



VII CONGRESO NACIONAL DE TECNOLOGÍA APLICADA A CIENCIAS DE LA SALUD

"GENERACION DE NUEVAS TECNICAS DE DIAGNÓSTICO Y TRATAMIENTO"

16-18
junio 2016

Unidad de Seminarios, BUAP



ANÁLISIS DE ENTROPÍA EN TACOGRAMAS COMO MÉTODO DE APOYO AL DIAGNÓSTICO DE INSUFICIENCIA CARDIACA.

Muñoz Diosdado Alejandro, Gálvez Coyt Gonzalo, Solís Montufar Eric Eduardo.
UPIBI, IPN

RESUMEN

Las señales fisiológicas tienden a presentar comportamientos complejos con características fractales y multifractales en estado de salud, y existe una pérdida en la multifractalidad en las series de tiempo de las variables fisiológicas de personas que se encuentran en procesos patológicos. Existe una relación entre la naturaleza multifractal y el concepto de entropía, al involucrarse ambos en la cuantificación de la complejidad característica de los sistemas, uno de los objetivos del presente trabajo es tratar de establecer cual es esta relación. La entropía es un método de análisis no lineal que permite estimar la irregularidad de un sistema, sin embargo, existen diferentes tipos y algoritmos para el cálculo de la misma. El método elegido fue el de entropía de muestreo, al demostrar una independencia mayor de la longitud de las series, pudiendo así, cuantificar la irregularidad de las señales sin verse comprometida por el número de datos. Se utilizó un algoritmo para la generación de series de tiempo fractales con un valor de exponente de potencia fractal β deseado para la caracterización del algoritmo de entropía, obteniendo una correlación significativa al relacionar los valores de β con los valores de entropía. Posteriormente, se empleó este método para analizar series de tiempo de interlatido cardiaco de una base de datos que comprende a 54 sujetos sanos y 44 pacientes de insuficiencia cardiaca congestiva. El cálculo de entropía de muestreo se realizó para registros de 24 horas y series de 6 horas de periodos de sueño y vigilia de los individuos. Una vez analizadas las series, se realizó una prueba estadística de los resultados, corroborando la hipótesis planteada al obtener una diferencia estadísticamente significativa entre las medias de entropía de muestreo de ambas poblaciones, por lo que este método puede ser un apoyo para auxiliar en el diagnóstico de cardiopatías.

1. INTRODUCCIÓN

Actualmente, las enfermedades cardiovasculares son la principal causa de muerte a nivel mundial y más personas mueren por esta razón que por cualquier otra causa [1]. El cuerpo humano tiende a la homeostasis en situaciones de alteración de sus condiciones normales, sin embargo, alteraciones radicales y la combinación de éstas, pueden generar daños severos en el funcionamiento de los sistemas que componen al organismo. Factores como la obesidad, el consumo de tabaco, la inactividad física, hipertensión, hiperlipidemia, diabetes, o el llevar una dieta inadecuada, son factores de riesgo para las enfermedades cardiovasculares, más aún cuando estos se presentan en conjunto [1]. La edad es un factor que afecta el funcionamiento cardíaco y de los vasos, así como de las estructuras y sistemas del cuerpo en general, de modo que favorece la aparición de enfermedades cardiovasculares.

La utilización de técnicas de análisis no lineal para el estudio de sistemas complejos y con propiedades fractales, permite describir de forma más acertada el comportamiento de estos. El cuerpo humano está formado por diferentes sistemas complejos que presentan comportamientos fractales y multifractales en estado de salud [2], y se ha reportado una disminución en estas tendencias fractales cuando los sistemas se encuentran en estado patológico [3].



VII

CONGRESO NACIONAL DE TECNOLOGÍA APLICADA A CIENCIAS DE LA SALUD

16-18
junio 2016

Unidad de Seminarios, BUAP

"GENERACION DE NUEVAS TECNICAS DE DIAGNÓSTICO Y TRATAMIENTO"



La insuficiencia cardiaca (IC) es un síndrome en el que los pacientes presentan las siguientes características: síntomas de IC, típicamente falta de aire o fatiga tanto en reposo como durante el ejercicio; signos de retención de líquidos, como congestión pulmonar o hinchazón de tobillos, y evidencia objetiva de una alteración cardiaca estructural o funcional en reposo [4].

