



HHS Public Access

Author manuscript

J Eng Res (Ponta Grossa). Author manuscript; available in PMC 2022 June 15.

Published in final edited form as:

J Eng Res (Ponta Grossa). 2022 ; 2(8): . doi:10.22533/at.ed.317282219056.

A MACHINE LEARNING-BASED APPROACH TO EPILEPTIC SEIZURE PREDICTION USING ELECTRO-ENCEPHALOGRAPHIC SIGNALS

Bruna Carolina Rebello,

Engenharia da Computação, Universidade Do Vale Do Itajaí, Santa Catarina, Brasil

Alejandro Rafael Garcia Ramirez,

Engenharia da Computação, Universidade Do Vale Do Itajaí, Santa Catarina, Brasil

Frances Heredia-Negron,

Center for Collaborative Research in Health Disparities - RCMI Program University of Puerto Rico Medical Sciences Campus, San Juan, Puerto Rico

Abiel Roche-Lima

Center for Collaborative Research in Health, Disparities - RCMI Program, University of Puerto Rico Medical Sciences, Campus, San Juan, Puerto Rico

Abstract

The brain is made up of billions of neurons, which control all actions performed by us. In epilepsy, the pattern order of brain signals is altered, causing epileptiform discharges in an individual's brain. Approximately 1% of the world population has epilepsy and, therefore, there is a need for studies that can help in the diagnosis and treatment of this disorder. The objective of this work is to develop a machine learning-based approach to predict epileptic seizures using non-invasive electroencephalography (EEG). Therefore, the classification of interictal and preictal states was performed using the CHB-MIT database. The algorithm was developed to predict epileptic seizures in multiple subjects using a patient-independent approach. The Discrete Wavelet Transform was used to perform the decomposition of the EEG signals in 5 levels and, as characteristics, the Spectral Power, the Mean and the Standard Deviation were studied, in order to analyze which one would present the best result and as a classifier, the Supported Vector Machine (SVM). The study achieved an accuracy of 92.30%, 84.60% and 76.92% for the Power, Standard Deviation and Mean characteristics, respectively.

Keywords

Epilepsy; electroencephalography; prediction of epileptic seizures

INTRODUCTION

Approximately 1% of the world's population has epilepsy and up to 5% of people may have at least one seizure during their lifetime (PEKER; SEN; DELEN, 2015). This disorder, when it occurs, ends up generating epileptiform discharges in the brain and the attack is usually

accompanied by episodes of seizure or even mental confusion, memory loss, depression and fainting.

The cerebral cortex is the main element in the generation of epileptic seizures. Elevated levels of cortical excitability play an important role in the initiation and spread of epileptic seizures (MEISEL; LODDENPKEMPER, 2019). This way, these temporary brain dysfunctions of a set of neurons can occur in one part of the brain or in several areas simultaneously.

There are several types of epilepsy and not all of them manifest in the same way, however, all of them can damage the quality of life of the individual who has the disorder. Given that the crisis can hardly be avoided, the prediction of epileptic seizures could help to alleviate some of the problems that accompany the crisis. There are two methods for capturing brain signals: the invasive method and the non-invasive method. The invasive method involves the insertion of electrodes directly over the brain and this implantation involves performing a surgery. The non-invasive method involves the superficial placement of electrodes directly on the person's scalp, without the need for surgery.

It is estimated that, in Brazil, there are 1.8 million patients with active epilepsy, and that at least 9 million people have had an epileptic seizure at some time in their lives (MARANHAO; GOMES; CARVALHO, 2011). The main manifestation of epilepsy is recurrent seizures. The epileptic seizures are caused due to the sudden development of neuronal synchronization in the cerebral cortex (TZALLAS et al., 2012).

It is this origin that distinguishes epilepsy from other cerebral crises such as partial or global ischemic, metabolic or psychogenic (LIMA, 2005). It is extremely important to understand about periods of crisis. The epileptic seizure has four periods. The ictal period, which is the exact moment of the epileptic seizure. The postictal period, which are clinical manifestations that follow on an electroencephalogram after the crisis. The interictal period, which corresponds to the normalization of the person's brain signals, and the post-ictal period, which is the moment immediately before the ictal period (PEKER; SEN; DELEN, 2015).

