**Universidad del Cauca**

**Facultad de Ingeniería Electrónica y Telecomunicaciones**

**Programas de Maestría y Doctorado en Ingeniería Telemática**

**Seminario de Investigación**

**Algoritmo para la detección automática de anomalías epileptiformes en señales EEG**

**Maritza Fernanda Mera Gaona**

Estudiante de Doctorado

6 de mayo de 2016

La presentación estuvo compuesta de los siguientes ítems:

# Introduction

La epilepsia es una enfermedad que afecta aproximadamente una de cada 100 personas alrededor del mundo (Song, Crowcroft, & Zhang, 2012). El diagnóstico es realizado a través de estudios de encefalografía, el cual es un método no invasivo de bajo costo empleado para examinar la actividad eléctrica del cerebro (Stam , etal, 1999). Los datos capturados durante la práctica de un encefalograma son mostrados al especialista en forma de onda. El estudio consiste en identificar actividad neuronal anormal en la onda que describe la señal EEG capturada.

La lectura que realizan los especialistas sobre el encefalograma es una tarea que implica mucho esfuerzo por parte del médico, debido a que debe hacer lecturas de grabaciones de señales EEG con duraciones entre 20 y 30 minutos y en algunos casos hasta de 48 horas (Garcés, 2015). Lo anterior representa una de las principales causas del alto costo de diagnóstico de epilepsia. Por otro lado, en países en vía de desarrollo la dificultad del diagnóstico también es afectada por la falta de personal médico; en países como Colombia se tiene un neurólogo por cada 200.000 habitantes, en consecuencia se dificulta garantizar el diagnóstico y atención oportuna a los pacientes (Javeriana, 2013). La situación es más preocupante en el caso de los pacientes que residen en áreas rurales debido a que los especialistas se encuentran ubicados en los centros clínicos de las grandes ciudades.

Teniendo en cuenta lo anterior, la detección automática de eventos epileptiformes en señales EEG surge como una alternativa para reducir los tiempos de lectura de una señal EEG, debido a que una vez identificadas las anormalidades sobre la señal el especialista solo tendría que confirmarlas. Esto representaría una herramienta útil para el soporte de toma de decisiones en el campo clínico.

Durante los últimos años se ha desarrollado un gran número de investigaciones sobre el diagnostico automático de Epilepsia (Jin, Dauwels , & Cash) (Chen, 2014). Algunas soluciones detectan automáticamente eventos epileptiformes e incluso, otras proponen clasificar el tipo de epilepsia que el paciente padece. La mayoria de estas soluciones obtienen resultados de sensibilidad o especificidad superiores a un 95 %. Sin embargo, en las evaluaciones realizadas no se refleja un análisis sobre el rendimiento computacional de la solución o los resultados de clasificación son preliminares (Gopika Gopan, 2013) (Sunil Kumar, 2015).

Los sistemas de diagnóstico automático de Epilepsia extraen un conjunto de valores de las señales EEG; los valores extraídos son la entrada de algoritmos de clasificación que determinan si la señal es normal o anormal. La precisión y eficiencia de los clasificadores depende principalmente de la capacidad de diferenciación de los descriptores calculados. Debido a esto, algunos estudios han empleado algoritmos de selección de características para eliminar descriptores que representen ruido para el clasificador con el fin de obtener mejores resultados en la precisión y exactitud (RAJENDRA, 2012) (Yanti, 2016).

Aquí debemos escribir los trabajos más relevantes en la detección automática de eventos epileptiformes e introducir la selección de características como un aspecto clave de esa detección. Luego mostrar cuales son los trabajos más relevantes en la reducción de dimensionalidad de EEGs(Uno o dos párrafos resumiendo la sección dos).

Este artículo ha sido organizado de la siguiente manera: La sección 2 describe el proceso de detección automática de eventos epileptiformes, basado principalmente en la extracción de características y clasificación. En la sección 3, los resultados obtenidos en el proceso de selección de características son descritos. En la sección 4, los resultados obtenidos son discutidos. Finalmente, algunas conclusiones son descritas en la sección 4.

# Métodos

Un sistema de detección automática de eventos epileptiformes está construido de acuerdo al esquema descrito en la Figura 1. Las etapas de adquisición de señales y procesamiento de la señal generalmente están asociadas al software incorporado en el dispositivo que soporta la captura de la encefalografía.

