

**CLASSIFICATION OF NEURODEGENERATIVE (EP, EH, EELA) DISEASES
USING LS-SVM AND ADABOOSTING LEARNING ALGORITHMS****CLASIFICACIÓN DE ENFERMEDADES NEURODEGENERATIVAS (EP, EH,
EELA) USANDO LOS ALGORITMOS DE APRENDIZAJE LS-SVM Y
ADABOOSTING****Ing. Julio Cesar Rico, MSc. Luis Enrique Mendoza, MSc. Hernando Velandia V.**

Universidad de Pamplona, Facultad de Ingeniería y Arquitectura.
Programa de Ingeniería electrónica e Ingeniería en Telecomunicaciones.
Grupo de Investigación en Ingeniería Biomédica GIBUP.
Tel.: (+577) - 568 5303, Fax: (+577) - 5685303, Ext. 144.
E-mail: julioceasrjcrs@gmail.com, luis.mendoza@unipamplona.edu.co.

Abstract: This paper present a method for multi-classification of neurodegenerative diseases known as Parkinson's (PD), Huntington's (HD) and Amyotrophic Lateral Sclerosis (EELA). We show the importance and performance of multi-classification algorithms known as LS-SVM and Adabosting, applied to EMG signals which characterize the neurodegenerative diseases. The mathematical tools used to pre-processing signals are Fourier transform (FT), cosine discrete transform (DCT), principal component analysis (PCA) and wavelet. The obtained results of this multi-classification system, we have a support tool for the specialist in the detection of EP, EH and EELA disease. Further we demonstrated that the EMG signals can be used to diagnose diseases was as: EH, EP EELA.

Keywords: LS-SVM, Adabosting, TF, DCT, PCA, wavelet.

Resumen: Este trabajo presenta un método para la multi-clasificación de enfermedades neurodegenerativas conocidas como: Parkinson (EP), Huntington (EH) y la Esclerosis Lateral Amiotrófica (EELA). Se muestra la importancia y resultados de los Algoritmos de multi-clasificación conocidos como Adabosting y LS-SVM aplicado a señales EMG las cuales caracterizan las enfermedades neurodegenerativas mencionadas. Las herramientas matemáticas usadas para acondicionar las señales son: transformada de Fourier (TF), transformada discreta del coseno (DCT), análisis de componentes principales (PCA) y wavelet. Con los resultados conseguidos de este sistema de multi-clasificación, se genera una herramienta de soporte para el especialista en la detección de las enfermedades EP, EH y EELA. Además se demostró que las señales electromiográficas pueden ser usadas para diagnosticar enfermedades como: EH, EP y EELA.

Palabras claves: LS-SVM, Adabosting, TF, DCT, PCA, wavelet.

1. INTRODUCCIÓN

En la actualidad las enfermedades neurodegenerativas afectan generalmente la población de edad avanzada y sus sintomatologías

se están ramificando provocando un crecimiento en sus diagnósticos, los estudios se centran en detener el desarrollo de estas enfermedades intentando lograr terapias celulares (Carlos III, 2010). De igual manera se realizan investigaciones que

permitan mediante el procesado de imágenes del cerebro el acertado y precoz diagnóstico de las enfermedades neurodegenerativas (SIPBA, 2012). Es conocido que estas enfermedades afectan la marcha y la movilidad, siendo necesario cuantificar y distinguir las características generadas de las señales electromiográficas mediante técnicas matemáticas, centro de estudio de este trabajo.

2. ENFERMEDADES NEURODEGENERATIVAS

El concepto "Enfermedades Neurodegenerativas" agrupa una serie de enfermedades que atacan el sistema nervioso del ser humano. Las más conocidas son la enfermedad de Alzheimer, el mal de Parkinson y la Esclerosis Lateral Amiotrófica (Enfermedad de Lou Gehrig o ALS). Típicamente, estas enfermedades se desarrollan en adultos y progresivamente deterioran sus funciones motoras o cognitivas (Citren, 2010).