Studies involving the prediction of epileptic seizures can be resolved using two approaches. Patient independent and patient dependent. In the patient-independent model, the studies aim to design a classifier that can recognize seizures in many different patients. In this model, the dataset is used, and the objective is to learn a global predictive function that has the ability to make predictions in different subjects (FAUBERT, et al., 2018).

In the diagnosis of epilepsy, the Electroencephalogram (EEG) is often used. Most of the works found involve the detection and classification of EEG signals, aiming at the improvement and agility in obtaining diagnoses. Despite this area of study being successful, prediction has not yet been studied in depth (DAOUD; BAYOUMI, 2019).

In this work we developed an algorithm to predict epileptic seizures in multiple subjects using a patient-independent approach. The classification of interictal and preictal states was performed using the CHB-MIT database (USMAN; KHALID; ASLAM, 2020). The

Discrete Wavelet Transform was used to decompose the EEG signals into 5 spectral levels and, as characteristics, the Spectral Power, Mean and Standard Deviation were studied and the Supported Vector Machine (SVM) was used as a classifier.

FUNDAMENTALS

The electroencephalogram (EEG) is a graphic record of electrical brain activity, figure 1, which measures the variations in voltages related to this activity, through several electrodes located in different predefined positions on the head (LOTTE, 2008). The EEG is widely used for the detection and analysis of epileptic seizures (ACHARYA et al., 2012).

Figure 1 shows an example of an EEG signal with the periods of a person with epilepsy broken down. In the Figure, one can observe the signal in relation to amplitude and time, resulting from this recording.

EEG signals can be obtained by electrodes, amplifiers, Analog-Digital Converters and Filters. In the non-invasive approach, the electrodes acquire the signal from the scalp, while the amplifiers process the analog signal to increase the amplitude, so that the converter can digitize the signal more accurately. In turn, the Filter allows you to reinforce the signal-to-noise ratio as well as remove information that is not part of the signal under study (NICOLAS-ALONSO; GOMES-GIL, 2012).

In this work, a bandpass type FIR filter was used to pre-process the signals, removing unwanted noise.

FEATURE EXTRACTION

According to Lotte (2008), there are three types of relevant information that can be extracted from EEG signals:

1. **Spatial Information:** Features that describe where the relevant signal is coming from. Algorithms are used to select specific EEG channels or assign more weight to some channels and less weight to others;
2. **Spectral Information:** These characteristics describe the variation of the signal amplitude in certain frequency bands that are relevant to the system;
3. **Temporal Information:** Characteristics that describe how the relevant signal varies over time. In practice, it consists of analyzing the values of the EEG signals at different times.

In the literature on prediction, classification or detection of epileptic seizures, the most used method for extracting features is the Discrete Wavelet Transform (DWT), which provides spectral information about the signal (PEKER; SEM; DELEN, 2015). Other works use the Fourier Transform, such as FFT or DFT.

DWT is used in signal preprocessing to represent frequency characteristics through its coefficients. The DWT algorithm decomposes a given signal into approximation and detail

coefficients, to obtain a desired level of decomposition (FAUST et al., 2015). In Figure 2 it is possible to observe the decomposition of a signal in 3 levels.

The input signal $a(m)$ is initially decomposed generating an approximation coefficient and a detail coefficient. Afterwards, the detail coefficient generates other approximation and detail coefficients, respectively. In the example of Figure 2, 3 approximation coefficients and 3 detail coefficients are generated, referring to the level used in the decomposition.

CLASSIFICATION

The classification assigns a class to a set (vector) of features extracted from the signal (LOTTE et al., 2008). Classification algorithms aim to classify items or samples according to observed characteristics.

The Support Vector Machine (SVM) was developed from the theory of computational learning because of its precision and ability to deal with large numbers of predictors, being successfully used in biomedical applications (USMAN, S. M.; KHALID, S.; ASLAM, 2020).

SVM uses hyperplanes to perform the classification. The selected hyperplanes maximize the margins, that is, the distance from the closest training points (BENNET; CAMPBELL, 2000). It builds a classifier according to a set of patterns identified by it in the training examples, where the classification is known. The results of applying this technique are comparable and many times superior to those obtained by other learning algorithms, such as Artificial Neural Networks (ANNs).

SVM uses a regularization called Parameter C (Cost), which allows outliers and errors to be in the training dataset. It is even possible to create non-linear decision delimiters, with little increase in classifier complexity, using the “kernel trick” technique. This technique consists of implicitly mapping the data into another space, usually of higher dimensionality, using a kernel function.

