

Comparación entre el procesamiento de señales por Wavelets y Fourier, aplicados al diagnóstico de cáncer de próstata por una red neuronal

Isidro Amaro¹, Jesus Celis²

¹ Doctorado en Sistemas Computacionales. Universidad Da Vinci
iamaro11@udavinci.edu.mx

² Departamento de Ingenierías Eléctrica y Electrónica del Instituto Tecnológico de Durango.
Durango, Dgo., México

(*Paper received on June 30, 2013, accepted on August 15, 2013*)

Abstract. The present study compares two powerful tools for signals analysis; the result is used in a backpropagation neural network, to diagnose prostate cancer. Obtaining as a result, the use of wavelet analysis as the best strategy for building a diagnostic model, which simplifies the representation of the information in the universe of study, making possible the development of a predictive model more simple, robust and efficient as the results show.

For these results, the database was subjected to an analysis through the Transform Fourier and one analysis through Wavelets, thereby applying a statistical method, non-parametric: Kruskal-Wallis, for dimensionality reduction. And finally submit the information to a neural network for classification, and therefore the diagnosis of the disease.

Keywords: Fourier, Wavelets, Neuronal Networks, Cancer, Prostate.

Resumen. El presente estudio, realiza un análisis experimental, el cual compara dos herramientas poderosas para el análisis de señales, cuyo resultado se utiliza en una red neuronal backpropagation, para el diagnóstico de cáncer de próstata. Obteniendo como resultado, el uso del análisis Wavelets como la mejor estrategia de construcción para el modelo de diagnóstico; el cual simplifica la representación de la información del universo de estudio, haciendo posible el desarrollo de un modelo de predicción más sencillo, robusto y eficiente, como los resultados lo demuestran.

Para tales resultados, la base de datos fue sometida a un análisis a través de la Transformada de Fourier y un análisis a través de Wavelets, para así aplicar un método estadístico, no paramétrico: Kruskal-Wallis, para una reducción de dimensionalidades. Y al final se sometió la información a una red neuronal para su clasificación, y por ende el diagnóstico de la enfermedad.

Palabras clave: Fourier, Wavelets, Redes Neuronales, Cáncer, Próstata.

1 Introducción

Para el análisis de señales se pueden enumerar un gran número de herramientas que se han ido desarrollando a lo largo del tiempo, aunque en este documento nos centraremos en dos de las más usuales, la Transformada de Fourier y el análisis Wavelets. Dichas herramientas ya han probado el arrojar excelentes resultados en el análisis de señales. El propósito de este estudio es la aplicación de estas dos herramientas, para así poder determinar la mejor estrategia de construcción de un modelo de diagnóstico de cáncer de próstata, basado en la espectrometría de masa de la proteína, obteniendo las mejores características aplicadas a una red neuronal. Obteniendo como resultado un modelo de diagnóstico automatizado no invasivo.

Se han encontrado algunos artículos que muestran de forma matemática las diferencias, ventajas y desventajas de la Transformada de Fourier y el análisis Wavelets, como en el artículo de Gilbert Strang², además han hecho trabajos sobre la comparación entre estas herramientas como en el análisis de reconocimiento de voz, eliminación de ruido, detección de vibraciones⁵ y comparación de espectros, y existe uno en específico, el de Petricoin E. F.⁶, donde se utilizan técnicas diferentes a las propuestas, usando en su trabajo un algoritmo genético para extraer las características más importantes de sus señales en crudo y una red de Kohonen perteneciente a la categoría de redes no supervisadas, para la clasificación de los datos, los cuales muestran sus resultados que se comparan al final del trabajo.

1.1 Análisis de Fourier

En la ingeniería, muy a menudo, se representan señales en función del tiempo, pero en muchos casos, se requieren dichas señales en el dominio de la frecuencia, y para hacer el cambio, se hace uso de la Transformada de Fourier, sin que haya pérdida de información en este proceso, aunque se tendrán enfoques distintos, véase figura 1.



