COMPARISION OF K-NN AND SVM CLASSIFIER FOR THE DISEASE CLASSIFICATION FROM EEG SIGNAL

V.Pavithra[$], R.Praveena[$], S.Sajeetha[$],Dr.R.Deepa*

[$]*Student, *Assistant Professor, Department of Electronics and Instrumentation Engineering*
Bannari Amman Institute of Technology
Sathyamangalam, Erode, Tamilnadu, India

Abstract

Brain signal is an electrical activity of neuron. These signals can be monitored and recorded by a system is known as Electro Encephalogram (EEG). Mostly the EEG signals are contaminated with noises. The most common noise in EEG signal occurs from muscle movement or electromyogram (EMG) and electrooculogram (EOG). These noises can be removed by using suitable filters. In this project, the disease is classified from the EEG signal based on features extraction and classifiers used.The aim of this work is to achieve an automatic patient's classification from the EEG biomedical signals involved in Alzheimer disease, epilepsy and Mild Cognitive Impairment in order to support medical doctors in the right diagnosis formulation. To analyse the biological EEG signal the suitable computer methods are used for pre-processing, feature extraction and classification process which will give a correct accuracy values to classify the diseases.

Keywords: K-NN, SVM algorithm.

Corresponding Author E-mail: **praveena.ei16@bitsathy.ac.in**

I. INTRODUCTION

Electro encephalography (EEG) is a process of copying the neural information from the brain's electrical potential. It deals with the recording and study of electrical activity of the brain by placing the electrodes on the scalp. The electrical activity of the brain exhibits significant complex with strong non-linear and dynamic properties. The amplitude of EEG signal measured from the scalp ranges from 10 µV to 100 µV and frequency is from 1 Hz to 100 Hz. EEG signal can be recorded in two ways i) Monopolar and ii) Bipolar. The voltage difference between an active electrode on the scalp and a reference electrode on the ear lobe is Monopolar recording whereas the voltage difference between two scalp electrodes is Bipolar recording.

EEG signals are characterised by four types of rhythms. They are alpha waves (α), beta waves (β), theta waves (θ), delta waves (δ). These commands are developed by using an algorithm on special features. EEG database contains states of emotions and diseases and they can be classified by using classifiers like K-NN algorithm and support vector machine (SVM) methods. Alzheimer disease (AD) is one of the major threats and it is the widespread form of dementia. Mild Cognitive Impairment (MCI) is the initial stage of Alzheimer's disease. These are the most widespread degenerative disorders. Epilepsy is a common brain disorder that has affected almost 60 million people around the world according to the estimate of the World Health Organization (WHO). The symptoms of Alzheimer's disease are loss of memory and Mild Cognitive Impairment affects social activities. Epilepsy is a symptom of brain damage and it may also occur due to brain tumour. It is caused due to defects in birth delivery or head injury due to accident.

Feature extraction can be done using different methods. In Fourier Transform only the time domain features can be extracted and the samples cannot be decomposed effectively so Wavelet Transform is used. In Wavelet Transform both time domain and frequency domain features can be extracted and the samples are decomposed effectively. The time domain features includes alpha waves (α), beta waves (β), theta waves (θ), delta waves (δ) and frequency domain features are mean, variance, standard deviation, maximum value, kurtosis.In this paper Support vector machine (SVM) and k-nearest neighbour (KNN) are used for classification.

II. LITERATURE SURVEY

[1]**Rajamanickam Yuvaraja et al (2018).,** "A deep Learning Scheme for Automatic Seizure Detection from Long-Term Scalp EEG" monitoring brain activity through electroencephalogram (EEG) is an established approach for epilepsy prognosis and for monitoring epilepsy patients. In this paper, we gift a method based totally on the deep convolutional neural networks (CNNs) perform unsupervised feature getting to know framework for automatic seizure onset detection.

[2] **Giulia Fiscon, Emanuel Weitschek et al (2014).,** "Alzheimer's disease patients classification through EEG signals processing" is the purpose to attain an automatic sufferers classification from the EEG biomedical alerts concerned in AD and MCI on the way to assist scientific researchers within the right analysis components. The evaluation of the biological EEG signals calls for effective and efficient computer technological know-how techniques to extract applicable information.

[3] **M. Rajya Lakshmi, Dr. T. V. Prasad et al (2013).,** "Survey on EEG Signal Processing Methods" aimed towards addressing the numerous methodologies required to be adapted in every section of mind sign processing. Prior to this survey, previous surveys have been listed numerous techniques, a few experimental consequences and in comparsion them. This paper shows clear and clean interpretation of every approach and their blessings and downsides along with the signal acquisition, signal enhancement, characteristic extraction and, signal category.