2.1 Parkinson (EP)

La enfermedad de Parkinson es un trastorno degenerativo del sistema nervioso central. Fue descrita por primera vez en 1817 por James Parkinson, un médico británico que publicó un artículo sobre lo que llamó "la parálisis temblorosa" (James, 1995). En este artículo, expuso los síntomas principales de la enfermedad que posteriormente llevaría su nombre.

2.2 Esclerosis lateral amiotrófica (ELA)

La esclerosis lateral amiotrófica, a veces llamada enfermedad de Lou Gehrig, es una enfermedad neurológica progresiva, invariablemente fatal, que ataca a las células nerviosas (neuronas) encargadas de controlar los músculos voluntarios. Esta enfermedad pertenece a un grupo de dolencias llamado enfermedades de las neuronas motoras, que son caracterizadas por la degeneración gradual y muerte de las neuronas motoras (neurological disorders, 2010).

2.3 Huntington (EH)

En 1872, el médico estadounidense George Huntington escribió sobre una enfermedad que él llamó "una herencia de generaciones muy atrás en el oscuro pasado." (Huntington, 1872).

3. MÁQUINAS DE APRENDIZAJE

Las Máquinas de Aprendizaje son una rama de la Inteligencia Artificial cuyo objetivo es desarrollar algoritmos capaces de generalizar comportamientos a partir de una información no estructurada suministrada en forma de ejemplos.

Según el modelado de funciones de densidad de probabilidad condicionadas podemos encontrar algoritmos de clasificación y de regresión. Los usados en este trabajo fueron los algoritmos de clasificación LS-SVM (Máquinas de Soporte Vectorial de Mínimos Cuadrados) y AdaBoosting.

3.1 Algoritmo de AdaBoost

El algoritmo AdaBoost, es introducido en 1995 por Freund y Schapire. AdaBoost es un algoritmo para la construcción de clasificadores fuertes de combinación lineal (Matas *et al.*, 2009; Cohen, 1995; Velandia *et al.*, 2013).

Notaciones:

y = Conjunto de etiquetas
 Dt = Distribución de los datos
 T = número límite de iteraciones
 h = Hipótesis
 at = Parámetro para actualizar pesos
 z_t = Factor de normalización
 e_i = Error
 H = Hipótesis final

El Algoritmo inicia con un conjunto de datos de entrenamiento $(X_1, Y_1), \dots, (X_m, Y_m)$, con m como la longitud de cada vector. En donde cada X_i pertenece a un dominio o una instancia del espacio X y cada etiqueta Y_i es alguna etiqueta del conjunto Y . Estos datos son etiquetados para su clasificación $x(\{-1, +1\})$ la forma más común es (-1) un clase y (+1) la siguiente clase.

AdaBoost hace un llamado reiterado al algoritmo de aprendizaje, estas rondas son definidas por el número de iteraciones T . Para $t = 1, \dots, T$.

Una de las ideas principales del algoritmo es mantener una distribución o conjunto de pesos sobre el conjunto de entrenamiento. El peso de esta distribución es denotado $Dt(i)$. Inicialmente todos los pesos se establecen por igual, Dt : Distribución de la probabilidad entre los datos. Se inicializa uniformemente:

$$Dt(i) = \frac{1}{N} \quad (1)$$

Las hipótesis $h_t : x \in \{-1, +1\}$ se crean minimizando la suma de los pesos de los ejemplos mal clasificados. De esta forma se consigue que el modelo aprendido en la siguiente iteración otorgue más relevancia a los ejemplos mal clasificados por los modelos anteriores. Las bondades de una hipótesis débil se miden por su error.