Figure 3 shows a 3D model of the functioning of an SVM. It is possible to observe two distinct classes, the circle and the x, and between the classes there is a hyperplane called the “optimal hyperplane”. This apparent hyperplane in the dimensional space is what performs the classification of a new sample in the space of classes.

TRABALHOS RELACIONADOS

A bibliographic research was carried out through the Univali Integrated Library System (SIBIUN), which performs a search in the Univali collection, CAPES Portal, EBSCO, Biblioteca A, Saraiva, Vlex, Scielo Livros, Scielo Periodicals and Open Access Directories. The search strings “Epilepsy prediction” AND “Seizure Detectio” were used, yielding 29 results.

After reading the abstracts, the three most relevant studies were selected, mainly involving the prediction of epileptic seizures, Table 1.

Table 1 informs the databases used, the processing techniques, the classifiers, the characteristics studied (features), the accuracy of the classifiers and the prediction time, when informed.

METHODOLOGY

The public data source CHB-MIT Scalp EEG Database was used in this work. It consists of EEG recordings of 23 pediatric patients with intractable seizures collected at Children's Hospital Boston. EEG signals were obtained at 256 samples per second with 16bit resolution and recorded simultaneously on 21–28 channels. The seizure events and the corresponding start and end times were identified by clinical analysts. The complete dataset has an average of 665 recordings, most of which are one hour long, while others are three to four hours long.

The intention of this work was to predict epileptic seizures using the patient-independent approach. Thus, an analysis of the complete dataset was performed and some restrictions were defined to be applied in the selection of subjects for this work. Among these:

The recording time of the subject's EEG signals: only subjects who had up to one hour of recording were used. The selection was made because many subjects had recordings of up to 4 hours. Thus, a standardization in the recording time was sought.

Files with seizures: Only files containing 5 or more epileptic seizure events were selected.

Time of the pre-ictal period: The selection of files with epileptic seizures was subjected to an analysis of the pre-ictal period and only files with seizures where there were 10 minutes of pre-ictal period were selected.

Gender of the individual to be analyzed: As a final step, after selection based on the first three parameters, only female individuals were selected for analysis. Sex selection was defined as a parameter because of the neurological reactions that different sexes can present during the interictal and pre-ictal period. Table 2 shows the list of subjects chosen for the job (chb01, 03, 05, 08 and 14).

In this work, the MNE-Python library was used, which is an open source library that allows exploring, visualizing and analyzing human neurophysiological data, such as MEG, EEG, sEEG, ECoG and others. Includes modules for data input/output, signal preprocessing, visualization, source estimation, time frequency analysis, connectivity analysis, machine learning and statistics. In the following link you will find the archives repository of this research.

<https://drive.google.com/drive/folders/1HnM0s7LR9SxuzrXnrq1wx19lw24Iy7EG?usp=sharing>

FILTERING AND FEATURE EXTRACTION

EEG data were initially prepared to perform feature extraction. EEG sampling needs to conform to the Nyquist criterion, so appropriate sample rates need to be selected during data acquisition to ensure that aliasing does not affect the signal of interest.

The dataset signals contain a sampling frequency of 256Hz. Considering Nyquist's theorem, the maximum frequency that can be represented in the frequency domain is 128Hz. In this work, only the 1–40Hz band was considered, so a bandpass filter was designed and applied to the selected files. Thus it was possible to remove any high and low frequency artifacts (including the DC signal).

Then, the Discrete Wavelet Transform was applied to decompose the signal in the selected electrodes. The Wavelet Daubechies db4 was the Wavelet mother applied to decompose the signals into 5 levels. As a resource for extraction, the Average of coefficient values, Average power of the coefficients and the Standard deviation of the coefficients were used.

Subsequently, the extraction of characteristics was performed using 52 samples, 26 samples referring to the interictal period (normal period) and 26 samples referring to the pre-ictal period (before the crisis).

CLASSIFIER

In the development of the algorithm based on the SVM, the Scikit-Learn library was used. In the training of the classifier, 75% of the samples were used, while for the validation, 25% were used. The proportion of classes was stratified to prevent the dataset from becoming unbalanced in relation to the classes. In addition, the cross-validation method was also used to ensure the generalization of the classification model.

