AN ALIVE ELECTROENCEPHALOGRAM ANALYSIS SYSTEM TO ASSIST THE DIAGNOSIS OF EPILEPSY

Malik Anas Ahmad [1, 2], *Waqas Majeed* [1], *Nadeem Ahmad Khan* [1]

[1]SIVPLab, EE, SBASSE, Lahore University of Management Sciences, Lahore, Pakistan
[2]RISELab, R&AI, SMME, National University of Sciences & Technology, Islamabad, Pakistan

ABSTRACT

Computer assisted electroencephalograph analysis tools are trained to classify the data based upon the "ground truth" provided by the clinicians. After development and delivery of these systems there is no simple mechanism for these clinicians to improve the system's classification while encountering any false classification by the system. So the improvement process of the system's classification after initial training (during development) can be termed as *'dead'*. We consider neurologist as the best available benchmark for system's learning. In this article, we propose an '*alive'* system, capable of improving its performance by taking clinician's feedback into consideration. The system is based on taking DWT transform which has been shown to be very effective for EEG signal analysis. PCA is applied on the statistical features which are extracted from DWT coefficients before classification by an SVM classifier. After corrective marking of few epochs the initial average accuracy of 94.8% raised to 95.12.

***Index Terms*—**Electroencephalography (EEG), Epilepsy, Computer Assisted Analysis, Machine Learning, Biomedical Signal Processing,

1. INTRODUCTION

Epilepsy is a neurological disease which is characterized by undesirable yield of excessive neural activity in the brain. According to World Health Organisation, one out of 100 people suffers from this disease [1]. This disease is more common among the habitant of third world country. Epilepsy related abnormal brain activity's detection and localization is very important for diagnoses and treatment of an epileptic disorder. Electroencephalography is used to record electrical activity along the surface of the brain. EEG signal is a representation of voltage fluctuations which are caused by the ionic current flow in the neurons. Epileptic disorders cause generation of unique patterns in the EEG. This is one of the main reasons behind the wide usage of EEG for the detection and localization of epileptic seizure and its location.

A diagnostic EEG recording duration can vary from few minutes to couple of days. It causes the generation of an immense amount of data to be manually inspected by the neurologist which could prove to be a daunting task.

Advancement in signal processing and machine learning techniques is making it possible to analyse EEG data automatically to detect epochs with epileptic patterns. A system based on these techniques can aid a neurologist by highlighting the epileptic parts of the EEG. Of course, the task of diagnosis should be left to the neurologist. However, the task of the neurologist become efficient as it reduces the data which is required to be analysed. Along with classification these analysis software can also provide simultaneous visualization of multiple channels which helps the clinician in differentiating between generalized and focal epilepsy.

It is these neurologists' markings of EEG data which are benchmarked to train these analysis software systems. There is no simple mechanism available in the currently available analysis systems to improve their classification after initial training by mentioning the wrong markings. In order to comprehend with this complex signal processing methods, clinician who lacks the prior understanding of signal processing algorithms hires technicians and relies on them. This makes that analysing procedure using that software prone to misinterpretation and over-interpretation because then the marking rely on the technician's expertise [2] and not on the clinician.

Upcoming section is about a brief discussion on some of the existing work in the field of computer aided EEG analysis of Epilepsy.

2. LITERATURE REVIEW

In a computer assisted EEG analysis system, usually the EEG signal is divided into multiple small chunks. Later signal processing and machine learning techniques are applied on these small chunks to classify them as epileptic or non-epileptic. These chunks of the signal are known as epochs.

It is well known that an epileptic seizure brings changes in certain frequency bands. That is why EEG signal's spectral content is commonly used while diagnosing an epileptic disorder [1]. These are identified as δ (0.4 – 4 Hz), θ (4 – 8 Hz), α (8 - 12 Hz) and β (12 – 30 Hz).

Noachtar et al. mention almost ten types of epileptic patterns. However, most of the existing work only focuses on one of the epileptic pattern, i.e. 3-Hz spike & wave which is a trademark for absence seizure. Other types of the patterns are rarely addressed [3].