Las etapas de extracción de características y clasificación describen el proceso mediante el cual se genera una representación de las ondas que observa el especialista para posteriormente a través de técnicas de Aprendizaje de Máquina o Inteligencia Artificial determinar si la señal es normal o anormal; en algunos casos, si un evento epileptiforme es identificado es posible clasificar el tipo de epilepsia que padece el paciente.



Procesamiento de Señal

Extracción de Características

Clasificación

Adquisición de señales EEG

**Figure 1.** Scheme of the automated epileptic seizure detection system.

##  Extracción de características

El proceso de extracción de características consiste en realizar una serie de operaciones matemáticas sobre las señales EEG para obtener un conjunto de descriptores capaces de representar la información contenida en las ondas. Algunas soluciones proponen realizar una tarea adicional en la etapa de extracción de características con el fin de validar la relevancia de los descriptores obtenidos. Este proceso es conocido como Selección de Características y es empleado para eliminar los descriptores que representan ruido o información redundante. Este último proceso es una tarea de optimización, ya que evita realizar operaciones para el cálculo de los descriptores no relevantes y en consecuencia, el clasificador tiene un mejor rendimiento.

### Experimental data

En este trabajo fue empleada la base de datos pública de señales EEG del Hospital Infantil de Boston. La base de datos contiene registros encefalográficos de 21 pacientes tomados empleando el sistema internacional 10-20 con una tasa de muestreo de 256 muestras por segundo (Shoeb, 2010). Cada registro de EEG contiene información de 23 canales y está bajo el formato estándar .edf (European Data Format). EDF es un formato digital empleado para intercambiar y almacenar señales multicanal de origen biológico y físico (Kemp, 1992). De cada paciente se cuenta con varios registros EEG, algunos de ellos presentan episodios epilépticos y los otros registran actividad cerebral normal.

Teniendo en cuenta lo anterior, se seleccionaron 52 segmentos de registros de señales EEG de la base de datos seleccionada, 26 de los segmentos registran actividad cerebral anormal y los otros 26 registran actividad cerebral normal.

### Extractores de Características

La extracción de características se describe como el proceso que emplea un conjunto de operadores para obtener un conjunto de descriptores capaces de representar la información contenida en la señal EEG. Teniendo en cuenta la literatura revisada se identificaron los siguientes descriptores

* Entropía
* Máxima Amplitud
* Mínima Amplitud
* Media
* Varianza
* Máxima Potencia
* Potencia Media

Cada registro multicanal de señal EEG contiene información en 23 canales, de tal manera que los 7 descriptores seleccionados fueron aplicados sobre cada canal dando como resultado un total de 161 características extraídas. Las características extraídas fueron codificadas con el prefijo ‘a’ y el sufijo determinado por el número de característica calculada.

## Clasificación

El proceso de clasificación emplea las características extraídas en la fase anterior para determinar si la señal es normal o anormal. En este trabajo se probaron los clasificadores Naive Bayes, Rule Induction, Arboles de decisión y KNN.

# Resultados

La selección de características (FS) es una técnica ampliamente usada para reducir la dimensionalidad de los datasets en problemas de aprendizaje de máquina, se realiza con el objetivo de eliminar información duplicada o ruido. A través de FS es posible mejorar la eficiencia y exactitud de los clasificadores, debido a que al eliminar la información duplicada o ruido se evitan datos que puedan confundir al clasificador. Generalmente, FS se realiza como una tarea de preprocesamiento para disminuir el costo computacional e incrementar la exactitud de la clasificación. El costo computacional se ve disminuido debido a que al reducir el número de características, se deben realizar menos operaciones en la caracterización de la instancia que se desea clasificar y las operaciones de comparación con las instancias del Dataset.

Para realizar el proceso de selección de características se empleó el software RapidMiner. Siguiendo el proceso de extracción de características descrito en la sección anterior, se extrajeron las características de 52 segmentos de señales EEG. Las características extraídas fueron exportadas a un documento CSV que fue empleado como fuente de datos de los algoritmos de selección de características seleccionados.

