Supervised Learning Approach for Predicting the Presence of Seizure in Human Brain

Sivagami P, Sujitha V
M.Phil Research Scholar
PSGR Krishnammal College for Women
Coimbatore, India
sivagamithiru@gmail.com, vsujitha1987@gmail.com

Vijaya MS
Associate Professor and Head
GRG School of Applied Computer Technology
PSGR Krishnammal College for Women
Coimbatore, India.
msvijaya@grgsact.com

Abstract— Seizure is a synchronous neuronal activity in the brain. It is a physical change in behavior that occurs after an episode of abnormal electrical activity in the brain. Normally two diagnostic tests namely Electroencephalogram (EEG) and Magnetic Resonance Imaging (MRI) are used to diagnose the presence of seizure. The sensitivity of the human eye in interpreting large numbers of images decreases with increasing number of cases. Hence, it is essential to automate the accurate prediction of seizure in patients. In this paper supervised learning approaches has been employed to model the prediction task and the experiments show about 94% high prediction accuracy.

Keywords—Seizure; Support vector machine; K-NN; Naïve Bayes; J48

I. INTRODUCTION

Seizure is defined as a transient symptom of "abnormal excessive in the brain". Seizures can cause involuntary changes in body movement or function, sensation, awareness, or behavior. It is an abnormal, unregulated electrical discharge that occurs within the brain's cortical grey matter and transiently interrupts normal brain function [1]. Based on the physiological characteristics of seizure and the abnormality in the brain, the kind of seizure is determined. Seizure is broadly classified into absence seizure, simple partial, complex partial and general seizure. Absence seizure is a brief episode of staring. It usually begins in childhood between ages 4 and 14. Simple partial seizure affects only a small region of the brain, often the hippocampus. Complex partial seizure usually starts in a small area of the temporal lobe or frontal lobe of the brain. General seizure affects the entire brain.

Various diagnostic techniques normally employed for patients are Computed Tomography (CT), Magnetic Resonance Imaging (MRI) and PET (Positron Emission Tomography). Magnetic Resonance Imaging (MRI) is used as a valuable tool and widely used in the clinical and surgical environment for seizure identification because of its characteristics like superior soft tissue differentiation, high spatial resolution and contrast. Magnetic Resonance Images are examined by radiologists based on visual interpretation of the films to identify the presence of seizure.

Machine learning is a technique which can discover previously unknown regularities and trends in diverse datasets [2]. Today machine learning provides several indispensable tools for intelligent data analysis. Machine learning technology is currently well suited for analyzing medical data and empirical results reveal that the machine learning systems are highly efficient and could significantly reduce the computational complexities.


The motivation behind the research reported in this paper is to predict the presence of seizure in human brain. Machine learning techniques are employed here to model the seizure prediction problem as classification task to facilitate physician for accurate prediction of seizure presence. In this paper supervised learning algorithms are made use of for the automated prediction of type of seizure.

II. PROPOSED METHODOLOGY

The proposed methodology models the seizure prediction as a classification task and provides a convenient solution by using supervised classification algorithms. Descriptive features of MRI image such as energy, entropy, mean, standard deviation, contrast, homogeneity of grey scale image have been extracted and used for training. The model is trained using training datasets and the trained model is built. Finally the trained model is used to predict the type of seizure.

The proposed model is shown in Figure.1.
A. Image Acquisition

A magnetic resonance imaging (MRI) scan of the patient’s brain is a noninvasive method to create detailed pictures of the brain and surrounding nerve tissues. MRI uses powerful magnets and radio waves. The MRI scanner contains the magnet. The magnetic field produced by an MRI is about 10 thousand times greater than the earth's. The magnetic field forces hydrogen atoms in the body to line up in a certain way. When radio waves are sent toward the lined-up hydrogen atoms, it bounces back and a computer records the signal. Different types of tissues send back different signals.

The MRI dataset consisting of MRI scans images of 350 patients of five types namely Normal, Absence Seizure, Simple Partial Seizure, Complex Partial Seizure and General Seizure are taken into consideration.