RESULTS ANALYSIS

The classification performance was evaluated in the test set in terms of Accuracy (Acc), Sensitivity (S) and Specificity (SP), Equations 1, 2 and 3, as in (KILL; CIARELLI; CÔCO, 2020).

$$Acc = \frac{TP + TN}{TP + FP + TN + FN} 100\% \quad \text{Equation 1}$$

$$S = \frac{TP}{TP + FN} 100\% \quad \text{Equation 2}$$

$$Sp = \frac{TN}{FP + TN} 100\% \quad \text{Equation 3}$$

Where TP (True Positive) and TN (True Negative) correspond to samples correctly classified into positive and negative classes. FP (False Positive) and FN (False Negative) are the

samples incorrectly classified by the model into positive and negative classes. It is possible to better observe these parameters by analyzing the confusion matrix summarized in Table 3, for each of the three characteristics used. The total shown in Table 3 refers to the number of samples that were used in the test stage.

As can be seen in Table 4, the characteristic that achieved the best result was the Spectral Power of each subband, with an accuracy of 92.30%, sensitivity of 83.33% and specificity of 100%. The results obtained are compatible with the results obtained in related works that used the SVM, being 92.70 and 97.29, respectively. In particular, the work by Jade Barbosa Kill et al. (2020) reports a sensitivity of 96.25% and a specificity of 98.33%.

CONCLUSIONS

In this work, a methodology for discriminating the interictal and pre-ictal period of an epileptic seizure was presented, making it possible to predict it. The database used in this work was the CHB-MIT, which was used by other authors mentioned in this work.

The results obtained were satisfactory. The accuracy obtained with the Spectral Power feature proved to be the most promising in the epileptic seizure prediction step, with 92.30%. The Standard Deviation feature also presented a good accuracy, with 84.60%. Despite the Average characteristic having the worst result, which was 76.92%, it is important to emphasize that it is still a valid characteristic and must be considered when performing new tests.

The highest accuracy obtained in the related works was 99.66%, however, spatial characteristics were used instead of spectral ones. The works that used spectral characteristics had an accuracy between 92.70% and 97.29%.

The diversity and characteristics of the patients and the presence of physiological artifacts present in the data may interfere with the results obtained, however, the work presented favorable results and must be continued and improved for a future application in prediction of epileptic seizures.

For future work, we suggest the use of a larger data set with more diversity among patients. The evaluation of results based on specific electrodes and optimization techniques could be useful to develop a prototype with a reduced set of electrodes, focusing on the development of a product. It would also be interesting to compare the analysis of the methods implemented in this work with other types of classifiers, in order to analyze which classifier has better accuracy. Considering other databases would be important, mainly, in the independent approach to the patient.

Despite the satisfactory results obtained in this work, the prediction of epileptic seizures has a wide range of opportunities challenging future works. This is a new area with potential for growth and study.

THANKS

This research was supported by Fundação de Amparo à Pesquisa e Inovação do Estado de Santa Catarina (FAPESC) grant TO2020TR729. and National Council for Scientific and Technological Development (CNPq) grant 315338/2018-0. It was also supported by RCMI grant U54 MD007600 (National Institute on Minority Health and Health Disparities) from the National Institutes of Health.