Fig. 1. Representación gráfica sobre el cambio de tiempo a frecuencia por la transformada de Fourier.

La Transformada de Fourier descompone mediante sus coeficientes una señal en senos y cosenos de diferentes frecuencias y amplitudes. Se puede ver este análisis como la división de la señal en pequeñas ventanas de tiempo, en las cuales se aplica la función de Fourier a cada una de ellas, el inconveniente es que una vez elegido el tamaño de ventana de tiempo, éste es el único que se le aplica a toda la señal.

1.2 Análisis Wavelets

El análisis por medio de Wavelets, se presenta como una extensión de Fourier, en el cual la información de la señal es convertida a números, los cuales son más fáciles de manipular, incluso al reconstruir la señal original.

Una de las ventajas de esta herramienta, es la posibilidad de mostrar para cada punto de la señal, de manera casi instantánea, la información en el tiempo y frecuencia.

El análisis Wavelet consiste en la descomposición de una señal en un conjunto jerárquico de aproximaciones y detalles. Su principal objetivo es encontrar discontinuidades o picos de corta duración en la señal original, como se observa en la figura 2.

La definición de Wavelet es una forma de onda de duración limitada que tiene un valor promedio de ceros.

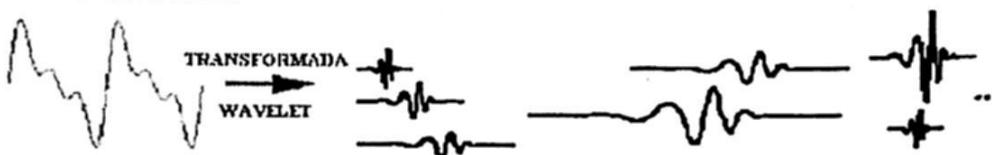


Fig. 2. División de la señal en un conjunto ondas wavelets con distinta posición (adelantar o retrasar) y escala (estirar o encoger) del wavelet original.

1.3 Cáncer de próstata

El cáncer de próstata es reconocido como el tumor maligno más frecuente del varón mayor de 50 años⁴. Tradicionalmente, el método de detección precoz del cáncer de próstata es el tacto rectal, en el que el médico introduce un dedo en el recto para poder palpar la glándula prostática para comprobar la existencia de nódulos sospechosos; ahora bien, después de este examen, se puede llevar a cabo una biopsia transrectal de próstata guiada con ultrasonido, esto debido a la detección de alteraciones en el tacto o en el antígeno prostático.

1.4 Prueba PSA

Para el diagnóstico del cáncer de próstata, desde hace unos años, se puede hacer una prueba de laboratorio llamada como la detección del antígeno prostático específico, conocido como PSA (sus siglas en inglés). Esta prueba inicialmente fue introducida como prueba de laboratorio para el diagnóstico y seguimiento del cáncer de próstata y posteriormente fue propuesta como prueba de diagnóstico precoz.

El PSA elevado (por encima del 4), puede indicar que hay un cáncer de próstata, pero también puede indicar otras patologías como la hiperplasia prostática benigna, o una infección.

2 Proceso

2.1 Descripción General

Una prueba de detección del cáncer de próstata realizada entre 1996 y 2001, fue aprobado por la Junta de Revisión Institucional (IRB) de la Universidad Católica de Chile (Santiago, Chile)⁶. Hombres asintomáticos de 50 años de edad o más sin antecedentes de cáncer de próstata fueron elegibles.

De dicho estudio, se cuenta con una base de datos con la lectura de señales de 322 pacientes basados en la espectrometría de masa de la proteína PSA, para revelar el nivel en que se encuentra dicha proteína, sobre hombres con un diagnóstico histopatológico de cáncer de próstata, de las cuales se tiene información sobre PSA, distribuidas de la siguiente manera:

- 63 pacientes que no muestran evidencia de cáncer o crecimiento benigno con un PSA menor a 1.
- Lecturas de 190 pacientes con crecimiento Benigno con un PSA mayor a 1 y menor que 4.
- 26 pacientes que tienen detectado cáncer de próstata con un PSA entre 4 y 10
- Y 43 pacientes también con diagnóstico de cáncer de próstata con un PSA mayor a 10.