Los algoritmos Forward Selection, Backward Elimination y Optimize Selection fueron empleados para el proceso de Selección de Características y los algoritmos Naive Bayes, Rule Induction, Decision Tree y KNN para la clasificación con cada subconjunto de características generado. Los resultados de las pruebas realizadas son descritos en la tabla 1.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Forward Selection** | **Optimize selection** |  | **Backward Elimination** |
| Naïve Bayes | 57.69% | 42.30% |  | 48.07% |
| Rule Induction | 70.00% | 63.46% |  | 40.38% |
| KNN | 71.11% | 42.30% |  | 40.38% |
| Decision Tree | 80.77% | 34.61% |  | 48.07 |

El algoritmo Forward Selection arrojó los mejores resultados en la clasificación con las características seleccionadas. En la figura 3 se muestra el diseño creado en RapidMiner para la selección de características.



Figure 3. Esquema de selección de características.

Forward Selection encuentra el conjunto mínimo de características capaz de clasificar correctamente una señal. El proceso de selección que realiza está basado en la conformación de subconjuntos de características de tamaño 1, 2, 3…,n. A partir de cada subconjunto emplea un algoritmo para clasificar las instancias del dataset empleando únicamente las características del subconjunto seleccionado. Al finalizar la ejecución del algoritmo, si existe, indica el subconjunto mínimo de características que logra clasificar correctamente las instancias del dataset de prueba.

En la figura 4 se muestra el esquema de clasificación empleado dentro del algoritmo Forward Selection.



Figure 4. Esquema de clasificación.

La tabla 2 describe los subconjuntos de características generados por Forward Selection de acuerdo a cada algoritmo de clasificación probado. Forward Selection reduce de 161 características a: (i) 4 empleando el clasificador Naive Bayes, (ii) 8 empleando el clasificador Rule Induction, (iii) 4 empleando arboles de decisión y (iv) 5 empleando KNN. Sin embargo, la mejor exactitud es obtenida utilizando el subconjunto generado de características generado para arboles de decisión (tabla 1).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Característica** | **Naive Bayes** | **Rule Induction** | **Decision Tree** | **KNN** |
| a6 |  |  | x |  |
| a26 | x |  |  | x |
| a27 |  |  | x |  |
| a32 |  | x |  |  |
| a42 |  |  |  | x |
| a45 | x |  |  |  |
| a51 | x |  |  |  |
| a32 |  | x |  |  |
| a69 |  | x |  |  |
| a81 |  | x |  | x |
| a82 |  |  |  | x |
| a83 |  | x |  |  |
| a94 |  |  | x |  |
| a121 |  |  | x |  |
| a129 |  | x |  |  |
| a136 |  | x |  |  |
| a137 |  | x |  |  |
| a160 | x |  |  | x |

Las características seleccionadas son a6, a27, a94 y a121. De acuerdo al esquema de codificación empleado para nombrar las características se puede inferir que la Potencia Media del primer canal, Potencia Media del segundo canal, la Mínima Amplitud del catorceavo canal y la Máxima Amplitud del dieciochoavo canal son las características que permiten determinar si en una señal se presenta un evento epileptiforme.

# Discusión

Teniendo en cuenta el alto costo del diagnóstico de la Epilepsia debido la poca disponibilidad de neurólogos y los largos tiempos de lectura de los encefalogramas, un conjunto de características para soportar la detección automática de eventos epileptiformes es propuesto. Esto contribuirá en el desarrollo de sistemas de soporte al diagnóstico automático de Epilepsia.

Considerando la literatura revisada, diferentes soluciones han sido desarrolladas para detectar automáticamente eventos epileptiformes en señales EEG. Sin embargo, la mayoría de las investigaciones no ha evaluado el rendimiento computacional de las herramientas desarrolladas ni la relevancia de las características calculadas. Por lo tanto, las características propuestas en este trabajo representan un avance para el desarrollo de sistemas eficientes de diagnóstico automático de Epilepsia.

Por otro lado, el uso de las características propuestas permitirá desarrollar sistemas eficientes y precisos que permitan disminuir los tiempos de lectura de encefalogramas. Por lo tanto, los médicos podrían incrementar la cantidad de exámenes inspeccionados.