AdaBoost emplea un parámetro:

$$\mathbf{a}t = \frac{\frac{1}{2} \ln((1 - e_t))}{e_t}$$

Al combinar las hipótesis h_t y determinar las distribuciones D_t . Por una parte, $\mathbf{a}t$ definido en función del error e_t , mide la relevancia de h_t en el Clasificador final:

$$H(x) = \text{sign} \left(\sum_{t=1}^T \mathbf{a}_t h_t(x) \right)$$

Nótese que $\mathbf{a}t \geq 0$ si $e_t \leq 0.5$ y que $\mathbf{a}t$ es mayor cuanto menor es e_t dando más importancia a aquellas hipótesis que comenten menos errores [Freud *et al.*, 1999].

3.2 Algoritmo LS-SVM

Las Máquinas de Soporte Vectorial de Mínimos Cuadrados (LS-SVM) son una reformulación de las SVMs de Vapnik, en las que la optimización lleva a resolver un sistema de ecuaciones lineales, más sencillo de utilizar que las soluciones a la programación cuadrática.

Una SVM primero mapea los puntos de entrada a un espacio de características de una dimensión mayor por ejemplo si los puntos de entrada están en \mathfrak{R}_2 entonces son mapeados por la SVM a \mathfrak{R}_3 y encuentra una serie de planos que puedan separar y maximizar el margen m entre las clases en este espacio como se aprecia en la Figura 1, estos planos se conocen como Hyperplanos (Cortes *et al.*, 1995).

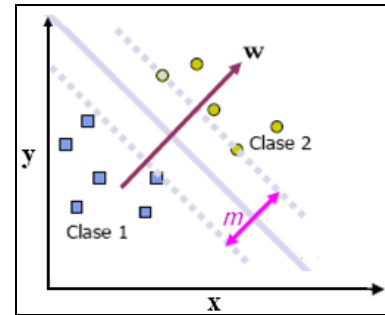


Fig. 1. Hyperplanos y margen m

Maximizar el margen m es un problema de Programación Cuadrática (QP) y puede ser resuelto por su problema dual introduciendo multiplicadores de LaGrange. Sin ningún conocimiento del mapeo, la SVM encuentra el hiperplano óptimo utilizando el producto punto con funciones en el espacio de características que son llamadas *Kernels*. La solución del hiperplano óptimo puede ser escrita como la combinación de unos pocos puntos de entrada que son llamados vectores de soporte (Vapnik, 1998).

4. METODOLOGÍA

La clasificación se realiza sobre las señales electromiográficas, sus formas de onda contienen alto grado de similitud (figura 2), estas no generan patrones característicos para su correcta clasificación, por eso el primer paso realizado fue acondicionar, normalizar las señales entre rangos de 0-1.

El siguiente paso es el procesamiento de las señales mediante herramientas matemáticas.

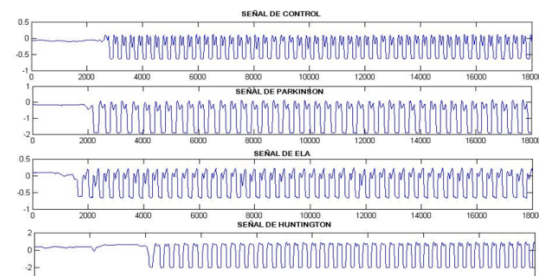


Fig. 2. Señales electromiográficas originales

Inicialmente se aplicó transformada de Fourier (TDF) y transformada discreta del coseno (DCT), para mapear las señales en el dominio de la frecuencia, en este proceso se seleccionaron sus respectivas zonas de interés como se ilustra la figura 3 y 4.

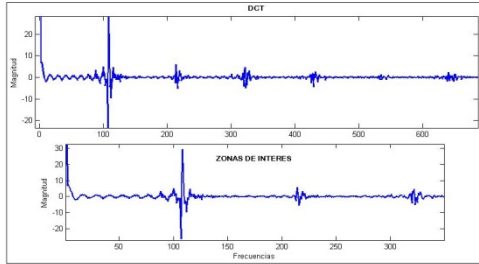


Fig. 3. Zonas de interés en DCT

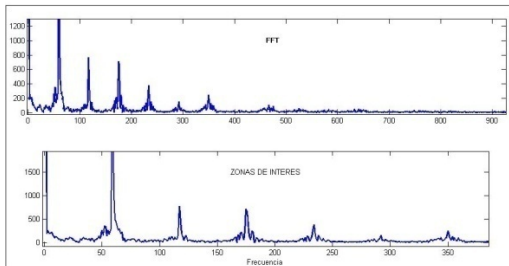


Fig. 4. TDF con Zonas de interés

Las anteriores técnicas se examinaron mediante el uso de la correlación para identificar cual facilitaría la tarea de clasificación.