2. TEORÍA

Una serie de tiempo es una secuencia de registros acomodados de acuerdo al tiempo en el que fueron medidos. En muchas ocasiones, la información obtenida de los dispositivos de medición proporciona señales temporales y se denotan por $x(i)$, donde $i = 1, \dots, N$. El registro temporal de los datos obtenidos a partir de mediciones periódicas permite predecir el comportamiento de los sistemas o la optimización del control de los mismos [5].

Las series de tiempo pueden caracterizarse por cuatro tipos básicos de variación, a los cuales se atribuyen los cambios observados en la serie en cierto periodo de tiempo, y confieren a la serie un comportamiento errático. Estos componentes son: Tendencia secular, que representa las variaciones propias de la serie a largo plazo, Variación estacional, que corresponde a la variabilidad de los datos debida a estaciones o periodos específicos de tiempo dentro de la serie, variación cíclica, que son variaciones alternantes que se presentan en periodos de tiempo similares y variaciones irregulares, que son factores a corto plazo, impredecibles y no recurrentes que afectan a la serie de tiempo.

En muchas ocasiones, una serie de tiempo parece no aportar información a vista superficial, pero un análisis más cuidadoso puede mostrar la existencia de correlaciones dentro de la serie, esto es, la existencia de alguna influencia entre el valor de un instante i con el valor en el instante j siempre y cuando $j > i$. Cuando no existe influencia entre el valor en el instante i con el valor en el instante j ($i \neq j$) se dice que se tiene una sucesión de muestras independientes, la serie de tiempo es llamada ruido blanco cuya correlación es cero. En el caso de series de tiempo de ruido blanco al realizar un ordenamiento aleatorio de sus componentes se tendría nuevamente un ruido blanco.

Cuando el valor de la variable i influye en el valor o valores futuros de $j > i$ se dice que hay dependencia y por lo tanto existe correlación. Por ejemplo, una serie de tiempo obtenida a partir de la acumulación de los eventos independientes de ruido blanco se le llama ruido o movimiento Browniano, en particular este tipo de movimiento corresponde a correlación de corto alcance [6].

Se utilizó una base de datos de la página Physionet [7], para la obtención de series de interlatido de 24 horas se utilizó la base *Normal Sinus Rhythm RR Interval Database*, compuesta por una población de 54 personas sanas (30 hombres de entre 28 y 71 años, y 24 mujeres de entre 58 y 73 años) y de dos poblaciones de pacientes de insuficiencia cardiaca congestiva, una de ellas correspondiente a la base de datos The BIDMC Congestive Heart Failure RR Interval Database, la cual consta de 15 pacientes (11 hombres de entre 22 y 71 años, y 4 mujeres de entre 54 y 63 años) clasificados en la escala NYHA 3-4, y una última base de datos llamada Congestive Heart Failure RR Interval, con 29 pacientes (8 hombres y dos mujeres de entre 34 y 79 años, y otros 17 individuos de género desconocido) de diferentes clasificaciones NYHA [8].

Las señales de 24 horas fueron segmentadas en series de 6 horas para cuando los sujetos estaban en periodos de sueño y señales de 6 horas para los periodos de vigilia. La discriminación fue posible gracias a la notoria disminución en la frecuencia cardíaca que sufren los individuos mientras se encuentran dormidos, y el contraste con el tiempo de interlatido del individuo despierto, el cual demanda un mayor flujo de sangre y esto se ve reflejado en el tiempo entre R y R [9].



Para calibrar el método se generaron series de tiempo de ruido con un valor determinado de potencia espectral β , estas fueron generadas empleando el método propuesto por Gálvez [6]. En este se genera un ruido gaussiano blanco con N datos, al que posteriormente se le aplica la transformada discreta de Fourier. El resultado es el espectro plano correspondiente a un espectro de potencia $\beta = 0$, donde la amplitud $|X(m)|$ es igual para los coeficientes con $m = 1, 2, 3, \dots, N$. Posteriormente se construyen nuevos coeficientes de Fourier $X'(m)$, a partir de los obtenidos de la transformada de Fourier del ruido blanco gaussiano mediante:

$$X'(m) = \left(\frac{m}{N}\right)^{-\beta/2} X(m) \quad (1)$$

Con $m = 1, 2, 3, \dots, N$. Ahora se toma la magnitud de los datos filtrados $|X'(m)|$ y se genera de forma uniforme y aleatoria una fase θ_m para cada una de ellas en el intervalo $[0, 2\pi]$. A continuación, se obtienen los coeficientes de Fourier para las magnitudes y fases previamente calculadas empleando:

$$X''(m) = |X'(m)|(\cos \theta_m + i \sin \theta_m), \quad n = 1, 2, 3, \dots, N \quad (2)$$