# Conclusiones

La Selección de Características permitió reducir el conjunto inicial de características extraídas en un 97.51%, lo cual demuestra que aunque es posible obtener diferentes datos de una señal EEG, no toda ésta es relevante para soportar un proceso de clasificación.

Los arboles de decisión arrojaron mejores resultados en la exactitud respecto a los otros algoritmos de clasificación empleados para evaluar los subconjuntos de características generados por el algoritmo de Selección de Características Forward Selection.

El cálculo de la Potencia Media del primer canal, Potencia Media del segundo canal, la Mínima Amplitud del catorceavo canal y la Máxima Amplitud del dieciochoavo en un registro EEG permiten detectar automáticamente un evento epileptiforme con una exactitud del 80.77%.

Los resultados en esta investigación representan un aporte para el proceso de extracción de características de señales EEG, lo que contribuye en el desarrollo de sistemas eficientes de detección automática de eventos epileptiformes. Teniendo en cuenta lo anterior, como trabajo futuro nosotros proponemos calcular y evaluar la relevancia de nuevas características y así mismo, probar nuevos algoritmos de selección de características y clasificación.

References

1. A.N. Author, *Book Title*, Publisher Name, Publisher Location, 1995.
2. A.N. Author, Article title, *Journal Title* **66** (1993), 856–890.
3. Y. Song, J. Crowcroft y J. Zhang, «Automatic epileptic seizure detection in EEGs based on optimized sample entropy and extreme learning machine,» Journal of Neuroscience Methods, vol. 210, pp. 131-146, 2012.
4. C. Stam , j. Pijin, P. Suffczynski y F. Lopez da Silva, «Dynamics of the human alpha rhythm: evidence for non-linearity.,» Clin Neurophysiol, vol. 110, nº 10, pp. 1801-1813, 1999.
5. B. Kemp, A. Varri, A. C. Rosa, K. D. Nielsen y J. Gade, «A simple format for exchange of digitized polygraphic recordings,» Electroencephalography and Clinical Neurophysiology, vol. 82, pp. 391-393, 1992.
6. J. Jin, J. Dauwels y S. Cash, «Automated localization of the seizure focus using interictal intracranial EEG.,» Conf Proc IEEE Eng Med Biol Soc, pp. 4439-4442, 2014.
7. K. Gopika Gopan, . A. Harsha, L. A. Joseph y E. S. Kollialil , «Adaptive neuro-fuzzy classifier for ‘Petit Mal’ epilepsy detection using Mean Teager Energy,» Advances in Computing, Communications and Informatics (ICACCI), 2013 International Conference on , pp. 752-75, 2013.
8. G. Chen, «Automatic EEG seizure detection using dual-tree complex wavelet-Fourier features,» Expert Systems with Applications, vol. 41, nº 5, pp. 2391-2394, Abril 2014.
9. T. Sunil Kumar, V. Kanhangad y R. Bilas Pachori, «Classification of seizure and seizure-free EEG signals using local binary patterns,» Biomedical Signal Processing and Control, vol. 15, pp. 33-40, Enero 2015.
10. U. RAJENDRA ACHARYA, P. CHUAN, R. YANTI y J. S. SURI, «Application of non-linear and wavelet based features for the automated identification of epileptic EEG signals,» International Journal of Neural Systems, vol. 22, nº 2, p. 1250002, abril 2012.
11. Yanti, R. Yanti, G. Swapna, V. Sree y J. Suri, «Automated diagnosis of epileptic electroencephalogram using independent component analysis and discrete wavelet transform for different electroencephalogram durations.,» Proc Inst Mech Eng H, vol. 227, nº 3, pp. 234-244, 2013.
12. A. Garcés Correa, L. Orosco, P. Diez y E. Laciar, «Automatic detection of epileptic seizures in long-term EEG records,» Computers in Biology and Medicine, vol. 57, pp. 66-73, February 2015.
13. P. U. Javeriana, «Estudio de Disponibilidad y Distribución de la Oferta de Médicos Especialistas, en Servicios de Alta y Mediana Complejidad en Colombia,» Pontificia Universidad Javeriana, Bogotá, 2013.
14. A. Shoeb y J. Guttag, «Detection of Epileptic Seizure Event and Onset Using EEG,» Proceedings of the 27th International Conference on Machine Learning , p. 975–982, 2010.