Abstract: Feature extraction of EEG signals is one of the major issues on EEG based brain mapping analysis. The classification of EEG signals has been performed using features extracted from EEG signals. Many features have proved to be unique enough to use in all brain related medical application. EEG signals can be classified using a set of features like Autoregression, Energy Spectrum Density, Energy Entropy, and Linear Complexity. However, different features show different discriminative power for different subjects or different trials. In this research, two-features are used to improve the performance of EEG signals. Neural Network based techniques are applied to feature extraction of EEG signal. This paper discusses extracting features based on Average method and Max & Min method of the data set. The Extracted Features are classified using Neural Network Temporal Pattern Recognition Technique. The two methods are compared and performance is analyzed based on the results obtained from the Neural Network classifier.

Keywords: EEG, Feature Extraction, Feature Classification, Temporal Pattern Technique.

I. INTRODUCTION

Analysis of brain signals that provides direct communication between the brain and a body can help patients who suffer from ill health and several psychic problems and severe motor impairments to improve their living quality [1-5]. The mental decision and reaction into control commands by analyzing the bioelectrical brain activity. A kind of analysis brain computer interface system based on analysis of EEG. Generally, the EEG has poor spatial resolution and low signal-to-noise ratio (SNR) of any evoked response embedded within ongoing background activity. To distinguish signals of interest from the background activity various feature extraction methods have been applied, including autoregressive models [6-8], phase [9-10], entropy [11], spatial filter [12-14], wavelet transform [15-16], etc. It is known that EEG signals under appropriate well designed experimental paradigms allow a subject to convey her/his intentions by e.g. motor imagery or executing specific mental tasks. Once the intentions have manifested themselves in brain activity and have been measured by EEG, the scene is set for advanced signal processing and machine learning technology. Feature vectors need to be extracted from the EEG signals, then this feature vectors are translated by machine learning techniques like linear discriminant analysis or neural networks. It’s helpful for classification that the EEG-features are extracted such that they hold the most discriminative information for a chosen paradigm. Several authors point out the potential gain in using all such features. However, investigations of feature combining were announced, but so far poorly covered in publications [17]. This paper describes a two-feature EEG signals. Our aim in this papers it to provide further perspective on the possibility of EEG.

II. DATA DESCRIPTION

EEG signals are extracted from sophisticated machines in highly secured and de-noised labs are prone to artifacts and several other type of non-separable noise. EEG signal when analyzed has a very low frequency in the range of hertz. These EEG signals can be classified based on their frequency bands. The classification is shown in Table.1 it also mentions the region of brain from where it is extracted.

Table.1 Classification of EEG Signals Based On Their Frequency

<table>
<thead>
<tr>
<th>Type</th>
<th>Frequency</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>Delta</td>
<td>up to 4</td>
<td>Frontally in adults, posteriorly in children; high amplitude waves</td>
</tr>
<tr>
<td>Theta</td>
<td>4 – 8</td>
<td>Found in locations not related to task at hand</td>
</tr>
<tr>
<td>Alpha</td>
<td>8 – 13</td>
<td>Posterior regions of head, both sides, higher in amplitude on non-dominant side</td>
</tr>
<tr>
<td>Beta</td>
<td>13 – 30</td>
<td>Both sides of Brain, symmetrical distribution, most evident frontally; low amplitude waves</td>
</tr>
<tr>
<td>Gamma</td>
<td>31 - 100</td>
<td>Somatosensory cortex</td>
</tr>
</tbody>
</table>

As we have discussed earlier it very difficult to extract EEG signal from the brain and separate the artifacts, based on the classification of their frequency we generates signals of those frequency. our data will be simulated EEG signals.