Se realiza la transformada inversa de Fourier a $|X''(m)|$

$$x(n) = \frac{1}{N\delta} \sum_{m=1}^N X''(m) e^{2\pi i n m / N}, \quad (3)$$

$$n = 1, 2, 3, \dots, N.$$

Para obtener un ruido gaussiano con un valor de β , se considera la parte real de la serie [6]. En la Fig. 1 se pueden apreciar las series autoafines con $N = 10000$ y valores de β de 0, 1, 1.5 y 2.0.

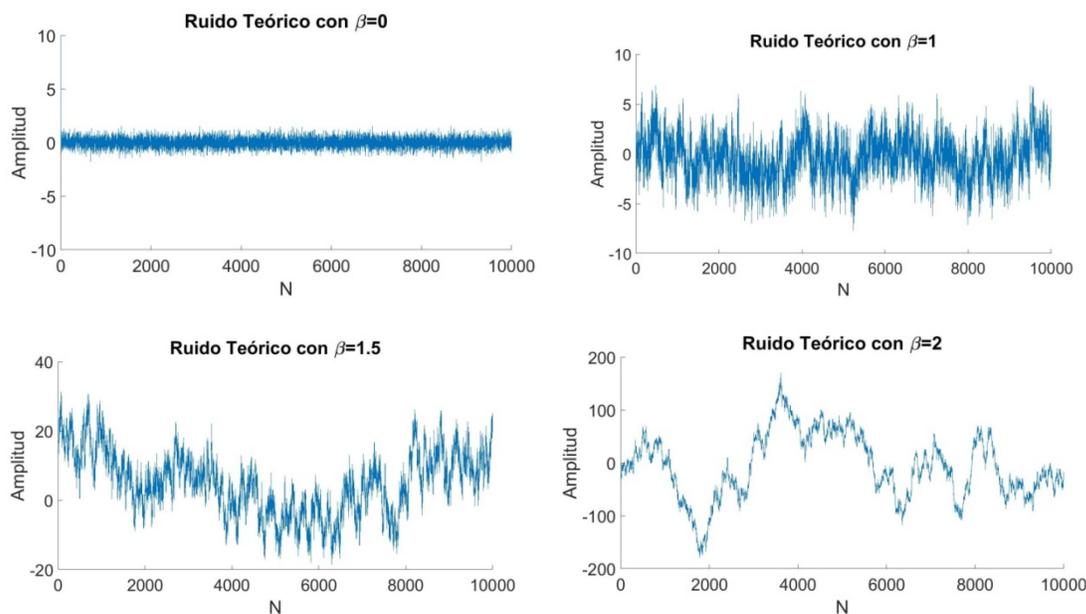


Fig. 1. Gráficas obtenidas de series de tiempo generadas por el algoritmo descrito con exponentes de potencia espectral $\beta=0, 1.0, 1.5$ y 2.0



Se implementó el método para generar un grupo de señales de un número de muestras $N = 1000$ con valores conocidos de β , que va desde $\beta = 0$ a $\beta = 2$ con incrementos de 0.1 entre cada serie. Para realizar el cálculo de la entropía de muestreo (SampEn), se empleó el método propuesto por Richman y Moorman [10]. Tomando en cuenta las series de tiempo, se definen vectores subsiguientes de longitud m definidos por:

$$X_m(i) = \{x(i), x(i+1), \dots, x(i+m-1)\}$$

Para $1 \leq i \leq N - m + 1$. Estos vectores representan m valores consecutivos de x partiendo del punto i . Posteriormente se define la distancia entre los vectores $X_m(i)$ y $X_m(j)$, como el absoluto de la máxima diferencia entre sus componentes escalares:

$$d_m[X_m(i), X_m(j)] = \max[X_m(i+k) - X_m(j+k)], 0 \leq k \leq m-1 \quad (4)$$

Posteriormente, para una $X_m(i)$ dada, se cuenta el número de j ($1 \leq j \leq N - m, j \neq i$), tal que la distancia entre $X_m(i)$ y $X_m(j)$ sea menor o igual a r . Tomando en cuenta $1 \leq i \leq N - m$:

$$B_i^m(r) = \frac{1}{N-m-1} B_i \quad (5)$$

Se define $B^m(r)$ como:

$$B^m(r) = \frac{1}{N-m} \sum_{i=1}^{N-m} B_i^m(r) \quad (6)$$

Posteriormente se realiza un incremento en la dimensión a $m + 1$ y se calcula $A_i^m(r)$ tal que:

$$A(r) = \frac{1}{N-m} \sum_{i=1}^{N-m} A_i^m(r) \quad (7)$$

En consecuencia, $B^m(r)$ se define como la probabilidad de que dos secuencias coincidan en m puntos, y así, $A^m(r)$ corresponde a la probabilidad de que dos secuencias coincidan en $m+1$. Así puede definirse la Entropía de muestreo como:

$$SampEn(m, r, N) = -\ln \left[\frac{A^m(r)}{B^m(r)} \right] \quad (8)$$

Los valores de r y m son determinantes para la precisión. En este caso, se eligieron valores de $r = 0.2$ y $m = 3$, los cuales han demostrado ser apropiados para este tipo de series de tiempo [9].

3. RESULTADOS

Para la caracterización del método de entropía de muestreo, se aplicó a las series de ruido teórico generadas previamente. Calculando el valor de entropía para series de $N = 10000$ y valores de exponente de potencia espectral desde $\beta = 0$ a $\beta = 2$ con incrementos de 0.1. En la Fig. 2 se muestra la gráfica obtenida.

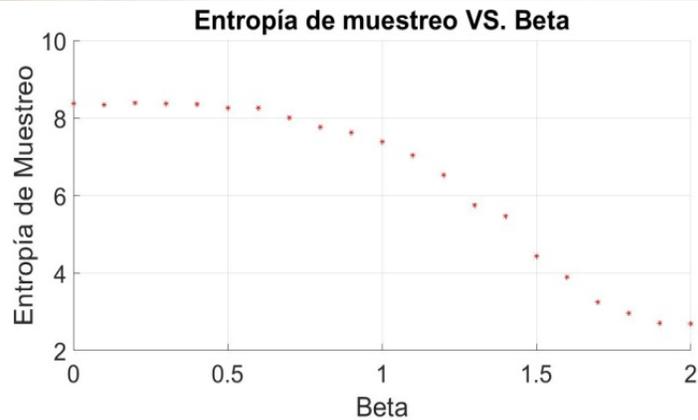


Fig. 2. Entropía de muestreo para diferentes valores de β

Una vez caracterizado el método con valores de β conocidos, se analizaron las series de tiempo de la base de datos de pacientes con IC y personas sanas para los tres tipos de series.

Se procedió a realizar el análisis con el método de entropía de muestreo, el cual fue caracterizado previamente con un grupo de series temporales con un valor de β conocido. Se realizó la representación gráfica de los valores de entropía de muestreo para cada una de las muestras, y el valor de promedio de ambas poblaciones como líneas continuas. En las gráficas de dispersión mostradas a continuación, el color azul corresponde a las personas sanas, y el color rojo a los pacientes de Insuficiencia cardiaca. Asimismo, se realizó la prueba estadística para cada tipo de estudio, es decir, series completas, series de 6 horas en periodo de sueño y series de 6 horas en periodo de vigilia empleando el método T-Student con un valor de significancia de 0.5 y las hipótesis que se utilizaron para el estudio anterior: la hipótesis nula $H_0: SE \text{ sanos} = SE \text{ congestivos}$ y la hipótesis alternativa $H_1: SE \text{ sanos} \neq SE \text{ congestivos}$.

En la comparación realizada en periodo de sueño, se aprecia una diferencia considerable entre ambos promedios, lo cual es corroborado con la prueba estadística, al obtener una diferencia significativa entre ambas medias, lo cual permite aprobar la hipótesis alternativa propuesta anteriormente. Mientras que, para las series en las que los individuos se encuentran despiertos, se puede observar una brecha muy pequeña entre ambas medias, e incluso, el promedio de las entropías de personas sanas, tiene un valor menor al de los pacientes con IC, lo cual no corresponde con los valores obtenidos para el periodo de sueño.