Esta investigación pretende identificar cual herramienta, entre Fourier y Wavelets, extrae las mejores características sobre el análisis de la base de datos, para después con el uso de una herramienta estadística reducir la dimensionalidad de la misma, y con ello alimentar una red neuronal de tipo backpropagation, para así obtener un modelo capaz de diagnosticar si una persona está enferma de cáncer, si tiene crecimiento benigno o se encuentra sana, esto por medio de la espectrometría de masa de proteínas de suero para la detección de cáncer de próstata.

3 Análisis con Fourier

Teniendo la base de datos compuesta de los 322 pacientes, con sus 15154 concentraciones, se realizó el análisis de cada uno de ellos por medio de la Transformada Rápida de Fourier, cabe mencionar que dicho análisis pone algunas limitaciones en la señal y en el espectro resultante. De las señales que se tienen deben consistir de un número de muestras igual a una potencia de dos, en este caso 16384, siendo ésta la más cercana al número de muestras obtenidas, las cuales son rellenasadas con ceros, lo que no implica desajustes.

Se realizaron los análisis correspondientes para poder ver el comportamiento de la señal con los distintos pacientes, como se observa en las figuras 3, 4 y 5, donde se aprecia el comportamiento de un paciente de control.

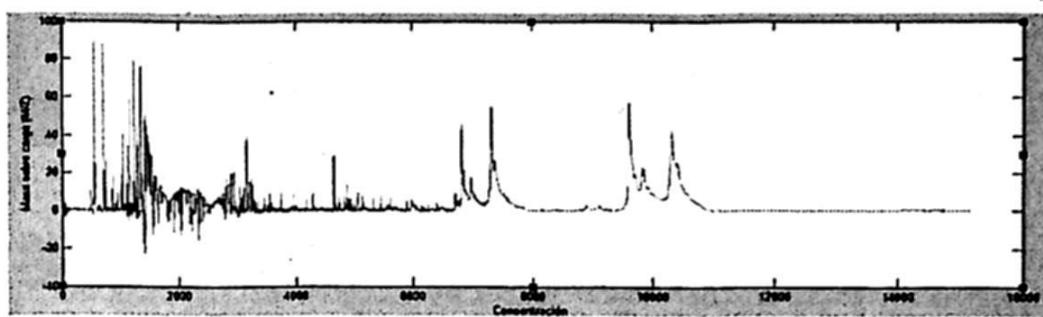


Fig. 3. Comportamiento de la señal de un paciente de control.

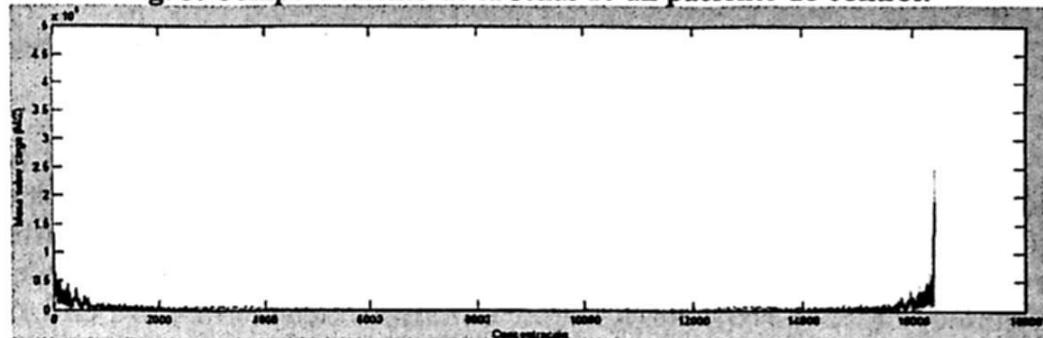


Fig. 4. Señal analizada con la Transformada Rápida de Fourier.