B. Feature Extraction

The purpose of feature extraction is to reduce the original data set by measuring certain properties or features that distinguish one input pattern from another. A brain MRI slices is given as an input. The various features based on statistical, grey level co-occurrence matrix and grey level run-length matrix from the MRI is extracted. The extracted features provide the characteristics of the input type to the classifier by considering the description of the relevant properties of the image into a feature space.

The statistical features based on image intensity are mean, variance, skewness and kurtosis. The grey level co-occurrence matrices (GLCM) features such as Contrast, Homogeneity, Correlation, Energy, Entropy and the features of grey level run length matrices (GLRLM) such as Short run emphasis, Long run emphasis, Grey level distribution, Run-length distribution, Run percentage, Low grey level run emphasis, High grey level run emphasis are used to investigate the adequacy for the discrimination of the presence of seizure. Table I shows the features of MRI of a human brain.

<table>
<thead>
<tr>
<th>Statistical Features</th>
<th>Grey Level Co-occurrence Matrix</th>
<th>Grey Level Run Length Matrix</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean Variance</td>
<td>Contrast</td>
<td>Short run emphasis</td>
</tr>
<tr>
<td>Skewness</td>
<td>Homogeneity</td>
<td>Long run emphasis</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>Correlation</td>
<td>Grey level distribution</td>
</tr>
<tr>
<td></td>
<td>Energy</td>
<td>Run length distribution</td>
</tr>
<tr>
<td></td>
<td>Entropy</td>
<td>Run percentage</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Low grey level run emphasis</td>
</tr>
<tr>
<td></td>
<td></td>
<td>High grey level run emphasis</td>
</tr>
</tbody>
</table>

1) Grey Level Co-occurrence Matrix(GLCM)
The GLCM is defined as a tabulation of different combinations of pixel brightness values (grey levels) occur in an image. The texture filter functions provide a statistical view of texture based on the image histogram. This function provides useful information about the texture of an image but does not provide information about shape, i.e., the spatial relationships of pixels in an image.

The features corresponding to GLCM statistics and their description are:
- **Contrast** - Measures the local variations in the grey-level co-occurrence matrix.
- **Homogeneity** - Measures the closeness of the distribution of elements in the GLCM to the GLCM diagonal.
- **Correlation** - Measures the joint probability occurrence of the specified pixel pairs.
- **Energy** - Provides the sum of squared elements in the GLCM. Also known as uniformity or the angular second moment.
- **Entropy** - Statistical measure of randomness.

2) Grey Level Run Length Matrix(GLRLM)
The GLRLM is based on computing the number of grey-level runs of various lengths. A grey-level run is a set of consecutive and collinear pixel points having the same grey level value. The length of the run is the number of pixel points in the run [7]. Seven features are extracted from this matrix.

C. Supervised Classification Algorithms

Supervised learning is a machine learning technique for deducing a function from training data. The training data consist of pairs of input objects and desired outputs. The output of the function can predict a class label of the input object called classification. The task of the supervised learner is to predict the value of the function for any valid input object after having seen a number of training examples i.e., pairs of input and target output. The supervised classification techniques namely, support vector machine, decision tree
induction, Naïve Bayes and k-nn are employed in seizure prediction modeling.

1) Support Vector Machine

The machine is presented with a set of training examples, \((x_i, y_i)\) where the \(x_i\) is the real world data instances and the \(y_i\) are the labels indicating which class the instance belongs to. For the two class pattern recognition problem, \(y_i = +1\) or \(y_i = -1\). A training example \((x_i, y_i)\) is called positive if \(y_i = +1\) and negative otherwise [6]. SVMs construct a hyper plane that separates two classes and tries to achieve maximum separation between the classes. Separating the classes with a large margin minimizes a bound on the expected generalization error.