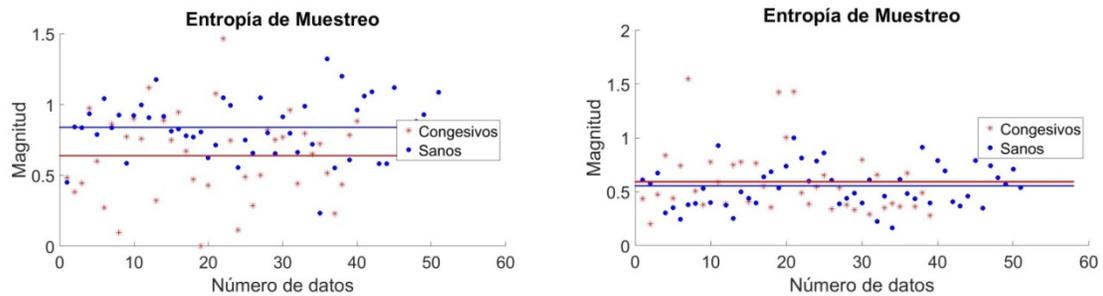


Fig. 3. Izquierda: Entropía de muestreo de personas sanas y pacientes con IC en series de 6 horas en periodo de sueño. Derecha: Lo mismo en 6 horas en periodo de vigilia



Una vez analizada la base de datos con el método de curvatura del espectro multifractal, se procedió a realizar el análisis de ellas, pero con el método de entropía de muestreo, el cual fue caracterizado previamente con un grupo de series temporales con un valor de Beta conocido. Se realizó la representación gráfica de los valores de entropía de muestreo para cada una de las muestras, y el valor de promedio de ambas poblaciones como líneas continuas. En las gráficas de dispersión mostradas a continuación, el color azul corresponde a las personas sanas, y el color rojo a los pacientes de Insuficiencia cardiaca. Asimismo, se realizó la prueba estadística para cada tipo de estudio, es decir, series completas, series de 6 horas en periodo de sueño y series de 6 horas en periodo de vigilia empleando el método T-Student con un valor de significancia de 0.5 y las hipótesis que se utilizaron para el estudio anterior: la hipótesis nula $H_0: SE sanos = SE congestivos$ y la hipótesis alternativa $H_1: SE sanos \neq SE congestivos$. En la comparación realizada en periodo de sueño, se aprecia una diferencia considerable entre ambos promedios, lo cual es corroborado con la prueba estadística, al obtener una diferencia significativa entre ambas medias, lo cual permite aprobar la hipótesis alternativa propuesta anteriormente. Mientras que, para las series en las que los individuos se encuentran despiertos, se puede observar una brecha muy pequeña entre ambas medias, e incluso, el promedio de las entropías de personas sanas, tiene un valor menor al de los pacientes con IC, lo cual no corresponde con los valores obtenidos para el periodo de sueño.

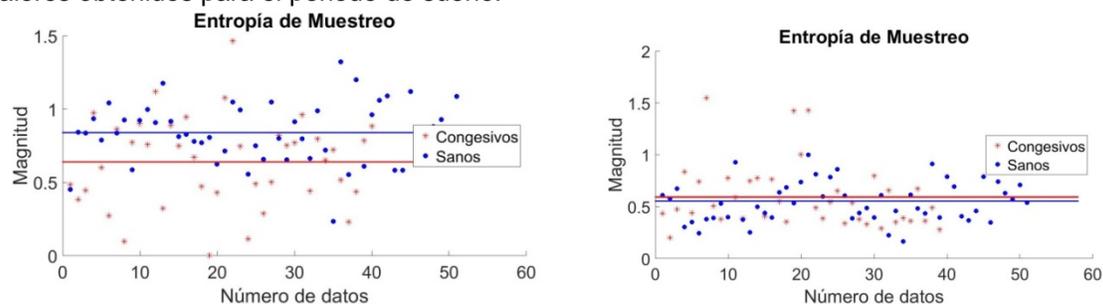


Fig. 4. Izquierda: Entropía de muestreo de personas sanas y pacientes con IC en series de 6 horas en periodo de sueño. Derecha: Lo mismo en series de 6 horas en periodo de vigilia

Posteriormente, se realizó el cálculo de la entropía de muestreo pero sólo para las series de 6 horas, ya que la mayor diferencia entre curvatura fue observada en las series obtenidas cuando los sujetos están dormidos, y tomando en cuenta que el método de la entropía de muestreo es muy demandante en cuanto tiempo de cómputo, se optó por realizar únicamente el estudio que proporcionara más información para realizar una discriminación entre sanos y enfermos, posteriormente se procedió a realizar el análisis estadístico para la entropía de muestreo, tomando en cuenta la siguiente hipótesis en los casos de series de dormidos y despiertos.