Usually the process of epileptic epoch detection is divided in three major stages 1) feature extraction, 2) feature pre-processing and 3) classification.

The Size of an epoch usually depends on the EEG signal's sampling frequency and signal processing techniques which are applied to extract features. Usage of different epoch size has been cited in the previous work. Epoch size as low as 0.3 sec [4] and as high as 23.6 sec [5] has been cited, but the most cited is 1 sec. Seng et al. compared the performance of different epoch size, their work resulted that among epoch size of 23.6, 11.5, 5.8, and 1 sec, 1 sec epoch size works best in terms of accuracy [6].

EEG signals are non-stationary signals [7]. For feature extraction different signal transforms for analysing non-stationary signals are applied to extract out the frequency related features.

The most commonly used signal transformation is Discrete Wavelet Transform (DWT) [8] but there are some exceptions where authors like Seng et al. did not apply any transform [6]. Empirical Mode Decomposition (EMD) method is used by Kaleem et al. [9] and Alam et al. [10] for feature extraction. Other than that multilevel Fourier transform (FT) is also used as in [5]. Abdullah et al. used the combination of both DWT and FT for extracting features from EEG data [11]. Guo et al. used the orthogonal matching pursuit method for extracting features [12].

Usually these transforms are followed by some mathematical operations which generate the statistical features of that transformed signal. Petersen et al. applied log-Sum Energy [8] over his DWT's detailed coefficients as it was suggested by Shoeb et al. [13]. Alam et al. took the variance, skewness and kurtosis of the EM decomposed signal as features. Murugavel et al. introduced a novel feature named as Combined Seizure Index which was calculated from the wavelet packets' coefficients [14]. Ocbagabir et al. used the energy, entropy and standard deviation of the filtered data as features [5]. But the most widely used features are energy, variance, mean, standard deviation and/or their small variants.

Lots of the proposed techniques do not resort to the extracted statistical feature. Feature reduction or feature extraction techniques are applied on these features, so that redundant and noisy data can be removed in order to facilitate the classifier. The application of Principal Component Analysis (PCA) by Luo et al. [15], Linear Discriminant Analysis by Subasi et al. [16] and Fast independent component analysis by Chang et al. [4] is few of the many examples of feature reduction application. Choi et al. applied feature selection using the sequential floating forward selection algorithm [7].

At the end these methods apply some classification method. There is whole versatile range of classifier used for this purpose. Alam et al. used Artificial Neural Network on their features, where as Subasi et al. used the most commonly used Support Vector Machine (SVM). Abdulah et al. used Hidden Markov Model. Quadratic Discriminant Function was found to best working by Choi et al. on his selected features.

Data fusion from multiple channels is not very common in most of these procedures. The independence of different channels especially in the case of localized epileptic disorder is ignored most of the time. Instead, signal data from all of the channels are processed in series, so as they are from one large EEG signal source, instead of multiple independent parallel signals.

Majority of the commercially available Neurophysiological Data Analysis software tools are quite generalized considering an epileptic disorder. These tools are a lot user dependant and they are not focused on any specific neurological disorder. Though these tools allow the neurologist to interactively apply multiple signal processing techniques on the EEG data but still the neurologist who lacks a proficient background in signal processing does not feel comfortable using them. None of these tools are intelligent as they don't learn or improve themselves as per neurologist's marking. Every time the neurologist went through a time wasting fatigue by monitoring lots of useless epochs of EEG data. A huge majority of these software tools are also hardware dependent. They usually come alongside the EEG equipment [1].

3. MATERIAL AND METHOD

It is neurologists' marking and labelling of the data which is benchmarked to develop an automated computer assisted analysis system. But after initial development these systems have no simple mechanism for these neurologists to improve system's classification while encountering a false classification by the system. So we have proposed a method by which system's classification can be improved by the user in an uncomplicated way.