señales DCT		señales TDF	
correlacionadas	Promedio	correlacionadas	Promedio
PP	0.3317	PP	0.5697
PE	0.3333	PE	0.5749
PC	0.2909	PC	0.5433
PH	0.2939	PH	0.5881
EP	0.3333	EP	0.5749
EE	0.4762	EE	0.6969
EC	0.3523	EC	0.5351
EH	0.3558	EH	0.6089
CP	0.2909	CP	0.5433
CE	0.3523	CE	0.5351
CC	0.3398	CC	0.6126
CH	0.3140	CH	0.5986
HP	0.2939	HP	0.5881
HE	0.3558	HE	0.6089
HC	0.3140	HC	0.5986
HH	0.3685	HH	0.6532

Fig. 5. Promedio de correlaciones en TDF y DCT

Como resultado de la comparación (figura 5), Se observó que Presenta mayor porcentaje de correlación en la comparación de las señales usando TDF ya que al menos en la comparación mutua de sus clases mostraba similitud facilitando la tarea del clasificador.

Teniendo en cuenta el costo computacional que conlleva la gran cantidad de muestras de las señales utilizadas (18000 Muestras), se mejoran los resultados con análisis de componentes principales (PCA), realizando los siguientes procesos:

Primero: Se intenta reducir la cantidad de muestras por promedio cada cuatro puntos, como resultado de 18.000 muestras por señal se reduce a

4500 de estas componentes se seleccionaron por el método de porcentajes 19 componentes.

Segundo: Para reducir las muestras se tomó el proceso de transformada de Fourier con la selección de sus zonas de interés, este otorga comprensión a 400 muestras.

En esta ocasión se seleccionaron 51 componentes principales, dado que la suma total de esta cantidad de porcentajes arrojaba un aproximado del 100% de la información.

Los resultados presentaron mejor funcionamiento con el uso de PCA-TDF, por el contrario los errores aumentan en las muestras reducidas por promedio y se entiende un mal funcionamiento en ese procedimiento.

El procesamiento termina con el uso de la transformada wavelet discreta (DWT).

Realizando DWT se obtiene una serie de coeficientes que se pueden interpretar gráficamente, así con los coeficientes de detalle puedo identificar componentes frecuenciales o sectores que me permitan seleccionar patrones para diferenciar las señales y con los de aproximación pueden arrojar mejorías en la respuesta de los clasificadores y ventajas en la disminución del tiempo computacional (Santafé et al, 2012).

5. RESULTADOS

La Tabla 1 muestra la comparación entre los clasificadores usados en este trabajo AdaBoost y LS-SVM, con los correspondientes porcentajes de errores de las enfermedades EP, ELA y EH. En esta comparación se tienen en cuenta los porcentajes tanto de las señales originales como de las dos técnicas que mejores resultados ofrecieron (TDF y TDF-PCA). En donde claramente se destaca un mejor rendimiento con el algoritmo AdaBoost, sus porcentajes de error en la mayoría de los casos son menores que LS-SVM.

Tabla 1. Porcentajes de error

Clasificadores	Señales originales (%)			TDF (%)			TDF-PCA (%)		
	EP	ELA	EH	EP	ELA	EH	EP	ELA	EH
AdaBoost	30.76	23.07	46.15	30.76	0.0	30.76	23.07	0.0	15.38
LS-SVM	53.84	46.15	53.84	53.84	46.15	53.84	15.38	7.69	53.84

Se creó una Interfaz en GUIDE, con la finalidad de ser presentada como una herramienta de soporte para que un especialista en neurología pueda manipularla y ofrecer un diagnóstico rápido y oportuno de EH, EP y EELA.