$$\begin{aligned}
 H_0: & \quad SampEn sanos = SampEn congestivos \\
 H_1: & \quad SampEn sanos \neq SampEn congestivos
 \end{aligned}$$

En este estudio, los promedios de entropía de muestreo resultaron no tener una diferencia estadísticamente significativa, lo cual lleva a la aceptación de la hipótesis nula H_0 y al rechazo de la hipótesis alternativa que distinguía a ambas poblaciones. Mientras que, para los promedios de entropía en sujetos dormidos se logró obtener lo que se esperaba, una diferencia estadísticamente significativa entre ambas medias, lo cual corrobora lo obtenido en el caso anterior, una discriminación entre ambas poblaciones que se agudiza en el periodo de sueño.



VII

CONGRESO NACIONAL DE TECNOLOGÍA APLICADA A CIENCIAS DE LA SALUD

16-18
junio 2016

Unidad de Seminarios, BUAP

"GENERACION DE NUEVAS TECNICAS DE DIAGNÓSTICO Y TRATAMIENTO"



4. CONCLUSIONES

El método de entropía de muestreo es un algoritmo muy demandante en cuanto a tiempo de procesamiento y memoria del ordenador, por lo que se recomienda el uso de diferentes computadoras procesando las series en paralelo para obtener resultados en menor tiempo. Como se mencionó previamente, el método de entropía de muestreo no presenta alteraciones en sus resultados por la longitud de las series, sin embargo, el tiempo de procesamiento aumenta de forma exponencial al incrementar el tamaño del vector de la serie, lo cual demora mucho al cómputo de las series de 24 horas, tanto para las bases de datos de sanos y pacientes de IC. El uso del valor $r = 0.2$, representa una escala del valor de desviación estándar de la serie total, por lo que el rango de comparación de la serie es $r \times DS$ el cual permitió obtener resultados fiables en el cálculo de la entropía de muestreo. Así también, se recomienda el uso del valor $m = 3$ en cual fue tomado de la bibliografía como se reporta previamente, el cual es reconocido como un valor que permite un estudio de entropía satisfactorio y aportando precisión al método, sin perder información.

BIBLIOGRAFÍA

1. Cardiovascular Diseases (CVDs). [Internet]. Enero 2015. Disponible en <http://www.who.int/en/>.
2. M. Dvorkin, D. Cardinali, R. Hermoli. Bases Fisiológicas de la Práctica Médica. Editorial Médica Panamericana. Buenos Aires, Argentina. 2010.
3. A. Muñoz Diosdado. E. E. Solís Montufar, A. M. Aguilar Molina, G. Gálvez Coyt. "Análisis de la curvatura del espectro multifractal de series de interlatido cardiaco en sujetos sanos y con insuficiencia cardiaca". XIX Reunión de Física y Matemáticas, México, D. F. 2014.
4. K. Dickstein. "Guía de práctica clínica de la Sociedad Europea de Cardiología (ESC) para el diagnóstico y tratamiento de la insuficiencia cardiaca aguda y crónica". Heart Failure Association (HFA). Disponible en www.revespcardiol.org . 2008.
5. Michael Falk." A. First Course on Time Series Analysis". Chair of Statistics. University of Würzburg. 2012.
6. G. Galvez Coyt. "Un análisis comparativo de métodos fractales en series de tiempo y una aplicación al estudio de sismogramas". Tesis de Doctorado. Escuela Superior de Física y Matemáticas. Instituto Politécnico Nacional. México. 2008.
7. PhysioBank, PhysioToolkit. and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals. Circulation. [Internet]. [Consultado agosto 2014]. Disponible en www.physionet.com.
8. American Heart Association. Classes of Heart Failure. [Internet]. Junio 2015. Disponible en <http://www.heart.org/HEARTORG/Conditions/HeartFailure>.
9. C. Virgilio Gonzáles, J. Zamora Justo. "Análisis con dinámica no lineal y diseño de un monitor de ECG ambulatorio para diagnosticar el daño cardiaco en pacientes con insuficiencia cardiaca y síndrome metabólico". Proyecto Terminal. Unidad Profesional Interdisciplinaria de Biotecnología. Instituto Politécnico Nacional. México. 2015.
10. J. S. Richman, J. Randall Moorman. "Physiological time-series analysis using approximate entropy and simple entropy". The American Journal of Physiology - Heart and Circulatory Physiology. Vol. 278. Pp 2039. 2000.