Exploring Complexity in Health: An Interdisciplinary Systems Approach
A. Hoerbst et al. (Eds.)

doi:10.3233/978-1-61499-678-1-722

Towards a Selection Mechanism of Relevant Features for Automatic Epileptic Seizures Detection

Maritza MERA-GAONA[a,1], Rubiel VARGAS-CANAS[b] and Diego M. LOPEZ[a]
[a] *Telematics Engineering Research Group, Universidad del Cauca, Colombia*
[b] *Dynamics systems, Instrumentation and Control Research Group, Universidad del Cauca, Colombia*

Abstract. Background: Epilepsy diagnosis is frequently confirmed using electroencephalogram (EEG) along with clinical data. The main difficulty in the diagnosis is associated with the large amount of data generated by EEG, which must be analyzed by neurologists for identifying abnormalities. One of the main research challenges in this area is the identification of relevant EEG features that allow automatic detection of epileptic seizures, especially when a large number of EEG features are analyzed. Objective: The aim of this paper is to analize the accuracy of algorithms typically used in feature selection processes, in order to propose a mechanism to identify a set of relevant features to support automatic epileptic seizures detection. Results: This paper presents a set of 161 features extracted from EEG signals and the relevance analysis of these features in order to identify a reduced set for efficiently classifying EEG signals in two categories: normal o epileptic seizure (abnormal). A public EEG database was used to assess the relevance of the selected features. The results show that the number of features used for classification were reduced by 97.51%. Conclusions: The paper provided an analysis of the accuracy of three algorithms, typically used in feature selection processes, in the selection of a set of relevant features to support the automatic epileptic seizures detection. The Forward Selection algorithm (FSA) produced the best results in the classification process, with an accuracy of 80.77%.

Keywords. Epilepsy, Epileptic seizure, EEG signal processing, Feature selection

1. Introduction

Epilepsy is a disorder which affects approximately one in every 100 people worldwide [1]. The diagnosis is frequently confirmed through electroencephalogram (EEG), which is a noninvasive and low-cost method used to examine electric activity of the brain [2]. The data captured during an EEG is represented as waveforms signals where a specialist can identify abnormal neuronal activity in the acquired signal. Inspection of EEG is a long-lasting task, because duration of typical Epilepsy EEG recordings are between 20 and 30 minutes and in some cases reaches up to 48 hours [3]. This represents one of the main reasons for the high cost of diagnosis and treatment of Epilepsy. In addition, in Low- Middle Income countries (LMIC) the difficulties and

[1] Corresponding Author: Maritza Mera-Gaona, PhD Student, Telematics Department, University of Cauca, Calle 5 No 4-70, Popayán, Colombia; email: maritzag@unicauca.edu.co

cost of neurological diseases arise due to the lack of trained physicians. In Colombia, for example, there is approximately one neurologist per 200.000 inhabitants, which is almost nine times lower than European Countries [4]. The situation is even worst in rural areas. In recent years, a large number of research has been developed on the automatic diagnosis of Epilepsy, e.g,[5][6]. Some solutions automatically detect epileptic seizures and even propose to classify the type of Epilepsy. Furthermore, the evaluations performed showed that most of these solutions have a sensitivity or specificity higher than 95%. Nonetheless, these studies do not provide any analysis of the computational performance, the classification results presented are preliminary [7][8], and the data used for evaluation are collected in non-clinical settings.

The aim of this paper is to analyze the accuracy of algorithms typically used in feature selection process, in order to propose a mechanism to identify a set of relevant features to support the automatic epileptic seizures detection.

2. Methods

An automatic epileptic seizures detection system was built according to the scheme described in Figure 1. It is wort mentioning that the stages of signal acquisition and signal processing components are generally included into the software that captures the EEG signals, therefore, they were not addressed in this study. Feature Extraction describes the process used to calculate a representation of the EEG signals while Classification describes how to determine whether the signal is normal or abnormal through Machine Learning and Artificial Intelligence algorithms. In addition, if an abnormality is identified, it is possible classify the signal according to the type of Epilepsy suffered by the patient.