[4] **D. Puthankattil Subha , Paul K. Joseph et al (2010).,** "EEG analysis survey on classification of diseases" describes the important features may be extracted for the diagnosis of different diseases the use of advanced sign processing strategies. In this paper the effect of different events on the EEG sign, and unique signal processing methods used to extract the hidden records from the sign are mentioned in detail.

[5] **Dewi Rahmawati, Riyanarto Sarno et al (2017)** "Classify Epilepsy and Normal EEG using wavelet transform and K-nearest neighbour" describes the stages of epilepsy and ordinary electroencephalogram (EEG) signal. Stages inside the selection-making become performed by way of using a characteristic extraction and mixed with Wavelet Transform (WT). The result from features extraction become applied dimension reduction approach with the aid of the usage of Principal Component Analysis (PCA) set of rules. KNN changed into carried out the usage of result from dimension reduction ranges as features.

[6] **Dayanand Vishwanath Dhongade, T.V.K.H Rao et al(2017).,** "Classification of sleep disorders based on EEG signals by using feature extraction technique with KNN algorithm" describes thediverse statistical measures like maximum cost, minimum value, imply value and fashionable deviation cost of DWT sub-bands are used as extracted functions for detection of various sleep problems.The results of these function reduction techniques are used to classify specific sleep disorders the usage of okay-nearest neighbor (KNN) classifier.

III. PROPOSED METHODOLOGY

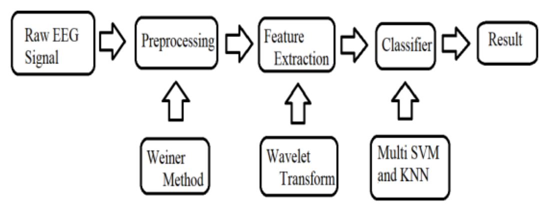

Figure 1 Block diagram of Classification of disease from EEG signal

The EEG signal dataset is collected. The collected EEG signal is then pre-processed to remove the noise using Weiner method which is more effective method in removing noise compared to other filters. Once pre-processing of the signal is done the time domain and frequency domain features are extracted by using Wavelet Transform method. After extracting the features, classifiers are used to classify the disease for EEG signal. SVM and KNN classifiers are used for classification. From the results of SVM and KNN the accuracy of both are compared. SVM method gives more accuracy when compared to KNN. Thus the EEG signals are classified and the result will be displayed in the output. The block diagram of the proposed method is shown in the figure 1.

PREPROCESSING

Pre-processing is a filtering process in which the unwanted noises can be removed significantly. In EEG signal, the noise could affect the accuracy of the signal. The noises in EEG signal are from eye movement and blinking and interference with other devices. Monitoring the EEG signal for classification of diseases without removing the noise could result in incorrect output. There are many types filtering techniques are used to remove the noises from the process like Median filter, butter worth filter, adaptive least mean square, adaptive wiener filter, band pass filter etc.

ADAPTIVE WIENER FILTER

An adaptive filter is a system with linear filter and have a controlled parameter and a mean value is adjusted with the help of algorithms. The adaptive wiener filter is a pre-processing method in which the signal can be processed by linear time invariant filter. In this process, the wiener filter can minimize the mean square error between the random and desired signal. It is the linear estimation of original image. The principle involved in the wiener filtering method in Fourier domain shows that

$$W(f_1,f_2) = \frac{H*(f_1,f_2)S_{xx}(f_1,f_2)}{|H(f_1,f_2)|^2 S_{xx}(f_1,f_2) + S_{nn}(f_1,f_2)}$$

Where $S_{xx}(f_1,f_2), S_{nn}(f_1,f_2)$ are the power spectra of additive noise and original image and $H(f_1,f_2)$ is the blurring noise.

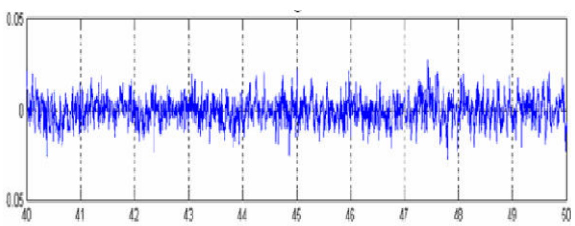

Figure 2 EEG signal after using wiener filter.