REFERENCES

- ACHARYA UR et al. Automated diagnosis of epileptic {EEG} using entropies. *Biomedical Signal Processing and Control*, v. 7, n. 4, p. 401–408, 2012. ISSN 1746–8094.
- BENNETT KP; CAMPBELL C. Support Vector Machines: Hype or Hallelujah? *SIGKDD Explorations*. Volume 2, Issue 2. (2000)
- DYER S, MARTIN J. AND ZULAUF J. (1995) “Motion Capture White Paper”, http://reality.sgi.com/employees/jam_sb/mocap/MoCapWP_v2.0.html, December.
- DAOUD H.; BAYOUMI MA Efficient Epileptic Seizure Prediction Based on Deep Learning. *IEEE Transactions on Biomedical Circuits and Systems*, vol. 13, no. 5, pp. 804–813, 10 2019. [PubMed: 31331897]
- FAUST O; ACHARYA UR; ADELI H. AMIR A. Wavelet-based EEG Processing for Computer-aided Seizure Detection and Epilepsy Diagnosis. January 2015. DOI: 10.1016/j.seizure.2015.01.012
- KILL JB; CIARELLI PM CÔCO KF. de Souza ML. Wavelet Analysis Applied on EEG Signals for Identification of Preictal States in Epileptic Patients. *Brazilian Applied Science Review*. Vol 4, No 3 (2020).
- LIMA José M. Lopes. Epilepsia - A abordagem clínica. *Revista Portuguesa de Medicina Geral e Familiar*, [S.l.], v. 21, n. 3, p. 2918, maio 2005. ISSN 2182–5181. Acesso em: 17 fev. 2021. DOI: 10.32385/rpmgf.v21i3.10141
- LOTTE F. Study of Electroencephalographic Signal Processing and Classification Techniques towards the use of BrainComputer Interfaces in Virtual Reality Applications. *Human-Computer Interaction [cs.HC]*. INSA de Rennes, 2008. English. fftel-.
- MARANHAO MVM; GOMES EA; CARVALHO PE Epilepsia e anestesia. *Revista Brasileira de Anestesiologia*, Scielo, v. 61, p. 242–254, 04 2011.
- MEISEL C.; LODDENPKEMPER T. 2019. Seizure prediction and intervention. *Neuropharmacology* 2020 Aug 1;172:107898. doi: 10.1016/j.neuropharm.2019.107898. Epub 2019 Dec 5. [PubMed: 31839204]
- NICOLAS-ALONSO LF; GOMEZ-GIL J. Brain Computer Interfaces, a Review. Department of Signal Theory, Communications and Telematics Engineering, University of Valladolid, Valladolid 47011, Spain, 2012.
- PEKER M.; SEN B.; DENLEN D. A Novel Method for Automated Diagnosis of Epilepsy Using Complex-Valued Classifiers. January 2015 *IEEE Journal of Biomedical and Health Informatics* 20(1). DOI: 10.1109/JBHI.2014.2387795.
- SMITH SW *The Scientist and Engineer’s Guide to Digital Signal Processing*. Second Edition. California Technical Pub, San Diego, 1997.
- TZALLAS AT et al. Automated epileptic seizure detection methods: A review study. In: *Epilepsy: Histological, Electroencephalographic and Psychological Aspects*. [S.l.]: InTech, 2012. p. 75–97.
- USMAN SM; KHALID S.; ASLAM MH Epileptic Seizures Prediction Using Deep Learning Techniques. February 2020. *IEEE Access* pp (99):1–1. DOI: 10.1109/ACCESS.2020.2976866
- ROY Y, BANVILLE H, ALBUQUERQUE I, GRAMFORT A, FALK TH; FAUBERT J, “Deep learningbased electroencephalography analysis: A systematic review,” p. 37, 8 2019. [Online]. Disponível em: 10.1088/174-12552/ab260c.

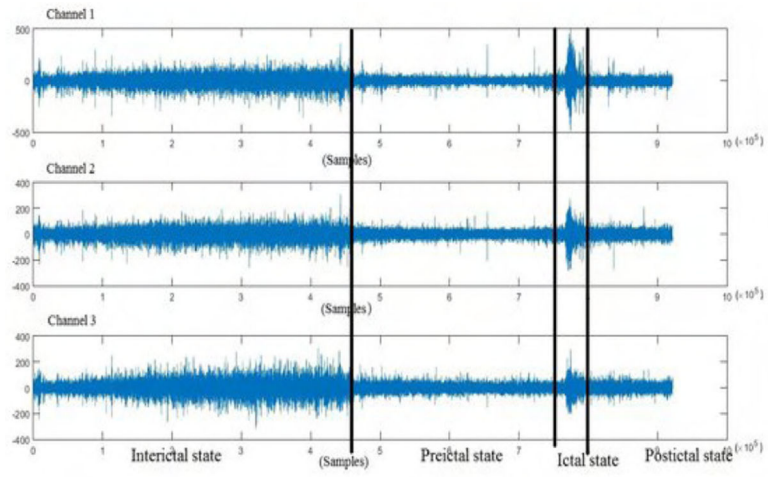


Figure 1. Example of EEG signal with periods of epilepsy broken down. Usman, Khalid, Aslan (2020).