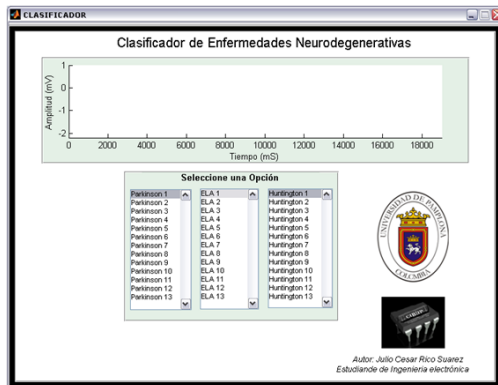


Fig. 6. GUI del Sistema clasificador

6. CONCLUSIONES

Se demuestra que las señales electromiográficas pueden ser utilizadas para la detección de las enfermedades neurodegenerativas. AdaBoost demuestra ser un gran clasificador para las señales electromiográficas de la pierna izquierda correspondientes a las enfermedades del Parkinson, ELA, y Huntington con un porcentaje adecuado de error.

Aplicando las herramientas matemáticas de TDF, DCT, PCA, el porcentaje de error en los clasificadores (AdaBoost y SVM) fue disminuyendo considerablemente. SVM ofrece buenos resultados en comparación de 2 clases, y tiene ventajas en el diagnóstico persona sana con enferma. El mejor comportamiento del clasificador ha sido con el resultado del procesamiento de la señal aplicando transformada discreta de Fourier, para luego seleccionar sus componentes principales y aplicarlas a AdaBoost Multi-clases. Finalmente, Se encontró dificultades para la validación debido a la cantidad reducida de datos de cada enfermedad.

REFERENCIAS

- Cortes, C. and Vapnik, V. Ñ. (1995). Support vector networks. *Machine Learning*, Vol. 20, pp. 273-297.
- Fundación Citren, [http:// citren.org/? q= node/2](http://citren.org/?q=node/2), 2010.
- Huntington, G. (1872). On Chorea, *Medical and Surgical Reporter*, 26, 320-321.
- Instituto de Salud Carlos III. (2010). Esfuerzo compartido contra las enfermedades neurodegenerativas, <http://www.elblogsaludable.es/archive/2010/07/26/esfuerzo-compartido-c>.
- Matas, J. and Sochman, J. (2009), Centre for Machine Perception, Czech Technical University, Prague, <http://cmp.felk.cvut.cz>.
- National Institute of Neurological Disorders and Stroke. (2010). Esclerosis lateral amiotrófica, http://espanol.ninds.nih.gov/trastornos/esclerosis_lateral_amiotrofica.htm.
- Parkinson, J. (1995). An essay on the shaking palsy.
- Santafe, Y.; Chaparro, B. y Franco, J. (2012). Detección de patrones característicos con transformadas wavelet en señales electromiográficas del cuádriceps. *Revista de Tecnologías de avanzada*. Vol 1, No. 21.
- SiPBA (2010). Procesado de Señal y Aplicaciones Biomédicas de la Universidad de Granada.
- Vapnik, V. Ñ. (1998). *Statistical learning theory*. New York: Wiley.
- Velandia, H.; Medica, R., y Mendoza, L. (2013). Estimación del movimiento de la pared del ventrículo izquierdo y cuantificación de su funcionalidad en imágenes ventriculográficas. *Revista de Tecnologías de Avanzada*, Vol. 1, No. 21.
- Cohen, W. (1995). Fast effective rule induction. In *Proceedings of the Twelfth International Conference on Machine Learning*, pp. 115-123.
- Freund, Y. and Schapire, R E. (1999). Short Introduction to Boosting, *Journal of Japanese Society for Artificial Intelligence*, 14(5):771-780, September.