FEATURE EXTRACTION

The EEG signal which is pre-processed is then used to extract the features by means of feature extraction. Feature extraction can be either time domain or frequency domain for EEG signal. Both time domain and frequency domain capabilities are extracted in EEG signal. Wavelet Transform (WT) is used to extract the features of the EEG signal. The time domain features are mean, variance, standard deviation, skewness, kurtosis and so on. The frequency area functions are alpha, beta, delta and theta. The time area functions are calculated via the usage of statistical method. Different techniques are used to extract the functions from EEG signal. Fast Fourier Transform (FFT) that is a software of Discrete Fourier Transform (DFT) is used extensively in characteristic extraction. As DFT cannot decompose the data it is not efficient, so Wavelet Transform (WT) is used. Wavelet Transform is the maximum efficient approach and widely used as it is able to decompose the information. The time domain and frequency domain features are explained in detail below.

WAVELET TRANSFORM

In time domain and frequency domain the Wavelet Transform (WT) has a high resolution which allows to know at which frequency the signal oscillates and at which time the oscillations occur. Wavelet Transform can be Discrete Wavelet Transform (DWT) or Continuous Wavelet Transform (CWT). The wavelet transform is the most recent technique for processing signals both in time domain and frequency domain.

TIME DOMAIN FEATURES

The change in signal at a certain period of time is referred to as time domain. With respect to time the mathematical functions, physical signals or a time series of a set of data are analyzed. The variation of amplitude of EEG signal with respect to time refers to time domain features. The time domain features are mean value, variance, standard deviation, skewness, kurtosis etc….

MEAN VALUE:

The most common and easily implemented feature of the time domain is mean. The mean value of EEG signal is constant. The range of EEG capacity is microvolts with time. The equation of mean is

$$mean(\mu_x) = \frac{1}{N}\sum_{i=1}^{N} x_i$$

VARIANCE

Variance is a measure of how far each value from the data set is from the mean. It is a statistical parameter which subtracts the mean value from each data. The subtracted value is then squared and the average of the value is taken. This average represents the variance. The equation for variance is given by

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

STANDARD DEVIATION

The square root of variance is standard deviation. It is a measure of the amount of variation or dispersion from the average. If the data points are further from the mean then there is a higher deviation and vice versa. The calculation of standard deviation for EEG signal is given by the formula

$$std(\sigma) = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(x_i - \mu_x)^2}$$

SKEWNESS

It is a measurement of the asymmetry of the threat distribution of an EEG signal. The mean value may be either positive, negative or undefined. In symmetrical distribution the mean value can be zero and in asymmetrical distribution the mean values are non-zero. The equation for calculating skewness is

$$Skew = \frac{\sum_{i=1}^{N}(x_i - \mu_x)^3}{(n-1)\sigma_x^3}$$

KURTOSIS

Kurtosis is measure of the height of probability distribution or degree of fourth order cumulative. The signal carries temporary spikes, remoted excessive-voltage wave organization. The kurtosis of it gives high positive features or negative qualities. These characteristics are visible when EEG with high or low recurrence and amplitude modulation is examined. It can be written as

$$Kurt = \frac{\sum_{i=1}^{N}(x_i - \mu_x)^4}{(n-1)\sigma_x^4}$$

FREQUENCY DOMAIN FEATURES

The frequency domain features of EEG signal are alpha, beta, delta and theta waves. It refers to the frequency range of the signal.

Rhythm	Frequency (Hz)	Amplitude (µV)	Recording and location
Alpha(α)	8-13	50-100	Adults, rest, eyes closed. Occipital region
Beta(β)	14-30	20	Adult, mental activity Frontal region
Theta(θ)	5-7	Above 50	Children, drowsy adult, emotional distress Occipital
Delta(δ)	2-4	Above 50	Children in sleep

CLASSIFIERS

Classification is an important task of EEG signal. It can perform primarily based on the feature extraction. The talents values are an enter and it can looking ahead to the magnificence for classifier method. A classifier has a number of parameters that want to be found from educated facts. The observed classifier is a model of the association between the capabilities and the classes. The extracted statistics from EEG datasets had been then categorised the use of linear and nonlinear classifiers model for the evaluation of better classifier for EEG facts magnificence. Many kind of classifiers are used. They are K-Nearest Neighbor (KNN), Support Vector Machine (SVM), Artificial Neural Network (ANN), Convolutional Neural Network (CNN) and decision trees. Of those classifiers K-Nearest Neighbor (KNN) and Support Vector Machine (SVM) are used in our challenge for class and their accuracy are in comparison. These are described below in brief.