In this proposed system we are processing each channel for each epileptic pattern exclusive to each other. This exclusive processing of each channel not only helps the user in diagnosing localized epilepsy but it also eases up the classifier's job. First we will explain feature extraction method then we will explain the feature standardization and their reduction, in the last we will explain the classification and retraining mechanism.

3.1. Classification Method

The time series signal from each channel was divided in small non overlapping epochs. At first DWT is applied on these non overlapping epochs. Then statistical features from the relevant detail coefficients are selected as the features. Then we applied data reduction using PCA. These reduced features are then fed into a classifier to classify that epoch as epileptic or not. Following are the details of each step.

3.1.1. Feature Extraction

The first important part of the feature extraction is epoch selection. In this system epoch size of 1 sec was selected as it yielded the most accurate results which re-established the work by Seng et al. [6]. Then DWT was applied on each epoch with Daubechies-4 (db4) as mother wavelet. The detailed coefficients levels of the DWT are determined with respect to sampling frequency.

The detailed levels of our interest were adjusted according to the sampling frequency such as that we may get if not exact than at least the closest separate δ (0.4 – 4 Hz), θ (4 – 8 Hz), α (8 - 12 Hz) and β (12 – 30 Hz) component of the signal. We discarded all the detail coefficient levels which were beyond the 0.4Hz to 30Hz range.

Then we took the statistical features of these selected detail coefficients by calculating mean, standard deviation and power of each epoch's selected DWT coefficients. These features are inspired form Subasi et al. work [16].

These statistical features were then standardized. During training time z-score standardization was applied on them. During testing time we used the blind data to test our system's performance. We assumes that the data trend do not change much, so we took the mean and standard deviation values from the training time and used them to standardized the features of the test data. We normalized them by subtracting and dividing it with mean and standard deviation respectively.