The simplest model of SVM called Maximal Margin classifier, constructs a linear separator (an optimal hyper plane) given by \(w^T x - y = 0\) between two classes of examples. The free parameters are a vector of weights \(w\) which is orthogonal to the hyper plane and a threshold value. These parameters are obtained by solving the following optimization problem using Lagrangian duality.

\[
\text{Minimize} = \frac{1}{2} \|w\|^2
\]

subject to
\[
D_{ii}(w^T x_i - \gamma) \geq 1, \ i = 1, \ldots, l.
\] (1)

where \(D_{ii}\) corresponds to class labels +1 and -1. The instances with non null weights are called support vectors. In the presence of outliers and wrongly classified training examples it may be useful to allow some training errors in order to avoid over fitting. A vector of slack variables \(\xi_i\) that measure the amount of violation of the constraints is introduced and the optimization problem referred to as soft margin is given below. In this formulation the contribution to the objective function of margin maximization and training errors can be balanced through the use of regularization parameter \(C\). The following decision rule is used to correctly predict the class of new instance with a minimum error.

\[
f(x) = \text{sgn}[w^T x - \gamma]
\] (2)

The advantage of the dual formulation is that it permits an efficient learning of non–linear SVM separators, by introducing kernel functions. Technically, a kernel function calculates a dot product between two vectors that have been (non- linearly) mapped into a high dimensional feature space [8]. Since there is no need to perform this mapping explicitly, the training is still feasible although the dimension of the real feature space can be very high or even infinite. The parameters are obtained by solving the following non linear SVM formulation (in matrix form),

\[
\text{Minimize } L(u) = \frac{1}{2} u^T Qu - c^T u
\]

\[
d^T u = 0 \quad 0 \leq u \leq C e
\] (3)

where \(K\) - the Kernel Matrix. \(Q = DKD\).

The Kernel function \(K(AAT)\) (polynomial or Gaussian) is used to construct hyperplane in the feature space, which separates two classes linearly, by performing computations in the input space.

\[
f(x) = \text{sgn}(K(x, xi^T)u - \gamma)
\] (4)

where \(u\) - the Lagrangian multipliers. In general larger the margins will lower the generalization error of the classifier.

2) Naïve Bayes

Naïve Bayes is one of the simplest probabilistic classifiers. The model constructed by this algorithm is a set of probabilities. Each member of this set corresponds to the probability that a specific feature \(f_i\) appear in the instances of class \(c\), i.e., \(P(f_i \mid c)\). These probabilities are estimated by counting the frequency of each feature value in the instances of a class in the training set. Given a new instance, the classifier estimates the probability that the instance belongs to a specific class, based on the product of the individual conditional probabilities for the feature values in the instance. The exact calculation uses bayes theorem and this is the reason why the algorithm is called a bayes classifier.

3) K-NN

K-nearest neighbor algorithms are only slightly more complex. The \(k\) nearest neighbor of the new instance is retrieved and whichever class is predominant amongst them is given as the new instance's classification. K-nearest neighbor is a supervised learning algorithm where the result of new instance query is classified based on majority of K-nearest neighbor category [9]. The purpose of this algorithm is to classify a new object based on attributes and training samples. The classifiers do not use any model to fit and only based on memory.

4) J48 Decision Tree Induction

J48 algorithm is an implementation of the C4.5 decision tree learner. This implementation produces decision tree models. The algorithm uses the greedy technique to induce decision trees for classification [10]. A decision-tree model is built by analyzing training data and the model is used to classify unseen data. J48 generates decision trees, the nodes of which evaluate the existence or significance of individual features.
III. EXPERIMENTAL SETUP

The seizure data analysis and Prediction has been carried out using WEKA and SVMlight for machine learning.

WEKA is a collection of machine learning algorithms for data mining tasks [11]. SVMlight provides the extensive support for the whole process of experiment including preparing the input data, evaluating learning schemes statistically and visualizing the input data and the result of learning.

The dataset is trained using SVM with most commonly used kernels linear, polynomial and RBF, with different parameter settings for d, gamma and C –regularization parameter. The parameters d and gamma are associated with polynomial kernel and RBF kernel respectively. Image processing toolbox of Matlab has been used for MRI feature extraction. The datasets are grouped into five broad classes namely Normal, Absence Seizure, Simple Partial Seizure, Complex Partial Seizure and General Seizure to facilitate their use in experimentally determining the presence of seizure in MRI. The seizure dataset has 17 attributes, there are 350 instances, and as indicated above, 5 classes. Supervised classification algorithms such as support vector machine, decision tree induction, naive bayes and K-NN are applied for training. Support vector machine learning is implemented using SVM light. Decision tree induction, Naïve Bayes and K-NN are implemented using WEKA. The performance of the trained models has been evaluated using 10 fold cross validation and their results are compared.