K-NEAREST NEIGHBOR

The k Nearest Neighbor algorithm is a simple machine learning algorithm which is easy to be implemented. It is used to solve both classification and regression problems. It is a non-parametric method or linear method. KNN algorithm assumes the new data and compares the new data with the available data and puts the new data into the category which it fits with the available data. Based on the similarity the new data is classified. Nearest neighbor classifiers are primarily based on gaining knowledge of by way of analogy, every sample represents a factor in an n-dimensional area. In this way, all the training samples are stored in an n dimensional sample space which is represented as red colour dots in the figure 3.9. When an unknown sample is given as input, the K-nearest neighbour classifiers compares the unknown sample with the patterned space of trained samples. The unknown samples are represented as blue colour dot and the closest values to the trained data should be close as the nearest value. Nearest neighbor classifiers are instance based or lazy learners in that they store all of the training samples and do not build a classifier until a new (unlabelled) sample needs to be classified.

SUPPORT VECTOR MACHINE (SVM)

A support vector machine (SVM) is a supervised machine learning method that uses the classification algorithm for two groups of classification problem. It can work with both linear and non linear process and it is easy to solve the classification problem. It can make a hyperplane which is used to divide the data into two groups. In machine learning, SVM or Support Vector Networks shown in the figure are supervised learning model which are associated with learning algorithms. The analysed data used for classification and analysis. SVM models are the representation of the examples as points in space, mapped so that these examples of the different categories are divided by a clear gap that is as wide as possible.

SVM is a maximum linear classifier with a maximum range which achieves maximum separation by using hyperplane to classify, it might end up closer to set of data set compared to others. The concepts of maximum margin classifier or hyper plane are an apparent solution.

IV. RESULTS AND DISCUSSION

In this project, the input EEG signal is taken from the dataset and the original signal is undergone for the pre-processing techniques. The pre-processed signals are trained with the help of available samples. The suitable methods for the feature extraction are used and to extract the data's related to the time domain and frequency domain specifications. Here, the K-nearest neighbour (K-NN) and the Multi Support Vector Machine(multi SVM) classifiers are compared and to get the finest result with the accuracy of classifiers. At last the disease is classified and the particular disease is displayed as output.

The features of the EEG signal is extracted by using the wavelet transform method where the frequency domain features of alpha ,beta, theta and delta coefficient values are extracted and which is shown as peak like EEG signal in the figure 3.

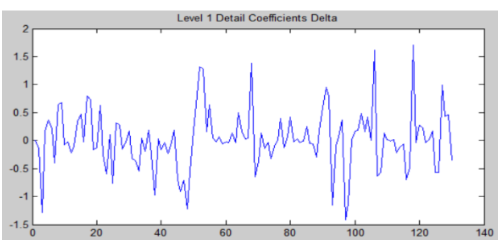

Figure 3 EEG signal with delta coefficient.

The coefficient alpha waves are shown in figure 4. The amplitude and the time frequency of the wave is graphed in level 2.

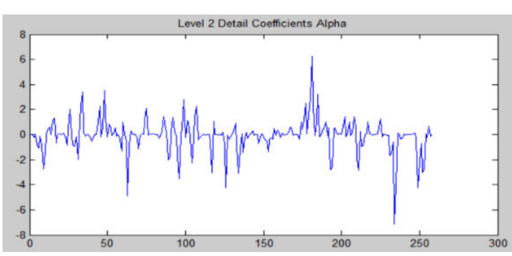

Figure 4 EEG signal with alpha coefficient

The beta waves are irregular waves and is seen in temporal and frontal lobe. The coefficient of the beta wave is shown in figure5 and the corresponding frequency for the given wave is shown.

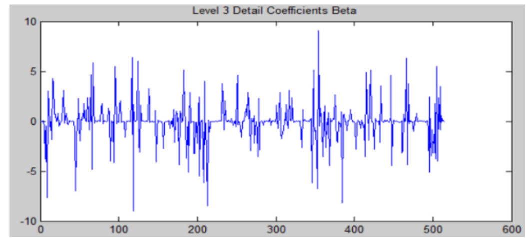

Figure 5 EEG signal with beta coefficient

The K-NN classifier is used to classify the disease from the EEG signal where the trained data are classified into the classes. The input data is assigned and move to the nearest neighbour classes and the exact disease is shown and the graphical view of classes is shown in figure6.

