Roberts, S. "Markov models for automated ECG interval analysis". Advances in Neural Information Processing Systems 16, MIT Press, Cambridge, 2004.
- Jager, F. "Feature extraction and shape representation of ambulatory electrocardiogram using the Karhunen-Loève transform". Elektrotehniski Vestnik, Vol. 69, N^o 2, 2002, pp. 83-89.
- Jain, A.; Duin, R. y Mao, J. "Statistical pattern recognition: a review". IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 22, N^o 1, 2000.
- Kunzmann, U. et al. "Parameter extraction of ECG signals in real-time". Biomedizinische Technik, Vol. 47, N^o 1, 2002, pp. 875-878.
- Ladino, M. "Análisis de los cambios presentados en el segmento ST relacionados con enfermedades coronarias". Tesis presentada a la Universidad Nacional de Colombia, sede Manizales, para optar al título de ingeniero electrónico, 2004.
- Lemire, D. et al. "Wavelet time entropy, T wave morphology and myocardial ischemia". IEEE Transactions on Biomedical Engineering, Vol. 47, N^o 7, 2000, pp. 967-970.
- Morales, L. A. "Segmentación de ECG normal con wavelets en tiempo real sobre DSP". Tesis presentada a la Universidad Nacional de Colombia, sede Manizales, para optar al título de ingeniero electrónico, 2003.
- Ojeda, F. "Extracción de características usando transformada wavelet en la identificación de voces patológicas". Tesis presentada a la Universidad Nacional de Colombia, sede Manizales, para optar al título de ingeniero electrónico, 2003.
- Olszewsky, R. "Generalized feature extraction for structural pattern recognition in time-series data". Tesis doctoral, School of Computer Science, Carnegie Mellon University, Pittsburg, PA 15213, 2001.
- Orozco, M. "Clasificación de arritmias cardiacas usando transformada wavelet y técnicas de reconocimiento de patrones". Tesis presentada a la Universidad Nacional de Colombia, sede Manizales, para optar al título de ingeniero electrónico, 2003.
- Ramírez, J. "Reducción en tiempo real de perturbaciones en señales de ECG empleando la transformada wavelet sobre DSP". Tesis presentada a la Universidad Nacional de Colombia, sede Manizales, para optar al título de ingeniero electrónico, 2004.
- Rencher, A. C. Methods of Multivariate Analysis, Wiley-Interscience, 1992.
- Reyna, M. y JANÉ, R. "Análisis multi-wavelet para la detección de conductividad ventricular anormal en señales ECG de alta resolución". Biomédica, Vol. 12, N^o 2, Abril/Junio 2001, pp. 98-110.
- Rosado, A. et al. "Enhancing feature extraction for VF detection using data mining techniques". The 29th Annual Conference of Computers in Cardiology, 2002.
- Schölkopf, B.; Smola, A. y Müller, K.-R. "Nonlinear component analysis as a kernel eigenvalue problem". Tech. Rep., Max-Planck-Institut für biologische Kybernetik, 1996.
- Schölkopf, B.; Smola, A. Learning with kernels support vector machines, regularization, optimization and beyond. Cambridge, MIT Press, 2002.
- Silipo, R. "Investigating electrocardiographic features in fuzzy models for cardiac arrhythmia classification". International Computer Science Institute, Berkeley, USA, 1999.
- Stamkopoulos, T. et al. "ECG analysis using nonlinear PCA neural networks for ischemia detection". IEEE Transactions On Signal Processing, Vol. 46, N^o 11, Nov., 1998, pp. 3058-3066.
- Suárez, J. et al. "Métodos multivariados en la selección efectiva de características para la clasificación de voces patológicas". Universidad Nacional de Colombia, sede Manizales, 2004.
- Vapnik, V. The nature of statistical learning theory, New York, Springer, 1995.
- Vila, J. "Análisis de la variabilidad de señales fisiológicas, integración en un sistema de monitorización inteligente". Tesis doctoral, Universidad de Santiago de Compostela, Departamento de Electrónica y Computación, 1996.
- Wolfe, P. "The simplex method for quadratic programming". Econometrica, Vol. 27, pp. 382-398, 1959.