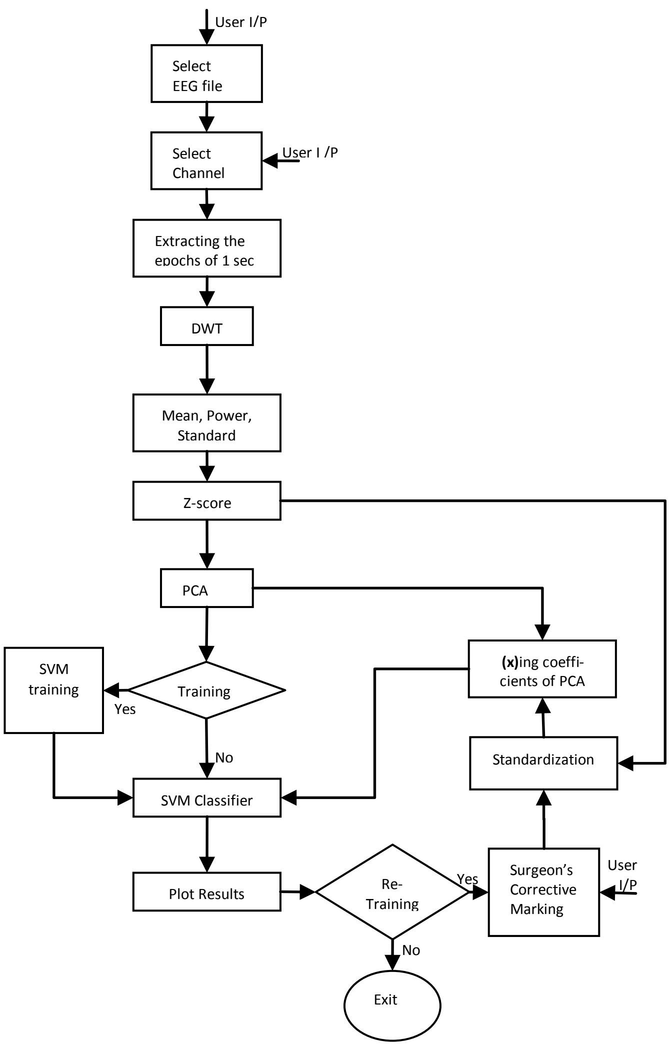

Fig. 1. Workflow

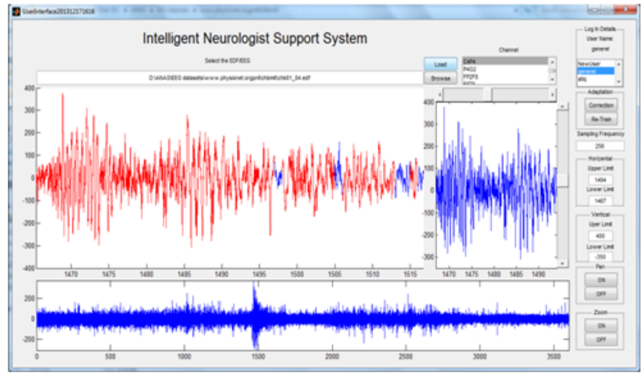

Fig. 2. GUI

3.1.2. Feature Reduction

During training phase PCA was applied on these features in order to reduce the redundant and/or noisy data. We kept the components which projected the approx 95% of the total variance. We reduced the 21 features into 9.

Again assuming that the data trend will not change much we used the coefficients matrix of the PCA output during the training phase and multiplied it with the standardized statistical features of the blind test data.

3.1.3. Feature Classification

These reduced features were then fed to a linear SVM classifier. All of these three processing phases were exclusive for each channel and each epileptic pattern. So like previous steps the classifiers were also trained and tested exclusively for each channel.

Our system requires individual labelling of channels. There is a separate classifier for each channel and for each epileptic pattern type. So it makes total of number of classifier as the product of number of channels by ten where ten is the number of epileptic patterns described by Noachtar et al. [3].

3.2. Retraining/User adaptation mechanism

A very important and novel part of our system is user adaptation mechanism. By adapting the user which is supposed to be a neurologist the system will try to improve its classification. It has been cited that some time even the expert neurologist have some disagreement over a certain observation of an EEG data. In this case over fitting by the classifier could constitute a major challenge, In order to keep the classifier improving its performance with the encounter of more and more examples we have introduced a user adaptive mechanism in our system.

Our system allows the user to interactively select epochs of his choice by simply clicking the "*correction*" button. These details will be saved in a log in the background and they will be used to retrain the classifier to improve its classification rate and adapt itself according to the user with the passage of time. When the user is going to select the retraining option in our system then classifiers will re-train themselves on the previous and the newly logged training examples.

4. EXPERIMENTATION

In this section we will discuss the results in detail after describing the testing condition.

4.1. Data set

The Datasets available to us for testing and validation were about generalised absence seizure which is identified by the 3Hz spike and wave epileptic pattern in almost each channel. We have classification results available only for one type of epilepsy which is absence seizure.

Scalp EEG database was used to test our technique. Children Hospital Boston provided this data which is freely available at physionet website without any charges [17] [13]. It contains 916 hours of 23 channel 256Hz sampled EEG recordings from 24 subjects. 129 out of 664 EEG recordings files contained one or more seizures.

4.2. Features

For CHB-MIT database we had to be train 220 classifiers. The 23^{rd} channel was same as 15^{th}. The frequency range of 0.3 to 30 Hz is important for us so detailed coefficients of level 3,4,5,6 and 7 were of our importance in 256 Hz sampled signal. The rest of the detailed coefficients were discarded.

4.3. Classification

We used the support vector machine classifier package available in Matlab Bioinformatics toolbox. Here we found *"Linear"* SVM kernel with 50 as box constraint to be best performing parameters in terms of classification rate. We used 10-fold validation method to validate the classification rate on blind data for each channel. 8736 epochs were used to validate our approach. They were randomly taken from whole of the dataset.

4.3.1 Results with Initial Training:

The average classifier performance on ten random distribution of the data set is reported. Due to unavailability of the data currently we have only classification rates for generalized absence seizure. On the initial training of the classifier average specificity was 95.7% and average sensitivity was 91.7%. Average accuracy stood at 94.8%. It can be noted that after the initial training our specificity is better than the Shoeb et al. and Nasehi et al. [13] [18].

We then tested the system on training the system on separate channel. The processing of the each channel exclusive to each other improved over average accuracy from approximately 91 % to approximately 95%. So there is a significant improvement of 4% by this change.

4.3.2 Results on Re-training

Table 1 is compiled to show the average initial classification and retrained classification results of our system for each channel. Following approach was adopted in re-training: Every user will have exclusive classifiers trained for him and his marking will not affect other user's classifier. For re-training in case of more than one example with same attributes but different labels: The classifier is trained to the most popular one. The user's marking will increase the examples of his choice thus making that classifier adapt itself to the user's choice in a simple way. In this system we have shown that after correction of few epochs there is visible improvement in the system's classification. The average accuracy of the system rose from 94.8% to 95.12%.

Channel	Accuracy after initial training (%)	No of epochs marked by the user	Accuracy after retraining (%)
'FP1F7'	94.2538	482	95.2082
'F7T7'	95.0086	482	95.7058
'T7P7'	94.2500	482	95.0423
'P7O1'	96.6401	269	96.7644
'FP1F3'	95.3026	482	96.0180
'F3C3'	92.8084	482	92.8892
'C3P3'	94.0857	482	94.1071
'P3O1'	93.6903	482	94.5970
'FP2F4'	95.3740	482	95.9311
'F4C4'	93.0260	482	93.2652
'C4P4'	97.9583	269	98.0277
'P4O2'	94.8940	269	94.9677
'FP2F8'	94.2318	482	94.5492
'F8T8'	96.3653	269	96.4675
'T8P8'	96.7284	482	97.6134
'P8O2'	97.3512	482	94.5511
'FZCZ'	89.1693	482	89.2381
'CZPZ'	92.8969	482	93.5356
'P7T7'	94.4767	482	95.1011
'T7FT9'	93.2681	482	93.4411
'FT9FT10'	96.5399	482	97.3534
'FT10T8'	97.8812	482	98.2949

Table 1. First column shows the channel label, the second column shows the initial training accuracy, and third one shows the marked correction by a neurologist and the last one shows the final accuracy.

5. DISCUSSION & FUTURE WORK

Epilepsy is a common neurological disease. Computer assisted EEG analysis for epileptic diagnosis significantly helps a neurologist. That is why usability, accuracy, robustness and being informative are very essential features of these systems. Addition of our suggested technique in the existing work will improve the robustness and the classification rate. Our result is tested on a versatile data set and its high average accuracy for different type of datasets clearly shows its robustness.

In future we are planning to make this a web based application in which a neurologist can login and consult each other's reviews about a particular subject. This will make our system experience a whole versatile of examples and learn from all of them.

We would also be investigating how much over fitting is an issue in the reported performances which are now even touching 100% based on some claims. There is a need of method/criteria which could limit these algorithms improving their detection on a limited number of available examples.

Furthermore, a neurologist should also be able to suggest correction while observing a wrong marking by the computer aided system in a way that the classifier can learn from those corrective marking. This can lead to a personalized Neurologist support system that is as per user's desire. This system is made keeping in mind that we have to facilitate the neurologist by supplementing him in the analysis of the EEG. We do not want to enforce the classification of the EEG data on a user.

REFERENCES

[1] M. A. Ahmad, W. Majeed and N. A. Khan, "Advancements in Computer Aided Methods for EEG-based Epileptic Detection," in *Proceedings of the 7th International Conference on Bio-inspired Systems and Signal Processing (BIOSIGNALS)*, Eseo, Angers, Lorie Valley, France, 2014.

[2] S. R. Benbadis and W. O. Tatum, "Overinterpretation of EEGs and misdiagnosis of epilepsy," *Neurophysiol,* vol. 20, pp. 42-44, 2003.

[3] S. Noachtar and J. Rémi, "The role of EEG in epilepsy: A critical review," *Epilepsy & Behavior,* pp. 22-33, 2009.

[4] H. Y. Chang, S. C. Yang, S. H. Lan and P. C. Chung, "Epileptic seizure detection in grouped multi-channel EEG signal using ICA and wavelet transform," in *International Symposium on Circuits and Systems (ISCAS)*, Paris, 2010.

[5] H. T. Ocbagabir, K. A. I. Aboalayon and M. Faezipour, "Efficient EEG analysis for seizure monitoring in epileptic patients," in *Long Island Systems, Applications and Technology Conference (LISAT)*, Farmingdale, New York, 2013.

[6] C. H. Seng, R. Demirli, L. Khuon and D. Bolger, "Seizure Detection in EEG Signals Using Support Vector Machines," in *Northeast Bioengineering Conference (NEBEC)*, Philadelphia, 2013.

[7] K. S. Choi, Y. Zeng and J. Qin, "Using sequential floating forward selection algorithm to detect epileptic seizure in EEG signals," in *International Conference on Signal Processing (ICSP)*, Beijing, 2012.

[8] E. B. Petersen, J. D. Henriksen, A. Mazzarettoa, T. W. Kjær, C. E. Thomsen and H. B. D. Sorensen, "Generic Single-Channel Detection of Absence Seizures," in *Annual International Conference of the EMBS*, Boston, 2011.

[9] M. Kaleem, A. Guergachi and S. Krishnan, "EEG Seizure Detection and Epilepsy Diagnosis using a Novel Variation of Empirical Mode Decomposition," in *International Conference of the EMBS*, Osaka, 2013.

[10] S. M. S. Alam and M. I. H. Bhuiyan, "Detection of Seizure and Epilepsy Using Higher Order Statistics in the EMD Domain," *IEEE Journal of Biomedical and Health Informatics,* vol. 17, no. 2, pp. 312-318, March 2013.

[11] A. A. Abdullah, S. A. R. and A. Ibrahim, "Development of EEG-based Epileptic Detection using Artificial Neural Network," in *International Conference on Biomedical Engineering (ICoBE)*, Penang, 2012.

[12] P. Guo, J. Wang, X. Z. Gao and J. M. A. Tanskanen, "Epileptic EEG Signal Classification with Marching Pursuit based on Harmony Search Method," in *International Conference on Systems, Man, and Cybernetics*, Seoul, 2012.

[13] A. Shoeb, "CHB-MIT Scalp EEG Database," 2000. [Online]. Available: http://physionet.org/pn6/chbmit/. [Accessed 11 November 2013].

[14] A. S. Muthanantha Murugavel, S. Ramakrishnan, U. Maheswad and B. S. Sabetha, "Combined Seizure Index with Adaptive Multi-Class SVM for epileptic EEG classification," in *International Conference on merging Trends in VLSI, Embedded System, Nano Electronics and Telecommunication System (ICEVENT)*, Tiruvannamalai, 2013.

[15] K. Luo and D. Luo, "An EEG Feature-based Diagnosis Model for Epilepsy," in *International Conforence on Computer Application and System Modeling (ICCASM)*, Taiyuan, 2010.

[16] A. Subasi and M. I. Gursoy, "EEG signal classification using PCA, ICA, LDA and support vector machines," *Expert Systems with Applications,* vol. 37, no. 12, p. 8659–8666, December 2010.

[17] A. L. Goldberger, L. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C. K. Peng and H. E. Stanley, "PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals," vol. 23, 2000, pp. 215-220.

[18] S. Nasehi and H. Pourghassem, "Epileptic Seizure Onset Detection Algorithm Using Dynamic Cascade Feed-Forward Neural Networks," in *International Conference on Intelligent Computation and Bio-Medical Instrumentation*, 2011.

[19] M. A. Ahmad, N. A. Khan and W. Majeed, "Computer Assisted Analysis System of Electroencephalogram for Diagnosing Epilepsy," in *proceedings of 22nd international conference on pattern recognition*, stockholm, 2014.