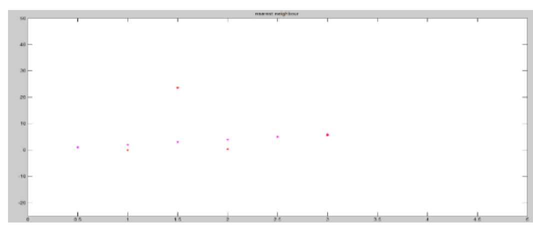

Figure 6 K-NN algorithm

Finally, the disease is classified and it isdisplayed whether the disease is Alzheimer, Epilepsy, Mild Cognitive and Normal is shown in figure 7.

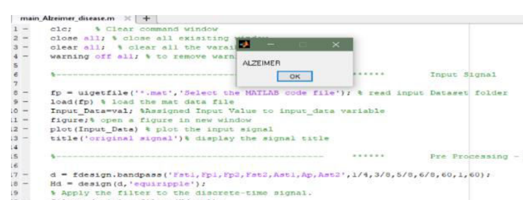

Figure 7 Output of disease classificationfrom EEG signal

V. CONCLUSION

The record of electrical activity of the brain is EEG (Electroencephalogram) signal. EEG helps physicians to diagnose the level of consciousness, sleep disorders, brain death, brain tumour and multiple sclerosis. EEG signal is very important in the diagnosis of epilepsy, Alzheimer's disease and mild cognitive disorder which is the initial stage of Alzheimer. In this paper patients with Alzheimer, epilepsy and mild cognitive disorder were investigated. Their EEG signals was analysed and pre-processed by using different filters and features are extracted by time domain and frequency domain and classifiers are used to classify the disease. The supervised learning methods such as SVM and KNN provided an accurate result for the classification. SVM provided a more accurate result compared to KNN. Thus with the given input of EEG signal, the disease in the patient could be identified.

REFERENCES

[1] Sachin M. Elgandelwar, Vinayak K. Bairagi, "Role of EEG for Diagnosis of Alzheimer Disease", International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-8 Issue-10, August 2019.

[2] Elgandelwar S.M. and Bairagi v. k., "Analysis of EEG Signals for Diagnosis of Alzheimer Disease". International Journal of Scientific & Engineering Research, 7(1), pp 529-532, 2016.

[3] HafeezUllah Amin, Aamir Saeed Malik, Rana Fayyaz Ahmad, NidalKamel NasreenBadruddin, Muhammad Hussain, Weng-Tink Chooi, "Feature extraction and classification for EEG signals using wavelet transform and machine learning techniques", Australasian Physical & Engineering Sciences in Medicine, vol. 38, no. 1, pp. 139-149, March 2015.

[4] O.D. Eva and A. M. Lazar,"Comparison of Classifiers and Statistical Analysis for EEG Signals Used in Brain Computer Interface Motor Task Paradigm", International Journal of Advanced Research in Artificial Intelligence (IJARAI), vol. 4, no. 1,pp 8-12, 2015.

[5] Kulkarni N, Bairagi V. K., "Diagnosis of Alzheimer disease using EEG signals". International Journal of Engineering Research & Technology (IJERT) 3(4):1835–1838, 2014.

[6]. Iscan Z, Dokur Z, Tamer D. Classification of electroencephalogramsignals with combined time and frequency features. Expert Systemswith Applications. 2011;38:10499–10505.

[7] C. Lehmann, T. Koenig, V. Jelic, L. Prichep, R. E. John, L.-O. Wahlund,Y. Dodge, and T. Dierks, "Application and comparison of classificationalgorithms for recognition of alzheimer's disease in electrical brain[activity (eeg),"Journal of neuroscience methods, vol. 161, no. 2, pp.342–350, 2007.

[8] Hazarika N, Chen JZ, Tsoi AC, Sergejew A, Classification of EEG signals using the wavelet transform, Signal Process 59(1):61, 2005.

[9]N. N. Kulkarni and V. K. Bairagi, "Extracting Salient Features for EEG based Diagnosis of Alzheimer Disease Using Support Vector Machine Classifier", Taylor and Francis - IETE Journal of Research, Vol 63, No 1, pp 11-22, Jan 2017. (Impact Factor 0.909)

[10]M. R. N. Kousarrizi, A. A. Ghanbari, M. Teshnehlab, M. Aliyari, and A. Gharaviri, "Feature extraction and classification of EEG signals using wavelet transform, SVM and artificial neural networks for brain computer interfaces," in Proceedings of the International Joint Conference on Bioinformatics, Systems Biology and Intelligent Computing (IJCBS '09), pp. 352–355, August 2009.