Como se puede ver en la figura 4, muestra la señal con Fourier, pero sólo la mitad es útil para el análisis, ya que a la mitad de la señal se repite la misma como si fuera un espejo, y haciendo un acercamiento a los puntos más importantes se logran distinguir ciertos elementos que tienen puntos diferenciables por su mayor concentración de carga, que son observables a la vista, pero que será necesario determinar si éstos son características posibles de utilizar en la aplicación de la red neuronal, véase la figura 5.

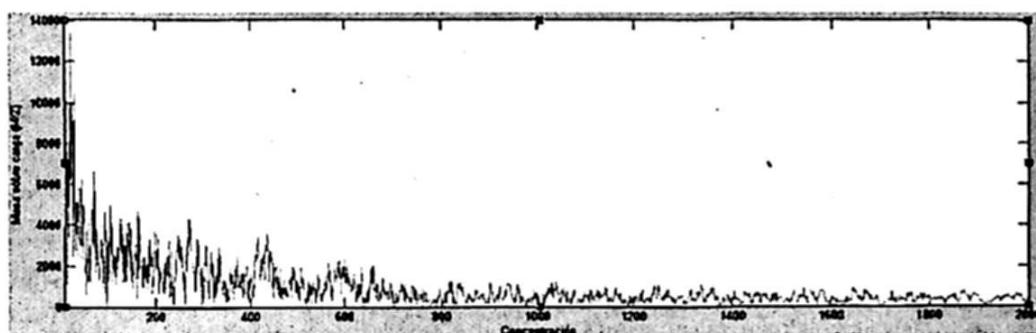


Fig. 5. Representación de la señal en el rango de 10 a 2000.

Al finalizar el análisis de la base de datos, se pueden comparar los tres niveles analizados como se observa en la figura 6, la primera gráfica muestra el análisis realizado a los pacientes de control, la segunda muestra el análisis a los pacientes con crecimiento benigno y en la última se observa a los pacientes con cáncer de próstata.

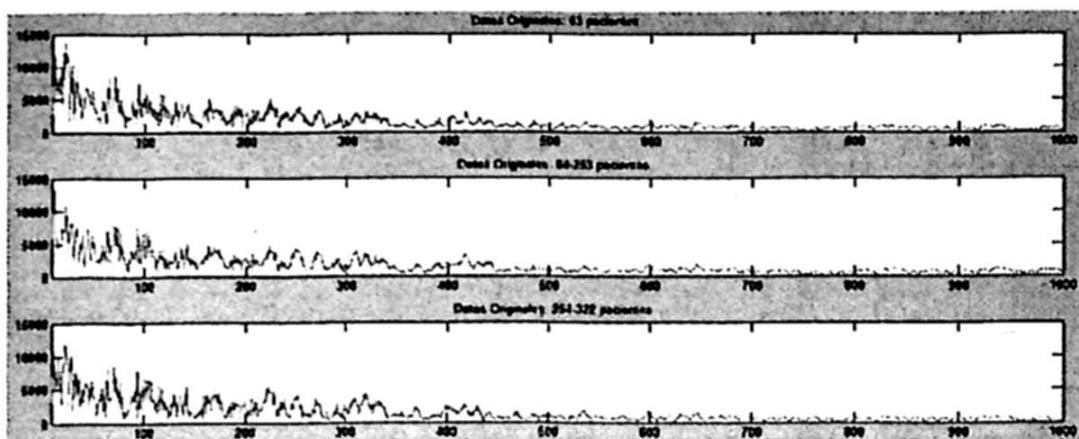


Fig.6. Gráfico de los tres tipos de pacientes.

Con los datos obtenidos, después de la aplicación de Fourier, se realiza la reducción de dimensionalidades por medio de la prueba estadística de Kruskal-Wallis, para así, de tener 16384 características, obtener las principales, que en este caso se extrajeron 505 más significativas, para usarlas en una red neuronal backpropagation.

La arquitectura de la red para construir el modelo de clasificación utilizó una red neuronal de una capa oculta con 15 neuronas con funciones tangente sigmoidea, y una capa de salida con 3 neuronas con funciones tangente sigmoidea.

Su aprendizaje redujo el error cuadrático medio a 0.00001 en 34 épocas, donde se usó el algoritmo de optimización gradiente descendente conjugado para la reducción del error, además de utilizar el algoritmo de detección temprana para evitar el sobre entrenamiento en la red.

1. Los datos de entrada hacia la red fueron 322 pacientes con 505 características significativas cada uno.
2. Para la validación y prueba de los datos se dividieron las 322 muestras de forma aleatoria, poniendo un 70% para el entrenamiento de la red, un 15% para la validación y un 15% para la prueba.
3. Para la arquitectura de la red, se utilizaron 15 neuronas en la capa oculta, con tres a la salida.

		Matriz de confusión de Prueba		
		1	2	3
Clases de Salida	1	10 20.8%	0 0.0%	0 0.0%
	2	0 0.0%	25 52.1%	1 2.1%
	3	1 2.1%	0 0.0%	11 22.9%
	90.9% 9.1%	100% 0.0%	91.7% 8.3%	95.8% 4.2%
		Clases Objetivo		

Fig. 7. Matriz de Prueba.

Después del procesamiento de los datos por medio de la red neuronal, se observó que en la matriz de entrenamiento existe un 100% de efectividad, lo cual indica que los patrones utilizados para entrenar la red fueron clasificados de manera exacta.

Sin embargo, la matriz de validación clasificó de manera correcta los pacientes del grupo de control y los de crecimiento benigno, pero tuvo dos elementos clasificados erróneamente, ya que se clasificaron como de crecimiento benigno, y éstos eran pacientes determinados con cáncer, lo que nos genera un 95% de efectividad en ese rubro.

En lo que respecta a la matriz de prueba, figura 7, se observa de manera clara que la red clasificó con gran acierto el grupo de control, en tanto en el grupo de crecimiento benigno se tuvo un error, al clasificar a un paciente diagnosticado con cáncer, y en la tercera parte de la gráfica, donde se concentran los pacientes con la enfermedad, se clasificó de forma incorrecta a un elemento, al colocarlo en el grupo de control. Tales datos da un 95.8% de efectividad en el total de la clasificación, nada despreciable.

4 Análisis con Wavelets

Haciendo uso de la misma base de datos, se realizó el análisis con wavelets, el cual fue utilizado para un análisis multiseñal de una dimensión.

Existen varias familias de wavelets (Haar, Daubechies, Coiflets,...), y por ende, la necesidad de saber cómo elegir la correcta para nuestro trabajo es una tarea difícil, sin embargo, no existen criterios para escoger una familia frente a otra, aunque hay varios trabajos que muestran ciertos puntos a notar a la hora de la elección.

En este trabajo, se llevó a cabo una descomposición wavelet con nivel 5 y la familia Symmlets de nivel 4, ya que tienen soporte compacto y máxima cantidad de momentos, además de que posee las características de ortogonalidad y biortogonalidad.

En la figura 8, la primera gráfica corresponde a las señales originales, la segunda gráfica ilustra el análisis con wavelets, según los criterios especificados con anterioridad, y en la tercera gráfica se muestra la matriz que contiene la suma de todos los detalles analizados.

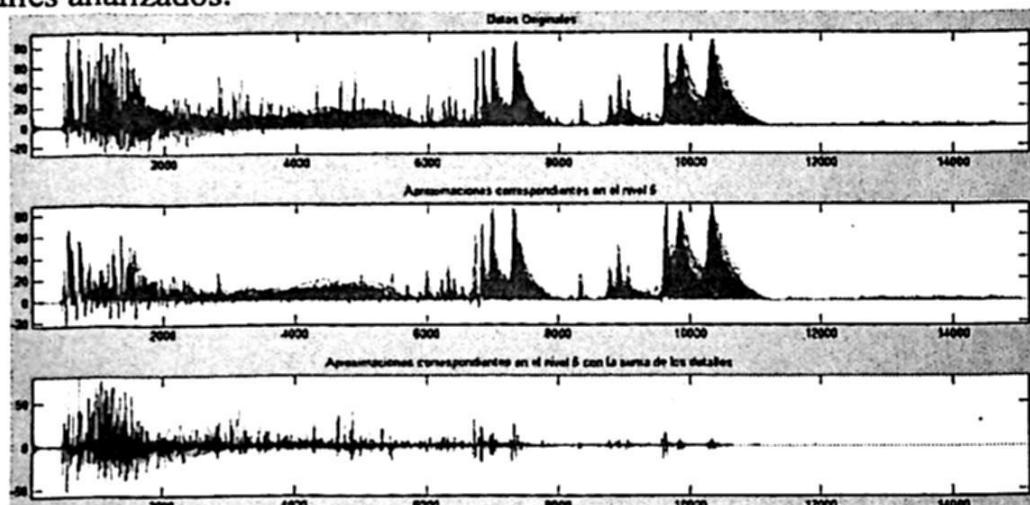


Fig. 8. Señales con Wavelets.

De la misma manera que en Fourier, a los datos obtenidos con el análisis Wavelets, se les realiza la reducción de dimensionalidades por medio de la prueba estadística Kruskal-Wallis, para así, de tener 15154 características originales, se extrajeron 907 que fueron las más significativas, las cuales fueron usadas en una red neuronal backpropagation, de las mismas características que el análisis anterior.

La arquitectura de la red para construir el modelo de clasificación utilizó una red neuronal de una capa oculta con 15 neuronas y una capa de salida con 3 neuronas con funciones tangente sigmoidea.

Su aprendizaje redujo el error cuadrático medio a 0.00001 en 31 épocas, donde se usó el algoritmo de optimización gradiente descendente conjugado para la reducción del error, además de utilizar el algoritmo de detección temprana para evitar sobreentrenamiento en la red.

1. Los datos de entrada hacia la red fueron 322 pacientes con 907 características significativas cada uno.
2. Para la validación y prueba de los datos se dividieron las 322 muestras de forma aleatoria, poniendo un 70% para el entrenamiento de la red, un 15% para la validación y un 15% para la prueba.
3. Para la arquitectura de la red, se utilizaron 15 neuronas en la capa oculta, con tres a la salida.

Después del procesamiento de los datos por medio de la red neuronal, se pueden observar datos importantes en cada una de las matrices obtenidas en el proceso; en la

matriz de entrenamiento se pudo observar un 99% de efectividad, esto debido a que la red clasificó a dos pacientes con cáncer como pacientes con crecimiento benigno.

En la matriz de validación se observó que la red clasificó de manera correcta los pacientes de control; en los de crecimiento benigno, uno lo clasificó como paciente con cáncer, y en los pacientes con cáncer, tres de ellos los clasificó como de crecimiento benigno, lo que nos generó un 91% de efectividad en ese rubro.

La matriz más importante que es la de prueba, véase la figura 9, explica de manera clara que la red no tuvo problemas con la clasificación de ningún paciente, esto nos da un 100% de efectividad en la clasificación.

		Matriz de confusión		
		1	2	3
Clases de Salida	1	9 18.8%	0 0.0%	0 0.0%
	2	0 0.0%	34 70.8%	0 0.0%
	3	0 0.0%	0 0.0%	5 10.4%
		100% 0.0%	100% 0.0%	100% 0.0%
		1	2	3

Fig. 9. Matriz de confusión de Prueba.

5 Comparación de Resultados

Se realizó un comparativo de dos técnicas de análisis de señales: la Transformada de Fourier y Wavelets, con el fin de simplificar la complejidad de la base de datos y extraer las mejores características, y así determinar cuál es la herramienta más adecuada para su uso en la construcción de un modelo de clasificación para un sistema automatizado de diagnóstico no invasivo de cáncer de próstata por medio de una red neuronal.

Estas técnicas hacen posible el desarrollo del modelo de predicción más sencillo, robusto y eficiente, ya que si se intentara trabajar la red neuronal con los datos totales que se tienen en la base de datos, las necesidades computacionales se verían afectadas creando la necesidad de máquinas más robustas de mayor procesamiento, por tratar de manejar miles de variables, que en este caso no sería posible.

En cuanto a los resultados obtenidos, primeramente se menciona el uso de las mismas características para ambos análisis, con la intención de tener una igual

comparación de datos, donde se observa que el porcentaje de clasificación es aceptable para ambos análisis realizados, Fourier y Wavelets, pero en números, se observa un mejor diagnóstico de la red haciendo uso de los datos previamente analizados con Wavelets.

Se obtuvo un 95% de efectividad usando el análisis de Fourier, y un 100% con Wavelets, como se mostró en cada análisis. Cabe hacer mención, que se realizaron varios ajustes a la red, trabajando con diferente números de neuronas y reentrenando en varias ocasiones, obteniendo resultados similares con ambas técnicas, lo que nos indica que Wavelets tiene un mayor poder de representación del problema sobre Fourier.

Con respecto a los resultados presentados en el artículo de Petricoin E. F.⁶, se muestra que ellos tienen un 95% de efectividad en su diagnóstico, mientras que en nuestro trabajo se obtuvo el mismo porcentaje usando Fourier, pero realizando el análisis Wavelets, se incrementó al 100%, lo que asevera una mejoría en los resultados siguiendo las técnicas aquí presentadas, sobre todo el análisis Wavelets.

Referencias

1. Cortés, Jimmy Alexander., Cano, Hugo., Chaves, José Andrés.: Del análisis de Fourier a las Wavelets - transformada continua Wavelet (CWT). Scientia et Technica Año XIII, No 37. Universidad Tecnológica de Pereira. ISSN 0122-1701. PAG 133. (2007)
2. Gilbert Strang. Wavelet transforms versus Fourier transforms. Bull. Amer. Math. Soc., 288-305. Primary 42C15; Secondary 65T20, 94A11. (1993)
3. Hernández Diaz, Marianito. Análisis Comparativo de Algoritmos para Reducción de Ruido en Señales Utilizando Wavelets. Universidad de las Américas Puebla. Escuela de Ingeniería. Departamento de Ingeniería Electrónica. Cholula, Puebla. (2003)
4. Jiménez et al, Cancerología 6 (2011): 13 - 18. Oncoguía: Cáncer de Próstata. Revista del Instituto Nacional de Cancerología. Volumen 6, Número 3. (2011)
5. Osorio Sánchez, A. 2006. Algoritmo para detección de vibraciones anormales en maquinarias utilizando la transformada wavelet. Tesis Maestría. Ciencias con Especialidad en Ingeniería Electrónica. Departamento de Computación, Electrónica, Física e Innovación, Escuela de Ingeniería y Ciencias, Universidad de las Américas Puebla. Mayo. Derechos Reservados © 2006.
6. Petricoin E. F., III, Ornstein D. K., Paweletz C. P., Ardekani A., Hackett P. S., Hitt B. A., Velassco A., Trucco C., Wiegand L., Wood K., Simone C. B., Levine P. J., Linehan W. M., Emmert-Buck M. R., Steinberg S. M., Kohn E. C., Liotta L. A. Serum proteomic patterns for detection of prostate cancer. Brief Communications. Journal of the National Cancer Institute, Vol. 94, No. 20. (2002).
7. Rufiner Di Persia, Hugo Leonardo. Comparación entre análisis Wavelets y Fourier aplicados al reconocimiento automático del habla. Universidad Autónoma Metropolitana – Iztapalapa. División Ciencias Básicas e Ingeniería. (1996)
8. The MathWorks, Inc, <http://www.mathworks.es/es/help/wavelet/ref/mdwtrec.html>
9. Triola, Mario F. Estadística. Décima edición. Editorial Pearson (2009)