



Prediction of Epileptic Seizure using Support Vector Machine and Genetic Algorithm

* Abisoye, O. A¹, Abisoye, B.O², Ekundayo Ayobami.³, & Ogunwede Emmanuel.⁴

¹ Computer Science Department, SICT, Federal University of Technology, PMB 65 Minna Niger State, Nigeria

² Computer Engineering Department, SEET, Federal University of Technology, PMB 65 Minna Niger State, Nigeria

³ Computer Science Department, SICT, Federal University of Technology, PMB 65 Minna Niger State, Nigeria

⁴ Computer Science Department, SICT, Federal University of Technology, PMB 65 Minna Niger State, Nigeria

*Corresponding author email: o.abisoye@futminna.edu.ng, +23460546074

ABSTRACT

Epilepsy is a condition defined by the occurrence of epileptic seizures. An epileptic seizure is a brief episode of symptoms caused by abnormal electrical activities in the brain. A common way to treat epileptic seizure is the use of medication. When medication fails, surgery is usually the proposed but surgeries have been found to fail in numerous cases leaving victims with no option than to manage their condition. This scenario, prompt the prediction of epileptic seizures earlier before its invasion so that appropriate precautions can be observed. This research proposes machine learning algorithm; support vector machine and genetic algorithm for the prediction of epileptic seizures. Genetic algorithm was adapted for feature selection while support vector machine was used in classifying EEG signals as seizure or non-seizure signals. The developed model generated accuracy 97.73%, sensitivity 97% and specificity 97%.

Keywords: classification, epileptic seizure, genetic algorithm, prediction, support vector machine.

1 INTRODUCTION

A study defined epilepsy as a diverse family of disorders having in common an abnormally increased predisposition to seizures (Fisher, et al., 2005). Generally epilepsy is first treated with medications but if the medications fail to treat the condition, doctors usually propose surgery or other type of treatment like therapies such as valgus nerve simulation and ketogenic diet. In determining the type of seizure a patient suffers from, a neurological exam to test behaviour and mental function is usually carried out (Betts & Boden, 1992).

Computational methods are usually employed in the detection of epileptic seizures. These techniques have been derived from both non-linear and linear analysis, morphological analysis, model based analysis and recently, there is a notion of selecting and combining the most robust features from various techniques and using them to reveal more striking features from the techniques for revealing various signal characteristics and making more reliable

assumptions. After the feature extraction, intelligent classifiers are then used to distinguish epileptic state from normal states. Some example of such classifiers is Bayesian networks, decision tree, support vector machine, and artificial neural network (Hosseini, Mohammed-Reza, & Mohammed-Bagher, 2013).

A study proposed a machine learning approach to patient-specific classifiers for detecting the onset of an epileptic seizure (Shoeb & Guttag, 2010). A data set of continuous scalp EEG sampled at 256khz which was recorded over 916 hours of pediatric patients. It was recorded that 173 events occurred during the recordings which were considered to be clinical seizures by medical experts. The data set was then divided into two classes; records containing a seizure and records without a seizure. A high performance machine learning algorithm (a support vector machine) was designed and when trained on 2 or more seizures per patients and tested on 916 hours long recording of



continuous EEG from 24 patients the algorithm was found to have had an accuracy of 96% (Hosseini, Akbarzadeh-T, & Naghibi-Sistani, 2013). Evaluated chaos dynamics for EEG signals of epileptic patients and applied non-linear analysis in understanding the role of chaos in the brain. This approach was found to be 98.6% accurate and they were able to prove that non-linear analysis could be used as an efficient tool in the detection of relative changes in complex brain dynamics. A study presented an overview of various methods for feature selection for seizure detection and prediction placing emphasis on information theory (Giannakakis, Sakkalis, Pediaditis, & Tsiknakis, 2014). Each method was subjected to an evaluation of performance; its ability to automatically detect/predict seizures was also assessed. A combination of various methods was employed in order to exploit efficiently the most robust features. The authors argued that when using support vector machines as a classifier, the selected features played a major role in the efficiency of the classifiers ability to accurately classify. Various results from various authors using the reviewed techniques were compared. A combination of genetic programming and K-nearest neighbours was found to be 99.2% accurate while a combination of PCA-FFT and AIRS classifier and ApEn with Elman network both returned an accuracy of 100%. The models were limited to small data sets and hence provided accurate results and there was no standard guidelines with reliable standardized data (Giannakakis, Sakkalis, Pediaditis, & Tsiknakis, 2014).

2 STATEMENT OF THE PROBLEM

Most people affected by epilepsy live in less developed of developing countries. Early diagnosis is quite useful better treatment of victims. When seizures cannot be predicted the activities of a victim is limited for the sake of safety. Surgery and medication are the common ways to treat seizures but surgeries have been found to fail in helping patients control seizures and medications may not always help patients manage their conditions, being able to predict seizures before occurrence becomes the best way of managing the condition when treatments fail. Due to varying experiences not all patients experience physical symptoms of the pre-ictal phase of a seizure hence the need for intelligent systems that could predict seizures

without a patient being admitted and examined for long hours. Previous works have been found to produce good accuracies, using a powerful optimization technique such as genetic algorithm can reduce the feature vector and hence increase prediction speed and also accuracy. This study examines the power of genetic algorithm and compares result to already existing methods.

3 SUPPORT VECTOR MACHINE

The support vector machine was invented in 1995 by Vapnik and Cortes for binary classification. It is a supervised learning model and although it has undergone many modification in its basic form the support vector machine classifies labelled data into two classes. Using separating hyperplane. The hyperplane is such that it is the farthest away from both classes of data to be classified.

Support vector machine is used for both classification and regression problem. The support vector machine is usually trained using a learning algorithm. The most common used algorithms include sequential minimal optimization and coordinate descent. Because of the success of support vector machines in classification, various derivation of the model that employ the same logic for other type of problems has emerged some of them include least square support vector machine(LS-SVM) used for classification and regression and support vector clustering used to perform cluster analysis.

A book by (Russell & Norvig, 2016) describes three characteristics that makes the Support vector machine attractive:

1. It constructs a decision boundary with the largest possible distance to example points (maximum margin separator) which helps it generalize well.
2. The possess the ability to embed data into a higher-dimensional space using something called the kernel trick
3. They combine the advantages of parametric and non-parametric models and thus have flexibility to represent complex functions.



4 GENETIC ALGORITHM

Genetic algorithm is a search optimization technique that is based on Darwinian principle of evolution and natural selection. A solution generated by genetic algorithm is called a chromosome, while collection of chromosome is referred as a population (Hermawanto, 2013). The chromosome is made up of genes whose value can either be numerical, binary, symbols or characters. The algorithm usually starts by generating a random population of parent chromosomes and defining a fitness function which is used to measure the fitness of each parent. The fit parents are usually selected to be members of the next generation of chromosomes. A new generation of chromosome is created by a process of selection mutation and cross over. The genetic algorithm will be used in feature extraction and Support vector machine will be trained to carry out binary classification of the data.

A description of the genetic algorithm as described by (Hermawanto, 2013) is as follows:

Step 1: Determine the number of chromosomes, generation, and mutation rate and crossover rate value

Step 2: Generate chromosome-chromosome number of the population, and the initialization value of the genes chromosome-chromosome with a random value

Step 3: Process steps 4-7 until the number of generations is met

Step 4: Evaluation of fitness value of chromosomes by calculating objective function

Step 5: Chromosomes selection

Step 6: Crossover

Step 7: Mutation

Step 8: Solution (Best Chromosomes)

5 METHODOLOGY

The process of detecting the epileptic seizure using the support vector machine classifier is broken down into phases of pre-processing, feature selection and classification of input vectors. Figure 2.1 shows a block diagram of the methodology



Figure 2.1: block diagram of proposed model

SVMs usually employ a learning algorithm for classification; our approach to classifying our dataset will involve the use of genetic algorithm for the selection of robust features for the SVM classifier. Our data set was obtained from UCI online data repository and consists of 5 classes of data subjects such that class 2-5 are subjects who did not have epileptic seizure and only subjects in class 1 have epileptic seizure. A binary classification of class 1 against the other classes will be performed. 2/3 of the data will be used in training the SVM and the remaining 1/3 of the data will be used to test the SVM classifier the result will then subjected to sensitivity specificity and accuracy test.

The parameters of the genetic algorithm are summarized in table 2.1 below.

| parameter | Parameter value |
|--------------------------|--|
| Probability of crossover | The probability that two population members will exchange gene |
| Probability of mutation | The probability that a gene will be flipped |
| Population size | The number of individuals in each generation |
| Number of generations | The maximum number of generations created by the algorithm |

Table 2.1: summary of parameters of Genetic algorithm

The population size is the total number of individuals in each generation. Each member is called a chromosome and each chromosome is made up of genes. The gene is represented using a binary number



of 0 or 1. The total number of gene in each chromosome equals to the total number of features in the feature vector. The number of generation is the maximum number of times the algorithm will generate a population to find the best individual. The probability of a crossover is the probability that two members of the population will exchange genes. If the probability is set to zero then there is no exchange of genes and hence crossover will not occur. The probability of mutation defines the possibility of random elements of chromosomes will be flipped, i.e. the probability that a gene with a current value of 1 will be flipped to 0 and vice versa.

The algorithm was implemented in weka (wekato environment) version 3.5 and a correlation based feature selection attribute evaluator was used to evaluate the subset of attributes. The objective function was therefore a measure of the correlation of each attribute to the target class and the attributes with higher correlation to the class labels and low correlation with other attributes was selected as fit by the algorithm.

After using genetic algorithm to select a subset of most fit features, SVM will then be used to classify the instances. The classifier's efficiency will be evaluated based on some performance metrics as summarized in table 2.2

| Performance metric | Brief description |
|--------------------|--|
| Accuracy | Ratio of instances correctly classified by the classifier |
| Sensitivity | The ratio of positive instances classified as positive by the classifier. Also called the true positive rate |
| Specificity | The ratio of negative instances classified as negative by the classifier also called the true negative rate. |

Table 2.2: summary of performance metrics used in the research

6 RESULTS AND DISCUSSIONS

Missing values were not found in the dataset hence for preprocessing the data, Z-score normalization was carried out. The preprocessed data was then passed to the genetic algorithm. The values of the genetic algorithm parameters are shown in Table 3.0

Table 3.0: Genetic algorithm parameter values

| parameter | Parameter value |
|-----------------------|-----------------|
| Crossover Probability | 0.6 |
| Mutation Probability | 0.033 |
| Population size | 100 |
| Number of generations | 100 |

The algorithm selected 100 out of the 178 original features, reducing the feature vector by 78 features. The features were then passed to the SVM classifier for classification. The hyper-parameters used in our SVM model are described in table 3.1 below

| Hyperparameter | Value |
|-----------------------------|--|
| Kernel | Gaussian kernel also known as the radial basis function(RBF) |
| Regularization parameter, C | 1000 |
| Gamma | 0.001 |

Table 3.1: Support Vector Machine hyper-parameters

The classifier achieved an accuracy of 97.39%. the performance of the classifier is summarized in table 3.2 below.



| Performance metric | Value |
|--------------------|--------|
| Accuracy | 97.73% |
| Sensitivity | 97% |
| Specificity | 97% |

Table 3.2: SVM classifier performance result

It is also noteworthy to know that most classifiers performance can be affected by many factors some of which are unique to the classifier, the quality of data or preprocessing techniques employed. It is therefore safe to say that feature selection employed in this research has pointed us in the right direction towards an improved prediction of epileptic seizures.

Figure 3.1 below shows the confusion matrix of the SVM classifier. A total of 3450 instances made up the test set. 2760 of them were labelled negative by the classifier, 2718 of them were correctly classified (true negative) while 42 was misclassified (false positive). 690 of them were classified as positive by the classifier, 642 were correctly classified (true positive) and 48 were misclassified (false positive).

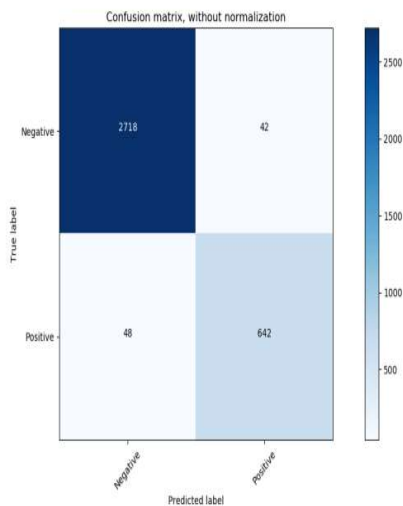


Figure 3.1: confusion matrix without normalization

7 CONCLUSION

From our research genetic algorithm was found to improve the classification accuracy by 0.36% and also improves the SVM sensitivity by 8.11% with respect to the existing models. We can therefore conclude that performing feature selection improved the classifier accuracy.

8 REFERENCES

- Betts, T., & Boden, S. (1992). Diagnosis, management and prognosis of a group of 128 patients with non-epileptic attack disorder. Part I. Seizure, 1(1), 19-26.
- Fisher, R. S., Boas, W. V., Blume, W., Elger, C., Genton, P., Lee, P., & Engel, J. (2005). Epileptic seizures and epilepsy: definitions proposed by the International League Against Epilepsy (ILAE) and the International Bureau for Epilepsy (IBE). 46(4), 470-472.
- Giannakakis, G., Sakkalis, V., Pedititis, M., & Tsiknakis, M. (2014). *Methods for seizure detection and prediction: an overview*. New York, NY: Humana Press.
- Hermawanto, D. (2013). Genetic algorithm for solving simple mathematical equality problem. arXiv preprint arXiv:1305.0001.
- Hosseini, S. A., Akbarzadeh-T, M. R., & Naghibi-Sistani, M. B. (2013). Qualitative and quantitative evaluation of eeg signals in epileptic seizure recognition. , 5(6), 41. *International Journal of Intelligent Systems and Applications*, 5(6), 41.
- Hosseini, S. A., Mohammed-Reza, A.-T., & Mohammed-Bagher, N.-S. (2013). Methodology for epilepsy and epileptic seizure recognition using chaos analysis of brain signals. *Computational Intelligence in Multi-Agent Systems: Theory and Practice*, (pp. 20 - 36).



**3rd International Engineering Conference (IEC 2019)
Federal University of Technology, Minna, Nigeria**



Russell, S. J., & Norvig, P. (2016). *Artificial intelligence: a modern approach*. Malaysia: Pearson Education Limited.

Shoeb, A. H., & Guttag, J. V. (2010). Application of machine learning to epileptic seizure detection. In Proceedings of the (pp. 975-982). *27th International Conference on Machine Learning (ICML-10)* (pp. 975 - 982). Proceeding.

Zhang, T., & Oles, F. J. (2001). Text categorization based on regularized linear classification methods. *Information Retrieval*, 4(1), 5 - 31.