IV. RESULTS

The results of the experiments are summarized in Table II. Prediction accuracy and learning time are the parameters considered for performance evaluation. Prediction accuracy is the ratio of number of correctly classified instances and the total number of instances. Learning time is the time taken to build the model on the dataset.

A. Classification using SVM

The performance of the three kinds of SVMs with linear, polynomial and RBF kernels are evaluated based on the prediction accuracy and the results are shown in Table II.

<table>
<thead>
<tr>
<th>SVM Kernels</th>
<th>C=1</th>
<th>C=2</th>
<th>C=3</th>
<th>C=4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear</td>
<td>74</td>
<td>76</td>
<td>72</td>
<td>79</td>
</tr>
<tr>
<td>Polynomial (d)</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>RBF (g)</td>
<td>92</td>
<td>94</td>
<td>93</td>
<td>92</td>
</tr>
</tbody>
</table>

Table III shows the average performance of the SVM based classification model in terms of predictive accuracy and depicted the same in Figure 2.

<table>
<thead>
<tr>
<th>Kernel Type</th>
<th>Prediction Accuracy(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear</td>
<td>75</td>
</tr>
<tr>
<td>Polynomial</td>
<td>80</td>
</tr>
<tr>
<td>RBF</td>
<td>94</td>
</tr>
</tbody>
</table>

![Figure 2. Comparing Prediction Accuracy of SVM Kernels](http://sites.google.com/site/ijcsis/)

The predictive accuracy shown by SVM with RBF kernel with parameter C=3 and g=2 is higher than the linear and polynomial kernel.

B. Classification using WEKA

The results of the experiments are summarized in Table IV and V.

<table>
<thead>
<tr>
<th>Classifiers</th>
<th>Evaluation Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Learning Time (secs)</td>
</tr>
<tr>
<td>Naïve Bayes</td>
<td>0.03</td>
</tr>
<tr>
<td>K-NN</td>
<td>0.02</td>
</tr>
<tr>
<td>J48</td>
<td>0.09</td>
</tr>
</tbody>
</table>
TABLE V. COMPARISON OF ESTIMATES

<table>
<thead>
<tr>
<th>Evaluation criteria</th>
<th>Classifiers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kappa Statistic</td>
<td>Naïve Bayes</td>
</tr>
<tr>
<td>Mean Absolute Error</td>
<td>0.7468</td>
</tr>
<tr>
<td>Root Mean Squared Error</td>
<td>0.2716</td>
</tr>
<tr>
<td>Relative Absolute Error</td>
<td>26.1099</td>
</tr>
<tr>
<td>Root Relative Squared Error</td>
<td>68.428</td>
</tr>
</tbody>
</table>

The performances of the three models are illustrated in Figure 3 and 4.

The time taken to build the model and the prediction accuracy is high in J48 when compared to other two algorithms in WEKA environment.

V. CONCLUSION

This paper describes the modeling of the seizure prediction task as classification and the implementation of trained model using supervised learning techniques namely, Support vector machine, Decision tree induction, Naïve Bayes and K-NN. The performance of the trained models are evaluated using 10 fold cross validation based on prediction accuracy and learning time and the results are compared. It is observed that about 94% high predictive accuracy is shown by the seizure prediction model. As far as the seizure prediction is concerned, the predictive accuracy plays major role in determining the performance of the model than the learning time. The comparative results indicate that support vector machine yield a better performance when compared to other supervised classification algorithms. Due to wide variability in the dataset, machine learning techniques are effective than the statistical approach in improving the predictive accuracy.

ACKNOWLEDGMENT

The authors would like to thank the Management and Acura Scan Centre, Coimbatore for providing the MRI data.

REFERENCES