III. FEATURES EXTRACTION

In pattern recognition, feature extraction is a special form of dimensionality reduction. When the input data to an algorithm is too large to be processed and it is suspected to be notoriously redundant (much data, but not much information) then the input data will be transformed into a reduced representation set of features (also named features vector). Transforming the input data into the set of features is called feature extraction. If the features extracted are carefully chosen it is expected that the features set will extract the relevant information from the input data in order to perform the desired task using this reduced representation instead of the full size input. Feature extraction involves simplifying the amount of resources required to describe a large set of data accurately. When performing analysis of complex data one of the major problems stems from the number of...
variables involved. Analysis with a large number of variables generally requires a large amount of memory and computation power or a classification algorithm which over fits the training sample and generalizes poorly to new samples. Feature extraction is a general term for methods of constructing combinations of the variables to get around these problems while still describing the data with sufficient accuracy. Fig 1 describes the flow model of the paper.

Fig 1: Proposed System Model

A. Average Method
The large data set is divided into 20 samples and average is computed for that set, then next 20 samples are taken and average is computed for that data set and process is repeated for all the samples and for the five set signals. The algorithm for computing average is given by

\[
\text{Average} = \frac{\text{sum}(f_{\theta_1}(k:k+19))}{20};
\]

\[
\text{Averages_signal2} = \left[ \text{averages_signal2}, \text{average} \right];
\]

Where K indicates start of the loop

B. Max Min method
The large data set is divided into 20 samples and maximum and minimum feature among 20 data samples are chosen as the and process is repeated for all the samples and for the five set signals. The principle in extracting max and min feature is given below:

\[
\text{Max_min_calculate} = \left[ \max(f_{\theta_1}(k:k+19)) \min(f_{\theta_1}(k:k+19)) \right];
\]

\[
\text{Max_min2} = \left[ \text{max_min2}, \text{max_min_calculate} \right];
\]

Then this process is repeated for all the available data set. All the average are stored in two columns one for maximum feature and the other for minimum feature.

IV. CLASSIFICATION

Neural network: The term neural network was traditionally used to refer to a network or circuit of biological neurons. The modern usage of the term often refers to artificial neural networks, which are composed of artificial neurons or nodes. Thus the term has two distinct usages:

Biological neural networks are made up of real biological neurons that are connected or functionally related in a nervous system. In the field of neuroscience, they are often identified as groups of neurons that perform a specific physiological function in laboratory analysis.

Artificial neural networks are composed of interconnecting artificial neurons. Artificial neural networks may either be used to gain an understanding of biological neural networks, or for solving artificial intelligence problems without necessarily creating a model of a real biological system.

V. NETWORK ARCHITECTURES FEED FORWARD NEURAL NETWORK

A feed forward neural network is an artificial neural network where connections between the units do not form a directed cycle. The feed forward neural network was the first and arguably simplest type of artificial neural network devised. In this network, the information moves in only one direction, forward, from the input nodes, through the hidden nodes (if any) and to the output nodes. There are no cycles or loops in the network. As shown in Fig 2.
Two layer feed forward network: A two-layer neural network capable of calculating XOR. The numbers within the neurons represent each neuron’s explicit threshold (which can be factored out so that all neurons have the same threshold, usually 1). The numbers that annotate arrows represent the weight of the inputs. This net assumes that if the threshold is not reached, zero (not -1) is output. Note that the bottom layer of inputs is not always considered a real neural network layer as shown in Fig 3.

![Fig 3: 2 Layer Feed Forward Neural Network](image)

VI. ALGORITHMS

Data division:

Dividing the Data: When training multilayer networks, the general practice is to first divide the data into three subsets. The first subset is the training set, which is used for computing the gradient and updating the network weights and biases. The second subset is the validation set. The error on the validation set is monitored during the training process. The validation error normally decreases during the initial phase of training, as does the training set error. However, when the network begins to overfit the data, the error on the validation set typically begins to rise. The network weights and biases are saved at the minimum of the validation set error. This technique is discussed in more detail in Improving Generalization. The test set error is not used during training, but it is used to compare different models. It is also useful to plot the test set error during the training process. If the error on the test set reaches a minimum at a significantly different iteration number than the validation set error, this might indicate a poor division of the data set.