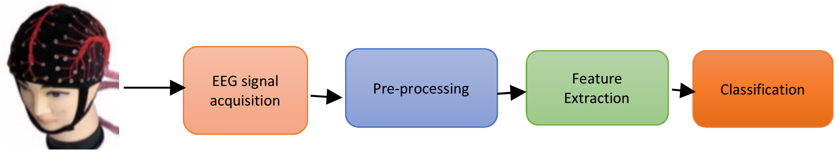

Figure 1. Scheme of the automatic epileptic seizures detection process.

2.1. Feature Extraction

The feature extraction process performs a series of mathematical operations on the EEG signals to obtain a set of descriptors which represent information in EEG signals. Some approaches propose an additional task in the feature extraction stage in order to validate the relevance of the obtained descriptors. This is known as Feature Selection (FS) [9] and is used to eliminate redundant information or noise from descriptors. Consequently, it is an optimization task, since it avoids performing operations for calculating irrelevant descriptors reducing the computational load and increasing the accuracy of the classification.

2.1.1. Experimental data

In this work, the CHB-MIT database recorded at Children's Hospital Boston was used [10]. The database contains recordings from 23 pediatric patients using the international 10-20 system of EEG electrode positions and nomenclature with a Sampling rate of 256 samples per second. Each EEG record contains information of 23 channels following the standard European Data Format (.edf) used to exchange and store multi-channel signals of biological and physical origin. In the database, each patient has several EEG records; some of them have epileptic seizures and other describe normal brain activity.

2.1.2. Feature Extractors

Feature extraction describes the process of applying a number of operators to obtain a set of descriptors. The descriptors represent the information contained in the EEG signal. According to the literature, the following descriptors are identified and calculated: Entropy [11], Maximum Amplitude, Minimum amplitude, Mean, Variance, Maximum Power and Mean Power [12]. The abovementioned features were computed for each EEG channel ending up in 161 descriptors, 23 EEG channels plus 138 computed features. Moreover, the Forward Selection, Optimize Selection and Backward Elimination algorithms were used for the Feature Selection process.

2.2. Classification

The classification process uses the features extracted in the previous phase to determine whether the signal is normal or abnormal. In this work, the algorithms Naive Bayes, Rule Induction, Decision Tree and KNN were evaluated.

3. Results

The obtained precision of the classification stage as a function of the FS process is described in Table 1.

Table 1. Accuracy of the Algorithms used in the Feature Selection process.

	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

The Forward Selection algorithm (FSA) produced the best results in the classification process with the subset of selected features.

The subsets of features generated by the FSA according to each classification algorithm tested are described in Table 2. The FSA reduced 161 features to: (i) 4 using Naive Bayes, (ii) 8 using Rule Induction, (iii) 4 using Decision Tree and (iv) 5 using KNN. However, the best accuracy was obtained using the subset of features generated using Decision Tree as classifier (Table 1).

The features selected were a6, a27, a94 and a121. According to the encoding scheme used to name the features, it was observed that the Mean Power of the first channel, Mean Power of the second channel, the Minimum Amplitude of the fourteenth channel and the Maximum Amplitude of the eighteenth channel are the features that determine whether a signal is normal or abnormal.

The obtained results became in a contribution for the automatic detection of epileptic seizure due to it describes a new set of features that can be extracted from an EEG signal for detecting some abnormalities in the cerebral activity of a person. In addition, our proposal was tested with data from a real database, i.e., data collected in a clinical setting, obtaining an accuracy of 80.77% using only 4 features whereas in the literature, others proposals achieved better results using much more than four features and using experimental data [12][13][14], i.e., data collected in non-clinical settings.

Table 2. Features selected using the FS Algorithm.

Feature	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

4. Discussion and Conclusions

This paper provided an analysis of the accuracy of three algorithms: Forward Selection, Optimize Selection and Backward Elimination, typically used in feature selection processes, in the selection of a set of relevant features to support the automatic epileptic seizures detection. The Forward Selection algorithm (FSA) produced the best results in the classification process, with an accuracy of 80.77%.

The results in this study represent a contribution to the process of feature extraction of EEG signals. The Feature selection process reduced the initial set of features extracted in a 97.51%. This shows that, although it is possible to obtain different data from an EEG, not all of them are relevant to support Classification.