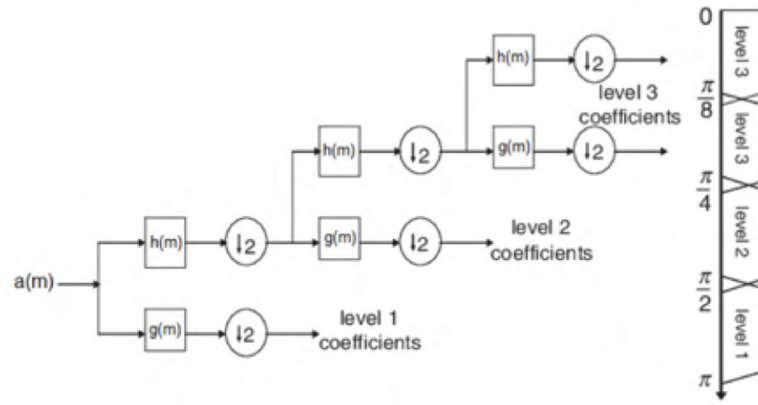


Figure 2. Signal decomposition using 3 levels. Kill et al (2020).

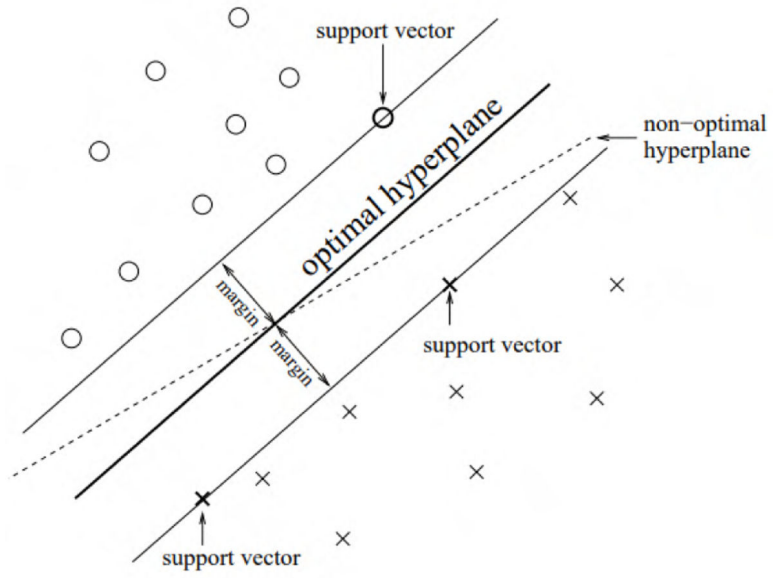


Figure 3.
Example of using SVM. Lotte (2008).

Table 1.

Related works.

| | Usman, Khalid e Aslam (2020) | Jade Barbosa Kill et al. (2020) | Daoud. Bayoumi (2019) |
|----------------------------|-------------------------------------|--|------------------------------|
| Data base | CHB-MIT | CHB-MIT | CHB-MIT |
| Processing | STFT | DWT | N/DA |
| Characteristic information | Spectral | Temporal and spectral | Spacial |
| Classifier | SVM | KNN e SVM | MLP, CNN, DCNN e DCAE. |
| Accuracy | 92,7% | 90,21% e 97,29% | 83,63%, 94,10% e 99,66%. |
| Prediction time | 21 minutes | N/DA | N/DA |

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 2.

Subjects chosen for the analysis.

| Subjects | chb01 | chb03 | chb05 | chb08 | chb14 |
|----------------|--------|--------|--------|--------|--------|
| Pre-ictal | 6 | 5 | 5 | 5 | 5 |
| Interictals | 6 | 5 | 5 | 5 | 5 |
| Recording time | 1 hour | 1 hour | 1 hour | 1 hour | 1 hour |
| Gender | F | F | F | F | F |

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 3.

Confusion matrix summarized.

| Features | TP | TN | FP | FN | Total |
|--------------------|----|----|----|----|-------|
| Standard deviation | 5 | 6 | 2 | 0 | 13 |
| Potency | 5 | 7 | 1 | 0 | 13 |
| Average | 5 | 2 | 1 | 5 | 13 |

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 4.

Analysis of results.

| Features | Accuracy | Sensibility | Specificity |
|---------------------------|-----------------|--------------------|--------------------|
| Standard deviation | 84.6% | 83.33% | 75% |
| Potency | 92.30% | 71.42% | 100% |
| Average | 76.92% | 50% | 66.66% |

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript