A SURVEY ON THE CLASSIFICATION TECHNIQUES FOR SEIZURE PREDICTION FROM EEG

Parvathy Prathap[1], Aswathy Devi T.[2]

[1] *M.Tech Signal Processing Student, Department of ECE, LBSITW, Kerala, India*
[2] *Assistant Professor, Department of ECE, LBSITW, Kerala, India*

ABSTRACT

Epilepsy is basically a neurological disorder which is characterized by epileptic seizures. These seizures can either be partial or generalized. Epileptic seizures are sometimes accompanied by heavy and vigorous shivering or shaking which can lead to physical injuries and unconsciousness. It can even get fatal if not properly attended. Timely Seizure prediction is important as far as the safety and health of an epileptic person is concerned. It helps in taking precautions like brain stimulation or appropriate medication before a seizure strikes a person. Here comes the relevance of an efficient and accurate seizure prediction algorithm. There are several classification techniques which can be used to predict the onset of seizure. This paper primarily focuses on a review of such available methods and their important features. Classification techniques such as the Support Vector Machines, Bayesian Classifier, Artificial Neural Networks, Adaboost, Adaptive Neuro Fuzzy Inference System etc. are being discussed here. It also discusses the pros and cons of each of them. The primary aim is to conquer a classification method which can effectively be used in predicting an ictal event. Different methods use different techniques of classification which can uniquely identify the onset of a seizure thereby reducing false positives.

Keywords: *Seizure, Ictal, Bayesian, Neural Network, Adaboost, Support vector Machine*

1. INTRODUCTION

Epilepsy is basically a neurological disorder which is characterized by epileptic seizures [1]. Epileptic seizures are sometimes accompanied by heavy and vigorous shivering which can lead to physical injuries, unconsciousness and even death .Seizure is the result of abnormal neural activity in the cortex of the brain. It can either be partial, wherein it affects only a part of the brain, or generalized, where the entire brain region gets affected. Partial seizures can get transformed into generalized ones as well. Seizure prediction can be helpful to provide timely medicine or brain stimulation so as to prevent an upcoming seizure from occurring. An implantable drug delivery system can also be used if proper seizure prediction is done. Thus timely and accurate seizure prediction has a great relevance. An effective algorithm with maximum sensitivity and minimum false positive rate is the requirement in such cases. Efficiency of most of the algorithms reduce when implemented for long term recordings. The methods

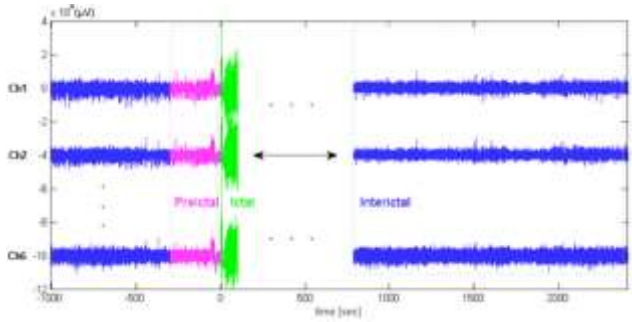

Fig 1. Stages in an EEG signal

discussed here uses various classification techniques which can classify and thereby predict an ictal event and hence make a seizure prediction. The problem mainly lies in distinguishing the preictal state (period preceding the onset of a seizure) from the interictal state (period between two successive seizures). An illustration of the various states of an EEG signal before, during and after a seizure has been shown in Fig 1.The seizure prediction system must be clearly defined to distinguish these states. The sensitivity and the false positive rate of the prediction system is important as far as the efficiency of the system is concerned. The computational complexity of the system should also be reduced as much as possible as it needs to be implemented on an implantable device.

2. BRIEFING OF THE VARIOUS CLASSIFICATION TECHNIQUES FOR SEIZURE PREDICTION

2.1 Classification using Support Vector Machines

Support vector machines can be used to train systems for proper classification [2]. SVM rely on preprocessing the data to represent patterns in a high dimension — typically much higher than the original feature space. With an appropriate nonlinear mapping ϕ () to a sufficiently high dimension, data from two categories can always be has been transformed to y = ϕ(x). For each of the n patterns, k = 1, 2, ..., n, we let z_k = +1 or -1, according to whether pattern k is in ω1 or ω2. A linear discriminant in an augmented y space is

$$g(\mathbf{y}) = \mathbf{a}^t\mathbf{y} \quad \text{… (1)}$$

A separating hyperplane ensures

$$z_k g(\mathbf{y}_k) \geq 1 \quad \text{… (2)}$$

Margin is said to be any positive distance from the decision hyperplane. The goal in training a Support Vector Machine is to find the separating hyperplane with the largest margin; it is expected that the larger the margin, the better the generalization of the classifier. Hence we need to find the weight vector that maximizes b in the equation

$$\frac{z_k g(\mathbf{y}_k)}{||\mathbf{a}||} \geq b \quad \text{… (3)}$$

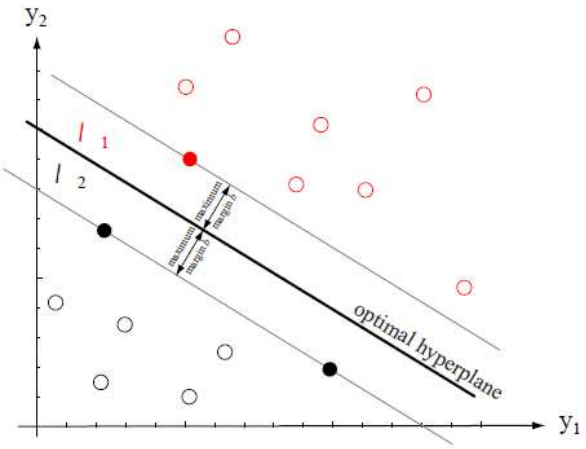

Fig 2. SVM Classifier

The hyperplane, margin and support vectors are illustrated in Fig 2. SVM can either be linear or nonlinear depending on the nature of data that needs to be classified. Nonlinear SVM can be Radial Basis Function SVM (RBF-SVM) which uses a kernel for classification.

2.2 Classification Using the Bayesian Classifier

The naive Bayes classifier is easy to build with no complicated iterative parameter estimation, which makes it particularly useful for hardware implementation [3]. The Bayesian classifier uses Bayes theorem to find out the probability of a data belonging to a particular class, given observations. For a set of feature vectors d and class Ci, the Bayes theorem is given as

$$P(c_i \mid d) = \frac{P(d \mid c_i)P(c_i)}{P(d)} \quad \ldots (4)$$

The best class to assign data is the one that maximizes this conditional probability out of all the classes. This can be represented by the following:

$$c = \operatorname{argmax}\ P(c_i \mid d)$$
$$c = \operatorname{argmax}\ P(c_i)\prod_x P(x \mid c_i). \quad \ldots (5)$$

Due to the fact that calculating the product of the above probabilities will lead to float point underflow, the product operation is converted into summation by using log. Thus, instead of choosing the class with the highest probability, we choose the one with the highest log score. Given that the logarithm function is monotonic, the decision remains the same. So above equation can be written as

$$c = \operatorname{argmax}\left[\log(P(c_i)) + \sum_x \log(P(x \mid c_i))\right] \quad \ldots (6)$$

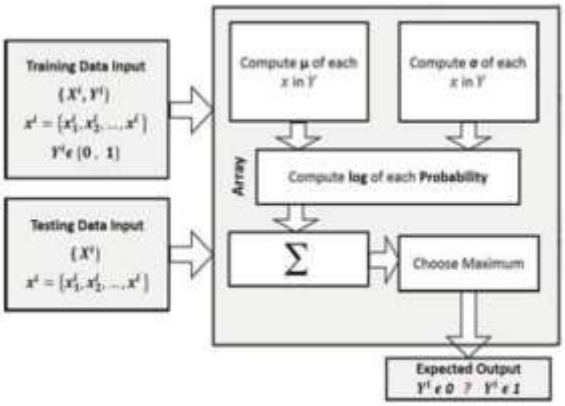

Fig 3. Naïve Bayes Classifier

The values of each feature vector associated with each class are distributed according to Gaussian. Hence, the likelihood conditional probability of a given value v from a feature vector x given a class C_i, P(v | Ci), is given as

$$P(x = v \mid c_i) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(v-\mu)^2/2\sigma^2}$$

$$\mu = \sum_{i=1}^{N} (x_i)$$

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^{N} (x_i - \mu)^2.$$

… (7)

The architecture of Naïve Bayes Classifier is shown in Fig 3. It basically involves computation of the mean and variance of the input data set, computing the logarithm of the probability values and finally doing the classification based on the maximum value of this function.

2.3 Classification Using Artificial Neural Network

Artificial Neural Networks are based on the neural structure of the brain [4]. They are crude electronic networks of neuron. They process records one at a time, and learn by comparing their classification of the record (i.e., largely arbitrary) with the known actual classification of the record. The errors from the initial classification of the first record is fed back into the network, and used to modify the networks algorithm for further iterations. A neuron in an artificial neural network is:

1. A set of input values (x_i) and associated weights (w_i).

2. A function (g) that sums the weights and maps the results to an output (y).

An illustration of the same is shown in Fig 4. Neurons are organized into layers: input, hidden and output. The input layer is composed not of full neurons, but rather consists simply of the record's values that are inputs to the next layer of neurons. The next layer is the hidden layer. Several hidden layers can exist in one neural network. The final layer is the output layer, where there is one node for each class. A single sweep forward through the network results in the assignment of a value to each output node, and the record is assigned to the class node with the highest value.

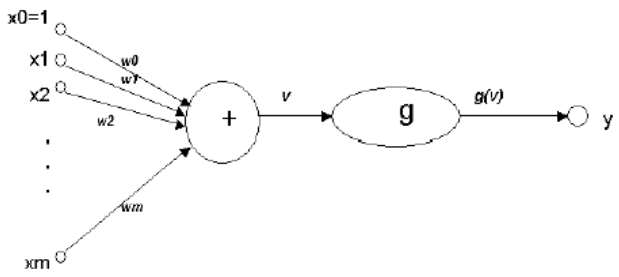

Fig 4. Neural Network Illustration

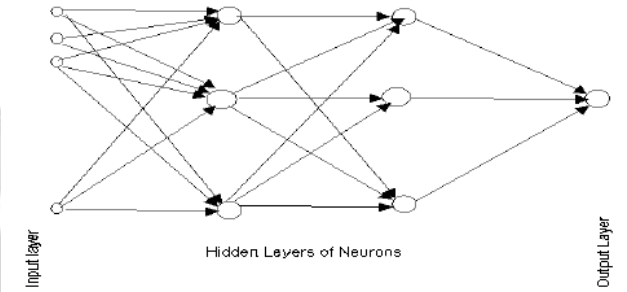

Fig 5. Neural Network Illustration of hidden layers

In the training phase, the correct class for each record is known (termed supervised training), and the output nodes can be assigned correct values -- 1 for the node corresponding to the correct class, and 0 for the others. (In practice, better results have been found using values of 0.9 and 0.1, respectively.) It is thus possible to compare the network's calculated values for the output nodes to these correct values, and calculate an error term for each node (the Delta rule). These error terms are then used to adjust the weights in the hidden layers so that, hopefully, during the next iteration the output values will be closer to the correct values. Illustration of the hidden layers of the neural network is shown in Fig 5.

2.4 Classification Using Adaboost

Adaboost is a machine learning algorithm which can be used as a classifier [5]. It includes a set of reliable features to form a strong classifier capable of correctly classifying the input values. The main challenge is to find these features. Adaboost can be used to find out these strong classifiers. After these features are found, a weighted combination of all these features is used in evaluating and deciding any given EEG segment is preictal or not. Each of these selected features are considered okay to be included if they can perform better than random guessing.(i.e. they can detect correctly more than half of the given input samples). Such classifiers can be termed as weak classifiers. The steps in this boosting algorithm can be briefed as follows. Initially, give equal weight to each training example. Then

1) Find best weak learner for current weighted training set

2) Raise the weights of training examples misclassified by current weak learner

3) Compute final classifier (strong classifier) as linear combination of all weak learners (weight of each learner is related to its accuracy)

A classifier is initially applied and the weights of the wrongly classified ones are increased. This process is repeated until a strong classifier (which can be expressed as a linear combination of weak classifiers) can be obtained.

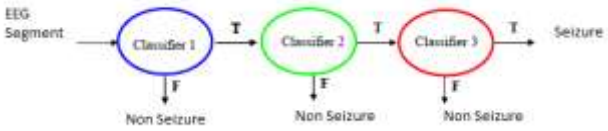

Fig 6. Cascading of Classifiers in Adaboost

The stages here are:

1) Start with simple classifiers which reject many of the negative sub-windows while detecting almost all positive sub-windows

2) Positive results from the first classifier triggers the evaluation of a second (more complex) classifier, and so on

3) A negative outcome at any point leads to the immediate rejection of the sub-window

This can be illustrated by Fig 6.

2.5 Classification Using Adaptive Neuro Fuzzy Inference System

ANFIS are a class of adaptive networks that are functionally equivalent to fuzzy Inference systems [6]. It uses a hybrid learning algorithm. It has added neural network learning capabilities. The antecedent or the premise part is linguistic in nature. Thus, the premise part performs qualitative fuzzy reasoning. The consequent parameter is a linear function of the input variables. Fuzzy if-then rules performs the logic "AND" operations on the inputs provided. The fuzzy if-then rules are defined as follows:

$$If\ (F_1\ is\ A_i)\ and\ (F_2\ is\ B_i)\ and\ (F_3\ is\ C_i)\ and\ (F_4\ is\ D_i)$$
$$then\ (f_i = p_iF_1 + q_iF_2 + r_iF_3 + s_iF_4 + t_i)$$

where Fi(i=1,2,…4) is the input, A_i, B_i, C_i and D_i are the fuzzy sets and p_i,q_i,r_i,s_i and t_i are the linear design parameters. The linear parameters are adaptable. The square nodes are adaptive whereas circular nodes are fixed. Fuzzifications of the input variables are performed in the first layer and all the nodes of the first layer are adaptive nodes. Fuzzy input membership function parameters and the design parameters were optimized. The outputs of the first layer are the fuzzy membership grades of the inputs. The membership grade parameters are used to adaptively estimate the membership grades during training to better map the input/output relationships. The second layer nodes perform the product operation (logic operation "AND") to calculate the firing strength of each rule. The third layer

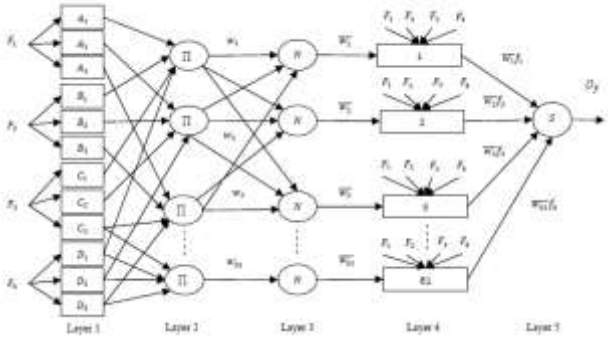

Fig 7. The ANFIS architecture for four inputs with three membership functions and one output

performs the data normalization. The fourth layer performs the following operation.

$$O_i^4 = \bar{w}_i (p_i F_1 + q_i F_2 + r_i F_3 + s_i F_4 + t_i)$$

Here the w_i term indicates the output of the previous layer, Fi(i=1,2,3…4) is the input and $\{p_i,q_i,r_i,s_i,t_i\}$ is the first order polynomial parameter set. The final layer consists of a single node which is responsible for performing the summation of all the incoming signals coming from previous layer. This fuzzy output variable which is a mapped output of all the input features can be used for issuing the seizure prediction alarm. The corresponding ANFIS architecture has been shown in Fig 7.

3. PERFORMANCE COMPARISON OF VARIOUS METHODS

The Support Vector Machine classifier is one of the most widely used classifier for seizure prediction. Among this, linear SVM is preferred when the key highlight is to reduce the computational complexity. But then, this choice also depends on the nature of features used. If the features chosen for seizure prediction are linearly separable, then linear SVM would be a good choice. But if the features are not linearly separable, then using a linear SVM would increase the misclassification rate. In such cases, nonlinear SVMs such as RBF-SVM can be used. But the issue here is that the computational complexity increases as the computation of RBF kernel itself is intensive in some cases. The Bayesian classifier has a major advantage that it is simple and easy to calculate. But the major drawback lies in the fact that it is effective only when the features are independent. ANN is nonlinear model that is easy to use and understand compared to statistical methods. ANN with Back propagation (BP) learning algorithm is widely used in solving various classification and forecasting problems. It has the ability to detect relationship between dependent and independent variables. Its major disadvantage is that it is a bit complex to understand and it might lead to over fitting when it has too many parameters compared to the number of observations. This might decrease the prediction capability. Adaboost classification technique is a bit complex compared to Bayesian classifier but it is a good choice when there are more number of features to be considered. ANFIS is actually a combination of neural networks and the fuzzy logic to inculcate the postives of both but this also makes it a bit more complicated to design since it involves an additional stage of fuzzification of the various layers of the neural network.

4. CONCLUSION

Classification is a major stage in seizure prediction wherein we finally say if the given EEG segment corresponds to a preictal event or not. Several algorithms are available for classification. The paper provides an overview of the major ones. The choice of a classifier mainly depends on the demands that are to be met by the seizure prediction algorithm. If complexity is the main concern, a linear classifier would be the best option to go for. If large number of features are considered, it is better to opt for algorithms like adaboost as it creates a strong classifier as a linear combination of various features. Hence, the classifier of choice varies depending on the features chosen, the feature selection steps taken, the quantity of data available for the training phase, the computational intensity and the memory requirements that can be tolerated by the system.

5. REFERENCES

[1] Fisher, Robert S., et al. "Epileptic seizures and epilepsy: definitions proposed by the International League Against Epilepsy (ILAE) and the International Bureau for Epilepsy (IBE)." Epilepsia 46.4 (2005): 470-472.

[2] Ma, Shuoxin, and Daniel W. Bliss. "Intra-patient and inter-patient seizure prediction from spatial-temporal EEG features." *Signals, Systems and Computers, 2014 48th Asilomar Conference on*. IEEE, 2014.

[3] Murphy, Kevin P. "Naive bayes classifiers." University of British Columbia(2006).

[4] Jain, Anil K., Jianchang Mao, and K. Moidin Mohiuddin. "Artificial neural networks: A tutorial." *Computer* 29.3 (1996): 31-44.

[5] Avinala, Manohar, and Keshab K. Parhi. "Low complexity algorithm for seizure prediction using Adaboost." *Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE*. IEEE, 2012.

[6] Rabbi, Ahmed F., Leila Azinfar, and Reza Fazel-Rezai. "Seizure prediction using adaptive neuro-fuzzy inference system." *Engineering in Medicine and Biology Society (EMBC), 2013 35th Annual International Conference of the IEEE*. IEEE, 2013.