Regarding the improvement of the accuracy in the Classification process, the Decision Tree algorithm showed better results compared to the other classification algorithms used to evaluate subsets of features generated by the Forward Selection algorithm (Naive Bayes, Rule Induction, and KNN). Therefore, it can also be concluded that the calculation of the Mean Power of the first channel, Power Media of the second channel, the Minimum Amplitude of the fourteenth channel and the Maximum Amplitude of the eighteenth channel in an EEG signal supports the automatic detection of epileptic seizures with an accuracy of 80.77%.

To the best of our knowledge, no single study has been conducted to provide an analysis of the computational performance and accuracy of relevant features selection in EEG signals, as the one presented in this study. The main limitation of this work is that, despite the proposed features have shown positive results to classify EEG signals as normal or abnormal, the results have not yet been evaluated to classify the abnormal signals according to the type of Epilepsy. This due to the EEG database used does not have this information.

The main contribution of this work to the medical field is the identification of the features that detect automatically an epileptic seizure in an EEG signal. This can decrease time of reading of an EEG signal and facilitate the diagnosis of the Epilepsy.

The accuracy analysis presented is relevant for the design of mechanisms to automatically identify a relevant features to support the automatic epileptic seizures detection, as well as for the proposal of new thechniques for automatic relevance analysis in EEG signals. As a further study, we propose to calculate and evaluate the relevance of new features in order to improve the accuracy of the results. In this context, it is very important to consider in the analysis new features, such as those extracted from electronic health records, in order to provide a more accurate diagnosis of Epilepsy. In addition, a data set combining EEG data and clinical information, as well as including information about the type of Epilepsy, have to be provided.

Acknowledgements

The work is funded by a grant from the Colombian Agency for Science, Technology and Innovation Colciencias - Call 647- 2015 Doctorados Nacionales.

References

[1] Y. Song, J. Crowcroft and J. Zhang, Automatic epileptic seizure detection in EEGs based on optimized sample entropy and extreme learning machine, Journal of Neuroscience Methods 210, (1993). 131-146.
[2] C. Stam , J. Pijin, P. Suffczynski and F. Lopez da Silva, Dynamics of the human alpha rhythm: evidence for non-linearity, Clin Neurophysiol 110 (1999), 1801-1813.
[3] M.E. Menshawy, A. Benharref and M. Serhani, automatic mobile-health based approach for EEG epileptic seizures detection, Expert Systems with Applications 42 (2015), 7157–7174.
[4] 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.
[5] WHO. Mental Health Atlas. Available http://www.who.int/mental_health/evidence/atlas/mental_health_atlas_2014(Last accessed: 8-feb-2016).
[6] J. Jin, J. Dauwels and S. Cash, Automated localization of the seizure focus using interictal intracranial EEG, Conf Proc IEEE Eng Med Biol Soc (2014), 4439-4442.
[7] G. Chen, Automatic EEG seizure detection using dual-tree complex wavelet-Fourier features, Expert Systems with Applications 41 (2014), 2391-2394.
[8] T. Sunil Kumar, V. Kanhangad and R. Bilas Pachori, Classification of seizure and seizure-free EEG signals using local binary patterns, Biomedical Signal Processing and Control 15 (2015), 33-40.
[9] H. Liu, H. Motoda and L. Yu, A selective sampling approach to active feature selection, Artificial Intelligence, 159 (2004), 49-74.
[10] N. Ahammad, T. Fathima and P. Joseph, Detection of Epileptic Seizure Event and Onset Using EEG, BioMed Research International, 2014 (2014), 1-7.
[11] Mathworks - Entropy, 03 March 2016. Available: http://www.mathworks.com/help/wavelet/ref/wentropy.html (Last accessed: 03-Mar-2016).
[12] M.A. Naderi and H. Mahdavi-Nasab, Analysis and classification of EEG signals using spectral analysis and recurrent neural networks, 17th Iranian Conference of Biomedical Engineering (ICBME) (2010), 1-4.
[13] G. Gopan, N. Sinha and D. Babu, Statistical Features based Epileptic Seizure EEG Statistical Features based Epileptic Seizure EEG, *Advances in Computing, Communications and Informatics (ICACCI)*, (2015), 1394-1398.
[14] A.M. Aldabbagh, Low computational complexity EEG epilepsy data classification algorithm for patients with intractable seizures, *Biomedical Engineering (ICoBE)* (2015), 1-4.