VII. TRAINING

Scaled conjugate gradient: As an illustration of how the training works, consider the simplest optimization algorithm — gradient descent. It updates the network weights and biases in the direction in which the performance function decreases most rapidly, the negative of the gradient. One iteration of this algorithm can be written as

\[ x_{k+1} = x_k - \alpha_k g_k \]

Where \( x_k \) is a vector of current weights and biases, \( g_k \) is the current gradient, and \( \alpha_k \) is the learning rate. This equation is iterated until the network converges.

Training stops when any of these conditions occurs:

- The maximum number of epochs (repetitions) is reached.
- The maximum amount of time is exceeded.
- Performance is minimized to the goal.
- The performance gradient falls below \( \text{min}_\text{grad} \).
- Validation performance has increased more than \( \text{max}_\text{fail} \) times since the last time it decreased (when using validation).

VIII. PERFORMANCE

Mean square error (MSE): is a network performance function. It measures the network’s performance according to the mean of squared errors. Mean squared error (MSE) of an estimator is one of many ways to quantify the difference between values implied by an estimator and the true values of the quantity being estimated. MSE is a risk function, corresponding to the expected value of the squared error loss or quadratic loss. MSE measures the average of the squares of the “errors.” The error is the amount by which the value implied by the estimator differs from the quantity to be estimated. The difference occurs because of randomness or because the estimator doesn’t account for information that could produce a more accurate estimate. As shown in Fig 4.

![Fig 4: Performance analysis](image)

IX. RESULT

After extracting the features from two methods Average method and Max_Min method. The comparison is done between these two models and performance is checked by classifying the data using this two methods the classifier work is done by Neural Network. The methods are compared for performance before that the data is trained by neural Network pattern recognition tool box. Error histogram is plotted and checked for the accuracy shown below in Fig 5.
In the field of artificial intelligence, a confusion matrix is a specific table layout that allows visualization of the performance of an algorithm, typically a supervised learning one. Each column of the matrix represents the instances in a predicted class, while each row represents the instances in an actual class. If a classification system has been trained to distinguish between cats, dogs and rabbits, a confusion matrix will summarize the results of testing the algorithm for further inspection. As shown in Fig 6.

In signal detection theory, a receiver operating characteristic (ROC), or simply ROC curve, is a graphical plot which illustrates the performance of a binary classifier system as its discrimination threshold is varied. ROC analysis provides tools to select possibly optimal models and to discard suboptimal ones independently from (and prior to specifying) the cost context or the class distribution. ROC analysis is related in a direct and natural way to cost/benefit analysis of diagnostic decision making. ROC is been used in medicine, radiology, biometrics, and other areas for many decades and is increasingly used in machine learning and data mining research. As shown in below Fig 7.

**X. CONCLUSION**

Features were extracted using Average method and Max.Min method. Two Features extraction methods are evaluated for their performance using Pattern Recognition tool box from the obtained results it has observed that the Max.Min feature extraction method gives better accuracy compared to the Average Feature Extraction Method and Accuracy of Max.Min method is 80%. Accuracy of Average method is 41%.

**REFERENCES**


AUTHOR’S PROFILE

Nandish M presently doing Master’s degree in Digital electronics and Communication from AMCEC Bangalore. Currently he is a student at AMCEC Bangalore.

Stafford Michahial is presently doing Master’s degree in Digital electronics and Communication from AMCEC Bangalore. Currently he is a student at AMCEC Bangalore. My area of interest includes Neural Network, Data Mining, Image Processing, Communication etc. I have presented a paper in one of the national conferences, and 3 international journals

Hemanth Kumar P. received the Master's degree in Digital Communication from National Institute of Technology (MANIT) Bhopal, Madhya Pradesh, India in the year 2010. Currently working as Assistant Professor in AMC Engineering College, Bangalore in Department of Electronics & Communication Engineering. I have cleared PhD entrance in the stream of Computer Science Engineering from Visvesvaraya Technological University. My area of interest includes Neural Network, Data Mining, Image Processing, Communication etc. I have two years of Teaching Experience and guided two M.Tech projects successfully.

Faizan Ahmed is presently doing B.E (Final year) in computer science from V.V.I.E.T Mysore.